UNIVERSITY *of the*
WESTERN CAPE

DEPARTMENT OF STATISTICS

MARKET SEGMENTATION AND FACTORS AFFECTING STOCK RETURNS ON THE JSE

MSc COMPUTATIONAL FINANCE PROJECT

MAY 15, 2008

STUDENT NAME : Artwell S.Chimanga

PROJECT SUPERVISOR: Prof. Danelle Kotze

Plagiarism Declaration

- I hereby declare that I know what plagiarism entails, namely to use another's work and to present it as my own without attributing the sources in the correct way.
- I know that plagiarism is a punishable offence because it constitutes theft.
- I understand the plagiarism policy of the Faculty of Natural Science of the University of the Western Cape.
- I know what the consequences will be if I plagiarise.
- I declare therefore that this project is all my work, and where I have made use of another's work, I will attribute the source in the correct way.

Signature: ....

Date: ....

This project has been submitted in partial fulfillment of the requirements of a "MSc Computational Finance" at the University of the Western Cape, South Africa.

Contents

List of Figures

List of Tables

Market Segmentation and Factors Affecting Stock Returns on the JSE *

Artwell S Chimanga

Department of Statistics

University of the Western Cape

South Africa

Abstract

This study examines the relationship between stock returns and market segmentation. Monthly returns of stocks listed on the JSE from 1997-2007 are analysed using mostly the analytic factor and cluster analysis techniques. Evidence supporting the use of multi-index models in explaining the return generating process on the JSE is found. The results provide additional support for Van Rensburg (1997)'s hypothesis on market segmentation on the JSE.

Keywords: Market-Segmentation, Principal Components, Cluster Analysis, Multifactor models, Co-variances, Abitrage Pricing Theory, Sector Indices, JSE.

*This project has been submitted in partial fulfillment of the requirements of a "MSc Computational Finance" at the University of the Western Cape, South Africa.

1 Introduction

This study seeks to identify the forces that drive stock market returns. The goal is to come up with factor models that completely capture all the systematic components of stock return co-variances, which is attributable to the differential rate of returns for different sectors or partial segmentation of financial markets. The research will also endeavour to show that multi-index models are appropriate in capturing the co-variation that is usually relegated to the residual errors in single index models.

Multivariate techniques are going to be empolyed to test the market hypothesis theory and come up with factors that explain returns. Factor analysis is applied to compounded monthly stock returns. The aim of the analysis is to discover and "explain" the degree of cross-sectional interdependence exhibited by the returns series. A desired result is to separate the large set of individual series into a smaller set of clusters of security price changes that tend to move as homogeneous groups. Conclusions are drawn concerning the degree of agreement between cluster analysis and the statistical factor analytic method.

The statistical factor analysis yielded a two factor model. The macro-economic identities of these factors provide interesting insights into the economic determinants of the JSE returns. The first factor can be represented by the Industrial Index and the second factor can be linked to the Gold Mining Index. These results emphasise the importance of the Market Segmentation Hypothesis on the JSE's return generating process. Results also show that the multi-index model provides a better representation of the determinants of stock returns than the Capital Asset Pricing Model (CAPM). The study also documents the links between the risk premia obtained in both approaches.

The research is organized as follows. The next section discusses the data and data manipulation used in this project. The third section presents the methodology, and the fourth section discusses the theoretical framework and reviews of other studies done on factor analysis.

The fifth section discusses the market segmentation hypothesis. The sixth section reports the estimation results on market segmentation and risk factor based on cluster and factor analysis methodology. The seventh section concludes the study.

2 Data Samples and Description

The JSE is the 14th largest equities exchange in the world, with a total market capitalisation of R5.7 trillion ($730bn). As is the case with most emerging markets, the JSE market is highly concentrated with the five largest companies accounting for almost 37% of the total market capitalisation. As a result, the remainder of the market suffers from severe liquidity problems, with some stocks being traded on a very irregular basis. The JSE market liquidity is currently around 30%.

To avoid the issue of liquidity the study will focus on portfolios representing various industry sub-sectors for cluster analysis, and the top 100 shares according to maket capitalisation for factor analysis. The data sample contains monthly share prices listed from January 1997 to December 2007. The choice of using a monthly frequency is dictated by the need to avoid serial correlation among the data. Statistical factor analysis assumes that the data has no serial correlation, and this assumption is often violated by financial data taken with a frequency less than or equal to a week. The tickers for the shares used in this study can be found in the appendix.

The data were cleaned, using the skipped Huber method. The method involves using confidence limits calculated from medians of share price series to identify extreme values. The only problem with the method is that it does not differentiate between observations that take on extreme values because of measurement errors and those that are real extreme values. Missing data were replaced using the mean of nearby data points.

In the next step the filtered and standardized data is then used to calculate the returns series. The returns on equity prices are calculated as log difference of the prices plus dividends *

$$R_{jt} = log_e((P_{jt} + D_{jt})/P_{jt-1})$$

where R_{jt} is the return on share j in month t; P_{jt} and P_{jt-1} are the prices of share j in month t and D_{jt} is the dividend paid on share j in month t. Since the computation of logarithmic returns involves the loss of the observation of price in the sample the price series is adjusted to have the same starting point as the log return series.

*see Barr (1990), R Mangani (2007) for the distributional properties of JSE prices and returns.

3 Methodology

3.1 Market Segmentation Hypothesis

Cluster analysis which is often referred to as market segmentation analysis, will be used to help establish market composition of the JSE, by sub-dividing it into discrete groups (known as 'clusters'). The advantage of using cluster analysis over other techniques used in most empirical research is that it is a classification technique.

In this study, an attempt is made to group stocks by sectors. Cluster analysis involves searching for natural groupings amongst objects and is thus a more inductive approach. Cluster analysis will be performed over two time-periods i.e. (1997-2002) and (2002-2007). The reason for splitting the 10-year data into two periods is to give a reflection on the developments within companies over the periods.

3.1.1 Similarity Measures

There are different methods for computing similarity measures between objects. A fundamental method is the shortest Euclidean distance method between two p dimensional observations. For each pair of shares, it is computed as:

$$distance(x, y) = [\Sigma(x_i - y_i)^2]^{\frac{1}{2}}$$

where x_i is the total return in month i from share x.
Cluster analysis is used to objectively group the stocks using a hierarchical agglomeration algorithm. The cluster analysis method uses as input the correlation coefficients in the form of a similarity matrix of the listed stocks, and sequentially merges the most similar cases. The mergers of clusters are represented visually via a tree diagram called a dendogram. This represents a hierarchical organisation of the relations between the data points. We propose the use of the average linkage method for calculating the distance between clusters. This method computes the average of stock returns, and joins stocks to a cluster depending on

this arithmetic average linkage. Its advantage lies in the fact that averaging returns across time compensates for the effects of the stochastic movement of JSE return series.

3.2 Identifying Factors

This study makes use of purely statistical factors to investigate the degree of commonality between JSE stock returns. A number of principal components will be extracted and regressed against returns for each market proxy. The technique allows us to compress a large set of correlated variables into a smaller set of principal components or factors which are mutually orthogonal and explain a significant proportion of the variability of the original set of variables. The factors of the Arbitrage Pricing Model (APT model) are the principal components of the space of security returns. Each variable is assumed to be dependent on a linear combination of the common factors, and the coefficients are known as loadings. Each measured variable also includes a component due to independent random variability, known as 'specific variance' because it is specific to one variable. Specifically, factor analysis assumes that the covariance matrix of the data is of the form.

$$\Sigma_x = \Lambda\Lambda^T + \Psi$$

where Λ is the matrix of loadings and the elements of the diagonal matrix Ψ are the specific variances. The decomposition of the index return space generally involves the rotation of the estimated factors so that the factor loadings are either very large or very small. The procedure used in order to perform this is known as the Varimax rotation, see Johnson and Wichern (1982). Orthogonalization eliminates any effects of causal interactions among factors. The actual estimates of the factors through time are known as the factor scores. These are simply the rotated factors at each point in time as calculated from the equation above. Factor rotation helps to simplify the structure in the loading matrix, to make it easier to assign meaningful interpretations to factors.

3.2.1 Number of Factors

The scree test will be used to determine the number of factors to retain in a factor analysis or principal component analysis. The scree test involves plotting the eigen-values in descending order of their magnitude against their factor numbers and determining where they level off. The break between the steep slope and leveling off indicates the number of meaningful factors. The technique is illustrated and compared with an alternative technique for determining the number of factors to retain. The theoretical framework behind this methodology is discussed in the literature review.

3.3 Validation of Risk Factors

After identifying the factors it is necessary to validate them. Here a technique proposed by Fama and MacBeth (1973) is empolyed. The data set is first divided into 120 periods to estimate the factor sensitivities. The sensitivities referred to here, are the coefficients obtained from the regression of our returns on the realizations as described by the equation below:

$$r_{it} = \beta_{i0} + \sum_{k=1}^{k} F_{ik}\beta_{kt} + \epsilon_{it}$$

where r_{it} represents the stock returns, F_{ik} represents the factors, β_{kt} represents the stock sensitivity, and e_{it} denotes residuals that are uncorrelated to the market. In the second step, we perform cross-sectional regression on our estimated sensitivities as described in the equation below:

$$r_{it} - r_{rfr} = \alpha_{0t} + \sum_{j=1}^{k} \alpha_{ik}\hat{\beta}_{kt} + \epsilon_{it}$$

where r_{rfr} is the risk free rate proxied by the call rate, $\hat{\beta}_{kt}$ is the estimated sensitivity of each individual stock to the ith factor. This procedure produces estimates of the risk premia for our factors for the 120 months under study. The next step is to test whether these estimates are different from zero, if the mean of these estimates are significantly different from zero, then it means our factors explain the return generating process.

The cross-sectional regression test is also performed on the estimated sensitivities taking into consideration the upward and downward markets. It is generally believed that there is a positive risk-return relationship in upward markets and a negative relationship in downward movement in the market. The relationship is described in the equation below.

$$r_{it} - r_{rfr} = \alpha_{0t} + \sum_{j=1}^{k} \alpha_{jt}^{+} \rho_j \hat{\beta}_{ij} + \sum_{j=1}^{k} \alpha_{jt}^{-} (1 - \rho_j) \hat{\beta}_{ij} + \epsilon_{it}$$

where α_{0t} is the constant term, $\alpha_{jt}^{+(-)}$ is the estimated risk premium for factor j conditional on a positive (negative) realization of factor j in month t, ρ_j is a dummy variable equal to 1 when factor j is positive and equal to 0 otherwise and ϵ_{it} represents the residual error term.

4 Arbitrage Pricing Theory

The APT of Ross (1976) provides a theoretical framework to determine the expected returns on stocks, but it does not specify the number of factors nor their identity. Hence, the implementation of this model follows two avenues: factors can be extracted by means of statistical procedures, such as factor analysis or principal component analysis, or be pre-specified using mainly macro-economic variables. This study will focus mainly on the statistical procedures.

4.1 Single-Index Model

This is the single-index covariance matrix of Sharpe (1963). Sharpe (1963)'s single-index model assumes that stock returns are generated by:

$$r_{it} = \alpha_i + \beta_i r_{0t} + \epsilon_{it}$$

where α_i and β_i are regression coeffients, and residuals e_{it} are uncorrelated to market returns r_{0t} and to one another. Also, within stocks the variance is constant, that is, $Var(e_{it}) = \sigma_{ii}$. The covariance matrix implied by this model is:

$$\mathbf{\Phi} = \sigma_{00}^2 \beta \beta^{'} + \mathbf{\Delta}$$

where σ_{00}^2 is the variance of market returns, β is the vector of slopes and $\mathbf{\Delta}$ is the diagonal matrix containing residual variances δ_{ii}. Let ϕ_{ij} be the $(i,j)-th$ entry of $\mathbf{\Phi}$. This model can be estimated by running a regression of stock $i's$ returns on the market. Let b_i be the slope estimate and d_{ii} the residual variance estimate. Then the single-index model yields the following estimator for the covariance matrix of stock returns:

$$\mathbf{F} = s_{00}^2 \mathbf{b}\mathbf{b}^{'} + \mathbf{D}$$

where s_{00}^2 is the sample variance of market returns, $\mathbf{b}$ is the vector of slope estimates and $\mathbf{D}$ is the diagonal matrix containing residual variance estimates d_{ii}. Let f_{ij} be the (i,j)-th entry of $\mathbf{F}$.

Two technical assumptions are made.

- Assumption 1. $\mathbf{\Phi} \neq \mathbf{\Sigma}$.
- Assumption 2. The market portfolio has positive variance, that is, $\sigma_{00}^2 > 0$.

There is a close relationship between the single index model and the CAPM. The only major difference between the two is that the exact composition of the market portfolio is not as critical here as it is for the CAPM (Roll, 1977). This means that any market index would do as long as it explains a significant part of the variance of most stocks.

4.2 Industry Factors

This refinement of the single-index model assumes that market residuals are generated by industry factors:

$$r_{it} = \alpha_i + \beta_i r_{0t} + \sum_{k=1}^{k} c_{ik} z_{kt} + \epsilon_{it}$$

where k is the number of industry factors, c_{ik} is a dummy variable equal to one if stock i belongs to industry category k, z_{kt} is the return to the k-th industry factor in period t and e_{kt} denotes residuals that are uncorrelated to the market, to industry factors and to each other. Industry factor returns are defined as the return to an equally-weighted portfolio of the stocks from this industry in our sample.

4.3 Statistical Factors

An alternative approach to multifactor models is to extract the factors from the sample covariance matrix itself using a statistical method such as principal components, see Chen et al., (1986). Statistical factor models treat the common factors as unobservable or latent variables to be estimated from return series. Since principal components are chosen solely for their ability to explain risk, fewer factors are necessary, but they do not have any direct economic interpretation. However, offering economic insight does not necessarily generate a valid model, see Fama (1991). The advantage of factor analysis is the absolute objectivity of the approach. Neither the sensitivities nor the factors are defined in advance, but rather are estimated based on the data.

Lets consider the return $\mathbf{r_t}'$ of k assets at time period t and assume that the return series $\mathbf{r_t}$ is weakly stationary with mean μ and covariance matrix $\mathbf{\Sigma_r}$. The statistical factor model is in the form

$$\mathbf{r_t} - \mu = \beta \mathbf{f_t} + \epsilon_t$$

where $\beta = [\beta_{ij}]_{k\times m}$ is the matrix of factor loadings, β_{ij} is the loading of the ith variable on the jth factor, and ϵ_{it} is the specific error of r_{it}. A key feature of the statistical factor model is that the factors f_{it} and the factor-loadings β_{ij} are unobsevable. Our factor model is orthogonal if it satifies the following assumptions,

- Assumption 1. $E(\mathbf{f}_t) = 0$ and $Cov(\mathbf{f}_t) = \mathbf{I_m}$, an $m \times m$ identity matrix;
- Assumption 2. $E(\epsilon_\mathbf{t}) = 0$ and Cov(ϵ_t) =D=diag($\sigma_1^2, ..., \sigma_k^2$)
- Assumption 3. $\mathbf{f}_t$ and ϵ_t are independent so that $Cov(f_t, \epsilon_t) = E(f_t \epsilon_t) = \mathbf{0_{m\times k}}$.

4.3.1 Estimation

The orthogonal factor model can be estimated by two methods. The first estimation method uses the principal component analysis. This method does not require the normality assumption of the data nor the prespecification of the number of common factors. It applies to both the covariance and correlation matrices. The second method is the maxium likelihood method that uses normal density and requires a pre-specification for the number of common factors.

Principal Component Method

Let $(\widehat{\lambda}_1, \widehat{e}_1), ..., (\widehat{\lambda}_k, \widehat{e}_k)$ be pairs of the eigenvalues and eigenvectors of the sample covariance matrix $\widehat{\mathbf{\Sigma}}_{\mathbf{r}}$, where $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \;\; \geq \widehat{\lambda}_k$. Let $m < k$ be the number of common factors. then the matrix of factor loadings is given by

$$\widehat{\beta} \equiv [\widehat{\beta}_{ij}] = [\sqrt{\widehat{\lambda}_1 \widehat{e}_1} | \sqrt{\widehat{\lambda}_2 \widehat{e}_2} | .... | \sqrt{\widehat{\lambda}_m \widehat{e}_m}].$$

The estimated specific variances are the diagonal elements of the matrix $\widehat{\mathbf{\Sigma}}_{\mathbf{r}} - \widehat{\beta}\widehat{\beta}'$. The error matrix derived by approximation is

$$\widehat{\mathbf{\Sigma}}_{\mathbf{r}} - (\widehat{\beta}\widehat{\beta}' + \widehat{\mathbf{D}}).$$

Ideally, this matrix should be close to zero. The estimated factor loadings based on the principal components method do not change as the number of common factors m is increased.

Maximum Likelihood Method

If the common factors $\mathbf{f_t}$ and the specific factor errors $\epsilon_{\mathbf{t}}$ are jointly normal, then $\mathbf{r_t}$ is multivariate normal with mean μ and covariance matrix $\mathbf{\Sigma_r} = \beta\beta' + \mathbf{D}$. The maximum likehood method can be used to obtain estimates of β and $\mathbf{D}$ under the constraint $\beta' \mathbf{D^{-1}} \beta = \mathbf{\Delta}$, which is a diagonal matrix. Here μ is estimated by the sample mean see, Johnson and Wichern (1982). In using the maximum likelihood method, the number of common factors must be given a priori. In practice, one can use a modified likelihood ratio test to check the adequacy of a fitted m-factor model. The test statistic is

$$LR(m) = -[T - 1 - \frac{1}{6}(2k+5) - \frac{2}{3}m](ln|\widehat{\mathbf{\Sigma_r}}| - In|\widehat{\beta}\widehat{\beta}' + \widehat{\mathbf{D}}|)$$

which, under the null hypothesis of m factors, is asymptotically distributed as a chi-squared with $\frac{1}{2}[(k-m)^2 - k - m]$ degrees of freedom.

4.3.2 Factor Rotation

As mentioned before, for any $m \times m$ orthogonal matrix $\mathbf{P}$,

$$\mathbf{r_t} - \mu = \beta \mathbf{f_t} + \epsilon_\mathbf{t} = \beta^* \mathbf{f_t^*} + \epsilon_\mathbf{t},$$

where $\beta^* = \beta\ \mathbf{P}$ and $\mathbf{f_t^*} = \mathbf{P}' \mathbf{f_t}$. In addition,

$$\beta\beta' + \mathbf{D} = \beta\mathbf{P}\mathbf{P}'\beta' + \mathbf{D} = \beta^*(\beta^*)' + \mathbf{D}$$

This result indicates that the commonalities and the specific variances remain unchanged under an orthogonal transformation. It is then reasonable to find an orthogonal matrix $\mathbf{P}$ to transform the factor model so that the common factors have simple interpretations. Denote the rotated matrix of factor loadings by $\beta^* = [\beta^*_{ij}]$ and the ith communality by c_i^2. Define $\widetilde{\beta}^*_{ij} = \beta_{ij}/c_i$ to be rotated coefficients scaled by the (positive) square root of communalities. The varimax procedure selects the orthogonal matrix $\mathbf{P}$ that maximises the quantity.

$$V = \frac{1}{k} \sum_{j=1}^{m} [\sum_{i=1}^{k} (\widetilde{\beta}^*_{ij})^4 - \frac{1}{k} (\sum_{i=1}^{k} \beta^{*2}_{ij})^2].$$

Maximising V corresponds to spreading out the squares of the loadings on each factor as much as possible. Consequently, the procedure is to find groups of large and negligible coefficients in any column of the rotated matrix of factor loadings.

4.4 Asymptotic Principal Component Analysis

The classic principal components analysis (PCA) discussed in section 4.3.1 only works in cases where the time series dimension T is greater than the cross-sectional dimension k (i.e. the number of stocks). In cases where $T < k$, the asymptotic principal component analysis (APCA) is used (Connor and Korajczyk (1993)). The method relies on the asymptotic results as the number of assets k increases to infinity. Thus, the APCA is based on eigenvalue-eigenvector analysis of the $T \times T$ matrix below

$$\widehat{\mathbf{\Omega}}_{\mathbf{T}} = \frac{1}{k-1}\sum_{i=1}^{k}(\mathbf{R_i} - \overline{\mathbf{R}} \otimes \mathbf{1}_{\mathbf{k}}^{'})(\mathbf{R_i} - \overline{\mathbf{R}} \otimes \mathbf{1}_{\mathbf{k}}^{'})^{'}$$

where $\mathbf{R_i}$ is the time series of the ith asset, $\overline{\mathbf{R}} = \frac{1}{k}\sum_{i}^{k}\mathbf{R_i}$ and $\mathbf{1_k}$ is the k-dimensional vector of ones.† Connor and Korajczyk (1993) proposed refining the estimation of $\widehat{\mathbf{f}}_{\mathbf{t}}$ (the first m eigenvectors of $\widehat{\mathbf{\Omega}}_{\mathbf{T}}$) as follows:

- use the sample covariance matrix to obtain an initial estimate of $\widehat{\mathbf{f}}_{\mathbf{t}}$ for $t = (1, ..., T)$.
- For each asset, perform the OLS estimation of the model $r_{it} = \alpha_i + \beta_i^{'}\widehat{f}_t + \epsilon_{it}$ where $t = (1, ..., T)$ and compute the residual variance $\widehat{\sigma_i^2}$.
- Form the diagonal matrix $\widehat{\mathbf{D}} = diag\{\widehat{\sigma}_1^2, ..., \widehat{\sigma}_k^2\}$ and rescale the returns.
- Compute the $T \times T$ covariance matrix using $\mathbf{R}_*$ the refined estimate of $\mathbf{R}_i$ as

  $$\widehat{\mathbf{\Omega}}_* = \frac{1}{k-1}\sum_{i=1}^{k}(\mathbf{R}_* - \overline{\mathbf{R}}_* \otimes \mathbf{1}_{\mathbf{k}}^{'})(\mathbf{R}_* - \overline{\mathbf{R}}_* \otimes \mathbf{1}_{\mathbf{k}}^{'})^{'}$$

  where $\overline{\mathbf{R}}_*$ is the vector of row averages of $\mathbf{R}_*$, and perform eigenvalue-eigenvector analysis of $\widehat{\mathbf{\Omega}}_*$ to obtain a refined estimate of $\mathbf{f_t}$, called $\mathbf{f_{t*}}$.

4.4.1 Selecting the Number of Factors

Determining the appropriate number of factors is crucial to APT analysis. Two methods are available to help select the number of factors in factor analysis. The first method proposed by Connor and Korajczyk (1993) makes use of the idea that if m is the proper number of common factors, then there should be no significant decrease in the cross-sectional variances of the asset specific error ϵ_{it} when the number of factors moves from m to $m+1$. The second method proposed by Bai and Ng (2002) adopts some information criteria to select the number of factors. This latter method is based on the observation that the eigenvalue-eigenvector analysis of $\widehat{\mathbf{\Omega}}_{\mathbf{T}}$ solves the least squares problem

†Note: As $k \to \infty$ eigenvalue-eigenvector analysis of $\widehat{\mathbf{\Omega}}_{\mathbf{T}}$ is equivalent to the traditional statistical factor analysis.

$$min_{\alpha,\beta,f_t} \frac{1}{kT} \sum_{i=1}^{k} \sum_{t=1}^{T} (r_{it} - \alpha_i - \beta_i' \mathbf{f_t})^2$$

where it is assumed that there are m factors so that $\mathbf{f_t}$ is m-dimensional. Let $\widehat{\sigma}_i^2(m)$ be the residual variance of the inner regression of the prior least squares problem for asset i. The study will be focusing mainly on the second methodology, as it is more intuitive and easy to implement.

5 Market Segmentation

5.1 An Overview of International Markets

In the international literature the segmentation of financial markets is mainly attributed to government controls and restrictions on international capital flows. All of the market segmentation attributes are predominantly found in emerging markets where governments still have major control over financial markets, see Choi and Rajan (1996). Empirical analysis indicates that national equity markets can be described as being partially segmented and partially integrated rather than a polar case of complete segmentation or complete integration. Some studies also suggest that market segmentation can be attributed increasingly to globalisation. For instance, Griffin and Karolyi (1998) find that industries with internationally traded goods are more sensitive to global industry factors than firms that produce goods that are only domestically traded.

5.2 A South African Perspective

The JSE comprises of 32% international shares and the rest are domestic shares. Firms with international traded goods are dependent on international politics and economic events that are usually different to developments in the South African economy. This brings another dimension to the pricing paradigm, as the stock returns are mostly likely to be affected by different underlying factors.

In 1979 Campbell pioneered work on the segmentation of the JSE. It was found that the CAPM beta of the industrial index was negatively related to beta of the Gold Index. This meant that the single index model was no longer sufficient to explain stock returns. It was also found that this non-stationarity of the beta was also predominant among individual shares within each index indicating that different macro-economic factors affected each set of shares over the ten year period under study. This study also revealed that individual share betas were more stable when measured against their respective sector indices.

Van Rensburg and Slaney (1997) documented that segmentation exist on the JSE. They suggested that the JSE Actuaries All Gold and Industrial Indices might be employed as observable proxies for the first two factors analytically extracted on the JSE. They went on to claim that a two factor model specified in this manner provides a more comprehensive explanation of the generating process operational on the JSE than the single index model of Markowitz (1959) and Sharpe (1963). The implication of their finding is that there is a separate pricing paradigm for mining and industrial assets. These results seem to support the two Security Market Line (SML) approach suggest by Campbell (1979) and Venter, Bradfield and Bowie (1992).

Contrary to the above finds Ward (1994) applied a clustering technique to the different sector indices; Ward found the use of the All Share Index rather than the relevant sector index an appropriate market index. Ward also found that there was not a statistical difference between the clusters observed on the JSE, supporting the notion of an integrated rather than a segmented market.

6 Analysis

6.1 Factor Analysis

The method proposed by Bai and Ng (2002) was initially used to determine the number of relevant factors. The method identifies factors by examining scree plots of eigenvalues against the factors; the break between the steep slope and levelling off indicates the number of meaningful factors. Judging from the scree plot and pareto chart, it can be seen that the factor structure can be explained by two factors. The method developed by Connor and Korajczyk (1993) also gives the same result as the Bai and Ng (2002) method. Table 1 reports the values of these statistics for the different number of factors. The statistic rejects the null hypothesis at $K = 3$, meaning only the first two factors are relevant. Figure 1 displays the pareto plot explaining variablity and figure 2 presents the scree plot of eigenvalues.

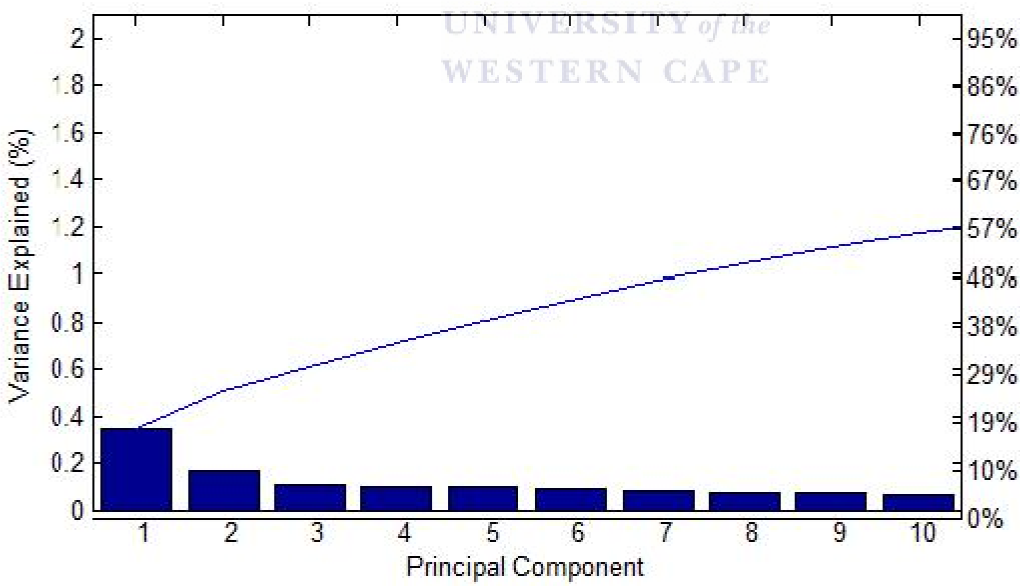

Figure 1: Pareto plot explaining variability.

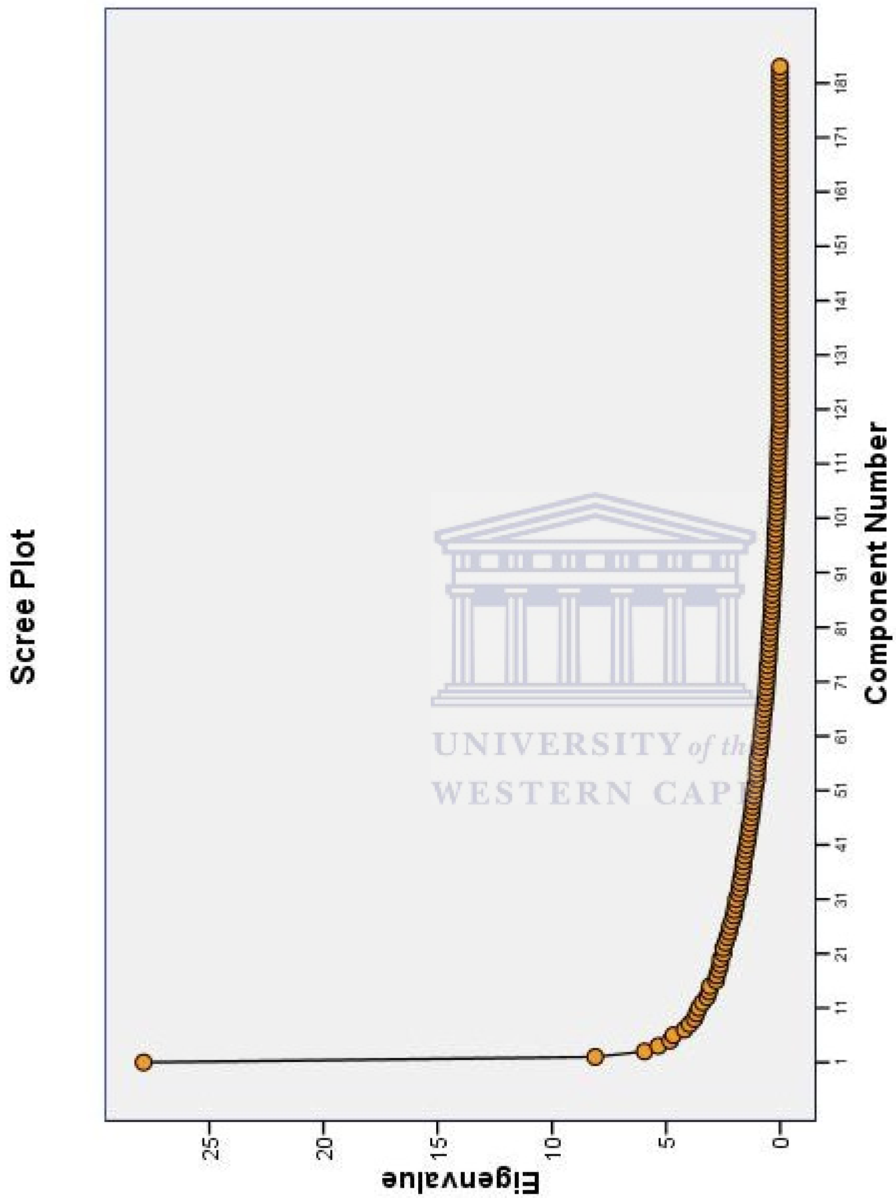

Figure 2: Scree plot of eigenvalues.

Table 1: Test of the number of statistical factors

Number of factors K	Asymptotic t-statistic($\overline{\Delta}_{K,K+1}$)
1	2.34
2	4.39
3	2.53*
4	0.65
5	2.42
6	1.28
7	3.65
8	4.52
9	3.27
10	1.23

* Null hypothesis of having K factors is rejected at the 5% level.

Table 2: Extraction sums of squared loadings.

Component	Total	% of Variance	Cumulative %
1	29.094	18.812	18.812
2	7.778	9.227	28.039
3	5.65	5.071	33.11
4	5.217	4.635	37.745
5	4.878	3.651	41.396
6	4.753	3.483	44.879
7	4.3	3.237	48.116
8	3.841	3.087	51.203
9	3.733	3.009	54.212
10	3.586	2.881	57.260

The principal component analysis is then run in order to determine how much each factor contributes to the JSE stock market returns. Table 2 shows the results for the principal component analysis. The first factor explains 18.812% of the variance, the second factor 9.227%, the third 5.071%, while the ten variables together explain 57.26% of the variance of the stock returns.

6.2 Factor Identification

In this section an attempt is made to give economic interpretation to our statistical factors identified above. The economic factors identified in this section are going to be used as the index proxies for the regression model. The factor variables are identified by correlating the factor scores against economic indicators that we suspect will explain the return generating process. Most of these economic factors have already been identified as being significant in explaining risk in the previous studies done on the JSE. Judging from the results in table 3 it can be seen that there is a very high correlation between Industrials and the first factor (0.891), which means Industrials explain the factor better than the other macro-economic factors. The second factor is best explained by the Gold mining index (0.689), an alternative index could be Resources as it explains almost the same variability (0.663).

6.3 Regression Results

This section test the hypothesis that returns are generated by the two-factor model below

$$r_{it} = \alpha_i + \beta_{iind} r_{indt} + \beta_{igold} r_{goldt} + \epsilon_{it}$$

where β_{iind} is the industrial index sensitivity measure, R_{indt} is the industrial index return, β_{igold} is the gold mines index sensitivity, R_{goldt} is the gold mines returns and and e_{kt} denotes residuals that are uncorrelated to the market, factors and each other.

From table 4 it is clear, that our model provides support for the APT. Thus, the inclusion of the two macro-economic variables gives us better explanation of equities' returns. The

Table 3: Correlations between factor scores and economic factors.

	REGR factor score 1	REGR factor score 2	REGR factor score 3
Resources (J258)	0.536(**)	.663(**)	.080
Indusrial(J520)	.891(**)	.092	-.023
Financials(J580)	.854(**)	-.063	-.007
Gold mines(J150)	.269(**)	.689(**)	-.051
Platinum mines(J153)	.324(**)	.571(**)	.148
Other mineral (J154)	.467(**)	.405(**)	.039
USDZAR	-.297(**)	.275(**)	.076
Commodities(WCPALL)	.310(**)	.121	.097
Long rate bond(RLRS)	-.558(**)	.167	-.038
S&P 500 (FSPI)	.548(**)	.013	.108

** Significant correlations at 5%.

Table 4: Results of the cross-sectional test of the statistical APT.

Average risk premia computed over the whole sample			
	Constant	Factor 1	Factor 2
Average risk premia	.00066	.045	.054
Positive factor realisation	-	0.245*	0.423*
Negative factor realisation	-	-0.757*	-0.564*
Risk premia difference	-	-0.512	-0.141

*Average risk premium is significant at the 5% level.

Table 5: Average R^2 results from regression.

	Gold and Industrial as proxies	Model with ALSI as proxy.
Average Adjusted R^2	.417	.262
Average mean of variation	.0808	.269

average adjusted R^2 is equal to 0.417 and the average cross-sectional R^2 is 0.6401. These levels of explanatory power are very high for cross-sections of stock returns. As the cross-sectional regressions are performed on excess returns, the constant term should be equal to zero. Table 3 indicates that the average risk premium on the intercept is close to zero and cannot be statistically distinguished from this value, which shows that the cross-sectional variation in excess returns is fully captured by the sensitivities to the two factors. It is noted that the risk premia values for both factors were not significant at the 5 % level. Their significance however improved when the positive and negative realisations of the factors were taken into account. It is also evident from table 4 that our two factor model explained 15.5% more of the average variation in individual share returns when compared to the CAPM.

6.4 An Analysis of Segmentation

This section gives a brief discussion of the market segementation results. Cluster analysis and the factor analytic method were used to test for market segmentation. Sector indices were included together with share returns for the factor analysis. Returns of the following indices were used in the analysis; Resources (J258), Industrial (J520), Financials (J580), Gold mines (J150), Platinum mines (J153), Other mineral (J154), USDZAR, Commodities (WCPALL), Long rate bond (RLRS) and SP 500 (FSPI). The varimax factor loadings are graphically presented in figure 3. ‡

The factor analytic results seem to suggest that the market is divided into two segments; the Gold mining/Resources for the smaller group of shares and Industrial/Financials for the larger group. These results give further evidence to Page (1996), Gilbertson and Goldberg (1981) and Van Rensberg and Slaney (1997)'s two index multi-market models that employ the All Gold and Industrial indices as proxies for the return generating process.

Cluster analysis results were inconclusive as clusters constituents kept changing over the two periods under study. Sokal and Rohlf (1962) suggested that in order to claim market segmentation, the dendogram should have the same cluster constituents over the periods under study. On the other hand to claim market integration it is imperative that the within cluster sum of squares be as close to zero as possible, so as to indicate a tight spread around the centroid which is not the case with our results. Hence we cannot make any meaningful conclusion from cluster analysis. Dendograms showing cluster analysis results are given in the appendix.

‡Verimax rotation was used in factors analysis to help simplify the structure in the loading matrix, to make it easier to assign meaningful interpretations to factors.Promax rotation results are given in the appendix.

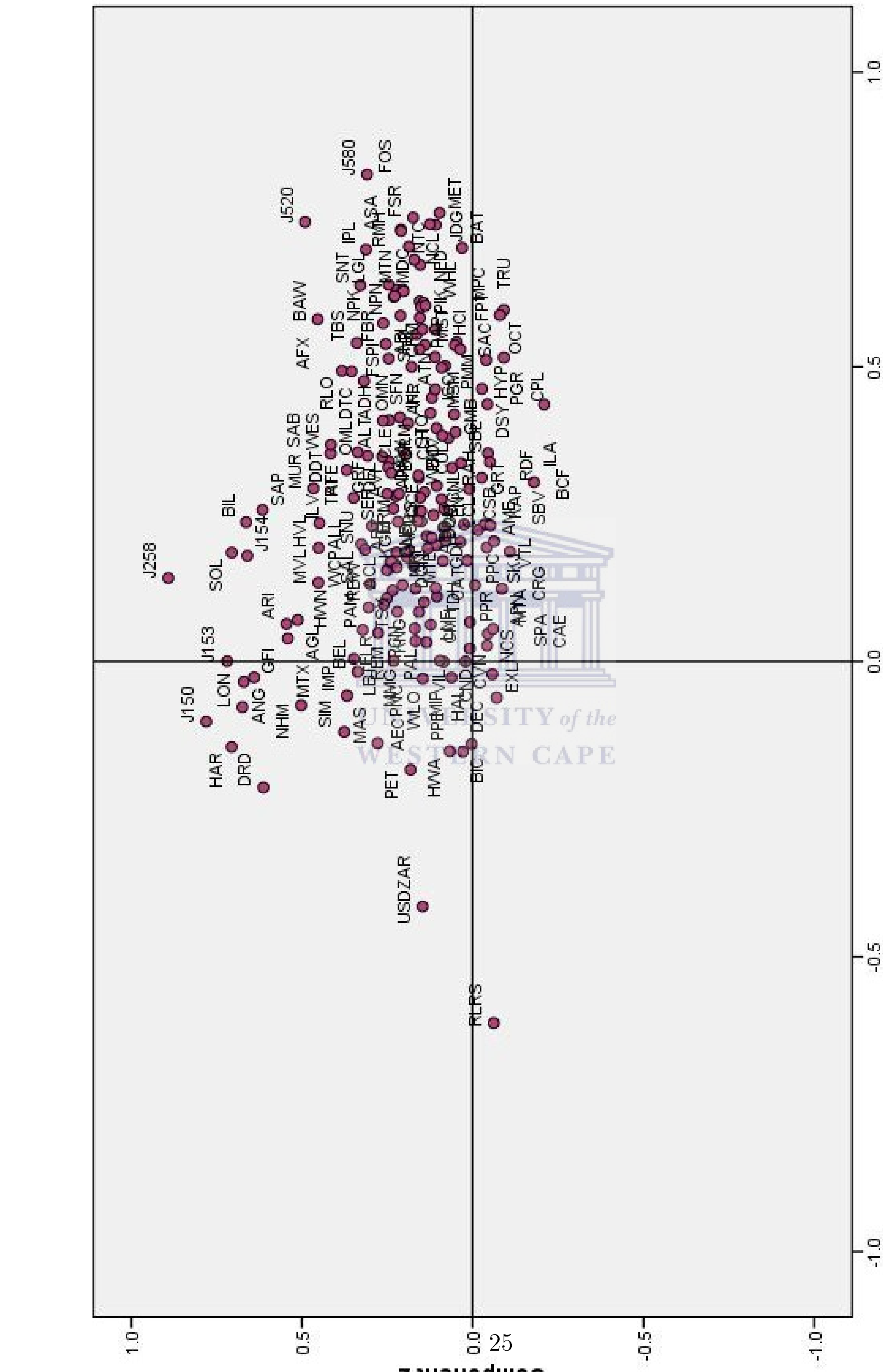

Figure 3: Factor pattern using varimax rotation.

7 Conclusion

This study proposed the use of multi-index models in explaining the return generating process on the JSE. The argument for the use of multi-index models is based on the APT which states that stock market returns can be determined by two or more factors and the fact that emerging markets are generally segmented. Our method involved extracting the model factors using purely statistical methods; principal components and factor analysis. The factors were chosen solely for their ability to explain risk; hence fewer factors were necessary.

We found evidence of partial segmentation on the JSE using the factor analytic method. The results seem to suggest that the Gold Mining and Industrial Indices can be used to explain returns for the JSE. Factor analysis gives further support to Van Rensberg and Slaney (1997)'s work on market segmentation. We also employed cluster analysis to give supporting evidence of market segmentation, but the results were inconclusive, as the cluster's constituents kept changing over the two periods under study.

The validity of our two factor model and the CAPM was tested, using a technique initially proposed by Fama and MacBeth (1973). The test shows that neither our model or the CAPM significantly explains the relationship between risk and return. The risk-returns for the two factor model only became significant when we considered the change in sign of our factor realisations. The result was however different for the CAPM whose risk -return relationship remained insignificant even after considering the positive and negative realisations. This result clearly confirms that APT better explains risk-relationship as compared to CAPM.

Future research should investigate whether financial markets in developed and other emerging markets are also segmented. It is possible that this phenomenon is not only unique to the JSE, but prevalent in most financial markets. Future research should also consider looking at the time varying components included in the factor models using methods, such as GARCH and the Kalman-filter method.

7.1 Limitations

During the course of implementing this research project several limitations became apparent. First, our dataset was reduced in the data cleaning process as cluster analysis data is required to have the same time series size. Sampling errors might have crept in. Second, sub-sector portfolios indices were created according to the FTSE/JSE global classification; this might present problems if there are stocks with a different natural classification.

Acknowlegments

Sincere thanks to all those who lent their time and support to this project. First and foremost among this group is my supervisor Professor Kotze. Thank you for your guidance, patience, and limitless enthusiasm for the project. Further recognition belongs to Ameer Amod; I would not have completed this project without his help. Thank you also for taking the time to meet with me weekly and encouraging me to continue to work to expand the scope of the investigation throughout the research. Finally, I'd like to thank my friends for all their support throughout my studies.

References

[1] Aber, J. (1976). Industry Effects and Multivariate Stock Price Behavior, The Journal of Financial and Quantitative Analysis, 11(4), 617-624.

[2] Bar, G.D.I. (1990). Macroeconomic Identification of the Pricing Factors on the JSE, Journal of Business Management, 21(1), 17-26.

[3] Bai, J. and Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models, Econometrica, 30(1), 191-221.

[4] Bowie, D.C and Bradfield D.J. (1993). A Review of Systematic Risk Estimation on the JSE, De Ratione, 7(1), 6-22.

[5] Bradfield, D.J. (1989). A Note on the Estimation Problems Caused by Thin Trading on the JSE, De Ratione, 22-25.

[6] Campbell, J.A. (1979). Risk and Return on the Johannesburg Stock Exchange, Unpublished MBA thesis, University of Cape Town.

[7] Chan, L.K.C., Karceski, J. and Lakonishok, J. (1998). The Risk and Return from Factors, Journal of Financial and Quantitative Analysis, 33(2), 159-188.

[8] Chen, N., Richard, R. and Stephen, R. (1986). Economic Forces and the Stock Market, Journal of Business, 59(1), 383-403.

[9] Choi, J.J. and Rajan, M. (1996). A Joint Test of Market Segmentation and Exchange Risk Factor in International Capital Markets, Journal of International Business Studies, 28(1), 29-49.

[10] Connor, G. and Korajczyk, R.A. (1993). A Test for the Number of Factors in an Approximate Factor Model, Journal of Finance, 48(1), 1263-1292.

[11] Domowitz, I. Glen, J. and Madhavan, A. (1997). Market Segmentation and Stock Prices: Evidence from an Emerging Market, Journal of Finance, 36(2), 1059-1085.

[12] Fama, F. and French, K. F. (1992). The Cross-Section of Expected Stock Returns, Journal of Finance, 47(2), 427-465.

[13] Fama, F. and MacBeth, J.D. (1973). Risk, Return and Equilibrium: Empirical tests, Journal of Political Economy, 81(3), 607-636.

[14] Gilbertson, B. and Goldberg, M.N. (1981). The Market model and the JSE, Investment Analyst Journal, 17(2), 40-42.

[15] Griffin, J.M. and Karolyi, A. G. (1998). Another Look at the Role of the Industrial Structure of Markets for International Diversification Strategies, Journal of Financial Economics, 50(3), 351-373.

[16] Grinold, R. C. and Kahn, R. N. (2000). Active Multivariate Portfolio Management: A Quantitative Approach for Producing Superior Returns and Controlling Risk, 2nd edition. McGraw-Hill, New York.

[17] Heaneye, H. and Gilbertson, K. (1997). Regional Integration of National Stock Markets, Investment Analysis Journal, 46(1), 40-42.

[18] Jay, J.C. and Rajan, M. (1997). A Joint Test of Market Segmentation and Exchange Risk Factor in International Capital Markets, Journal of International Business Studies,28(1), 29-49.

[19] Johnson, R.A. and Wichern, D.W. (1982). Applied Multivariate Statistical Analysis, Prentice Hall International,Inc.

[20] King, B. (1966). Market and Industry Factors in Stock Price Behavior, The Journal ofBusiness, 39(3), 139-140.

[21] Haugen, R. and Baker, N. (1996). Commonality in the Determinants of Expected Stock Returns, Journal of Financial Economics, 401-439.

[22] Markowitz, H.M. (1959). Portfolio Selection: Efficient Diversification of Investments, John Wiley Sons.

[23] Mangani, R. (2007). Distributional Properties of JSE Prices and Returns, Journal of Investment Analysis, 16-42.

[24] Roll, R. (1977). A Critique of the Asset Pricing Theory's Tests, Journal of Financial Economics, 4, 129-176.

[25] Page, M.J. (1986). Emprical testing of the Arbitrage Pricing Theory using Data form the JSE, South African Journal of Business Management, 17(1), 38-42.

[26] Sharpe, W.F. (1963). A Simplified Model for Portfolio Analysis, Journal of Management Science, 9, 277-293.

[27] Sokal, R. and Rohlf, J. (1962). New tests of the APT and their Implications, Journal of Finance, 40(3), 659-674.

[28] Sokal, R. and Michener, C.D. (1958). A Statistical Method for Evaluating Systematic Relationship, University of Kansa Scientific Bulletin , 6(2), 1409-1438.

[29] Venter, J.J., Bradfield, D.J. and Bowie, D.C. (1992). An Analysis of Market Related Risk Factors, Technical Report, University of Cape Town.

[30] Van Rensburg, P. (1998). Unifying the Factor Analytic and Pre-specified Variable Approaches to APT Factor Identification on the JSE, Journal of Business Management, 12(1), 15-45.

[31] Van Rensburg, P. and Slaney, K. (1997). Market Segmentation the Johannesburg Stock Exchange, Journal of Studies in Economics and Econometrics, 21(3), 1-20.

[32] Ward, M.J. (1994). Risk on the Johannesburg Stock Exchange, De Ratione, 8(2), 99-113.

8 Appendix

Matlab code for cluster analysis

```
tic %keeps track of time
clear all;clc;
%Imports data from excel
[numeric,txt,raw]=xlsread('D:\Dat\clusters2.xls',-1);
%Converts data to a financial time series
% n = datenum(datestr(datenum(text(1,2:end), 'dd.mm.yyyy'),1);
n=datenum(txt(1,2:end))
Tickseries=fints(n,numeric',txt(2:end,1)','monthly','Sector returns');
returns=fts2mat(Tickseries);
returns=numeric'
%splits returns set from Jan 1998-Nov 2002
category1 = returns(:,1:end/2);
%%splits returns set for Dec 1998- Dec 2007
category2 = returns(:,end/2+1:end);
%cluster analysis
%pdist caculates the Euclidean distance.
cluster1=pdist(category1);
cluster2=pdist(category2);
%linkage takes the distance information generated by pdist and links
pairs of objects that are close together into binary clusters
cluster1a = linkage(cluster1,'average');
cluster2b=linkage(cluster2,'average');
%Compute Spearman's rank correlation between the
%dissimilarities and the cophenetic distances
[c,D] = cophenet(cluster1a,cluster1);
```

```
r = corr(cluster1',D','type','spearman')
[c,E] = cophenet(cluster2b,cluster2);
v= corr(cluster2',E','type','spearman')
dendrogram(cluster1a,'colorthreshold','default');
title('Cluster analysis from 1998-2002')
subplot(2,1,1)
dendrogram(cluster2b,'colorthreshold','default');
title('Cluster analysis from 2002-2007')
subplot(2,1,2)
toc % Keep track of time
```

Matlab code for factor analysis

```
tic % Keep track of time
%Clear Workspace
clear all;clc;
%Imports Data from excel
[numeric,txt,raw]=xlsread('D:\CLEANEDDATA.xls',-1);
%Converts my raw data to continuously compounded return series.
% n = datenum(datestr(txt(1,3:end), 'dd.mm.yyyy'),1));
n=datenum(txt(1,3:end));
Tickseries=fints(n,numeric',txt(2:end,1)','monthly','JSE returns');
returns = tick2ret(fts2mat(Tickseries),[], 'Continuous');
[SUCCESS,MESSAGE]=XLSWRITE('E:\CLEANEDDATA.xls',returns,'economicdata','a1')
%Principal component analysis
[pc, score, latent, tsquare] = princomp(returns);
pareto(latent)
xlabel('Principal Component')
ylabel('Variance Explained (%)')
```

```
%Factor Loadings-Returns
returns=returns(all(~isnan(returns),2),:);
for i=1:7
[Prices,specVar,T,stats,F] = factoran(returns,7,'scores','regr')
end
%Factor pattern using verimax rotation.
biplot(Prices,'scores',F,'varlabels',num2str((1:end)'))
subplot(1,1,1);
xlabel('Component 1');
ylabel('Component 2');
zlabel('Component 3');
axis square;
view();
toc % Keep track of time
```

Cluster analysis results

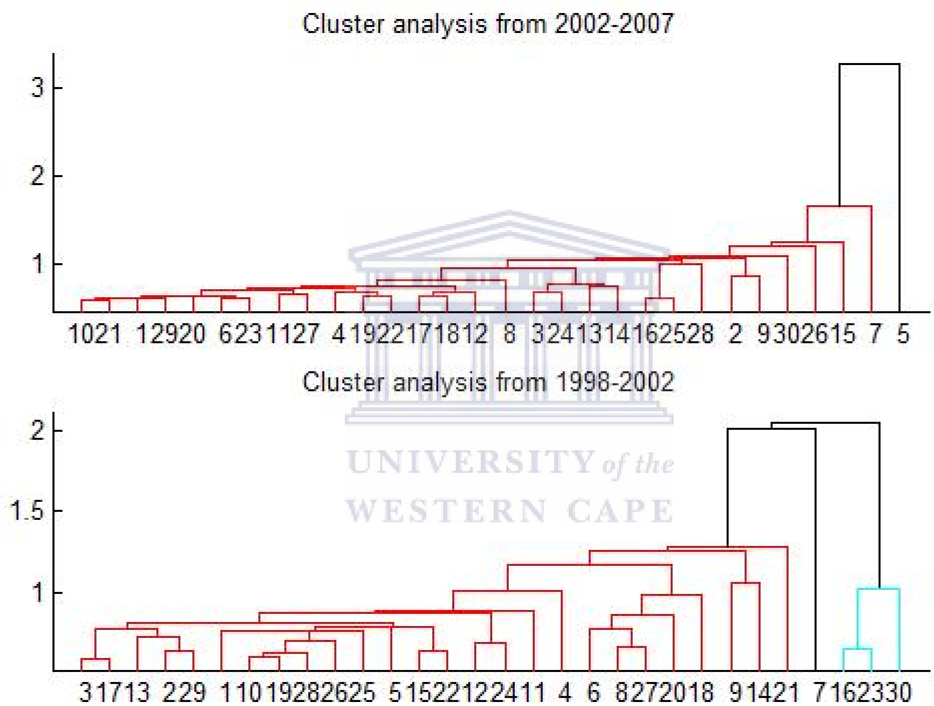

Figure 4: Dendograms showing clusters

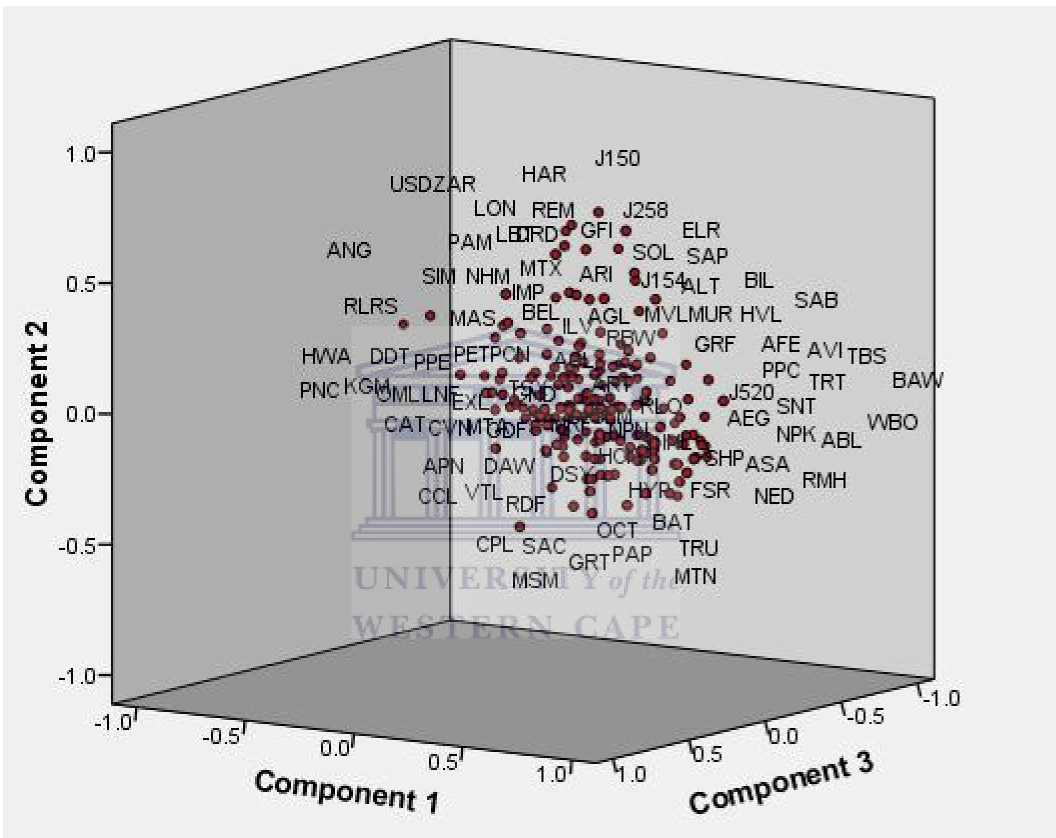

Figure 5: Factor pattern using promax rotation.

	Ticker	Ticker Sub-Sector Classification	Market Capitalization
1	AGL	Metals & Minerals	686834.02
2	BIL	Metals & Minerals	630291.74
3	MTN	Wireless Telecom Services	280710.86
4	SOL	Oil - Integrated	280022.85
5	SAB	Beverages - Brewers	266146.49
6	IMP	Platinum	202105.21
7	OML	Life Assurance	105191.1
8	ACL	Steel	104306
9	REM	Diversified Industrials	89643.57
10	FSR	Banks	88219.33
11	LON	Platinum	74049.68
12	ANG	Gold Mining	73050.43
13	NPN	Broadcasting Contractors	66949.36
14	GFI	Gold Mining	65253.85
15	ASA	Banks	64645.41
16	ARI	Metals & Minerals	54542.43
17	NED	Banks	53385.95
18	LBT	Real Estate Investment Trusts	51517.35
19	SLM	Life Assurance	44690.16
20	HAR	Gold Mining	35895.04
21	RMH	Banks	29816.69
22	MUR	Other Construction	29657.92
23	AEG	Other Construction	25269.67

	Ticker	Ticker Sub-Sector Classification	Market Capitalization
25	TBS	Food Processors	24998.91
26	SHP	Food & Drug Retailers	23722.88
27	BAW	Diversified Industrials	22609.82
28	PPC	Building & Construction Materials	21472.24
29	LGL	Life Assurance	21255.62
30	ABL	Consumer Finance	20933.16
31	HVL	Steel	17351.27
32	IPL	Shipping & Ports	16334
33	GRT	Real Estate Holdings & Development	16165.29
34	NTC	Hospital Management &Long Term Care	16065.88
35	NHM	Platinum	15728.02
36	PIK	Food & Drug Retailers	14981.56
37	DSY	Life Assurance	14502.85
38	MSM	Retailers - Multi Department	14444.8
39	INL	Investment Banks	13308.88
40	MVL	Metals & Minerals	12693.5
41	GND	Marine Transportation	12176.41
42	APN	Pharmaceuticals	12126.52
43	TRU	Retailers - Soft Goods	11895.59
44	MDC	Hospital Management &Long Term Care	11575.63
45	DDT	Computer Services	11237.47
46	RLO	Electrical Equipment	11188.31
47	NPK	Containers & Packaging	10851.17
48	ILV	Food Processors	10838.34
49	WHL	Retailers - Multi Department	10655.41

	Ticker	Ticker Sub-Sector Classification	Market Capitalization
50	SNT	Insurance - Non-Life	9915.2
51	HCI	Investment Companies	9183.32
52	FOS	Retailers - Soft Goods	8898.43
53	AFX	Chemicals - Speciality	8605.61
54	MRF	Metals & Minerals	8551.09
55	MTX	Metals& Minerals	8393.09
56	WBO	Other Construction	8314.68
57	AFE	Chemicals - Speciality	7969.01
58	MET	Life Assurance	7053.65
59	PAM	Nonferrous Metals	7008.94
60	CAT	Publishing & Printing	6740.7
61	JDG	Retailers - Hardlines	6570
62	HYP	Real Estate Holdings & Development	6187.72
63	SAC	Real Estate Investment Trusts	6163.9
64	GDF	Gaming	6073.4
65	GRF	Other Construction	5950.8
66	ALT	Wireless Telecom Services	5832.71
67	RDF	Real Estate Holdings & Development	5529.08
68	FPT	Real Estate Investment Trusts	5528.04
69	AVI	Food Processors	5019.65
70	DTC	Computer Services	4619.53
71	NCL	Retailers - Multi Department	4507.99
72	RBW	Farming& Fishing	4495.07
73	SIM	Gold Mining	4364.95
74	SNU	Metals & Minerals	4322.63

	Ticker	Ticker Sub-Sector Classification	Market Capitalization
75	MPC	Retailers - Soft Goods	4179.41
76	BEL	Leisure Equipment	4078.95
77	PAP	Real Estate Holdings & Development	4017.08
78	ATN	Electrical Equipment	3957.31
79	PSG	Investment Banks	3713.86
80	OMN	Chemicals - Speciality	3659.37
81	PGR	Investment Banks	3113.96
82	CLH	Hotels	3105.59
83	DAW	Building & Construction Materials	2680.5
84	PET	Metals & Minerals	2610.61
85	HDC	Engineering - General	2594.39
86	CPL	Real Estate Investment Trusts	2564.35
87	AFR	Farming & Fishing	2429.66
88	CLE	Life Assurance	2426.25
89	OCE	Farming& Fishing	2353.79
90	BAT	Investment Banks	2320.23
91	MTA	Auto Parts	2135.45
92	WES	Automobiles	2102.01
93	BSR	Other Construction	2049.84
94	DRD	Gold Mining	2003.1
95	KGM	Broadcasting Contractors	1735.6
96	ART	Diversified Industrials	1717.53
97	RAH	Investment Companies	1672.91
98	ADH	Specialised Consumer Services	1661.27
99	TRT	Leisure Facilities	1610.88
100	ILA	Builders Merchants	1594.66