UNIVERSITY OF CAPE TOWN

Hedge Fund of Funds Investment Process

A South African Perspective

Student Name: Mahzabeen Natasha HOSSAIN

Student Number: HSSMAH004

Supervisor: Associate Professor Sugnet LUBBE

DISSERTATION PRESENTED FOR THE DEGREE OF MASTER OF PHILOSOPHY IN MATHEMATICAL FINANCE

DEPARTMENT OF COMMERCE

UNIVERSITY OF CAPE TOWN

14 May 2014

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

PUBLICATION

I hereby grant the University free license to publish this dissertation in whole or part in any format the University deems fit.

PLAGIARISM DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is my own.
2. I have used the APA referencing guide for citation and referencing. Each contribution to, and quotation in this dissertation from the work(s) of other people has been contributed, and has been cited and referenced.
3. I know the meaning of plagiarism and declare that all of the work in the dissertation, save for that which is properly acknowledged, is my own.

Signature : Signed by candidate

Date : 17 February 2014

Acknowledgements

This dissertation would not have been possible without the support and encouragement of many individuals.

I would like to thank my supervisor, Professor Sugnet Lubbe, for her tremendous support and invaluable assistance at the most critical points and till conclusion.

Thank you to my family for pushing me, in particular my brother for extending just the right amount of pressure and my sister for the support.

A special thanks also to my friends, RG for your never faltering aid wherever possible, AR for your continued faith in me and NS for always coming to my tech rescue.

I could not have completed this without each of your contributions.

Abstract

The objective of this dissertation is to develop and test an investment process for hedge fund of funds (HFoFs) in South Africa. The dissertation proposes a three tiered process, adapted from the works of Lo (2008).

Step one of the process involves the categorisation of hedge funds into broadly defined groups based on predefined factors. Two classification methodologies are examined herein to determine optimal category definitions. These are 1) an adaption of the classification developed by Schneeweis and Spurgin (2000), based on the correlation of hedge funds to an appropriate benchmark and the returns offered by these hedge funds, and 2) classification by cluster analysis.

Once a finite set of classification is defined, step two of the process uses a minimum variance optimisation, based on forward-looking parameter estimates of return and co-variance to compute the optimal capital allocation to these categories.

The final stage of the process employs a mixture of quantitative and qualitative analysis to allocate capital within categories to individual hedge funds.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Hedge funds have historically referred to large unregulated pools of liquidity, managed by skilled professionals with great flexibility in the investment tools at their disposal (Stulz, 2007) These grew popularity because of their low correlations to conventional bond and equity markets. Despite being classified as ‘alternative investments’, these investment vehicles are becoming an industry standard asset class over the last decade (Gregoriou, Sedzro and Zhu, 2004) due to the pursuit of absolute returns which are largely accredited to their ability to take short positions. Hedge funds are able to generate positive returns even in volatile or downward sloping markets, while long-only investments only profit in rising markets (Strachman and Bookbinder, 2010). Despite increased commonality, access to hedge fund returns remains restricted, with three avenues of access (Nicholas, 2004):

1. Investing directly into a single or self customized portfolio of hedge fund,
2. Investing through an index fund, or
3. Investing in a hedge fund of funds (HFoFs).

An investor pursuing option one faces the challenge of investigating and choosing optimal hedge funds, strategies and investment instruments. The investor pursuing option two foregoes the opportunity to achieve returns in excess of the market or benchmark index, commonly known as alpha. Fund of funds (FoFs) provide a solution to these issues by offering investors a single point of entry into a diversified portfolio of hedge funds. As FoF managers are also skilled professionals, exposure to hedge funds through FoFs also improve the likelihood of making a successful investment. Investors purchase a portion of the FoF, these purchase considerations or investments are then consolidated and re-deployed into individual hedge funds according to an investment process (Nicholas, 2004).

The investment process followed by the manager is integral to the ultimate performance of any FoF as significant performance differentials may exist between managers pursuing similarly defined strategies. These varying returns can be attributed to loosely defined strategies, lack of appropriate benchmarks, incorrect assessment of available data and mismanaged risks (Nicholas, 2004). One of the primary risks in FoF fund selection is concentration risk i.e. the risk of large exposures to one trading strategy or a group of correlated funds (Koh, Lee and Fai, 2002). This is particularly problematic in South Africa where the domestic hedge fund industry is in its infancy, largely comprising of Long Short Equity funds.

Before allocating capital, the FoFs manager investigates the suitability of an investment into a particular hedge fund. This due diligence process is inseparable from the investment process as ongoing investment decisions are affected by insights obtained through due diligence efforts (Nicholas, 2004). While it is common for FoFs to employ some mix of quantitative and qualitative procedures, there is great variation between strategies employed by different FoFs as there is yet to be a consistent methodology accepted in the industry.

1.2 Objective

The objective of this dissertation is to develop an investment process for the FoFs industry, with particular reference to South Africa where hedge fund industry standards are still developing and many FoFs employ arbitrary selection procedures.

This dissertation proposes a three tiered investment process:

- *Stage one*: hedge funds are divided into groups based on predefined characteristics. This is the most critical stage of the process as the allocation decision to member funds is a function of the groups identified. This dissertation explores two separate classification schemes to examine the relative benefits of each, and determine which grouping methodology yields optimal results.

- *Stage two*: a minimum variance optimisation is conducted on predicted estimates of future returns of the groups identified in *Stage one*, to establish capital allocation to each classification. Return estimates used in the process are predicted using regression analysis.
- *Stage three*: sub-allocations to individual hedge funds or managers are determined through a combination of qualitative and quantitative analyses. Qualitative features of individual funds in each grouping are ascribed a rating by the FoF manager based on his discretion. Thereafter a second minimum variance optimisation is conducted on the ratings.

The process outlined above is a modification of that described by Lo (2008) in which he groups similarly classified hedge funds into ‘asset classes’. Historical returns of these asset classes are then used in a data set over which Lo (2008) conducts a two stage optimisation as described in *Stage two* and *Stage three* above. As Lo (2008) uses manager defined strategies such as ‘Long Short Equity’ and ‘Market Neutral’ to group asset classes his process does not require separate asset class discovery. This dissertation differs by adding an additional step, *Stage one* above, to begin the investment process under which individual hedge funds are examined against predefined properties and separated into groups, or asset classes, based on these characteristics. Two classification methodologies are considered for application at this stage, named below, after which the *Stage two* and *Stage three* optimisations are implemented.

1.3 Data and analysis

This proposal has been developed and tested using South African hedge fund return data gathered from individual hedge funds, data providers and hedge fund dedicated publications. The data spans a four year period from July 2006 to June 2010 and covers 46 funds from a spectrum of strategies.

Data analysis and implementation of the above discussed process is conducted using Microsoft Excel and the statistical package R (R Development Core Team, 2012), using various packages for subsets of analysis.

1.4 Notation

This dissertation employs a number of financial models and statistical techniques. The specific notation used is denoted in this section.

$\boldsymbol{Y}$: $n \times p = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_p]$ where $\boldsymbol{y}_i$ are n independent multivariate observations; in this analysis n = 46 funds and $p = 7$.

τ_k is the probability that an observation $\boldsymbol{y}_i$ belongs to group k.

f_k is the probability density function for the k-th group of hedge funds. In this dissertation only the p-variate normal distribution is used.

θ_k denotes parameters of the multivariate normal distribution viz.$(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ where $\boldsymbol{\mu}_k$ represents the mean vector and $\boldsymbol{\Sigma}_k$ the covariance matrix.

l_i is an indicator variable; $l_i = k$ if $\boldsymbol{y}_i$ is in group k.

$\boldsymbol{Z}: n \times G$ with $z_{ik} = \begin{cases} 1 & if\ \boldsymbol{y}_i\ belongs\ to\ cluster\ k \\ 0 & otherwise \end{cases}$.

r_{ij} = return on fund i during the j-th period.

r_i = unadjusted realised return of fund i.

r_m = return on the passive market portfolio.

r_f = return on a risk free asset.

r_0 = target return to be achieved by asset allocation to the set of classifications.

σ_i = standard deviation of unadjusted returns on fund i.

σ_m= standard deviation of returns on the market portfolio.

β_{ij} = risk exposure or sensitivity of fund i to factor j.

$\boldsymbol{\rho}: n \times n$ is a correlation matrix with the ij-th element being the correlation between returns of funds i and j.

$\boldsymbol{w} = [w_1, \dots, w_i]'$ where w_i is the optimal proportions of total capital allocated to asset class *i.*

K = Total capital available for allocation.

δ = constant parameter.

$\boldsymbol{\gamma}_{ik} = \left[\gamma_{ik_1}, \dots, \gamma_{im_j}\right]'$ where γ_{im_j} is the optimal proportions of capital allocated to manager m_j.

$_k$ is the constant of proportionality which describes the covariance matrix ($_k$) parameterisation and represents the volume of the *k*-th cluster.

$\boldsymbol{D}_k$ is the orthogonal matrix of eigenvectors which controls the orientation of the *k*-th cluster.

$\boldsymbol{A}_k = diag(1, \frac{2k}{k}, \frac{3k}{k}, \dots, \frac{pk}{k})$, the diagonal matrix whose elements are proportional to the eigenvalues and controls the shape of the cluster.

$\boldsymbol{T}: p \times p = \boldsymbol{Y}\,\boldsymbol{Y}$ is the total sums of squares and cross products.

$\boldsymbol{B}: p \times p = \overline{\boldsymbol{Y}}\,(\boldsymbol{Z}\,\boldsymbol{Z})^{-1}\overline{\boldsymbol{Y}}$ is the matrix of between cluster sums of squares and cross products.

$\overline{\boldsymbol{Y}}$: $G \times p$ is the matrix of cluster means.

$\boldsymbol{W}: p \times p = \boldsymbol{Y}\,\boldsymbol{Y} - \overline{\boldsymbol{Y}}\,(\boldsymbol{Z}\,\boldsymbol{Z})^{-1}\overline{\boldsymbol{Y}}$ is the matrix of within cluster sums of squares.

$\boldsymbol{m}$ is an eigenvector of $\boldsymbol{W}^{-1}\boldsymbol{B}$ associated with the largest eigenvalue $\hat{\pi}_1$.

$\boldsymbol{\pi} = [\hat{\pi}_1, \dots, \hat{\pi}_s]'$, $s = min(p, G\text{-}1)$ where $\boldsymbol{\pi}$ denotes non-zero eigenvalues in decreasing order yielding *s* linear combinations of $\boldsymbol{Ym}$.

K is the total amount of capital available to the FoF for allocation.

$K_i^* = w_i^* K$, where K^* is the optimal capital allocation to fund *i*.

s_k is the sum of ratings across each qualitative criterion for manager *k*.

S_{ik} is the relative score of manager *k* to other managers in asset class *i*.

δ is a constant parameter that determines the weighting for the relative scores in *Stage three*.

1.5 Layout

The remainder of the dissertation is structured as follows. Chapter 2 presents a description of the mathematics of the statistical processes involved in the clustering classification. Chapter 3 describes and discusses the various options considered for *Stage one* of the investment process. Chapter 4 describes and presents results of *Stage two* of the investment process. Chapter 5 describes *Stage three* of the process. Chapter 6 conducts an out of sample test to examine the validity or efficiency of the proposed investment process. Finally Chapter 7 concludes this dissertation and offers recommendations for further study.

Chapter 2

Theory of Clustering and Associated Statistics

2.1 Clustering

Cluster analysis is a statistical process that identifies groups of observations with distinct properties which separate them from other groups.

The most commonly applied method of clustering is hierarchical agglomerative clustering. This is an iterative process in which two groups, chosen to optimise a specific criterion, are merged at each stage of the algorithm. Popular criterion include the within-group sums of squares (Ward, 1963), and the shortest distance between groups, which underlies the single-linkage method. A non-hierarchical commonly applied method is based on iterative relocation, in which data points are moved from one group to another until the objective criterion is at its maximum. Iterative relocation with the sum of squares criterion is often called *k*-means clustering (MacQueen, 1967). *K*-means clustering requires the number of groups to be specified upfront, however the properties that define these groups remain undefined. Hierarchical and non- hierarchical methods are both examples of heuristic clustering as there is no modelling involved, the solution is thus regarded as a good approximation rather than an optimum solution.

Applying a model-based method provides a statistical approach to discovering the true number of groups. This dissertation employs a 'finite mixture method' algorithm developed in Fraley and Raftery (2002) based on multivariate models where cluster covariance matrices are parameterized by eigen-decompositions. Under this strategy each cluster is described by a multivariate probability distribution with unknown parameters. The data is generated by a classification likelihood (Murtagh and Raftery, 1984; Banfield and Raftery, 1993) and maximum likelihood estimation of the multivariate mixture models is conducted via the expectation-maximisation (EM) algorithm (McLachlan and Basford, 1988; Celeux and Govaert, 1992). The process is sequential, first implementing model-based hierarchical agglomeration which produces an initial partition for applying the EM algorithm to estimate the classification likelihood. Then significant factors with an appropriate information criterion, such as the Bayesian Information Criterion (BIC) approximation (Schwarz, 1978) are used to determine the number of groups.

This dissertation practically implements the model-based clustering discussed above using the R package mclust (Fraley, Raftery and Scrucca, 2012). This is used extensively in the analysis, specifically the Mclust function is used in code to select the most appropriate model (e.g. ellipsoidal or spherical) and optimal number of groups as discussed in section 2.3.

The likelihood function for a mixture model with independent multivariate observations $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ and G groups is

$$L_{MIX}(\theta_1, \dots, \theta_G; \tau_1, \dots, \tau_G | \boldsymbol{y}) = \prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k f_k(\boldsymbol{y_i} | \theta_k) \tag{2.1}$$

where f_k is the density function, θ_k the parameters of the k-th component in the mixture and τ_k is the probability that an observation belongs to the k-th group. This differs from the classification likelihood in that each group is weighted by the probability that an observation belongs to that group, whereas the classification likelihood (2.2) introduces a combinational aspect that makes exact maximization impractical (Fraley and Raftery, 2002).

$$L_{CL}(\theta_1, \dots, \theta_G; l_1, \dots, l_n | \boldsymbol{y}) = \prod_{i=1}^{n} f_i(\boldsymbol{y_i} | \theta_i) \tag{2.2}$$

where l_i indicates a unique classification of observations; $l_i = k$ if $\boldsymbol{y}_i$ belongs to the k-th grouping.

Generally, f_k is a multivariate Gaussian density, parameterised by mean $\boldsymbol{\mu}_k$ and covariance matrix ${}_k$,

$$f_k(\boldsymbol{y}_i|\mu_k,\ {}_k) \equiv \frac{1}{\sqrt{(2\pi)^N \det(\ {}_k)}} e^{-\frac{1}{2}(y_i-\mu_k)'\ {}_k^{-1}(y_i-\mu_k)} \tag{2.3}$$

The covariance matrix in (2.3) can be parameterised as shown in (2.4) to define the geometric properties, or type of clustering model i.e. the shape, orientation and volume of the clusters (Fraley and Raftery, 2002) as follows

$$ {}_k = \ {}_k\boldsymbol{D}_k\boldsymbol{A}_k\boldsymbol{D}_k' \tag{2.4}$$

The covariance parameterisation is described by ${}_k$ which is the largest eigenvalue of ${}_k$ and acts as a constant of proportionality which controls the volume of the k-th cluster, $\boldsymbol{D}_k$ the orthogonal matrix of eigenvectors which controls the orientation of the cluster, and $\boldsymbol{A}_k$ = *diag*(1, $\frac{2k}{k}$, $\frac{3k}{k}$,..., $\frac{pk}{k}$) the diagonal matrix whose elements are proportional to the eigenvalues and controls the shape of the cluster.

2.1.1 Models and Procedure

There are various models that might be determined. These include the following models with the associated restrictions on parameters in (2.4) (Banfield and Raftery, 1993):

- Equal volume spherical model

$$\boldsymbol{D}_k = I \text{ and } \lambda_1 = \cdots = \lambda_G \tag{2.5}$$

- Unequal volume spherical model

$$\boldsymbol{D}_k = I \tag{2.6}$$

- Equal shape ellipsoidal model

$$\boldsymbol{D}_1 = \cdots = \boldsymbol{D}_G \tag{2.7}$$

- Equal volume ellipsoidal model

$$\boldsymbol{D}_1 = \cdots = \boldsymbol{D}_G \text{ and } \lambda_1 = \cdots = \lambda_G \tag{2.8}$$

- Unequal volume ellipsoidal model, which has no restriction on parameters

The procedure begins by treating each observation as a single cluster, then successively merging pairs of clusters corresponding to the greatest increase in the classification likelihood (2.2) among all possible pairs. Each stage of merging corresponds to a unique number of clusters and partition of data. A given partition can be transformed into indicator variables (2.10), used as conditional probabilities in an M-step of the EM algorithm for parameter estimation (Fraley and Rafetery, 2002).

2.1.2 The Expectation-Maximisation Algorithm for Mixture Models

The EM algorithm (Dempster, Laird and Rubin, 1977; McLachlan and Krishnan, 1997) is a general approach to maximum-likelihood estimation for analysis where the complete dataset comprises several (*n)* multivariate observations $\boldsymbol{y_i}$ (Fraley and Rafetery, 2002). For $\boldsymbol{y_i}$ which are independent and identically distributed (*iid*), the complete-data likelihood is given by:

$$L_C(\boldsymbol{y_i}|\theta) = \prod_{i=1}^{n} f(\boldsymbol{y_i}|\theta) \tag{2.9}$$

The complete data set consists of both observable $\boldsymbol{y_i}$ and unobservable $\boldsymbol{z}_i = (z_{i1}, \dots, z_{iG})$ data with

$$z_{ik} = \begin{cases} 1 & if\ \boldsymbol{y_i}\ belongs\ to\ group\ k \\ 0 & otherwise \end{cases} \tag{2.10}$$

Assuming that each $\boldsymbol{z}_i$ is *iid* from a multinomial distribution of one draw from *G* categories, with probabilities $\tau_k, k = 1, \dots, G,$ and that the density of an observation $\boldsymbol{y}_i$ given $\boldsymbol{z}_i$ is given by $\prod_{k=1}^{G} f_k(\boldsymbol{y}_i|\theta)^{z_{ik}}$, the resulting complete data log-likelihood is

$$l(\boldsymbol{y}_i, \tau_k z_{ik}|\theta) = \sum_{i=1}^{n} \sum_{k=1}^{G} z_{ik} \log\,[\tau_k f_k(\boldsymbol{y_i}|\theta_k)] \tag{2.11}$$

The EM algorithm alternates between two steps, an 'E-step', in which the conditional expectation of the complete data log-likelihood (2.11) given the observed data and the current parameter estimates are computed, and an 'M-step' in which parameters that maximize the expected log-likelihood from the E-step are determined (Fraley and Raftery, 2002).

The E-step algorithm for mixture models is given by

$$\hat{z}_{ik} \leftarrow \frac{\hat{\tau}_k f_k(y_i|\hat{\theta}_k)}{\sum_{j=1}^{G} \hat{\tau}_j f_j(y_i|\hat{\theta}_j)} \tag{2.12}$$

while the M-step involves maximizing (2.11) in terms of τ_k and θ_k with z_{ik} fixed at the values computed in the E-step.

Under mild regularity conditions, EM can be shown to converge to a local maximum of the observed-data likelihood and has been widely used for maximum likelihood estimation for mixture models with good results (Fraley and Raftery, 2002).

2.1.3 Model Selection

Each partition of data points ($\boldsymbol{y_i}$) is transformed into indicator variables as per (2.10), which is then used as conditional probabilities from the E-step, in the M-step for parameter estimation. The EM algorithm is concluded for the one-cluster case of each model, and for the mixture model with the optimal parameters from the EM estimation for 2 to M clusters (Fraley and Raftery, 2002). Mclust uses an identifier for each possible parameterisation of the covariance matrix that has three letters: E for "equal", V for "variable" and I for "coordinate axes". The first identifier refers to volume, the second to shape and the third to orientation (Fraley, Raftery and Scrucca, 2012). In each iteration the BIC is calculated such that the chosen model and associated number of clusters is deemed appropriate when a maximum BIC criterion is obtained (Fraley and Raftery, 2002).

2.2 Variable Identification

Applying clustering separates funds into homogenous groups. However, the procedure does not disclose the identifying characteristics or properties which distinguish the groups. To determine these properties a linear discriminant analysis (LDA) in conducted. While both cluster and discriminant analysis classify objects into categories, discriminant analysis requires existing knowledge of group membership, thus cluster analysis is performed first to identify groups.

LDA, also widely known as canonical variate analysis (CVA) is performed to identify linear combinations of properties that characterise multiple groups or classes. The procedure optimally separate the class means, relative to the within class dispersion. Then a leave-one-out cross validation (CV) forward stepwise process is conducted where variables contributing to a lowering of the CV error rate are identified as significant in describing the grouping.

The observations can be represented in the $n \times p$ matrix $\boldsymbol{Y}$ and the clustering in $n \times G$ indicator matrix $\boldsymbol{Z}$. Gower and Hand (1996) shows that the sums of squares and cross product matrices yield the following decomposition

$$\boldsymbol{T} = \boldsymbol{B} + \boldsymbol{W} \tag{2.13}$$

where $\boldsymbol{T} = \boldsymbol{Y}\,\boldsymbol{Y}$ is the $p \times p$ matrix of total sums of squares and cross products, $\boldsymbol{B} = \overline{\boldsymbol{Y}}\,(\boldsymbol{Z}\,\boldsymbol{Z})^{-1}\overline{\boldsymbol{Y}}$ is the $p \times p$ matrix of between cluster sums of squares and cross products, $\overline{\boldsymbol{Y}}$ is the $G \times p$ matrix of cluster means and $\boldsymbol{W} = \boldsymbol{Y}\,\boldsymbol{Y} - \overline{\boldsymbol{Y}}\,(\boldsymbol{Z}\,\boldsymbol{Z})^{-1}\overline{\boldsymbol{Y}}$ is the $p \times p$ matrix of within cluster sums of squares.

Gower and Hand (1996) assert that the solution is achieved when a linear combination $\boldsymbol{Ym}$ of p variables maximises the between to within-groups variance ratio, i.e. (2.14) when $\boldsymbol{m}$ is an eigenvector of $\boldsymbol{W}^{-1}\boldsymbol{B}$ associated with the largest eigenvalue.

$$max\,\frac{\boldsymbol{m'Bm}}{\boldsymbol{m'Wm}} \text{ such that } \boldsymbol{m'Wm} = 1 \tag{2.14}$$

This results in $\boldsymbol{m'Bm} = diag(\boldsymbol{\pi})$ where $\boldsymbol{\pi} = [\hat{\pi}_1, \dots, \hat{\pi}_s]$, $s = min(p, G\text{-}1)$ and $\boldsymbol{\pi}$ denote s non-zero eigenvalues yielding s linear combinations of $\boldsymbol{Ym}$ Using the linear combinations $\boldsymbol{Ym}$, each row of $\boldsymbol{Y}$ is assigned to the cluster k with the minimum squared distance from $\boldsymbol{y}_i$ to cluster mean $\overline{\boldsymbol{y}}_k$ (Johnson and Wichern, 2007; Sylvain and Celisse, 2010).

Variables are chosen per the stepwise algorithm as follows (Johnson and Wichern, 2007; Sylvain and Celisse, 2010):

1. Let $q = 1$ and successively perform the LDA with each of the p variables. Let $\boldsymbol{y}_j^{-i}$ represent the j-th variable where $j = 1,..., p$ and the i-th observation was excluded. When performing the LDA on variable j, repeat for $i = 1, \dots, n$ perform the LDA on $\boldsymbol{y}_j^{-i}$ and predict the class membership of y_{ij}. The CV error rate is calculated as the average number of incorrect classifications
2. Select the variable j_1 with the smallest CV error rate and set $j = 1$.
3. Set $q = q + 1$. Perform LDA with variables $j_1, \dots, j_J$ and each of the remaining variables and compute the leave one out CV error rate associated with each of the remaining variables computing $\hat{y}_i^i$, as the fitted value for each i observation.

4. If the minimum CV error rate is less than that from the previous round, set $j = j + 1$ and the variable associated with the minimum CV error to variable j_J and return to step 3, else terminate the process and use the variables identified.

2.3 Cluster Descriptions

A CVA biplot is used to separate the clusters found during the cluster analysis graphically, based only on the variables identified by LDA. The CVA biplot is very useful in representing multivariate data in which the observer retrieves information on both samples and the variables of the data matrix simultaneously in multi-dimension. Funds are differentiated into clusters and represented as points dimensioned by variables which are represented by axes. The CVA biplot axes are used to visually identify to what extent significant variables impact the respective cluster (Darlington, Weinburg and Walberg, 1973). The CVA biplots are constructed in R with the package UBbipl (Le Roux and Lubbe, 2011) which accompanies the book "Understanding Biplots" by Gower, Lubbe and le Roux (2011).

Chapter 3

Stage One of the Investment Process

Stage one is the most critical step of the investment process as the individual hedge funds which are ultimately selected for investment, and as such the achievement of the target return, is a direct result of how the asset classes are defined. There are numerous ways by which to define asset classes; at present they are often defined at the FoF manager's discretion using industry knowledge and experience. However such discretion leads to increased volatility of returns around the target.

This study explores two different classifications to determine the most appropriate asset class definitions. Classification1 is adapted from the works of Schneeweis and Spurgin (2000) and Classification2 is based on statistically defined groups identified by cluster analysis, as described in Chapter 2.

3.1 Classification by Schneeweis and Spurgin

Schneeweis and Spurgin (2000) offers a hedge fund classification methodology founded on the concept that hedge funds were established as an additional asset class to be included in traditional bond or equity portfolios. They suggest that absolute hedge fund performance becomes of secondary importance to the risk return benefits available to be achieved in an existing investor portfolio on inclusion of hedge funds.

The classification uses annualized returns of hedge funds and the correlation of individual funds to a particular benchmark, to measure the impact of individual funds on an existing portfolio. The benchmark chosen must be representative of a portfolio of typical investments. Schneeweis and Spurgin (2000) construct a passive benchmark reference, equally weighted between equity and debt portfolios represented by the S&P 500 index and the Lehman Brothers Government and Corporate Bond Index.

Funds are categorized into four groups, defined as follows, and summarised in Table 3.1 below:

- Return Enhancers (RE) are funds with a high return and high correlation to the benchmark. A high return is classified as that above 9% and high correlation above 0.16.
- Risk Reducers (RR) have a lower return, between 4% and 9%, and a low correlation to the benchmark, between -0.16 and 0.16.
- Total Diversifiers (TD) have high returns similar to Return Enhancers but a low correlation to the benchmark like the Risk Reducers.
- Pure Diversifiers (PD) have low or even negative returns, below 4% but negative correlation to the benchmark, below -0.16.

Table 3.1: Schneeweis and Spurgin (2000) Classification

	Correlation		**Return**	
RE	High	0.16 < x	High	0.09 < x
RR	Low	-0.16 < x < 0.16	Lower	0.04 < x < 0.09
TD	Low	-0.16 < x < 0.16	High	0.09 < x
PD	high negative	x < -0.16	low/negative	x < 0.04

Schneeweis and Spurgin (2000) do not specify ranges or absolute levels of returns or correlations by which to separate funds into groups. It is left to the manager to examine funds' returns and correlation data and employ his experience to apply the rules specified in the methodology. The return and correlation levels employed in this analysis are determined after examining the dynamics of the given sample. This dissertation finds that in applying both criteria simultaneously, there are holes in the classification where some funds in the sample remain unclassified resulting in the manager having to 'manually' select a classification or grouping to apply to specific funds at his discretion. As an example under this scenario, as depicted in the chart to follow, a hedge fund that is highly correlated to the Benchmark but offers return less than 9% falls outside the given classifications.

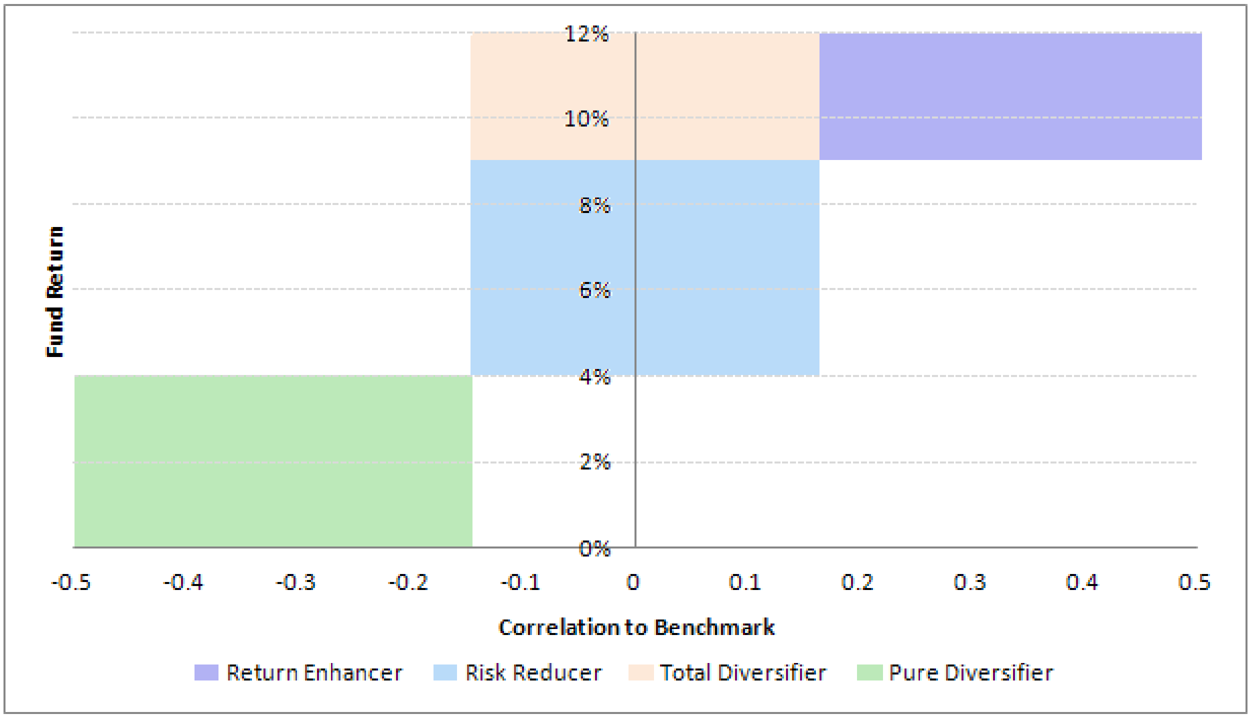


Figure 3.1: Map depicting restrictions of the Schneeweis and Spurgin (2000) Classification

This analysis constructs a similar benchmark portfolio to Schneeweis and Spurgin (2000), adopting the JSE All Share Index (ALSI) as reference equity index and the All Bond Index (ALBI) as reference bond index. The benchmark portfolio is created as an equally weighted portfolio of these indices. As the bulk of funds in this analysis operate in either or both the equity and fixed income asset classes, this is a representative benchmark to simulate passive returns that could be generated over the period. The classification is practically implemented by implementing the given algorithm as a function in Visual Basic within Microsoft Excel.

To bridge the gaps depicted in Figure 3.1, this dissertation adjusts the Schneeweis and Spurgin (2000) classification by applying the constraints sequentially rather than simultaneously. The manager should first apply a correlation restriction across all funds in line with Table 3.2, using the correlation of individual funds to the benchmark portfolio. This distinguishes funds as being RE, PD or one of either RR or TD as funds classified as either RR or TD are similarly correlated to the benchmark portfolio. The secondary restriction is applied to separate funds as either RR or TD classifications using the individual hedge fund returns.

Schneeweis and Spurgin (2000) construct the classification methodology on the basis that hedge funds are alternative investments to be considered for diversification, in addition to an existing base portfolio. This dissertation considers this observation as support for the adjustment to the original methodology by proposing that when considered in addition to an existing portfolio, correlation to an existing portfolio should be the primary consideration as the additional investment is considered for its diversification properties.

In this adapted methodology, hereafter referred to as "modified S&S", each classification retains the same correlation properties as described in Table 3.1. The classifications with correlation within the overlapping range, i.e. Risk Reducers and Total Diversifiers are considered further. Funds in these two categories are then divided across return, with Total Diversifiers being funds generating returns over 15.5% and the remainder Risk Reducers. The return criterion is increased from 9% to 15.5% to ensure clear distinction between the groupings as this adaption employs return as a secondary differentiator. The absolute levels of returns used to separate funds are again selected based on market environment and sample data under examination.

Modified S&S restrictions are summarised in Table 3.2 below:

Table 3.2: Modified S&S Classification

	Correlation		**Return**	
RE	High	0.16 < x	N/A	N/A
RR	Low	-0.16< x < 0.16	Lower	x < 0.155
TD	Low	-0.16< x < 0.16	High	0.155 < x
PD	high negative	x < -0.16	N/A	N/A

The dataset is segmented into four one year periods (Year1 – Year4). Thereafter, both the original Schneeweis and Spurgin (2000) and modified S&S methodologies are applied to each of Year1 to Year3 to identify hedge fund classifications in each year and to examine if fund strategies are consistent across different periods. These results follow in Table 3.3 – 3.5, in which column 5 represents groupings from the original Schneeweis and Spurgin (2000) methodology and column 6 represents groupings based on the modified S&S methodology.

Table 3.3: Year1 Classification Results

	Annualised Return	Annualised Stdev	Correlation with Benchmark	Classification	Modified Classification
X36One	0.127	0.014	0.214	RE	RE
AbsA	0.562	0.068	0.496	RE	RE
AllanG	0.101	0.011	0.094	RR	RR
Bacci	0.101	0.011	0.094	RR	RR
BadgQ	0.413	0.116	0.388	RE	RE
BigRock	0.095	0.062	0.070	RR	RR
CatACP	0.095	0.062	0.070	RR	RR
ClearH	0.491	0.139	0.586	RE	RE
CorCap	0.491	0.139	0.586	RE	RE
CorGran	0.491	0.139	0.586	RE	RE
CorMSA	0.491	0.139	0.586	RE	RE
CorPres	0.269	0.065	0.810	RE	RE
CorSA.A	0.269	0.065	0.810	RE	RE
Craton	0.269	0.065	0.810	RE	RE
CredS	0.213	0.020	0.200	RE	RE
Finch	0.213	0.020	0.200	RE	RE
Foord	0.340	0.065	-0.048	RR	TD
GEN.X	0.140	0.036	-0.021	RR	RR
Gryphon	0.139	0.034	-0.165	Undefined	PD
Hollard	0.054	0.026	-0.225	Undefined	PD
Intern	0.229	0.025	0.243	RE	RE
InvesAQ	0.153	0.090	0.691	RE	RE
InvesFI	0.153	0.090	0.691	RE	RE
InvesPA	0.153	0.090	0.691	RE	RE
Khul	0.512	0.068	0.285	RE	RE
Marco	0.501	0.093	0.194	RE	RE
Mango	0.121	0.026	0.512	RE	RE
MayfI	0.501	0.093	0.194	RE	RE
MergAR	0.426	0.062	0.357	RE	RE
MergU	0.316	0.046	0.689	RE	RE
Oakmon	0.316	0.046	0.689	RE	RE
OMI AR	0.366	0.068	0.495	RE	RE
OMI Cap	0.366	0.068	0.495	RE	RE
OMI MS	0.366	0.068	0.495	RE	RE
OMI N	0.366	0.068	0.495	RE	RE
Oryx	0.134	0.032	0.661	RE	RE
Peregrin	0.071	0.070	-0.260	Undefined	PD
Pardus	0.482	0.072	0.700	RE	RE
Prasid	0.396	0.036	0.225	RE	RE
PSG SE	0.396	0.036	0.225	RE	RE
RMBAM	0.165	0.037	-0.366	Undefined	PD
Tantalum	0.129	0.032	-0.439	Undefined	PD
Tempero	0.089	0.059	0.502	Undefined	RE
TrendI	0.089	0.059	0.502	Undefined	RE
VolArb	-0.045	0.075	0.566	Undefined	RE
X.Cheq	0.127	0.014	0.214	RE	RE

Table 3.4: Year2 Classification Results

	Annualised Return	Annualised Stdev	Correlation with Benchmark	Classification	Modified Classification
X36One	0.091	0.021	0.703	RE	RE
AbsA	0.054	0.158	0.701	Undefined	RE
AllanG	0.068	0.026	-0.410	Undefined	PD
Bacci	0.068	0.026	-0.410	Undefined	PD
BadgQ	0.109	0.105	0.386	RE	RE
BigRock	0.113	0.031	-0.069	RR	RR
CatACP	0.113	0.031	-0.069	RR	RR
ClearH	-0.135	0.053	0.743	Undefined	RE
CorCap	-0.135	0.053	0.743	Undefined	RE
CorGran	-0.135	0.053	0.743	Undefined	RE
CorMSA	-0.135	0.053	0.743	Undefined	RE
CorPres	-0.036	0.102	0.919	Undefined	RE
CorSA.A	-0.036	0.102	0.919	Undefined	RE
Craton	-0.036	0.102	0.919	Undefined	RE
CredS	0.159	0.003	-0.236	Undefined	PD
Finch	0.159	0.003	-0.236	Undefined	PD
Foord	0.300	0.087	0.341	RE	RE
GEN.X	0.050	0.085	0.296	Undefined	RE
Gryphon	0.141	0.022	-0.711	Undefined	PD
Hollard	0.138	0.025	-0.064	RR	RR
Intern	0.204	0.069	0.675	RE	RE
InvesAQ	0.119	0.092	0.584	RE	RE
InvesFI	0.119	0.092	0.584	RE	RE
InvesPA	0.119	0.092	0.584	RE	RE
Khul	0.280	0.120	0.700	RE	RE
Marco	0.211	0.140	0.690	RE	RE
Mango	0.022	0.034	0.021	RR	RR
MayfI	0.211	0.140	0.690	RE	RE
MergAR	0.149	0.101	0.750	RE	RE
MergU	0.030	0.084	0.649	Undefined	RE
Oakmon	0.030	0.084	0.649	Undefined	RE
OMI AR	0.025	0.089	0.836	Undefined	RE
OMI Cap	0.025	0.089	0.836	Undefined	RE
OMI MS	0.025	0.089	0.836	Undefined	RE
OMI N	0.025	0.089	0.836	Undefined	RE
Oryx	-0.020	0.059	0.316	Undefined	RE
Peregrin	-0.016	0.119	0.274	Undefined	RE
Pardus	0.066	0.088	0.010	RR	RR
Prasid	0.213	0.175	0.352	RE	RE
PSG SE	0.213	0.175	0.352	RE	RE
RMBAM	0.152	0.061	-0.416	Undefined	PD
Tantalum	0.237	0.047	0.008	RR	TD
Tempero	0.077	0.068	-0.117	RR	RR
TrendI	0.077	0.068	-0.117	RR	RR
VolArb	0.210	0.179	-0.338	Undefined	PD
X.Cheq	0.091	0.021	0.703	RE	RE

Table 3.5: Year3 Classification Results

	Annualised Return	Annualised Stdev	Correlation with Benchmark	Classification	Modified Classification
X36One	0.062	0.039	0.738	Undefined	RE
AbsA	0.067	0.098	0.359	Undefined	RE
AllanG	0.118	0.039	-0.163	Undefined	PD
Bacci	0.118	0.039	-0.163	Undefined	PD
BadgQ	0.187	0.087	0.328	RE	RE
BigRock	0.211	0.046	0.022	RR	TD
CatACP	0.211	0.046	0.022	RR	TD
ClearH	-0.084	0.132	0.712	Undefined	RE
CorCap	-0.084	0.132	0.712	Undefined	RE
CorGran	-0.084	0.132	0.712	Undefined	RE
CorMSA	-0.084	0.132	0.712	Undefined	RE
CorPres	0.082	0.098	0.929	Undefined	RE
CorSA.A	0.082	0.098	0.929	Undefined	RE
Craton	0.082	0.098	0.929	Undefined	RE
CredS	0.146	0.009	-0.260	Undefined	PD
Finch	0.146	0.009	-0.260	Undefined	PD
Foord	-0.372	0.363	0.577	Undefined	RE
GEN.X	0.009	0.079	0.287	Undefined	RE
Gryphon	0.129	0.042	-0.160	Undefined	PD
Hollard	0.143	0.015	-0.307	Undefined	PD
Intern	0.005	0.059	0.367	Undefined	RE
InvesAQ	-0.164	0.107	-0.184	PD	PD
InvesFI	-0.164	0.107	-0.184	PD	PD
InvesPA	-0.164	0.107	-0.184	PD	PD
Khul	-0.372	0.456	0.609	Undefined	RE
Marco	-0.563	0.354	0.582	Undefined	RE
Mango	0.120	0.083	0.800	RE	RE
MayfI	-0.563	0.354	0.582	Undefined	RE
MergAR	-0.062	0.192	-0.161	PD	PD
MergU	-0.002	0.068	0.489	Undefined	RE
Oakmon	-0.002	0.068	0.489	Undefined	RE
OMI AR	0.001	0.092	0.563	Undefined	RE
OMI Cap	0.001	0.092	0.563	Undefined	RE
OMI MS	0.001	0.092	0.563	Undefined	RE
OMI N	0.001	0.092	0.563	Undefined	RE
Oryx	0.010	0.179	0.691	Undefined	RE
Peregrin	0.157	0.063	-0.034	RR	TD
Pardus	-0.011	0.061	0.091	RR	RR
Prasid	-0.149	0.137	0.533	Undefined	RE
PSG SE	-0.149	0.137	0.533	Undefined	RE
RMBAM	0.215	0.116	0.004	RR	TD
Tantalum	-0.067	0.134	-0.078	RR	RR
Tempero	0.127	0.020	0.081	RR	RR
TrendI	0.127	0.020	0.081	RR	RR
VolArb	0.138	0.204	-0.399	Undefined	PD
X.Cheq	0.062	0.039	0.738	Undefined	RE

Comparison of the original and modified classifications finds the categorizations of funds in a particular grouping in the original Schneeweis and Spurgin (2000) methodology rarely changes in the modified classification. However those funds uncategorised under the original methodology are encompassed in the analysis using the modified S&S classification.

Application of these classifications to subsequent sample periods reveals that a considerable number of funds migrate between categories across periods, indicating a need to re-evaluate classifications on an ongoing basis under this classification algorithm. This is evident in Tables 3.3 – 3.5 by funds such as Peregrin, Pardus and RMBAM.

Summarised results are presented in Table 3.6.

Table 3.6: Number of Funds per Category by Original and Modified Classifications

	Original			Modified		
	Year1	**Year2**	**Year3**	**Year1**	**Year2**	**Year3**
RE	32	13	5	34	31	28
RR	4	8	10	3	7	4
TD	0	0	0	1	1	6
PD	0	0	2	8	7	8
Undefined	10	25	29	0	0	0

The proportion of funds in each category across sample is relatively constant, despite the movement of individual funds between the classifications.

3.2 Classification by Clustering

Clustering offers another alternative by which to categorise funds into groups that are best representative of a particular style. It applies distinction between groups based on the relative impact of a set of factors or metrics on the fund's performance, for example Group A can be separated from Group B by clustering because the returns of hedge funds in Group B are more sensitive to Factors 1 and 2 whilst the returns of hedge funds in Group A are more sensitive to Factors 3 and 4.

Similarly to section 3.1 the classification, as described in Chapter 2, is applied to Years 1 to 3 sequentially, to examine the robustness of the classification. It is possible to shorten each period of examination to three or six month periods, should more flexible commitment periods of investment in the underlying hedge funds be expected. However this is outside the scope of this dissertation, where only one year periods are evaluated.

3.2.1 Variables for Cluster Analysis

The selection of factors or metrics used in clustering has direct bearing on the resulting classifications. Research on the topic is plentiful and varied, however, resounding themes that emerge suggest the importance of including both asset based style (ABS) factors which proxy hedge fund styles, and implicit factors such as risk and return. Common amongst most researchers is the use of between four and seven factors in the analysis, evident in Conor and Korajczyk (1993), Amenc, Curtis and Martellini (2004), and Fung and Hsieh (2004).

While Fung and Hsieh (2004) employ only significant ABS risk factors, including market risk and credit spread other research such as Amenc, Curtis and Martellini (2004) find results across different factor models based on ABS factors alone to be widely varied. They find the inclusion of implicit factors in the explanatory models to be fundamental to the analysis. Thus, this study employs seven metrics, both ABS and those implicit to hedge fund absolute returns and alpha. These are described as follows:

- Volatility (σ) – expressed as the standard deviation of returns, is widely accepted as a default measure of risk. Inherent in hedge fund classification is a fund's risk-return characterisation as this often defines the investment strategies/products available. Studies reflect that hedge funds as an investment class have marked exposure to volatility (Agarwal, Bakshi, and Huij, 2008).
- Alpha (α) – measures active return, i.e. return generated over the market portfolio, or widely accessible passive risky benchmark. Historical studies have found significant dispersion in alphas across a range of hedge funds (Agarwal, Bakshi, and Huij, 2008) motivating the use of this metric in our clustering analysis. This study uses the ALSI as a proxy for the market portfolio.
- Beta (β) – measures the fund's sensitivity to the market. This assumes values between -1 and 1, calculated as $\frac{Cov(r_i, r_m)}{Var(r_m)}$ or by using linear regression of fund returns on market risk premium, where beta is the gradient (and alpha the intercept).
- Average Risk Adjusted Return (RAP) – hedge funds operate using leverage, however the level of gearing used by individual funds varies across managers. Returns are therefore not directly comparable unless adjusted to an equal risk concentration. This dissertation

uses a Risk Adjusted Performance (RAP) metric developed by Modigliani and Modigliani (1997) calculated as

$$\text{RAP} = \frac{(r_i - r_f)}{\sigma_i} \times \sigma_m + r_f. \tag{3.1}$$

where: r_i = unadjusted return on the fund,

r_f = risk free rate,

σ_i = standard deviation, dispersion or volatility, of unadjusted returns on the fund,

σ_m= standard deviation, dispersion or volatility, of market return

This study uses the three month Johannesburg Interbank rate (3mJibar) as a proxy for the risk free rate.

- Agarwal, Bakshi, and Huij (2008) also show that hedge funds are significantly impacted by factors such as skewness (3.2) and kurtosis (3.3). Significant skewness indicates a greater propensity for tail events i.e. positive skewness would suggest more positive returns while negative skewness suggests a greater number of negative returns. Higher kurtosis suggests a substantial portion of the variation in returns is due to extreme swings.

$$\text{skewness} = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n} \left(\frac{r_{ij} - r_i}{\sigma}\right)^3 \tag{3.2}$$

$$\text{kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n} \left(\frac{r_{ij} - r_i}{\sigma}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \tag{3.3}$$

- Tracking Error (TE) – measures deviation of the hedge fund returns from the market i.e. σ_α, and provides an indication of the extent of active management.
- Information Ratio (IR) – measures active return per unit of variation in active returns. It acts as a proxy for manager skill and is calculated by $\frac{\alpha}{\sigma_\alpha}$.

As detailed in Chapter 2, the type of model and number of clusters is determined on completion on the clustering process, once the BIC is maximised.

3.2.2 Results

The analysis suggests that optimal results are consistently achieved in by an ellipsoidal model with equal shape or equal volume. The number of clusters however varies from period to period, with nine categories of hedge funds identified in Year1 and five categories in Year2 and Year3, as summarised in Table 3.7.

Table 3.7: Summary of Cluster Analysis

	Number of Clusters	Model
Year1	9	Ellipsoidal, Equal Shape
Year2	5	Ellipsoidal, Equal Shape, Equal Volume
Year3	5	Ellipsoidal, Equal Shape

As illustrated in Figure 3.2, discriminate analysis on these clusters revealed a varying number of significant factors per period with only the Information Ratio common to all three periods.

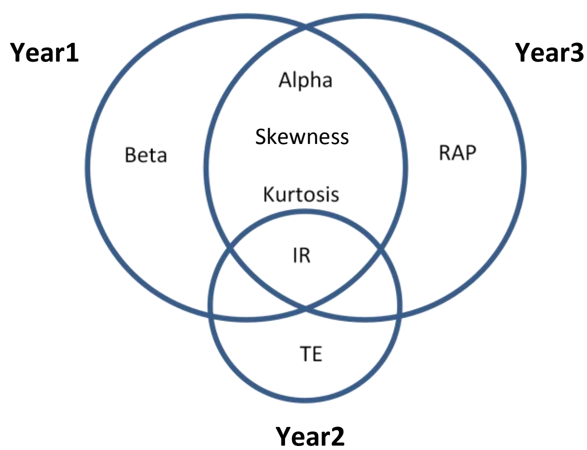


Figure 3.2: Significant Factors per Period

Clusters identified in Year1 were characterised by five significant factors. In Year2, where alphas were generally lower across all funds, the number of significant factors fell to two but reverted to five significant factors in Year3, with prevalent factors in Year1 and Year3 differing only by Beta and RAP.

Figures 3.3 – 3.5 below illustrate the forward stepwise process which determines the significant factors characterizing the identified clusters in Year1 - Year3.

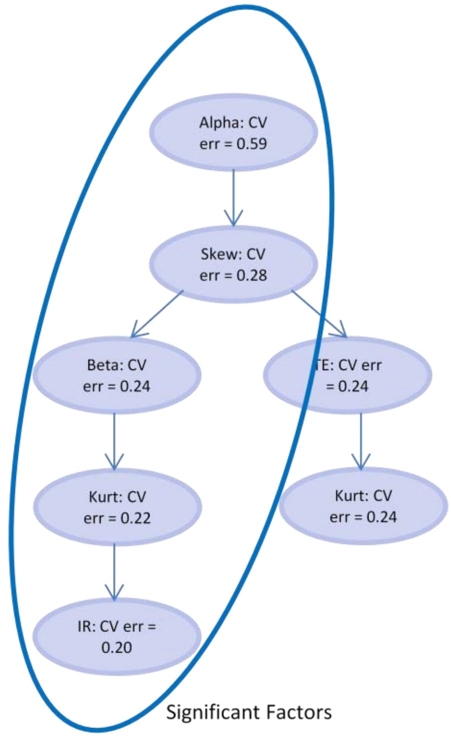


Figure 3.3: Significant Factors in Year1

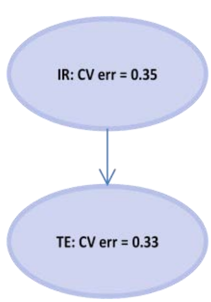


Figure 3.4: Significant Factors in Year2

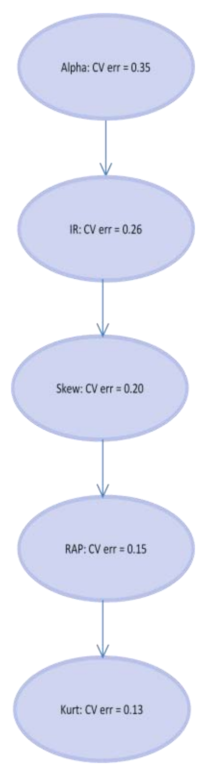


Figure 3.5: Significant Factors in Year3

Thereafter, based on the variables identified, CVA biplots, constructed using the UBbipl package (Gower, Lubbe and Le Roux , 2011) within R and presented in Figures 3.6 – 3.8, are constructed to graphically examine how the clusters are characterised.

Table 3.8 provides a summary of clusters identified within each period and the related factor properties.

Table 3.8: Categories by Clustering

	Significant Factors : Alpha, Skew, Beta, Kurt, IR	**Significant Factors : IR, TE**	**Significant Factors : Alpha, Skew, RAP, Kurt, IR**
	Year1	**Year2**	**Year3**
Red	High beta, alpha, IR. Negative skew	High TE	High kurtosis. Low alpha, IR, RAP, skew
Yellow	High beta, alpha, IR. Low skew	NA	NA
Orange	High beta, alpha, IR. Low kurtosis	Small TE to IR ratio	High RAP, skew, kurtosis
Purple	Positive alpha & skew. Negative kurtosis	High IR. Low TE	High IR. Low Kurt, RAP, skew
Blue	High beta. Low negative skew, kurtosis	High TE to IR ratio	Low RAP, skew, kurtosis
Grey	Low: beta, alpha, IR. Higher kurtosis than green	NA	NA
Pink	High skew, kurtosis. Low beta, alpha, IR	NA	NA
Green	Low beta, alpha, IR	Low IR, TE	High RAP, skew
Brown	Low beta, alpha, IR. High negative skew, kurtosis	NA	NA

The CVA biplots imply that a significant number of funds are characterised by different properties between periods. This could be as a result of shifting market conditions, a change in hedge fund manager, mandate or strategy.

Figure 3.6: Year1 CVA Biplot

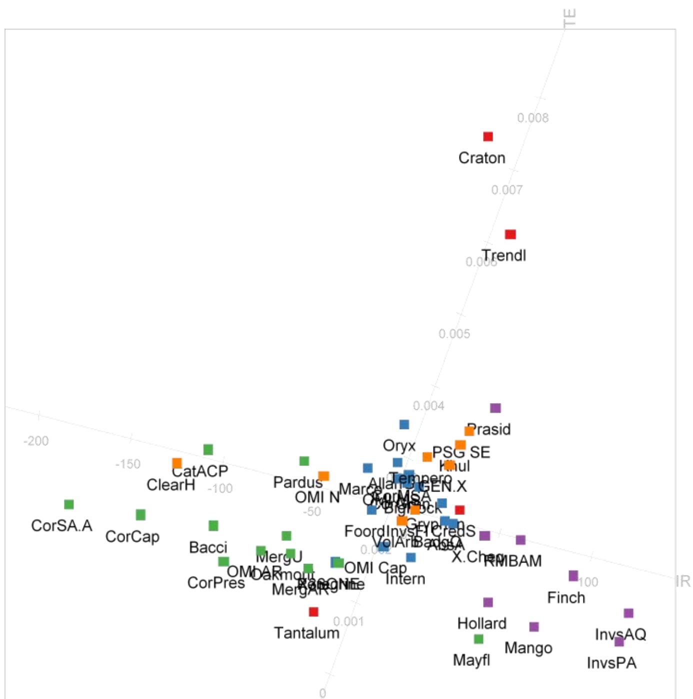


Figure 3.7: Year2 CVA Biplot

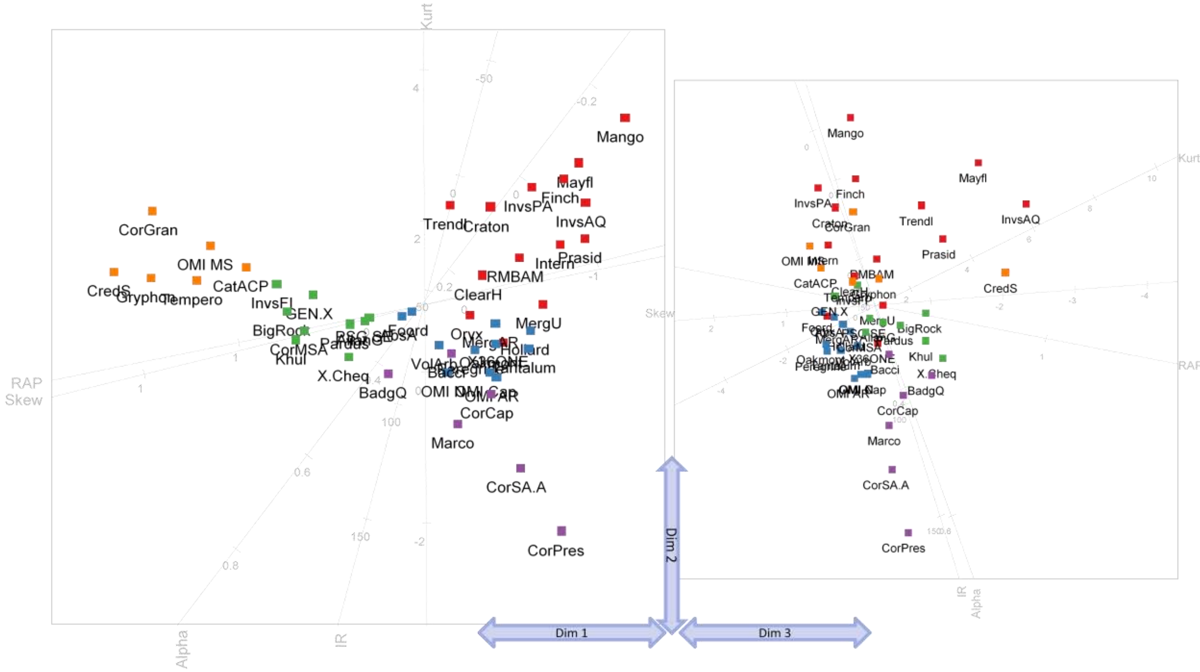


Figure 3.8: Year3 CVA Biplot

Unlike in section 3.1, classifications identified by clustering and described by CVA are non-static, i.e. hedge fund groupings differ between periods. Many within period classifications are very similar, for example the Red and Yellow clusters in Year1 are defined by the same factors differing in that Red exhibits negative skewness while Yellow exhibits low, but positive skewness. The biplots suggest that both categories are highly correlated to the market and exhibit high active returns, on both absolute and risk adjusted basis. However, funds in the Red cluster have a greater propensity for negative returns while the Yellow contains funds whose returns are more stable. These categories are both similar to the Orange cluster, differing in that funds contained therein may exhibit uncommonly repeated extreme returns (represented by kurtosis).

As there are only two significant factors identified in Year2, categories between this period and Year1 are not directly comparable. The analysis finds categories identified within Year2 have subtly differentiating factors, for example, funds within Orange exhibit a low proportion of active management relative to risk weighted active returns while funds in Blue exhibit lower risk weighted active returns relative to the extent of active management.

Groups of funds with the distinct characteristics discussed are identified in this stage of the process are used to determine broad capital allocation as described in Chapter 4.

Chapter 4

Stage Two of the Investment Process

Once a finite set of hedge fund groups or asset classes have been defined, the FoF manager next determines the proportions of capital to allocate to each grouping. At this stage FoF managers are able to make trade-offs between risk and return (Lo, 2008). Classification weightings are computed based on a minimum variance portfolio return optimization. Computing the minimum variance asset allocations (Lo, 2008) begins by setting the target return for the FoFs (r_0). Practically, this should be the annualised minimum return as per the FoFs mandate. The optimisations must be conducted using estimates of forward-looking mean and variance parameters as r_0 is a target for the next period.

The expected return for each asset class can be computed using a choice of factor models such as the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Linter, 1965), Arbitrage Pricing Theory (APT) (Ross, 1976) or non-linear models which entail some judgement in factor selection. Lo (2008) finds that the most relevant factors in explaining hedge fund returns to include risk premium, volatilities, liquidity, and investment cycle. This dissertation uses CAPM to compute expected return, defined as follows

$$E(r_i) = r_f + \beta_{i1}(r_m - r_f) \tag{4.1}$$

where r_f is the return earned on a risk free asset, r_m the return of the market or benchmark portfolio and β_i represents the risk exposure or sensitivity of the asset class to $\left(r_m - r_f\right)$ or market risk premium (MRP).

The sensitivities used in (4.1) are estimated by ordinary least squares (OLS) regression (4.2) using the plm (Croissant and Millo, 2008) package in R. The plm package enables the use of lm function, used in this dissertation, to compute ols regressions within panel data as presented here.

$$y_{it} = \alpha + \beta_i\left(r_{mt} - r_{ft}\right) + e_{it}. \tag{4.2}$$

The asset class covariance matrix (4.3) is estimated using historical data, by first estimating the correlation matrix ($\boldsymbol{\rho}$), as significant empirical evidence suggests that correlations are more stable over time than covariances (Lo, 2008). Estimates of σ_i and σ_j in (4.3) should represent target volatilities for category i and *j*. A good starting point is to compute the average realised volatilities and adjust these estimates to reflect changes in market conditions (Lo, 2008).

$$_{ij} = \rho_{ij} \times \sigma_i \times \sigma_j \tag{4.3}$$

Given the above defined parameters, asset class expected returns ($\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]'$) and asset class covariance matrix ( ), an optimisation function is constructed to compute the optimal weighting ($\boldsymbol{W}$) per group that minimises the portfolio or FoF variance providing that the expected return is greater than or equal to r_0 as defined below

$$min_w \tfrac{1}{2}\boldsymbol{w}' \quad \boldsymbol{w} \text{ subject to } \boldsymbol{w}'\boldsymbol{\mu} \geq r_0 \text{ and } \boldsymbol{w}'\mathbf{1} = 1 \tag{4.4}$$

The optimal capital allocation (K_i^*) is then calculated as $K_i^* = w_i^* K$, where K is the total amount of capital available to the FoFs for allocation.

4.1 Parameter Estimation

Results of section 3.1 and section 3.2 are used to calculate the required parameters and estimate the optimal allocations per classification. The allocation algorithm is conducted in Microsoft Excel using the Solver package. For purposes of this analysis, target volatilities are taken to equal estimates calculated from historical data without adjustment and expected returns are calculated by (4.1) using sensitivities obtained by the regression analysis in (4.2). Respective regression results follow in Table 4.1 and 4.2.

Table 4.1: Modified S&S Regression Results

Factor	**RE**	**RR**	**TD**	**PD**
		Year1		
Intercept	0.03473	0.010462	-	0.005366
	(2E-16)***	(0.0785).	-	(0.161)
MRP	0.27360	0.02396	-	-0.086263
	(3.29E-11)***	(0.258)	-	(0.0726).
R^2	0.15	0.02	-	0.05
		Year2		
Intercept	0.040104	0.006374	-	0.002235
	(2E-16)***	(0.0543).	-	(0.4566)
MRP	0.321616	-0.009167	-	-0.092658
	(2E-16)***	(0.319)	-	(0.0361)*
R^2	0.35	0.02	-	0.1
		Year3		
Intercept	0.029279	0.009463	0.018201	-0.0004634
	(2.62E-13)***	(0.119)	(0.000571)***	(0.589)
MRP	0.236457	0.039124	0.039838	-0.02255796
	(2E-16)***	(0.105)	(0.174).	(0.2)
R^2	0.25	0.05	0.05	0.02

Significance codes: 0 '***' 0.0001 '**' 0.01 '*' 0.05 '.'

Regression output, specifically the low R^2 criterion, suggests that the CAPM model does not significantly represent the variation in returns for all categories in either modified S&S or clustering classifications. Assessment of the best factor model to be used in this analysis is outside the scope of this dissertation, however such an analysis is recommended to FoF managers before commencing *Stage two* of the investment process.

The optimisation in this study uses a cash + 3% target return for the FoF, in line with various domestic FoF mandates. Thus, the target return adopted in this dissertation is 9%, based on an estimate of cash rates equal to 6% which is representative of such short term interest rates during the sample period under examination.

Table 4.2: Clustering Regresion Results

Factor	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8	Cluster9
Year1									
Intercept	0.050235	0.07555	0.060392	0.033349	0.056697	0.028048	0.00829	0.002998	0.014068
	(1.64E-11)***	(0.000164)***	(3.92E-8)***	(9.75E-10)***	(1.22E-09)***	(8.44E-16)***	(0.0323)*	(0.349)	(0.0491)*
MRP	0.29710	0.74963	0.481377	0.212959	0.598996	0.275469	-0.068098	-0.095811	0.10718
	(.00712)**	(0.009323)**	(.00106)**	(.00369)**	(1.48E-6)***	(1.73E-8)***	(0.133)	(0.0785).	(0.203)
R^2	0.2	0.15	0.25	0.15	0.55	0.25	0.1	0.082	0.074
Year2									
Intercept	0.04237	0.042148	0.015065	0.008329	0.043955	-	-	-	-
	(.0214)*	(2E-16)***	(3.38E-8)***	(.067).	(1.3E-10)***	-	-	-	-
MRP	0.30107	0.405082	0.087027	-0.02668	0.233767	-	-	-	-
	(.0573).	(2E-16)***	(0.000177)***	(.234)	(2.59E5)***	-	-	-	-
R^2	0.15	0.52	0.15	0.02	0.24	-	-	-	-
Year3									
Intercept	0.03258	0.014711	0.016606	-0.0004634	0.023355	0.032964	-	-	-
	(0.00234)**	(.0475)*	(4.54E-7)***	(0.589)	(1.30E-10)***	(7.25E-13)***	-	-	-
MRP	0.39521	0.002373	0.032234	-0.02255796	0.172901	0.181381	-	-	-
	(5.17E8)***	(0.542)	(.0785).	(0.2)	(1.06E-12)***	(2.78E-10)***	-	-	-
R^2	0.21	0.05	0.067	0.02	0.45	0.56	-	-	-

Significance codes: 0 '***' 0.0001 '**' 0.01 '*' 0.05 '.'

4.2 Modified S&S Allocations

If there are fewer than three funds in a category, this dissertation proposes that the category be excluded from *Stage two* of the process. At this point the FoFs manager should rely on his qualitative skills to either omit the fund from the analysis or re-assign it to another category. Only one fund was categorised as a Total Diversifier in Years 1 and 2 from section 3.1. This category was thus omitted from the capital allocation optimisation.

Results of the respective optimisations are presented in Tables 4.3 to 4.5 below.

Table 4.3: Category Allocations in Year1

	Year1					
	Risk free rate	**Beta**	**Risk Premium**	**Category return**	**Category Volatility**	**Weighting**
RE	0.089287	0.273606	-0.54089	-0.0587	0.032933	0.00
RR	0.089287	0.02396	-0.54089	0.076327	0.039873	0.311751
PD	0.089287	-0.08626	-0.54089	0.135945	0.011837	0.688249

Table 4.4: Category Allocations in Year2

	Year2					
	Risk free rate	Beta	Risk Premium	Category return	Category Volatility	Weighting
RE	0.111783	0.321616	-0.73207	-0.12366	0.067394	0.00
RR	0.111783	-0.00917	-0.73207	0.118494	0.029342	0.504922
PD	0.111783	-0.09266	-0.73207	0.179616	0.031711	0.495079

Table 4.5: Category Allocations in Year3

	Year3					
	Risk free rate	Beta	Risk Premium	Category return	Category Volatility	Weighting
RE	0.103882	0.236457	-0.80909	-0.08743	0.103527	0.00
RR	0.103882	0.039124	-0.80909	0.072227	0.030882	0.328297
TD	0.103882	0.039838	-0.80909	0.071649	0.080963	0.32812
PD	0.103882	-0.02258	-0.80909	0.122151	0.038467	0.343583

The high domestic interest rate environment combined with declining returns of risk asset during the period under examination resulted in a negative market risk premium in Years 1, 2 and 3. Due to the high correlation to benchmark property of the Return Enhancer category, expected returns across each period were negative for this asset class. As such, this category is excluded from the capital allocation process for this.

In Year1, the analysis finds that the optimal allocation to Pure Diversifiers is twice that to Risk Reducers. This is because both risk and return properties of Pure Diversifiers are superior to Risk Reducers in the sample. However, the Risk reducer category is not omitted as it reduces the overall risk of the portfolio or FoF, providing diversification benefits.

Years 2 and 3 reflect near equal division of capital across categories. Optimal allocation in Year2 is driven by lower risk in favour of higher return. In Year3, a larger allocation is made to the Pure Diversifier category which displays greater returns and marginally higher volatility than the Risk Reducer category.

In summary these results reflect a tendency towards risk aversion, with greater allocations made to categories with greater risk mitigation ability. In a market climate where risk assets are declining in value as in this sample period, capital allocation should be weighted more toward the Pure Diversifier category which has negative correlation with the market.

4.3 Clustering Allocations

Some groupings identified by clustering also result in categories with negative expected returns. These are categories identified by high betas, thus considerably affected by the negative market risk premium in the discussed market environment. Funds within these categories are largely consistent with those identified as Return Enhancers in section 3.1.

Table 4.6: Cluster Allocations in Year1

	Year1						
	No. Funds	**Risk free rate**	**Beta**	**Risk Premium**	**Category return**	**Category Volatility**	**Weighting**
Cluster1 (Red)	5	0.089287	0.2970	-0.54089	-0.0714		
Cluster2 (Yellow)	4	0.089287	0.7496	-0.54089	-0.3161		
Cluster3 (Orange)	4	0.089287	0.4813	-0.54089	-0.1710		
Cluster4 (Purple)	6	0.089287	0.2129	-0.54089	-0.0259		
Cluster5 (Blue)	3	0.089287	0.5989	-0.54089	-0.2347		
Cluster6 (Grey)	11	0.089287	0.2754	-0.54089	-0.0597		
Cluster7 (Pink)	4	0.089287	-0.0681	-0.54089	0.1261	0.014952	0.83874
Cluster8 (Green)	6	0.089287	-0.0958	-0.54089	0.1411	0.058244	0.14238
Cluster9 (Brown)	3	0.089287	0.1071	-0.54089	0.0313	0.026443	0.01888

Table 4.7: Cluster Allocations in Year2

	Year2						
	No. Funds	**Risk free rate**	**Beta**	**Risk Premium**	**Category return**	**Category Volatility**	**Weighting**
Cluster1 (Red)	4	0.111783	0.3010	-0.73207	-0.1086		
Cluster2 (Green)	12	0.111783	0.4050	-0.73207	-0.1847		
Cluster3 (Blue)	17	0.111783	0.0870	-0.73207	0.0480	0.02280	0.49633
Cluster4 (Orange)	5	0.111783	-0.026	-0.73207	0.1313	0.02887	0.50368
Cluster5 (Purple)	8	0.111783	0.2337	-0.73207	-0.0593		

Table 4.8: Cluster Allocations in Year3

	Year3						
	No. Funds	**Risk free rate**	**Beta**	**Risk Premium**	**Category return**	**Category Volatility**	**Weighting**
Cluster1 (Red)	13	0.103882	0.3952	-0.80909	-0.2158		
Cluster2 (Orange)	6	0.103882	0.0022	-0.80909	0.1020	0.042823	0.503214
Cluster3 (Green)	10	0.103882	0.0322	-0.80909	0.0778	0.022453	0.496787
Cluster4 (Blue)	11	0.103882	0.1729	-0.80909	-0.0360		
Cluster5 (Purple)	6	0.103882	0.1813	-0.80909	-0.0428		

As expected in a negative MRP sample period, identified clusters included for the optimisation in Year1 are characterised by low betas. This analysis finds that in general, the optimal allocation during the period favours lower risk (volatility) to higher return. However, the greatest proportion of capital is allocated to Cluster7 which also exhibits high skewness suggesting a greater propensity for positive returns. The second largest allocation is directed towards Cluster8 which displays greater volatility and although this is in contrast to the concluded investment philosophy of the period, it is ascribed to the absence of the negative skewness property displayed by Cluster9.

As identified in section 3.2.2, clusters in Year2 are characterised by two main factors, both driven primarily by volatility. Allocations here are broadly equal across included clusters, however, there is a slight bias for return over lower risk; more specifically, these allocations are skewed towards higher active return per unit of volatility. Allocations identified in Year3 are also broadly equal. Both categories included in the optimisation equation display high RAP and skewness, however, optimal allocations in this sample favour higher return over lower volatility. The differentiating factor between Cluster3 and 4 is high kurtosis, as such the higher volatility can be ascribed to extreme observations in the data.

In summary, although allocations were weighted more heavily toward the lowest beta categories, this analysis does not infer any particular factor by which to measure allocations by cluster analysis. This is due to the inherent ability of clustering to differentiate groups on marginal divides. Chapter 5 uses the groups identified here to optimise capital allocation to individual managers with the aim of achieving the estimated next period or targeted group returns.

Chapter 5

Stage Three of the Investment Process

In *Stage two* of the process, this dissertation follows the approach of Lo (2008), assuming equal capital allocation to individual funds within each asset class to calculate historical return and covariance matrices. *Stage three* of the investment process discriminates between individual hedge funds to distribute the capital allocated to each group K_i^* amongst individual funds, or managers, within each asset class. At this stage a combination of both quantitative and qualitative measures is employed. The manager begins with the *Stage two* assumption of equal allocation to funds within each asset class, i.e. $\gamma_{ik} = \frac{1}{m}$ where there are m funds within asset class i and γ_{ik} is the portion of funds allocated to fund k.

The next step is to develop a score card, where the FoF manager identifies important criteria by which to assess the individual hedge funds, for example:

- Past performance,
- Manager experience,
- Fund life,
- Investment strategy,
- Risk and Transparency

Hedge funds under evaluation should be rated between 1 and 5 for each criterion. For example, a hedge fund manager with twenty years of experience would receive a score of 5, whilst a manager with only two or three years of experience would receive a 1. Similarly a hedge fund which has historically reflected low risk characteristics (determined by volatility or another chosen measure) would receive a 5 relative to a fund with inconsistent risk results which may be rated 1 or 2. Scores will vary for different FoFs as this is largely dependent on the FoF manager's skill, experience and often relationship with individual hedge fund managers. For example, there are two views regarding the life of a hedge fund – one being that fund returns deteriorate with the passage of time, so a fund in operation for many years will achieve lower returns than a newly established fund. Another view is that long life is indicative of good market practice or manager skill. Thus different FoF managers may differ in the relative rating per this criterion depending on their view. The selection of criteria and associated ratings constitutes the qualitative process of *Stage three*.

The ratings are added to calculate a total score per manager (s_k) (Lo, 2008) and then used to calculate relative scores as follows

$$S_{ik} = \frac{s_k}{s_1 + \cdots + s_m} \tag{5.1}$$

The allocation to individual funds is then calculated as a weighted average of allocation based on equal distribution and allocation based on the relative score (S_{ik}) as follows

$$\gamma_{ik} = (1 - \delta) \times \frac{1}{m_i} + \delta \times S_{ik} \tag{5.2}$$

where δ is a parameter that determines the weighting for the relative scores (Lo, 2008).

The parameter δ can be chosen purely at the FoF managers discretion or by mathematically solving for the optimal value of δ such that $\sum_k \gamma_{ik} = 1$ and resulting expected return ($\tilde{\mu}_i$) and volatility ($\tilde{\sigma}_i$) in (5.3) are equal to those calculated in *Stage two*.

For a given set of fund allocations $\boldsymbol{\gamma}_{ik} \equiv \left[\gamma_{ik_1}, \ldots, \gamma_{im_j}\right]'$ in asset class i, the implied expected return and volatility of the asset class are given by

$$\tilde{\mu}_i = \gamma_i' v_i \quad \tilde{\sigma}_i = \sqrt{\gamma_i' \quad {}_i\gamma_i} \tag{5.3}$$

Finally the capital allocations to each fund can be calculated as

$$K_{ik}^* = K_i^* \times \gamma_{ik} \tag{5.4}$$

As qualitative ability of the FoFs manager is entrenched in this stage of the process, computation of individual fund allocations or weightings is beyond the scope of this dissertation. As such, the out of sample tests conducted in Chapter 6 are based on equal capital allocation to the individual managers which comprise the groups identified in Chapter 4.

Chapter 6

Out of Sample Analysis

To evaluate the investment process, and determine the relative effectiveness of different classification methodologies, this chapter applies the group weightings estimated in Chapter 4 in an out of sample test.

This investment process is based on categorisations identified and forward estimates calculated using statistical methods employed on historical data. These yield weightings and capital amounts to be allocated to hedge funds with the objective of earning a target return in the next period. In examining classification effectiveness, this dissertation uses classifications identified in each adjacent sample period to accordingly segment the next period returns of funds, into the associated categories. Thereafter, allocations per category achieved in each year are applied to the respective funds realised returns in the subsequent year to examine if the identified allocations result in the expected or target HFoFs return in that period.

6.1 Modified S&S Tests

Recall that in Year1 five hedge funds were identified as Risk Reducers and five as Pure Diversifiers. The Year2 results of these funds were grouped together, on the basis of equal allocation to individual funds within asset class as assumed in *Stage two*, to calculate the annualised return for respective asset classes as presented in Table 6.1. Allocations derived from Year1 data are then applied to these calculated returns to estimate the FoFs return that could have been achieved in Year2 if this investment process had been employed. These results indicate that the target return would have been achieved in Year2 by a 0.1% margin.

Similarly, allocations derived from Year2 and Year3 are applied to the realised returns of Year3 and Year4 respectively. These results are presented in Tables 6.2 and 6.3.

Table 6.1: Application of Year1 Allocations

		RR					**PD**				
		AllanG	**Bacci**	**BigRock**	**CatACP**	**GEN.X**	**Gryphon**	**Hollard**	**Peregrin**	**RMBAM**	**Tantalum**
Monthly Returns	Jul-07	0.0017	0.0036	-0.0020	0.0090	0.0082	0.0071	-0.0141	-0.0325	0.0158	0.0124
	Aug-07	0.0053	0.0193	0.0198	0.0566	0.0180	0.0194	0.0070	-0.0102	0.0057	0.0219
	Sep-07	0.0002	0.0133	0.0000	0.0294	0.0001	0.0050	0.0266	-0.0545	0.0027	0.0018
	Oct-07	0.0034	0.0293	0.0178	0.0441	0.0075	0.0013	0.0344	0.0187	0.0112	0.0305
	Nov-07	0.0244	-0.0072	0.0040	-0.0099	0.0124	0.0171	0.0169	-0.0581	0.0299	-0.0163
	Dec-07	0.0071	0.0097	-0.0040	-0.0176	0.0122	0.0097	0.0149	-0.0028	0.0092	-0.0112
	Jan-08	-0.0029	-0.0785	0.0247	-0.0637	0.0242	0.0055	-0.0284	-0.0168	0.0118	-0.0694
	Feb-08	0.0000	0.0330	0.0090	0.0455	0.0054	0.0202	0.0421	0.0551	0.0422	0.0758
	Mar-08	0.0133	-0.0187	0.0090	-0.0431	0.0144	0.0141	0.0159	0.0313	0.0114	0.0039
	Apr-08	0.0004	-0.0120	0.0080	-0.0630	0.0145	0.0021	0.0257	0.0231	0.0235	0.0393
	May-08	0.0109	0.0248	0.0100	-0.0621	0.0091	0.0220	0.0247	-0.0173	0.0228	0.0101
	Jun-08	0.0039	-0.0418	0.0090	-0.0736	0.0141	0.0133	0.0060	0.0152	0.0425	-0.0248
Annualised Return		**0.0695**	**-0.0308**	**0.1099**	**-0.1502**	**0.1492**	**0.1452**	**0.1832**	**-0.0539**	**0.2531**	**0.0691**
Average Group return		**0.0295**					**0.1193**				
Group Allocation		**0.3118**					**0.6883**				
Portfolio Return		**0.0913**									
Target Return		**0.09**									

Table 6.2: Application of Year2 Allocations

		RR							PD						
		BigRock	**CatACP**	**Hollard**	**Mango**	**Pardus**	**Tempero**	**TrendI**	**AllanG**	**Bacci**	**CredS**	**Finch**	**Gryphon**	**RMBAM**	**VolArb**
Monthly Returns	Jul-08	0.0060	0.1472	-0.0315	0.0270	0.0157	0.0235	-0.1655	0.0166	-0.0011	0.0147	-0.0135	0.0174	-0.0414	0.0094
	Aug-08	0.0129	0.1614	-0.0243	0.0222	0.0073	0.0073	-0.0157	0.0099	0.0098	0.0137	-0.0322	0.0171	0.0185	0.0094
	Sep-08	0.0100	-0.1014	-0.0274	-0.0225	0.0020	0.0103	0.0557	0.0223	0.0007	0.0148	-0.0234	0.0087	-0.0104	-0.0110
	Oct-08	0.0411	-0.0864	-0.0050	0.0634	0.0247	0.0071	0.0813	0.0158	-0.0120	0.0155	-0.3025	0.0216	-0.0131	-0.0129
	Nov-08	0.0227	-0.0038	0.0050	-0.0239	-0.0252	0.0109	-0.0096	0.0177	0.0175	0.0145	-0.0808	0.0131	-0.1002	-0.0040
	Dec-08	0.0178	-0.0492	0.0050	0.0305	-0.0056	0.0213	0.0372	0.0085	0.0146	0.0136	-0.0426	0.0091	-0.0492	0.0240
	Jan-09	-0.0131	0.0587	0.0129	0.0294	0.0153	0.0152	0.0187	0.0242	0.0067	0.0131	-0.1101	0.0140	0.0209	0.0046
	Feb-09	0.0218	0.0154	0.0040	0.0213	0.0111	0.0142	0.0227	0.0050	-0.0038	0.0121	-0.1428	0.0111	-0.0232	0.0006
	Mar-09	0.0266	0.0438	0.0149	0.0289	0.0276	0.0099	-0.0392	0.0020	0.0177	0.0116	-0.0060	0.0122	0.0172	0.0115
	Apr-09	0.0237	0.0596	-0.0050	0.0348	0.0461	0.0071	-0.0165	-0.0096	0.0142	0.0104	0.0959	0.0093	0.0070	0.0066
	May-09	0.0119	-0.0141	0.0208	0.0260	0.0244	0.0092	0.0019	0.0216	0.0100	0.0100	0.1023	0.0084	0.0220	0.0213
	Jun-09	0.0159	-0.0034	0.0040	0.0198	0.0182	0.0076	-0.0074	-0.0061	0.0102	0.0073	0.0768	0.0095	0.0409	0.0102
Annualised Return		**0.2152**	**0.2103**	**-0.0279**	**0.2854**	**0.1721**	**0.1531**	**-0.0570**	**0.1347**	**0.0874**	**0.1622**	**-0.4355**	**0.1622**	**-0.1135**	**0.0712**
Average Group return		**0.1359**							**0.0098**						
Group Allocation		**0.5049**							**0.4951**						
Portfolio Return		**0.0735**													
Target Return		**0.09**													

Table 6.3: Application of Year3 Allocations

		RR				**TD**				**PD**										
		Pardus	**Tantalum**	**Tempero**	**TrendI**	**BigRock**	**CatACP**	**Peregrin**	**RMBAM**	**AllanG**	**Bacci**	**CredS**	**Finch**	**Gryphon**	**Hollard**	**InvesAQ**	**InvesFI**	**InvesPA**	**MergAR**	**VolArb**
Monthly Returns	Jul-09	-0.010	0.056	0.007	0.011	0.001	0.063	-0.010	0.018	-0.001	0.009	-0.061	-0.033	0.008	0.022	0.024	0.007	0.030	0.013	0.017
	Aug-09	0.022	0.017	0.005	-0.001	0.003	0.089	0.022	-0.016	0.006	0.013	0.006	-0.020	-0.004	0.001	-0.011	0.005	0.010	0.016	0.013
	Sep-09	0.033	0.017	0.006	-0.022	0.021	0.029	0.033	0.011	0.004	0.003	0.005	0.060	0.007	0.015	0.013	0.007	0.073	0.001	0.008
	Oct-09	0.000	0.025	0.007	0.004	0.023	0.030	0.000	0.017	0.009	0.010	0.004	0.016	0.019	0.029	0.030	0.013	0.005	0.023	0.008
	Nov-09	-0.030	0.014	0.004	-0.004	0.011	-0.010	-0.030	0.020	0.003	0.011	0.005	0.012	-0.001	-0.001	0.000	0.001	-0.009	0.015	0.010
	Dec-09	0.006	0.017	0.006	-0.034	0.015	0.017	0.006	0.012	0.004	0.012	0.004	-0.018	0.004	0.025	0.005	0.012	0.016	0.012	0.008
	Jan-10	-0.001	0.014	0.004	-0.013	0.015	0.012	-0.001	0.011	0.004	-0.009	0.010	0.008	0.005	-0.015	-0.003	-0.002	0.049	-0.013	0.007
	Feb-10	0.004	0.010	0.006	-0.013	-0.010	0.014	0.004	0.016	0.003	0.005	0.013	-0.003	0.004	0.024	0.017	0.017	-0.003	0.009	0.003
	Mar-10	0.029	0.022	0.013	-0.042	0.019	-0.004	0.029	-0.001	0.000	0.020	0.012	0.078	0.006	0.026	0.008	0.015	0.076	0.033	0.012
	Apr-10	0.006	-0.004	0.006	0.055	0.001	0.015	0.006	0.013	0.013	0.006	0.012	0.000	-0.001	0.002	-0.010	0.010	-0.004	0.001	0.008
	May-10	-0.019	-0.002	0.014	0.027	0.004	-0.015	-0.019	0.011	0.008	-0.020	0.010	-0.012	0.008	-0.030	-0.016	0.007	-0.046	-0.020	-0.001
	Jun-10	-0.003	0.000	0.008	0.092	0.024	-0.003	-0.003	0.013	0.011	-0.011	0.013	-0.003	0.001	-0.009	0.008	0.007	0.008	-0.011	0.004
Annualised Return		**0.0365**	**0.2017**	**0.0903**	**0.0547**	**0.1325**	**0.2577**	**0.0365**	**0.1304**	**0.0686**	**0.0494**	**0.0312**	**0.0818**	**0.0581**	**0.0891**	**0.0636**	**0.1055**	**0.2180**	**0.0792**	**0.1014**
Average Group return		**0.0958**				**0.1393**				**0.0860**										
Group Allocation		**0.3283**				**0.32812**				**0.34358**										
Portfolio Return		**0.106705962**																		
Target Return		**0.09**																		

The above results indicate that using the modified S&S methodology developed in this dissertation within the investment process is successful in yielding target returns for FoFs. Using this classification, the target return is achieved and exceeded two out of three times.

6.2 Clustering Tests

As in section 6.1, allocations derived from clustering in Years 1 to 3 are applied to the realised returns of Years 2 to 4 respectively; results following in Tables 6.4 to 6.6.

Table 6.4: Application of Year1 cluster Allocations

		Cluster 7				Cluster 8						Cluster 9		
		CredS	**GEN.X**	**PSG SE**	**RMBAM**	**AllanG**	**BigRock**	**Khul**	**Pardus**	**VolArb**	**X.Cheq**	**CorGran**	**Gryphon**	**InvesAQ**
Monthly Returns	Jul-07	0.012	0.008	0.023	0.016	0.002	-0.002	0.005	-0.033	0.008	0.020	0.002	0.007	0.042
	Aug-07	0.015	0.018	0.007	0.006	0.005	0.020	0.024	-0.010	0.010	0.025	0.011	0.019	-0.023
	Sep-07	0.013	0.000	0.003	0.003	0.000	0.000	-0.015	-0.054	0.007	-0.015	0.011	0.005	0.018
	Oct-07	0.013	0.007	0.015	0.011	0.003	0.018	0.026	0.019	0.009	0.014	0.001	0.001	0.054
	Nov-07	0.014	0.012	0.036	0.030	0.024	0.004	0.002	-0.058	0.007	0.005	0.002	0.017	0.035
	Dec-07	0.013	0.012	0.007	0.009	0.007	-0.004	0.001	-0.003	-0.003	0.009	0.011	0.010	0.011
	Jan-08	0.013	0.024	0.001	0.012	-0.003	0.025	0.012	-0.017	0.003	0.013	0.017	0.005	0.032
	Feb-08	0.012	0.005	0.004	0.042	0.000	0.009	0.000	0.055	0.018	0.046	0.007	0.020	0.069
	Mar-08	0.014	0.014	0.024	0.011	0.013	0.009	0.047	0.031	0.001	0.036	0.003	0.014	0.038
	Apr-08	0.014	0.014	0.016	0.024	0.000	0.008	0.018	0.023	0.017	0.003	0.005	0.002	0.038
	May-08	0.014	0.009	-0.019	0.023	0.011	0.010	0.008	-0.017	0.010	0.022	0.011	0.022	0.002
	Jun-08	0.013	0.014	0.047	0.042	0.004	0.009	0.024	0.015	0.007	0.007	0.007	0.013	-0.001
Annualised Return		**0.1716**	**0.1492**	**0.1759**	**0.2531**	**0.0695**	**0.1099**	**0.1620**	**-0.0539**	**0.0987**	**0.2019**	**0.0913**	**0.1452**	**0.3591**
Average Group return		**0.1874**				**0.0980**						**0.1985**		
Group Allocation		**0.83874**				**0.14238**						**0.01888**		
Portfolio Return		**0.1749**												
Target Return		**0.09**												

Table 6.5: Application of Year2 cluster Allocations

		Cluster 3																	Cluster 4				
		AbsA	**AllanG**	**BadgQ**	**BigRock**	**ClearH**	**CorGran**	**CorMSA**	**Intern**	**InvesFI**	**Marco**	**OMI Cap**	**OMI MS**	**OMI N**	**Oryx**	**Peregrin**	**Tempero**	**VolArb**	**Foord**	**GEN.X**	**Gryphon**	**Khul**	**PSG SE**
Monthly Returns	Jul-08	0.007	0.017	0.027	0.006	-0.017	0.006	0.038	-0.046	0.057	0.022	-0.028	-0.004	0.050	0.013	0.016	0.024	0.009	0.027	-0.006	0.0174	0.0270	-0.0549
	Aug-08	-0.021	0.010	0.018	0.013	-0.008	0.012	0.038	-0.026	0.016	0.017	0.019	0.004	0.003	-0.013	0.007	0.007	0.009	0.016	0.020	0.0171	0.0222	0.0434
	Sep-08	-0.011	0.022	-0.007	0.010	-0.033	0.011	0.010	0.009	0.020	-0.018	-0.022	0.079	-0.011	0.010	0.002	0.010	-0.011	-0.023	0.009	0.0087	-0.0225	0.0300
	Oct-08	-0.031	0.016	0.022	0.041	-0.088	0.010	0.006	-0.029	-0.016	-0.008	0.007	-0.008	-0.079	-0.022	0.025	0.007	-0.013	-0.005	0.014	0.0216	0.0634	-0.0196
	Nov-08	-0.014	0.018	0.002	0.023	-0.015	0.011	0.006	-0.008	0.023	0.007	0.010	0.000	0.008	-0.024	-0.025	0.011	-0.004	0.011	0.000	0.0131	-0.0239	0.0667
	Dec-08	0.081	0.008	0.008	0.018	0.015	0.014	-0.012	0.003	0.026	0.029	0.018	0.063	-0.016	0.001	-0.006	0.021	0.024	-0.022	0.034	0.0091	0.0305	0.0099
	Jan-09	0.033	0.024	0.020	-0.013	-0.015	0.034	0.022	0.010	0.008	-0.010	0.001	0.001	-0.062	0.009	0.015	0.015	0.005	0.016	0.003	0.0140	0.0294	0.0116
	Feb-09	-0.012	0.005	0.029	0.022	-0.029	0.019	0.006	0.006	-0.011	-0.032	-0.010	-0.001	-0.024	0.016	0.011	0.014	0.001	-0.015	0.016	0.0111	0.0213	0.0434
	Mar-09	0.032	0.002	0.048	0.027	-0.008	0.009	0.025	-0.058	0.006	0.033	0.017	0.005	0.114	0.017	0.028	0.010	0.012	0.047	0.000	0.0122	0.0289	-0.0014
	Apr-09	-0.058	-0.010	0.022	0.024	0.015	0.005	0.016	-0.081	0.015	0.047	0.012	0.004	0.028	-0.023	0.046	0.007	0.007	-0.022	0.006	0.0093	0.0348	-0.0204
	May-09	0.044	0.022	0.053	0.012	0.077	0.008	0.015	0.024	0.010	0.024	0.033	0.004	0.033	0.027	0.024	0.009	0.021	-0.011	-0.003	0.0084	0.0260	0.0203
	Jun-09	0.000	-0.006	-0.042	0.016	0.000	0.005	0.001	-0.029	0.005	0.025	-0.008	0.023	0.015	-0.009	0.018	0.008	0.010	0.016	0.021	0.0095	0.0198	0.0108
Annualised Return		**0.0431**	**0.1347**	**0.2128**	**0.2152**	**-0.1075**	**0.1552**	**0.1841**	**-0.2078**	**0.1684**	**0.1406**	**0.0472**	**0.1790**	**0.0464**	**0.0007**	**0.1721**	**0.1531**	**0.0712**	**0.0338**	**0.1212**	**0.1622**	**0.2854**	**0.1425**
Average Group return		**0.0946**																	**0.1490**				
Group Allocation		**0.49632**																	**0.50368**				
Portfolio Return		**0.1220**																					
Target Return		**0.09**																					

Table 6.6: Application of Year3 cluster Allocations

		Cluster 2						Cluster 3									
		CatACP	**CorGran**	**CredS**	**Gryphon**	**OMI MS**	**Tempero**	**AllanG**	**BigRock**	**CorMSA**	**GEN.X**	**InvesFI**	**Khul**	**Oakmon**	**Pardus**	**PSG SE**	**X.Cheq**
Monthly Returns	Jul-09	0.063	0.007	-0.061	0.008	0.002	0.007	-0.001	0.001	0.027	0.015	0.007	-0.009	0.042	-0.010	0.009	0.018
	Aug-09	0.089	0.006	0.006	-0.004	0.002	0.005	0.006	0.003	0.008	0.004	0.005	-0.027	0.010	0.022	0.010	0.012
	Sep-09	0.029	0.006	0.005	0.007	0.052	0.006	0.004	0.021	-0.017	0.001	0.007	0.011	0.014	0.033	0.006	0.016
	Oct-09	0.030	0.006	0.004	0.019	0.001	0.007	0.009	0.023	0.026	-0.004	0.013	0.034	0.015	0.000	0.002	0.019
	Nov-09	-0.010	0.011	0.005	-0.001	-0.001	0.004	0.003	0.011	0.018	0.011	0.001	-0.002	-0.001	-0.030	0.006	0.004
	Dec-09	0.017	0.006	0.004	0.004	0.034	0.006	0.004	0.015	0.003	0.009	0.012	0.008	0.021	0.006	0.016	0.015
	Jan-10	0.012	0.006	0.010	0.005	-0.002	0.004	0.004	0.015	0.021	0.011	-0.002	-0.005	0.008	-0.001	0.000	0.001
	Feb-10	0.014	0.008	0.013	0.004	-0.001	0.006	0.003	-0.010	0.013	0.021	0.017	0.011	0.015	0.004	0.024	0.011
	Mar-10	-0.004	0.017	0.012	0.006	0.039	0.013	0.000	0.019	0.020	0.020	0.015	-0.004	0.013	0.029	0.013	0.009
	Apr-10	0.015	0.007	0.012	-0.001	0.001	0.006	0.013	0.001	0.001	0.014	0.010	0.012	0.004	0.006	0.017	0.009
	May-10	-0.015	0.007	0.010	0.008	0.006	0.014	0.008	0.004	0.013	0.005	0.007	0.002	-0.004	-0.019	0.025	0.004
	Jun-10	-0.003	0.003	0.013	0.001	-0.013	0.008	0.011	0.024	0.009	0.010	0.007	0.016	0.011	-0.003	-0.010	0.005
Annualised Return		**0.2577**	**0.0938**	**0.0312**	**0.0581**	**0.1259**	**0.0903**	**0.0686**	**0.1325**	**0.1517**	**0.1248**	**0.1055**	**0.0484**	**0.1568**	**0.0365**	**0.1232**	**0.1303**
Average Group return		**0.1095**						**0.1078**									
Group Allocation		**0.50321**						**0.49679**									
Portfolio Return		**0.1087**															
Target Return		**0.09**															

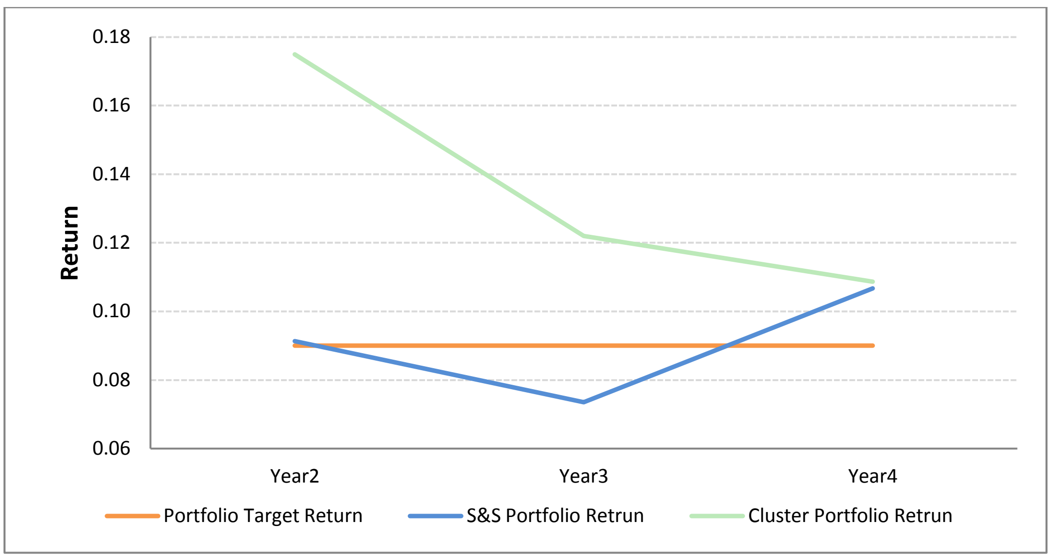
The above results show that when using clustering as the classification method, the target return is exceeded in every sample period.

6.3 Comparison

Comparison of the two classification methodologies finds a stronger case for the use of clustering as a classification tool relative to the modified S&S methodology, as FoFs returns simulated when using clustering is consistently above the target return. The modified S&S methodology results in inconsistent achievement of target returns. In addition, the marginal rate by which target returns are exceeded is consistently higher by clustering.

This is illustrated in Graph 6.1.

Graph 6.1: Classification Relative Returns



In summary, the above results indicate merit to the investment process outlined with five out of the six tests reflecting a portfolio return achieved which was greater than the stipulated target return. We thereby concluded that using a combination of structured, quantitative methods and qualitative knowledge can benefit the investment process.

Chapter 7

Conclusion and Recommendations

This Chapter summarises the observations of Chapters 3, 4 and 6 with reference to the objective defined in Chapter 1, in a concluding manner. It also discusses prevalent limitations and finalises the dissertation with some recommendations for further research.

7.1 Concluding Remarks

- The objective of this dissertation was to develop an investment process for FOFs suitable to the South African hedge fund industry. The proposal stated in Chapter 1 suggested a three tiered process, which was developed and tested in Chapters 3 to 6.

- The objective of *Stage one* was to identify appropriate asset classes within the universe of hedge funds. Two classification methodologies were investigated in Chapter 3 - the modified S&S and cluster analysis, both of which saw funds migrate between categories across periods indicating that, due to their inherent flexibility, hedge fund classifications are not static. Thus classification simulations must be conducted as part of the investment process on an ongoing basis.

- The objective of *Stage two* was to evaluate optimal allocations of capital for the identified asset classes. Optimisations based on forward-looking parameters were implemented in Chapter 5, for each classification methodology, estimated by regression analysis. These results suggested that in a high interest rate and downward trending risk asset market environment, capital allocation should be geared towards categories defined by low or negative betas with greater value placed on lower volatility than higher return.

- Chapter 7 conducted tests of the optimal allocations computed in stage two, for each stage one classification methodology by applying the allocations achieved from each sample period investigated and applying the results to realised returns of the subsequent period in accordance with the identified group for the previous year. This chapter found both methodologies to yield target returns in majority. However, on comparison, the modified S&S classification resulted in target returns being achieved in two out of three years while clustering achieved the target in each sample period. As such this dissertation finds cluster analysis to be the more efficient hedge fund classification methodology.

7.2 Limitations

The proposed investment process outlined herein relies on verified historical return information for the pool of hedge funds under consideration. Due to the limited regulation nature of the industry, historically these metrics have often been unaudited or unavailable which can lead to some selection bias.

7.3 Recommendations

The investment process as proposed in this dissertation considers the broad three step procedure for implementation, focusing primarily on classification methodologies. The following are three aspects relating to the investment process there are suggested for further research:

- **Shorter Investigation Period**

  This analysis is based on one year periods. An examination of the results of shorter periods on groupings and resultant allocations may highlight optimal allocation and re-evaluation periods.

- **Best Fit Factor Model**

  This analysis is based on CAPM. While this is the starting point for factor models, an examination of multiple factor models to select an appropriate model accounting for a majority of variation in returns may have significant impact on forward looking parameter estimates on which allocations are ultimately computed.

- **Individual Manager Selection**

  Due to the qualitative nature of allocation across individual managers, this is beyond the scope of this dissertation. However, during implementation of the investment process practising professionals should investigate the relative efficiency of allocating capital equally across hedge funds comprising an asset class and differential allocations based on the score card.

Bibliography

Agarwal, V., Bakshi, G. and Huij, J. (2008). *Implications of Volatility, Skewness, and Kurtosis Risks for Hedge Fund Performance.* Georgia: Georgia State University.

Amenc, N., Curtis, S. and Martellini, L. (2004). *The Alpha and Omega of Hedge Fund Performance Measurement.* EDHEC Risk And Asset Management Research Center, Singapore.

Banfield, J. D. and Raftery, A. E. (1992). Ice Floe Identification in Satelite Images Using Mathematical Morphology and Clustering About Principle Curves. *Journal of the American Statistical Association,* **87**, 7-16.

Banfield, J. D. and Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics,* **49**, 803-821.

Celeux, G. and Govaert, G. (1992). A Classification EM Algorithm for Clustering and Two Stochastic Versions. *Computational Statistics and Data Analysis* **14**, 315-332.

Conor, G., and Korajczyk, R. A. (1993). A test for the number of factors in an approximate. *Journal of Finance,* **48**, 1263-91.

Croissant, Y., and Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, **27(2)**, 1-43

Darlington, R. B., Weinberg, S. L., and Walberg, H. J. (1973). Canonical Variate Analysis and Related Techniques. *Review of Educational Research,* **43(4)**, 433-454.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Liklihood for Incomplete Data via EM Algorithm. *Journal of the Royal Statistical Society*, *Series B, **39**,* 1-38.

Efron, B. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Fraley, C., and Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association,* **97**, 611-631.

Fraley, C., Raftery, A. E., and Scrucca, L. (2002). Package 'mclust'. URL http://cran.r-project.org/web/packages/mclust/mclust.pdf.

Fung, W., and Hsieh, D. A. (2004). Hedge fund benchmarks: A risk based approach. *Financial Analysts Journal,* **60**, 65-80.

Gower, J. C., and Hand, D. J. (1996). *Biplots.* London: Chapman and Hall.

Gower, J.C., Lubbe, S. and le Roux, N.J. (2011) *Understanding Biplots*. Chichester: John Wiley.

Gregoriou, G. N., Sedzro, K., and Zhu, J. (2004). Hedge fund performance appraisal using data envelopment analysis. *European Journal of Operational Research,* **164(2)**, 555 - 571.

Johnson, R. A., and Wichern, D. W. (2007). *Applied Multivariate StatistIcal Analysis.* Prentice Hall International: Englewood Cliffs, NJ.

Koh, F., Lee, D., and Fai, P.K. (2002). Investing in Hedge Funds: Risk, Return and Pitfalls. Working Paper.

Le Roux, N.J. and Lubbe, S. (2011) UBbipl: R package accompanying Understanding Biplots by Gower, J. C. Lubbe, S. and le Roux N.J. (2011). URL www.wiley.com/go/biplots.

Linter, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, **47(1)**, 13–37.

Lo, A. W. (2008). *Hedge Funds: An Analytic Perspective.* Princton, New Jersey: Princton University Press.

MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *5th Berckley Symposium on Mathematical Statistics and Probability* (pp. 281-297). Berkeley, CA: University of California Press.

McLachlan, G. J., and Basford, K. E. (1988). *Mixture Models: Inference and Applications to Clustering.* New York: Marcel Dekker.

McLachlan, G. J., and Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: Wiley.

Modigliani, F., and Modigliani, L. (1997). Risk-adjusted Performance. *The Journal of Portfolio Management,* **23(2)**, 45-54.

Murtagh, F., and Raftery, A. E. (1984). Fitting Straight Lines to Point Patterns. *Pattern recognition,* **17**, 479-483.

Nicholas, J. (2004). *Hedge Fund of Funds Investing: An Investors Guide.* New Jersey: Bloomberg Press.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Ross, S. A. (1976). The Arbitrage Pricing Theory of capital asset pricing. *Journal of Economic Theory*, **13**, 341-360.

Schneeweis, T., and Spurgin, R. (2000). Hedge Funds: Portfolio Risk Diversifiers, Return Enhancers or Both ? *AIMA Newsletter*.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annuals of Statistics,* **2**, 461-464.

Sharpe, W. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance,* **19(3)**, 425–442.

Strachman, D., Bookbinder, R. (2010). *Fund of Funds Investing: A Roadmap to Portfolio Diversification.* Hoboken, New Jersey: John Wiley and Sons, Inc.

Stulz, R. (2007). Hedge Funds: Past, Present, and Future. *The Journal of Economic Perspectives,* **21(2)**, 175 - 194.

Sylvain, A., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys,* **4**, 40-79.

Ward, J. H. (1963). Hierachical Groupings to Optomize an Objective Function. *Journal of the American Statistical Association,* **58**, 234-244.