Single manager hedge funds – aspects of classification and diversification

by
Florian Martin Böhlandt

Dissertation presented for the degree of Doctor of Philosophy at
Stellenbosch University

Promoter: Prof Eon Smit, USB Graduate School of Business
Prof Niel Krige, USB Graduate School of Business

DECLARATION

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

August 2013

ACKNOWLEDGEMENTS

Access to various databases and additional financial pricing information in this dissertation was provided by Firstrand Alternative Investment Management (FRAIM). Special thanks go to Braam Jordaan, Bradley Anthony and Kaimini Naidoo for securing data access via the London office.

I would like to thank Laurent Favre for his information on pricing models for hedge funds and portfolio diversification as well as Professor Meshach Jesse Aziakpono (USB) for his quantitative insights on factor models and cointegration. Thanks to Professor Deirdre McCloskey (University of Illinois at Chicago) who provided great awareness for the potential pitfalls of statistical and econometric methods.

My supervisors, Professor Eon Smit (USB) and Professor Niel Krige (USB), gave great support throughout my studies and assisted this research with insights as well as financial backing. Also, my sincere appreciation goes to Marietjie van Zyl, Jeanne Kuhn and Francis Meyer for their administrative efforts.

Special thanks to my parents, Dr. Dietmar Böhlandt and Barbara Böhlandt, for their continuous moral and financial support, Dr. Margot Flint for loving support and tremendous patience and Dr. Gavin George for a warm handshake for finishing second.

ABSTRACT

A persistent problem for hedge fund researchers presents itself in the form of inconsistent and diverse style classifications within and across database providers. For this paper, single-manager hedge funds from the Hedge Fund Research (HFR) and Hedgefund.Net (HFN) databases were classified on the basis of a common factor, extracted using the factor axis methodology. It was assumed that the returns of all sample hedge funds are attributable to a common factor that is shared across hedge funds within one classification, and a specific factor that is unique to a particular hedge fund. In contrast to earlier research and the application of principal component analysis, factor axis has sought to determine how much of the covariance in the dataset is due to common factors (communality). Factor axis largely ignores the diagonal elements of the covariance matrix and orthogonal factor rotation maximises the covariance between hedge fund return series.

In an iterative framework, common factors were extracted until all return series were described by one common and one specific factor. Prior to factor extraction, the series was tested for autoregressive moving-average processes and the residuals of such models were used in further analysis to improve upon squared correlations as initial factor estimates. The methodology was applied to 120 ten-year rolling estimation windows in the July 1990 to June 2010 timeframe. The results indicate that the number of distinct style classifications is reduced in comparison to the arbitrary self-selected classifications of the databases. Single manager hedge funds were grouped in portfolios on the basis of the common factor they share. In contrast to other classification methodologies, these common factor portfolios (CFPs) assume that some unspecified individual component of the hedge fund constituents' returns is diversified away and that single manager hedge funds should be classified according to their common return components. From the CFPs of single manager hedge funds, pure style indices were created to be entered in a multivariate autoregressive framework.

For each style index, a Vector Error Correction model (VECM) was estimated to determine the short-term as well as co-integrating relationship of the hedge fund series

with the index level series of a stock, bond and commodity proxy. It was postulated that *a)* in a well-diversified portfolio, the current level of the hedge fund index is independent of the lagged observations from the other asset indices; and *b)* if the assumptions of the Efficient Market Hypothesis (EMH) hold, it is expected that the predictive power of the model will be low. The analysis was conducted for the July 2000 - June 2010 period. Impulse response tests and variance decomposition revealed that changes in hedge fund index levels are partially induced by changes in the stock, bond and currency markets. Investors are therefore cautioned not to overemphasise the diversification benefits of hedge fund investments. Commodity trading advisors (CTAs) / managed futures, on the other hand, deliver diversification benefits when integrated with an existing portfolio.

The results indicated that single manager hedge funds can be reliably classified using the principal factor axis methodology. Continuously re-balanced pure style index representations of these classifications could be used in further analysis. Extensive multivariate analysis revealed that CTAs and macro hedge funds offer superior diversification benefits in the context of existing portfolios. The empirical results are of interest not only to academic researchers, but also practitioners seeking to replicate the methodologies presented.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

a priori	in advance / *ex ante*
ABS	Asset-Based-Style
(A)DF	(Augmented) Dickey-Fuller test
APCA	Asymptotic PCA
APT	Arbitrage Pricing Theory
AR(MA)	Autoregressive (Moving Average)
AUM	Assets-under-Management
BABDIDX	Barclays Aggregate Bond Index
CAPM	Capital Asset Pricing Model
BXM	BuyWrite Index Monthly
CBOEVIX	Chicago Board Options Exchange Volatility Index
CISDM	Center for International Securities and Derivatives Management Database
Ceteris paribus	*lat.* All other conditions being equal
CFP	Common Factor Portfolio (≠ FRP)
CTA	Commodity Trading Advisor
Df	Degrees of freedom
e.g.	*exempli gratia* (*lat.* for example)

EACM	Evaluation Associates Capital Markets
ED	Event Driven
EH	Equity Hedge
EHf	Equity Hedge – Finance Sector
EHg	Equity Hedge – Fundamental Growth
EHv	Equity Hedge – Fundamental Value
EM	Emerging Markets
EMH	Efficient Market Hypothesis
et al.	*et alia* (*lat.* and others)
etc.	*et cetera* (*lat.* and so forth)
FoHF	Funds of Hedge Fund
FRAIM	Firstrand Alternative Investment Management
FRP	Factor Replicating Portfolio (≠ CFP) →PCA
GARCH	Generalised Autoregressive Conditionally Heteroskedastic Model
GOLDUSDL	Price change in Gold quoted in USD (London PM Fix)
GRG	Generalized Reduced Gradient
GSCI	S&P Goldman Sachs Commodity Total Return Index
HAC	Heteroskedasticity-Autocorrelation-Consistent
HFN	HedgeFund.Net Database

HFNI	HedgeFund.Net indices
HFR	Hedge Fund Research Database
HFRX	Hedge Fund Research benchmark indices
HML	High Minus Low Portfolio (Fama-French)
i.e.	*id est* (*lat.* that is)
i.i.d.	independent and identically distributed
IFC	International Finance Corporation
KPSS	Kwiatkowski-Phillips-Schmidt-Shin test
KKT	Karush-Kuhn-Tucker
L	Long Only / Designated Long
LM	Lagrange Multiplier
LS	Long/Short Equity
LTCM	Long Term Capital Management
M	(Global) Macro – Systemic Diversified
MA	Moving Average
(M)AIC	(Modified) Akaike Information Criterion
MAR	Managed Account Reports Database
(M)HQIC	(Modified) Hannan-Quinn Information Criterion
ML	Maximum Likelihood

MS	Markov switching model
(M)SBIC	(Modified) Schwarz's Bayesian Information Criterion
MSCI	Morgan Stanley Capital International
MSCIEM	MSCI Emerging Markets Index
MSCIEXUS	MSCI World Index excluding USA
MSCIW	MSCI World Index US-Dollar
MSE	Mean Square Error
MSR	Mean Square due to Regression
OLS	Ordinary Least Squares
P	Proportion (in tables only)
p.a.	*per annum* (*lat.* for per year)
PCA	principal component analysis
per se	(*lat.* by itself)
PTFS	primitive trend following strategy
PTFSBD	PTFS – Bond
PTFSFX	PTFS – Foreign Exchange
PTFSCOM	PTFS – Commodity
PTFSIR	PTFS – Short-term Interest Rates
PTFSSTK	PFTS – Stock Index

ROI	Return-on-Investment
RV	Relative Value
SIC	Schwarz Information Criterion
SMB	Small Minus Big (Fama-French)
SP500 (S&P 500)	Standard & Poor's 500
SSBWIDX	Salomon Smith Barney World Bond Index
SSE	Sum Square for Error
TASS	Lipper Tass Database
TED	Treasury-Eurodollar (spread)
US	United States
US10Y3M	Spread between 10-year US Treasury Bill and 3-month Treasury
USBAA10Y	Spread between Moody's yield on seasoned corporate bonds – all industries – Baa and 10-year Treasury Bill
USD (US-Dollar)	United States Dollar
USDIDX	Federal Reserve Traded Weighted Index of the US Dollar
USMO10Y	Spread between the contract rate on 30-year, fixed-rate conventional home mortgage commitments (US) and 10-year Treasury Bill
USTB3M	Yield on US Treasury Bill 3 month maturity
VAR(M)	Vector Autoregressive (Model)

VEC(M)	Vector Error Correction (Model)
WML	Winners Minus Losers (Momentum)

CHAPTER 1: INTRODUCTION

Forty years after the inception of the first hedge funds, despite many claims to the contrary, the alternative investment industry is still very much alive. As of April 2013, total industry assets are estimated at 2.375 trillion United States Dollars (USD). Investors attributed 15.2 billion USD net capital to hedge funds in the first quarter of 2013. In the past four years, hedge funds have received capital inflows in all quarters bar one. In total, assets-under-management increased by 122 billion USD in the first quarter of 2013. This marks the largest increase since the fourth quarter of 2010 (Hedge Fund Research, 2013).

Hedge funds are however not free from criticism. Lack of transparency, high management and performance fees, lock-up periods and minimum investment amounts, as well as a lack of regulatory oversight, increase the perceived risks associated with hedge fund investments. It is often held that incentive fees and high water marks lead managers to take unacceptable levels of systemic risk. The collapses of the Long-term Capital Management (LTCM) fund in 1998 and the Amaranth fund in 2006 show that high leverage ratios can increase the shortfall and default risks. Many criticised hedge funds' undue exposure to collateralised debt obligations and their susceptibility to the subsequent sub-prime lending crisis. The failure of three Bear Sterns funds in 2007 shook investor confidence in hedge funds' ability to offer superior downside protection in an adverse market environment. Some researchers argue that even though hedge funds did not cause the financial crisis of 2007, they exacerbated the situation by acting as an intermediary between investors seeking and banks creating high yield bearing securities (Lysandrou, 2011).

The diversity and complexity of the industry has brought with it the need for due diligence processes in the assessment of hedge fund investments. Institutional as well as individual investors are desperately seeking methods that allow them to sift through the vast universe of hedge fund investments. In particular, investors hope to lift the veil surrounding the myriad of style classifications and investment strategies. Additionally, there is an increasing demand to identify those hedge fund investments that offer diversification benefits in the context of traditional asset portfolios.

Hedge fund time series used in this dissertation were derived from two databases: Hedge Fund Research (HFR) and Hedgefund.Net (HFN). With respect to style classification, HFN distinguishes between 32 different main strategic categories, many of which contain only a handful of highly specialised hedge funds. HFR, on the other hand, classifies single manager funds according to four main strategies and several sub-strategies. Other than trading strategy, hedge funds are also classified according to geographic or sector focus. The classifications within and across database providers are far from consistent. It is questionable whether hedge funds belonging to the same strategy represent a homogenous group of funds representative of a distinct investment style and focus. Furthermore, the persistence of such classification is questionable in the presence of style drift, performance-smoothing and phase-locking behaviour.

The increasing number of hedge fund classifications has spawned numerous style indices to assess the portfolio diversification benefits of hedge funds. Besides the aforementioned difficulties concerning precise classification, hedge fund indices carry another heavy burden: the reconciliation of representativeness and investability. In addition, they suffer from several biases as a result of index composition and construction: survivorship, instant-history and self-selection / database selection. The performance of indices from database vendors and index providers differ on account of constituent weighting and re-balancing intervals as well as selection criteria. The range of style indices is as diverse as the strategic classifications of single manager funds.

In the past, research focused on establishing asset-class-pricing models for hedge funds in the form of adaptations of the original Sharpe (1992) model and its application to mutual funds, identifying sources of systematic risk and assessing exposure of hedge funds to different markets. Later extensions included the Fama-French and momentum factors as proxies for higher moments and non-linear risk exposure (e.g. Chan, Getmansky, Haas & Lo, 2006). Other researchers thought to replicate the return profiles of hedge fund investments by simulating primitive trend-following strategies (PTFS).

Recently, statistical factor models identified unobservable factors accounting for a significant proportion of the variation in hedge fund returns. Fung and Hsieh (1997a, 2002a) are amongst the most prominent researchers of statistical factor models and their application to hedge funds. The differentiation between a fund manager's 'location' and 'style' choice greatly improved upon the economic interpretability of factors extracted from principal component analysis (PCA). Additionally, PCA allowed for the construction of factor replicating portfolios representative of pure investment styles.

Inspired by earlier research using PCA, this study has sought to conduct factor analysis using an orthogonal factor model. Rather than seeking to explain the total variance in hedge fund returns, the focus of this study lay with maximising the explained covariance and, hence, communalities of hedge fund return series. As a result of the study, the common factor model introduced helps to shed some light on the somewhat arbitrary style classifications of hedge funds. It was implied that covariance across hedge fund returns can be explained by a limited number of unobserved random or common factors, whilst a proportion of the variance in return is attributable to additional noise or specific factors. Classifying hedge funds according to their rotated factor loadings yields an unbiased estimator of the strategic clusters of single manager funds. Hedge funds attributed to these common factor portfolios (CFP) made up the constituents of the pure style indices used in further analysis.

The results from the common factor model were used as indicators to group hedge funds together based on the similarities of their past performance. These portfolios could be thought of as clusters of single manager hedge funds that were invested into the same markets and adhered to a similar investment style as evidenced by the likeness of their past performance. The performance of every single manager hedge funds was then attributed to two components: the broad strategic theme shared amongst hedge fund within a strategic cluster and their individual return component, which is not shared with any other hedge fund of the sample. An index was created from the weighted series of each cluster, or CFP, and that index series was designated as the pure style index representation of a particular hedge fund strategy (the nomenclature for the indices was adopted from existing strategy indices and database classifications).

Portfolio diversification results from combining assets with lowly or negatively correlated returns. It has been argued in the past that hedge funds exhibit insignificant correlation coefficients with other asset classes and are thus ideal complements to existing portfolios. However, in efficient portfolios it is expected that this relationship will hold in a multivariate autoregressive framework. For this study, the question to be answered was whether past changes in stock, bond or commodity indices induce changes in hedge fund index levels (or vice versa). In well-diversified portfolios, it is expected that shocks to one asset class will have no discernible impact on the future value of another. Similarly, contemporaneous observations for any series are independent of the lagged series. Assuming that the Efficient Market Hypothesis (EMH) held, it was expected that the predictive power of the multivariate autoregressive model would be low. A number of tests were employed to detect any lead-lag relationships between standard assets and the hedge fund indices, including impulse response and variance decomposition. It was decided to use a Vector Error Correction Model (VECM) to allow for co-integrating relationships between the variables.

The VECM can be thought of as a system of equations that seeks to establish a characteristic function for the three asset classes as well as each of the hedge fund indices created. Due to the complexity of the model and requirements for the number of estimated coefficients, it was decided to determine a separate VECM for each distinct hedge fund index. It was assumed that a particular index, hedge fund or otherwise, could be described in sufficient detail by considering its own past performance as well as the performance across the other three indices. In contrast to asset-class factor models, all coefficients of the multivariate model are estimated simultaneously, since all model variables are regressands and predictor variables at the same time. The estimation technique is unbiased since there are no *a priori* assumptions about the causal relationship between the variables. However, the interpretation of coefficients and the impact from shocks to the system can only be assessed under *ceteris paribus* conditions.

The rest of this dissertation is structured as follows: Chapter 2 lists the contributions to existing research and Chapter 3 reviews the available literature on hedge funds. The focal points of research into hedge funds are as follows: the application of factor models, alternative investments and portfolio diversification, hedge fund indexation, quantification of attrition rates and survivorship bias, and the estimation of downside risk and value-at-risk. Concepts deemed relevant in the context of this study are addressed. Chapter 4 outlines the methodology and follows from the literature review. The principal factor axis methodology and VECM are discussed in detail. Because most statistical software makes provision for the estimation of the specified models, the reader interested in practical application or empirical results may skip the technical discussions. Chapter 5 introduces the two hedge fund databases and explains the data sampling process. Application of the factor model and empirical results thereof follow in Chapter 6. The statistical properties of hedge fund time series as well as data bias are accounted for to the degree possible. Chapter 7 estimates VECMs for several hedge fund styles and discusses the results and implications for EMH. Chapter 8 concludes the discussion and refers to potential future avenues of research.

CHAPTER 2: CONTRIBUTION TO EXISTING KNOWLEDGE

2.1 Introduction

This section outlines the distinct contributions to the research on hedge funds and provides the rationale for the methods introduced in Chapter 4. The point of departure for this research was the work by Fung and Hsieh (1997a, 2001). In order to differentiate between distinct hedge fund classifications and investment styles, statistical factor analysis categorises single manager hedge funds on the basis of their past performance rather than self-reported strategy. Contrary to the use of principal components and their application to hedge fund style indices in earlier research (see for example Amenc & Martellini, 2001; Kugler, Henn-Overbeck & Zimmermann, 2010), single manager hedge funds were evaluated in a principal factor axis framework. All data used are monthly observations.

2.2 Dimensionality reduction

As a dimensionality reduction technique, PCA yields promising results in the context of hedge fund indices. This is attributable to the small sample of variables entering the factor model and relatively long track records of indices. The same does not hold for factor analysis at the single manager fund level and the shortcomings of PCA and the application thereof in past research are discussed below. At every stage, the methodologies and models to date were compared to the adaptations and extensions suggested in this research. The following is an abbreviated list of the differences between PCA and the factor axis methodology suggested here:

1. Considering the diversity of investment philosophies, location choices, asset composition and dynamic trading strategies, it becomes increasingly unlikely to identify a limited number of factors explaining a significant proportion of variation in returns across many hedge funds. In the example of Fung and Hsieh (1997a), the five extracted style factors accounted for only 43 percent of the return variance across 409 hedge funds, despite a relatively short observation window of 36 months. For the 253 hedge funds of the joint HFR/HFN sample in this

study, 64 factors were extracted to explain 90 percent of the return variation. One problem results from small yet significant eigenvalues associated with the extracted eigenvectors. Considering the marginal differences between the eigenvalues, the truncated component model is arbitrary and of little statistical significance. Principal factors, on the other hand, acknowledge that part of a hedge fund's return variation is attributable to a unique component and seek to extract the communalities as defined by the covariance instead, thus reducing the number of extracted (common) factors.

2. In the broken-stick methodology, additional extracted factors are discarded if their inclusion does not significantly improve the explanatory power of the model. One shortcoming of this approach results from discarded factors being jointly significant. Consequently, the dimensionality reduction comes at the cost of the lack of representativeness. This research seeks to explain the common variance in hedge funds returns. This is achieved by a much smaller number of factors without having to discard potentially significant factors.
3. In an iterative framework using factor rotation, comparing the number of extracted factors to the number of simulated factors (parallel analysis) at every step, the resulting strategic clusters of hedge funds are comprised of single manager funds loading on a single common factor only. The resulting common factor portfolios are homogenous representations of distinct hedge fund classifications.
4. Much of the past research into hedge funds and the application of factor models relied on short track records. This posed problems with respect to cases where the number of hedge funds in the sample, and consequently the maximum number of extracted principal components, exceeded the number of return observations. The asymptotic properties of estimators from PCA may not hold where the number of factors is larger than the number of observations ($k > n$). Here, including 120 observations increased the confidence in establishing the long-term communalities between single manager hedge funds across periods of financial distress (LTCM, internet bubble bust, subprime lending crisis) and subsequent economic recovery. Thus, the results were expected to be robust in

the presence of phase-locking behaviour and style drift (see sections 3.3.4 and 3.4.1). By including observations up to June 2010, the impact of the world financial crisis was discernible.

5. Factor models need to account for the statistical properties of hedge fund return series, specifically non-normality of the frequency distribution and serial correlation. The upshot of using factor axis was that the abnormal return component of hedge funds was subsumed in the unique component of the model. In consequence, non-normality was less of a concern in comparison to other approaches requiring explicit assumptions about the frequency distribution (e.g. maximum likelihood). However, serial correlation in consecutive returns could still have distorted the initial covariance estimates for samples of single manager funds. Using residuals of autoregressive moving-average models eliminated the problems from serial dependence in estimation of the covariance matrix. Most of the autocorrelation at lags one to three months, and most of the autocorrelation at higher lags, was removed prior to factor modelling.
6. The factor replicating portfolios in Fung and Hsieh (1997a) and Amenc and Martellini (2003) were composed using an optimisation algorithm. Factor replicating portfolios seek to maximise the correlation with extracted components by adjusting the portfolio weights of hedge funds within factor portfolios. Consequently, the factor portfolios are near perfect representations of the principal components themselves. However, due in part to short-sale constraints, a large proportion of sample hedge funds' performance will have no bearing on the performance of the factor replicating portfolio. Factor rotation introduced in section 4.3.5 on the other hand, differentiates between positive as well as negative factor loadings. As in Fung and Hsieh (1997a), the optimisation in section 6.3 chose the portfolio weights of index constituents to maximise correlation with the extracted common factor. The portfolio weights were chosen based on non-linear optimisation and adjusted every 12 months.
7. Index construction from CFPs was unbiased (composition and constituent weighting are discussed in detail in section 3.4.4). The indices were built from hedge funds reporting to two databases (instead of one), they include defunct

and defaulted funds and the indices were weighted / continuously re-balanced to provide an unbiased estimate of the performance of a particular investment style at any given time.

8. Rolling-window analysis between July 1995 and June 2010 allowed for testing of the robustness of the factor portfolio composition. Funds from the HFR graveyard database were included to account for survivorship bias. It is shown that ten classifications suffice to describe the prevailing investment strategies between 1995 and 2010.

The classification and clustering methodology improved upon the prevalent methodology used in related research for several reasons. It ackowledged the diversity and complexity of dynamic hedge fund trading strategies and it was not reliant on observable asset-based factors. It was conducted for rolling-windows to allow for adjustments over time and included both a larger sample of single manager hedge funds as well as longer time series. As a consequence, the results were robust and stress-tested across different economic cycles.

2.3 Multivariate modelling

The second part of the dissertation explains how hedge fund style indices constructed from CFP constituents were tested for their susceptibility to changes in three standard asset classes: equity, bonds and commodities. It was assumed that if the EMH held, equity, bond and commodity indices should prove to be poor proxies for the performance of hedge fund indices (i.e. hedge fund index levels are independent of lagged observations of the other three series). Some research suggests that a lead-lag relationship exists between the return on broad asset indices and hedge fund returns due to illiquidity and managed prices and that the inclusion of lagged asset factors improves the predictive power of the asset-class factor models. This research suggests that a multivariate structural framework best accounts for the potential relationships between hedge funds and other assets. In turn, the benefits of the approach advocated are described below:

1. In vector models, the number of required lagged coefficients is formally addressed using multivariate information criteria. No significant relationship between variables is ignored. Although some *a priori* expectations exist with respect to the scale and direction of the relationships, in a balanced model all possible associations are considered. In a VECM of the general form, no endogeneity or exogeneity specification is required since all variables are endogenous. However, due to the number of required coefficients to be estimated and the rapidly decreasing degrees of freedom, the maximum number of variables in the model is restricted (limited for each estimated model to one hedge fund index series and three asset indices). It is argued that the benefits from correctly specifying the (lagged) relationship between variables outweigh the cost of excluding additional asset classes as factors.
2. One aim of the multivariate framework was to establish the portfolio diversification benefits of hedge funds. If the claims of hedge fund managers held, all lagged coefficients for hedge funds would be insignificant. Two phenomena can be differentiated between: significant coefficients for the lagged dependent variable indicate a tendency of past performance linking with present and future performance, suggesting that the alternative investment market is illiquid and inefficient. Lagged predictors from the other three asset classes imply that publicly available information is not priced in a timely manner and that hedge funds perform poorly in the context of portfolio diversification.
3. It is postulated that alternative investments are not completely independent from other investment markets. More to the point, at certain times this relationship might be reversed (see for example the impact of George Soros's Quantum fund on international currency markets in 1992 and 1997). In a simultaneous equation framework, the direction of that relationship is not dictated. Formal tests such as Granger causality, impulse response and variance decomposition allow for a systematic assessment of the relationship between hedge funds and traditional investments.
4. In a VECM, using the levels of the time series rather than the returns, it is possible to differentiate between the short-run relationship, long-term relationship

and speed of adjustment back to equilibrium for all variables. Accounting for the cointegrating relationship, ordinary least Squares (OLS) and standard procedures for statistical inference are valid in the presence of trend-deterministic processes. The structural model is useful in identifying breaks in the relationship of model variables that result from disruptive events impacting on different markets simultaneously.

No existing research could be identified applying a VECM framework to pure style index representations of hedge fund classifications in the context of a traditional portfolio consisting of equity, bond and commodity investments. The results from Chapter 7 shed light on the implications for the underlying assumptions of the EMH and provided some guidelines as to the diversification benefits of different hedge fund styles. Whilst the results confirmed some of the findings of earlier research, the VECM more accurately described the interdependencies between hedge funds and other forms of investments.

In summary, the contribution to science is two-fold: firstly, the creation of unbiased and representative hedge funds indices and, secondly, quantifying the diversification benefits and efficiency of hedge fund markets in a multivariate framework. It is expected that the results can be easily replicated for other databases such as TASS and CISDM/MAR.

CHAPTER 3: LITERATURE REVIEW

3.1 Introduction

This chapter comprises of the following: In section 3.2, a brief review of the application of asset pricing and statistical factor models to hedge funds is given. Next, section 3.3 discusses some of the unique statistical properties of hedge fund investments including non-normal return distributions, autocorrelation and phase-locking behaviour. Examples of univariate and multivariate autoregressive models in hedge fund return series are discussed. Subsequently, section 3.4 seeks to establish the differences between single manager funds, managed futures and funds of hedge funds (FoHFs). A brief introduction to research into hedge fund indexing is provided. Lastly, section 3.5 explains various data bias effects including survivorship, database selection and instant history bias. The literature review provides the rationale for the data sourcing and treatment as well as the applied methodologies of Chapter 4.

3.2 Factor models

Recent years have seen a surge in factor models trying to explain the returns in hedge fund investments, inspired by similar research into mutual funds (e.g. Sharpe, 1992). Much of the initial research is focused on linear factor models, while more sophisticated models account for the nonlinearities in style factors and asset markets. In the following, differentiation between existing research into asset-based factor models and statistical factor models is made.

3.2.1 Hedge funds and CAPM

Some early research dealt with the application of the Capital Asset Pricing Model (CAPM; Jensen, 1968) to hedge funds. Brown, Goetzmann and Ibbotson (1999) examined the performance of offshore hedge funds and found high Jensen's alphas consistent with positive risk-adjusted performance. They addressed the limitations of selecting a broad equity index as benchmark for performance in market neutral funds and acknowledged that hedge funds actively shift their exposure to risk factors.

Ackermann, McEnally and Ravenscraft (1999) used single-factor models to estimate exposure to risk factors for hedge fund samples from the HFR and MAR databases[1]. However, they pointed out that low systematic risk claims of hedge funds are problematic in the context of simple pricing models. Agarwal and Naik (2000b) estimated hedge fund returns from a single-factor model and used regression alphas as well as appraisal ratios to determine persistence in hedge fund performance.

Other examples include Gregoriou and Rouah (2001), who determined hedge fund top performers by estimating excess returns over the S&P 500 and MSCI World indices. Capocci and Hübner (2004) also considered a single-factor CAPM, amongst other multifactor models, to estimate Jensen's alphas for hedge funds in the MAR and HFR databases. The explanatory power of the model was found to be particularly low for market neutral funds. Applying the CAPM to hedge funds, Chen and Passow (2003) found that most funds exhibit positive and significant alphas. The results have been confirmed for extended multifactor models. Single-factor regression results for hedge funds often have a low explanatory power and, despite their ease of computation and interpretation, have given way to multifactor models accounting for various sources of risk exposure. In the following sections, it was differentiated between asset-based and statistical multifactor models.

Since there is no clear agreement on the terminology in literature to differentiate between multifactor and multivariate models, definitions are offered here. Multifactor models are defined as estimation models regressing a financial time series against a number of asset-based factors including passive indices and simulated portfolios representative of particular investment strategies. Multivariate models, on the other hand, consist of a set of structural equations that estimate the dependent variables simultaneously. The outcome of the variables is determined by looking at their own past

---

[1] Note that the Managed Account Reports database was named Center for International Securities and Derivatives Markets / Managed Account Reports (CISDM/ MAR) in 2001.

as well as the past values of some (in the presence of strictly exogenous variables) or all other model variables and an error term.

3.2.2 Asset-based factor models

In the literature, there is little agreement over the number of relevant factors required to accurately describe hedge fund performance. Goodworth and Jones (2007), for example, stated that a satisfactory level of representativeness for broad-based FoHFs may be achieved considering up to 100 factors descriptive of various equity, bond, commodity and foreign exchange markets. The dimensionality of the model may be reduced by removing co-linear factors in a stepwise process. However, for the regression model, the intercept can be overstated when the model is not correctly specified with respect to systematic risk exposure. For all underspecified models there is the risk of beta being disguised as managerial skill or a hedge fund's alpha (e.g. Jaeger & Wagner, 2005: 11; Kat & Miffre, 2008).

Fung and Hsieh (1997a) pioneered the development of factor models for hedge funds. As in a standard Arbitrage Pricing Theory (APT) framework according to Ross (1976), they assumed that a limited number of observable factors explain a significant proportion of the variation in hedge fund returns, where the error term of the regression function denotes idiosyncratic risk. The asset classes used in explaining hedge fund performance include three broad stock indices, two bond-market proxies, a trade-weighted dollar index and the gold price. Goodness-of-fit for regression (R^2) is found to be below 25 percent for nearly half the hedge funds and managed futures in the sample.

Schneeweis and Spurgin (1998) increased the number of factors by including a commodity index as well as intra-month volatility indices to account for hedge funds and managed futures taking up long and short positions. They establshed that hedge funds and managed futures derive different sources of return and systemic risk compared to mutual funds: the explanatory power of models including intra-month volatility confirmed the exposure of managed futures to intramonth movements (replacing the passive equity index with the intra-month volatility produced comparable model goodness-of-fit).

The results are similar for the inclusion of a futures-based commodity proxy in the estimation of hedge fund returns.

In a similar vein, Agarwal and Naik (1999) applied an asset class factor model to hedge funds in the spirit of the Sharpe (1992) 12-factor model and its application to mutual funds. They imposed the sum-of-coefficients constraint to ease the interpretation of factor loadings as portfolio weights. Imposing no constraints may be referred to as 'weak style analysis' and imposing the sum-of-coefficient constraint as 'semi-strong style analysis'. Imposing both the sum-of-coefficients constraint as well as non-negativity constraint constitutes 'strong style analysis' (Horst, Nijmen & Roon, 2004). Since hedge funds use shorting techniques to limit their exposure, the non-negativity constraint is often relaxed when generalised style models are applied to hedge funds. To limit multicollinearity between the regressors, they employed a stepwise regression algorithm.

Similarly, Liang (1999) employed stepwise regression to identify factor loadings on equity, fixed income, commodity and cash proxies. Edwards and Caglayan (2001) employed a six-factor model including the Fama and French (1992) High-minus-Low (HML) and Small-Minus-Big (SMB) portfolios, the Carhart (1997) Winners-minus-Losers (WML) portfolio, as well as a yield curve proxy to determine hedge fund alphas.

Analogous approaches to estimate hedge fund risk factors include: Boyson (2003) on multifactor models using standard asset indices, HML and SMB portfolios and a momentum factor; Teo, Koh and Koh (2003) explaining returns in Asian hedge funds replacing US Equity and Bond proxies with regional indices; Harri and Brorsen (2004) and Hasanhodzic and Lo (2007) on linear 6-factor models based on broad asset indices; Capocci, Corhay and Hübner (2005) combining the factors from previous research including Agarwal and Naik (2004); Ammann and Moerth (2008b) working on asset class factor models for FoHFs; Eling (2009) comparing several factor models including CAPM and the Fama-French / Momentum extension; Eling and Faust (2010) constructing asset-class factor models for emerging markets hedge funds using various equity and bond proxies.

Using look-back straddles on a number of standard asset indices, Fung and Hsieh (2002a) showed that primitive trend-following strategies (PTFS) can explain the returns in trend-following hedge funds. The PTFS subsumed the non-linear relationship between the hedge fund style factors and the markets in which hedge funds trade. In a similar approach, to account for the nonlinearities in the relationship between hedge funds and risk factors, Agarwal and Naik (2004) extended their original model by incorporating option-based risk factors. Other risk factors included the Fama-French SML and HML factors, the Carhart momentum factor, as well as a commodity proxy. The R^2 varied between 44 percent for the HFR Event Arbitrage index and 92 percent for the Equity Non-Hedge index. However, some market neutral strategies like Fixed Income Arbitrage or Equity Market Neutral were not represented. Related research indicated that the inclusion of option-based risk factors did not significantly improve upon the results for market neutral strategies (e.g., Fung & Hsieh, 2001: on the risk in fixed-income based hedge fund styles).

An extension of asset-class factor modelling is observed in models including asset-based style (ABS) factors as described in Fung and Hsieh (2003, 2011). The four equity ABS factors included the S&P 500, and emerging market index as well as Small Cap – Large Cap stock and Value- Growth stock proxies. The proxies for fixed income hedge funds included various yield curve spreads.[2] The risk factors for hedge funds depended on the prevailing underlying strategy: Directional, event driven, market neutral / relative value. ABS factors aid investors in identifying (portable) alphas adjusted for systematic style risks (see also Fung & Hsieh, 2004a).

Extensive research has been conducted with respect to factor models considering the option-like payoff of hedge fund investments. Mitchell and Pulvino (2001) found that the return profile of risk arbitrage funds correlates with that of selling uncovered index put options. Kouwenberg (2003) accounted for nonlinearities by considering the exposure of

[2] For a detailed description of the ABS factors refer to section 4.4.2.

hedge funds to two option strategy portfolios, the first one selling one-month put options and the second portfolio selling one-month call options on the S&P 500. A similar approach was used by Jaeger and Wagner (2005) employing the Chicago Board of Trade's BuyWrite Monthly Index (BXM), which mimics a covered-call-writing strategy using the S&P 500. An updated application of the PTFSs described in Fung and Hsieh (2001) can be found in Kosowski, Naik and Teo (2005). The hypothetic hedge funds used in Levchenkov, Coleman and Li (2009) to compare various approaches to hedge fund return modelling are option based market timing dynamic strategies. Some researchers suggested using non-linear asset pricing models to account for hedge funds' exposure to higher moments of market indices (e.g. Ranaldo & Favre, 2005; Ding & Shawky, 2007).

Aragon (2007) argued that an *ex ante* estimation of an appropriate model to describe the systematic risk of hedge funds may be difficult to achieve. Four models were considered for hedge funds of the HFN database: A lagged market model including contemporaneous as well as lagged terms for the value-weighted market index to account for illiquidity; a broad market model including passive equity, fixed income and commodity benchmarks; an option model accounting for the dynamic market risk exposure as presented in Fung and Hsieh (2001) and Agarwal and Naik (2004); a Fama-French four factor model including a momentum factor and a market index. The quality of the regression models was found to be comparable across all four models.

3.2.3 Statistical factor models

Whilst analysing the factor exposure of hedge fund indices to option strategies has somewhat improved the predictive quality of asset factor models, it still proves difficult to estimate relevant factor loadings for single manager funds employing more specific strategies. Some research has been conducted using statistical factor models and, more specifically, principal component analysis (PCA). The result is a number of unobservable orthogonal factors explaining a significant proportion of the variation in hedge fund time series.

The first to use PCA in factor modelling for single manager funds were Fung and Hsieh (1997a).They identified five principal components which account for 43 percent of the variation in 409 hedge fund return series. These five components were found to be representative of five distinct investment styles. While they acknowledged that five components may not suffice to cover the broad universe of hedge fund investments, they allowed for differentiation of a hedge fund manager's 'location choice' (where to invest) and the 'trading strategy' (how to invest) component. Although the initial style factors bear no economic interpretation, regression against asset indices revealed systematic exposure of the style factors to different asset classes and, hence, allowed for labelling of the distinct strategies. Fung and Hsieh (2002a, 2004b) provided extensions of their initial research using a larger dataset as well as a broader standardized framework of the asset-based style factors identified to account for multi-strategy hedge funds in particular. The asset-based style factors were compared against conventional models including asset-class indices (Fung & Hsieh, 2004b)

General Style Classification estimated hedge fund styles from *ex post* returns on individual funds (Brown & Goetzmann, 2003). Statistical groupings of hedge funds were identified by minimisation of a within-group sum-of-squares criterion. Contrary to asset-class-factor models, the benefit of this type of classification is that homogenous groups of hedge funds could result from either similar asset composition or from dynamic portfolio strategies. A generalised least square procedure allows for time-varying and fund-specific residual return variance. For individual funds, Brown and Goetzmann (2003) found that eight general style classifications do a better job of predicting the variability in subsequent returns than the 17 categories of the TASS database. Style differences accounted for approximately 20 percent of the cross-sectional hedge fund return dispersion.

Barès, Gibson and Gyger (2003) also relied on principal components to determine factors driving hedge fund performance. However, they refrained from adding any economic interpretation to the factors. The estimation of principal components is in line with related research: estimation of eigenvalues and corresponding eigenvectors from the correlation matrix of the standardised hedge fund return series and subsequent calculation of principal components as a function of the original series and the eigenvector matrix. Rather than using five factors as in Fung and Hsieh (1997a), Barès *et al.* (2003) compromised on eight factors explaining 60 percent of the variations in returns.

Christiansen, Madsen and Christensen (2004) extracted five principal components from hedge fund return series of the CISDM/MAR database in a similar way to Fung and Hsieh (1997a), allowing for a statistical classification of the dominating components in various strategies. They regressed the five components against broad market indices and passive option strategies (compare to Agarwal & Naik, 2000a). For 185 funds with a continuous track record of 37 months, Christiansen *et al*. (2004) found that the first five components explained more than 60 percent of the total variance. Although the first component explained a larger proportion of the total variance in comparison to the Fung and Hsieh (1997a) research, this was mainly due to a smaller sample of funds.

Amenc and Martellini (2001, 2003) considered principal components in the context of constructing equally weighted portfolios of competing hedge fund indices. PCA was used to extract the best possible one-dimensional representation of competing indices. The portfolio weights were chosen so as to capture the largest possible fraction of information contained in the original index series. The first principal component captured a large proportion of the cross-sectional variation considering that competing indices are at least somewhat positively correlated. The resulting index had a higher degree of representation of hedge fund performance than any individual index (Lhabitant, 2004). In a similar approach, Goltz, Martellini and Vaissié (2007) constructed factor replicating portfolios from a small number of individual hedge funds. They extracted the first k principal components and formed style portfolios from hedge funds that were highly correlated with the kth principal component. The resulting continuously rebalanced portfolios could be thought of as investable pure style indices.

Kugler *et al.* (2010) tested the consistency of style classifications across database providers. Using PCA, they identified considerable heterogeneity of index returns within the same classification. They analysed the series of 78 hedge fund indices pooled from seven different index providers. Indices for the same classification from different providers showed similar loading for the first five components, indicating homogeneous style characteristics across different providers. However, the cumulative proportion of variance explained by the first five components was below 80 percent. Additionally, no explanation was given to the somewhat arbitrary attribution of single manager funds to style classifications. Considering that common classifications of hedge funds were used in the analysis, single manager funds could be expected to report under the same classification across database providers. Thus, certain homogeneity in index return series was expected.

In this section it was differentiated between the following factor models: simplistic factor models based on a single benchmark in the spirit of the CAPM, multifactor extensions of the CAPM in the form of asset-based factor models and statistical factor models explaining the variation in hedge fund returns based on a number of unobservable factors. In summary, the research presented supports the notion that asset-based factor models explain only a small proportion of the overall variation in hedge fund returns. Conversely, statistical factor models improve upon the overall explanatory power but forego economic interpretability of the coefficient estimates. However, factor extraction allowed for differentiation between a hedge fund manager's location choice on the one hand, and trading style choice on the other hand. Interpretation of results from existing research into economic models for hedge funds suffers from the same limitations: the unique properties of the time series of hedge fund returns are not adequately addressed. The literature review continues by outlining the research pertaining to hedge fund time series properties and the consequences for the estimation of linear models.

3.3 Statistical properties of time series

The following section addresses existing research into some common traits of hedge fund time series and their implications for modelling hedge fund returns. In particular, the following themes are discussed in detail: non-normality of the return distribution, autocorrelation of the returns (and the consideration of lagged factors in univariate and multivariate time-dependent models), as well as phase-locking behaviour and synchronisation of hedge fund returns.

3.3.1 Non-normality

Substantial research was conducted into the applicability of mean-variance analysis in hedge fund assessment. In particular, hedge fund return distributions were found to exhibit significant deviations from a normal distribution as measured by the third and fourth moment.[3] For the HFR database, Agarwal and Naik (2000c) demonstrated significant skewness and kurtosis. Kat and Miffre (2008) commented on the overstatement of hedge fund alphas and the risks from non-normality of the return distribution. Similarly, Chan *et al.* (2006) established higher moments and nonlinear risk exposure for classifications of the TASS database.

Analysing hedge fund data from TASS and HFN for the July 2005 to September 2010 timeframe, Anand, Maier, Kutsarov and Storr (2011) found that the probability of delivering absolute returns as expressed by skewness of the t-distribution of returns differed greatly depending on the time regimes considered (i.e. pre-crisis, crisis, post-crisis). Positively skewed returns were observed for all hedge fund strategies in the pre-crisis and post-crisis state. However, during the crisis the return distribution became negatively skewed due to greatly amplified returns. Similarly, the standard deviation of returns increased significantly during the crisis and did not revert to its pre-crisis levels during the recovery regime. Relative value hedge funds offered the best risk-return ratio during the crisis.

[3] Skewness is a measure of asymmetry of the return distribution function. A right-skewed or right-tailed function indicates that the mass of the distribution is concentrated on the left of the figure, a left-skewed or left-tailed function that the distribution is clustered on the right (the heavy left-tailed function is associated with an increased likelihood of observing extreme losses). Kurtosis is the relative 'peakedness' or flatness of the distribution function in relation to the normal distribution function. Skewness and kurtosis are the third and fourth standardized moment of the distribution function.

Four potential concerns have been addressed in the literature. Firstly, deviation from normal distributions in hedge fund returns leads to non-linear risk factor exposure. Consequently, the normality and constant variance conditions for the error variable in OLS are violated ($\varepsilon \sim N(0, \sigma)$). Secondly, investors may place a price tag on higher moments of the distribution (i.e. investors favour negative skewness and platykurtic distributions). Thus, mean-variance analysis fails to account for the investor's utility function. Thirdly, the assumption of asymptotic normality in maximum likelihood estimates may be violated. Lastly, the estimation of shortfall risk in a Value-at-Risk framework relies upon normality.

Fung and Hsieh (1999) argued that most hedge funds employ dynamic trading approaches, whereas mutual funds often follow a static buy-and-hold strategy. They reasoned that trend-following hedge funds and CTA return distributions are bound to exhibit fat tails not captured by a stationary distribution. Additionally, the asymmetric performance fees that managers receive resemble an embedded put option on fund performance. Thus, the distribution of returns of hedge funds is fundamentally different from a normal distribution.

A number of practitioners have tried to improve the reliability, accuracy, and appropriateness of pricing models by incorporating the unique pay-out structure of hedge funds. For the HFR database, Agarwal and Naik (2000c) established negative skewness and significant kurtosis for all hedge fund classifications but Short Bias and Global Macro. They argued that the presence of fat tails in the return distribution results in non-linear option-like exposure that may be accounted for with simple option writing/buying strategies. Fung and Hsieh (2002a) modelled PTFS portfolios using lookback straddles on five asset classes to account for the non-linearities in trend-following hedge funds in the TASS database. Huber and Kaiser (2004) also advocated using option-like structures in regression models for Standard & Poor's hedge fund indices. They differentiated between option strategies for CTAs (Fung & Hsieh, 1997a: long lookback straddle), fixed income hedge funds (Fung & Hsieh, 2002c: short straddle), and merger arbitrage funds (Mitchell & Pulvino, 2001: shorting uncovered index put options).

A logical extension of the mean-variance framework is to consider a quadratic or cubic utility function accounting for the third and fourth co-moment of the distribution.[4] Ranaldo and Favre (2005) applied a four-moment CAPM to hedge funds in the HFR database. They argued that non-normality in returns (skewness and excess kurtosis) may be due to dynamic trading strategies, illiquidity, lack of divisibility, and low information transparency, all of which applies to hedge funds.

Alternatively, Chung, Johnson and Schill (2006) considered Fama-French factors as additional risk factors in their pricing model. They observed high correlations between higher order co-moments and Fama-French factor loadings. Regressing SMB and HML on second- to tenth-order co-moment estimates revealed highly significant goodness-of-fit statistics (R^2 between 81% and 93%). The explanatory power of co-moments seems to suggest that the Fama-French factors merely proxy for higher systematic moments. Empirical tests seemed to confirm this notion: for monthly returns, factors SMB and HML factors were significant only when few higher-order co-moments were included in the regression. In addition, the tests showed that systematic co-moments reduced the significance of Fama-French factors, whereas standard moments did not.

For the impact of non-normality in the context of Value-at-Risk estimation, one may refer to Favre and Galeano (2002), Kooli, Amvella and Gueyié (2005) or Liang and Park (2007) on the use of the Cornish-Fisher approximation (Cornish & Fisher, 1938) in Value-at-Risk estimation, Füss, Kaiser and Adams (2007) on $GARCH$-type Value-at-Risk and Goodworth and Jones (2007) on factor decomposition and non-parametric Value-at-Risk estimation using Monte Carlo simulation.

[4] In this context the co-moments are defined as relative skewness (co-skewness) / peakedness (co-kurtosis) of hedge fund returns in comparison to a standard asset index.

3.3.2 Serial correlation and autoregression

Extensive research has been conducted on hedge funds to identify autocorrelation and to quantify the effects of serial correlation in time series.[5] For example, Asness, Krail and Liew (2001) found significant exposure of hedge fund returns to lagged market betas. They argued that stale or managed prices may prevent hedge funds that are closely correlated with the market from moving in tandem with their benchmark in the same month. Thus, the hedge funds' true directional exposure to broad markets is disguised by lagged reported returns.

Lo (2001) argued that autocorrelation coefficients are a good proxy for illiquidity in hedge funds. Because hedge funds engage in illiquid, non-publicly traded securities and make use of extensive leverage, they are exposed to considerable liquidity risks. In efficient markets, it is assumed that price changes cannot be forecast and follow a random walk. In reality, however, market frictions such as transaction costs, borrowing constraints, and short sales contribute to the possibility of serial correlation that cannot be arbitraged away. Hence, correlation coefficients can be regarded as an estimator of market frictions, of which illiquidity is one of the most influential. Lo's (2001) research showed that eight out of 12 hedge fund strategies displayed significant autocorrelation coefficients up to the 6th month lag.

Hayes (2006) showed that illiquidity in hedge funds has serious implications for the estimation of downside risk. He advocated a maximum-drawdown-at-risk framework to account for illiquidity in alternative investment funds. Estimating the impact from systematic liquidity risk, Brandon and Wang (2013) showed that the performance of equity-driven hedge funds (in particular Event Driven, Long/Short Equity and Emerging

---

[5] Serial correlation, also referred to as autocorrelation, is defined as correlation of a time series with itself and the lag specification denotes the time intervals considered. For financial time series, the autocorrelation coefficient may be regarded as an estimator of the ability to predict future prices from past observations,

Markets) is substantially reduced when illiquidity is incorporated into the hedge fund performance evaluation framework of Hasanhodzic and Lo (2007).

Kat and Lu (2002) found evidence of both non-normality as well as autocorrelation of hedge funds reporting to TASS. Skewness, excess kurtosis and first-order autocorrelation were found to be significant across all style classifications, whilst merger arbitrage, distressed securities, convertible arbitrage and emerging markets exhibited the highest, positive autocorrelation coefficients. Kat and Lu (2002) suggested using the Geltner (1991, 1993) ‘unsmoothing’ technique to remove first-order autocorrelation from the observed time series. Unsmoothing the return series causes significant increases in return standard deviations in direct proportion to the degree of autocorrelation present.

Okunev and White (2002) found those first-order autocorrelation coefficients are systematically significant in alternative strategies (higher order coefficients are only significant for convertible arbitrage and fixed income strategies). Amenc, Malaise, Martellini and Vaissié (2004) recommended the Herfindahl index or, alternatively, Ljung-Box statistic (Ljung & Box, 1978) to determine the significance of cumulative autocorrelation coefficients and, thus, the liquidity risk in hedge funds.

Okunev and White (2002) expanded on the simple method developed by Geltner (1991, 1993) to correct time series for autocorrelation of the k^{th} order. The return at time t is assumed to be the result of a linear combination of the real return recorded at t and the return observed at $t-1$. In an iterative process, one can account for autocorrelation at higher lags by identifying the remaining autocorrelation after adjusting the time series for serial correlation at the preceding lag.

The magnitude of price distortions owed to illiquidity of assets, stale prices and performance-smoothing is not identical across different hedge fund classifications. Loudon, John, Okunev and White (2006) found significant first and second order autocorrelation coefficients for both high-yield as well as mortgage-backed securities hedge fund indices. Diversified hedge funds, on the other hand, showed no evidence of serial correlation. The authors recommended eliminating the first two autocorrelations in an adjusted time series. They found that risk, defined as return standard deviation, increases between 37 and 64 percent for adjusted returns. Further evidence of conditional return-smoothing can be found in Bollen and Pool (2009): the return smoothing process is conditional on the current performance of the fund (i.e. performance smoothing occurs during periods of large negative returns).

The presence of autocorrelation has some implications for the modelling of hedge fund returns and the use of autoregression. According to Miura, Aoki and Yokouchi (2009), monthly individual hedge fund returns cannot be treated as independent and identically distributed observations (i.i.d.). Rather, current observations are dependent on lagged return observations plus an error term and are expressed in an autoregressive model of order p ($AR(p)$). Using rolling-window observations, Miura *et al.* (2009) inferred that for most hedge funds, returns are autoregressive at times, but not autoregressive in all sub-periods. In addition, risk-adjusted returns are higher for some hedge funds series described by an $AR(p \geq 1)$ process, for the long-short equity and managed futures classification in particular.

Bollen and Whaley (2009) confirmed that 30 percent of single manager funds and 37.7 percent of FoHFs feature significantly positive coefficients in an $AR(1)$ process. They identified three sources of high autocorrelation: trading in illiquid assets and lagged response times to system shocks, deliberately inflated Sharpe ratios (performance smoothing), and performance measurement bias at the single manager level. For CTAs, no such evidence of illiquidity or performance manipulation could be discerned. Bollen and Pool (2008) addressed the discontinuity in hedge fund return distributions. They found that discontinuities did not exist during the three months preceding the audit of a hedge fund and also did not prevail in the subset of funds trading in highly liquid securities (e.g. CTAs). This suggests that hedge fund returns may be temporarily overstated.

Getmansky, Lo and Makarov (2004) considered a moving average process ($MA(q)$) to describe hedge fund returns as a linear combination of white noise processes, where the sum of the MA coeffcients is equal to one (i.e. smoothing takes place over only the most recent $q+1$ observations).[6] They found the $MA(2)$ specification to be a reasonable specification for hedge fund returns (i.e. the coefficient estimates are significant for all 908 funds of the TASS subsample bar one).

Autoregressive Moving Average models ($ARMA(p,q)$) are a natural extension when the current return observation depends linearly on both previous return observations as well as a combination of current and previous values of a white noise error term. Implementation of an $ARIMA(p,1,q)$ model for the HFR index series may be found in Lòpez de Prado and Peijan (2004).[7] In a stepwise procedure, the relevant factors are selected up to an $ARMA(3,3)$ process. No $ARMA$ time-dependence can be identified for equity market neutral funds, market timers, short sellers and managed futures.

Non-linear models to account for conditional variances were explored in Blazsek and Downarowicz (2008): The conditional mean of hedge funds was modelled as an $ARMA$ process, while the conditional volatility was specified as a generalised autoregressive conditionally heteroskedastic model ($GARCH$). A Markov switching model (MS) accounted for regime switches. From the empirical results it can be inferred that different combinations of AR, MA, $GARCH$ and MS may be required to forecast the time series of different HFR indices. The $ARMA-GARCH$ implementation was repeated by

---

[6] Note that as long as a $MA(q)$ process is invertible, it can be expressed as an $AR(\infty)$ process. Getmansky *et al.* (2004: 555) pointed out that maximum likelihood estimates need not yield an invertible $MA(q)$ process. However, they imposed additional restrictions on the parameter estimates so that the estimated process would be invertible.

[7] Autoregressive Integrated Moving Average models can be thought of as an adaptation for integrated autoregressive processes (e.g. the characteristic equation of the process has a unit root). An $ARMA(p,q)$ model in the differenced variable (d times) is equivalent to an $ARIMA(p,d,q)$ model on the original data (Brooks, 2008: 233).

Giamouridis and Ntoula (2009) for the HFR daily index return database. The results for the specified $ARMA(p,q) - GARCH(r,s)$ provide some evidence against the i.i.d. hypothesis, in particular for macro, relative value arbitrage, convertible arbitrage, and merger arbitrage classifications.

In the following, the analysis was limited to linear models. It is the opinion of the author that modelling conditional variance for hedge fund returns may be misleading in a forecasting framework that does not consider potential spill-over effects between markets or hedge funds. Multivariate extensions of the $GARCH$ considering both individual variances as well as covariances between time series may prove to be promising extensions of the research into univariate models.

3.3.3 Multivariate time series models

Amenc, El Bied and Martellini (2003) estimated hedge fund returns from a number of return series of equity and bond indices, Treasury bill rates, as well as commodity proxies. They postulated that moving averages for the S&P 500 and MSCI World index, lagged parameters for the oil price, and changes in the Treasury bill rate could induce changes in hedge fund returns. From the coefficient estimates, Amenc, Martellini and Vaissié (2003a) estimated portfolio diversification estimates of hedge fund investments. Contrary to designated multivariate Vector Autoregressive Models (VAR), no causality tests, formal multivariate statistics to decide on model scale, or autoregressive factors were used.

An application of VAR in the context of hedge fund investments can be found in Viebig and Poddig (2010) who, in order to determine whether contagion exists between equity markets and hedge funds, used VAR on return series of various hedge fund styles and equity proxies. Beltratti and Morana (2008), on the other hand, imposed a VAR structure to estimate the lagged linkages between asset flows and hedge fund returns. They found that asset flows positively depend on contemporaneous as well as lagged returns, and *vice versa*.

Whilst, to the author's knowledge, no extensive research has been conducted into assessing linkages between hedge fund investments and traditional portfolios in a VAR framework, the notion of a simultaneous equation framework applied to alternative investments and standard assets is well established. Seck (1996) provided an example from research into Real Estate Investment Trusts (REITs) addressing the substitutability of real estate assets that consider a VAR framework with several macroeconomic variables including the return on the S&P 500 index as well as the 5-year mortgage rate for existing homes. He found the stock market index and the term structure to be good price predictors of REITs and stock indices of the home building industry. In a similar vein, Liow and Yang (2005) assessed diversification benefits and substitutability of real estates and common stocks in a Vector Error Correction Model (VECM). Using the Johanson cointegration framework, they found evidence of at least one cointegrating relation establishing a long-run equilibrium between real estate investments, stock prices, GDP, inflation, money supply and short-term interest rates.

3.3.4 Phase-locking behaviour

Phase-locking can be thought of as previously uncorrelated systems becoming synchronised shifting to a collective behaviour. Phase-locking describes a link that is absent most of the time, but can be present in low-probability extreme down-markets. In terms of financial econometrics, it describes the behaviour of assets shifting correlations with broad indices in times of financial turmoil. Lo (2001) was amongst the first to characterise phase-locking behaviour in hedge funds. The default in Russian government debt and the subsequent demise of the LTCM fund in 1998 mark a period where correlations between hedge funds and standard asset indices became heavily correlated. The easiest way to account for such behaviour in financial markets is a two-factor model allowing for different states or regimes (also see Chan *et al.,* 2006 for a more recent application).

Spurgin, Martin and Schneeweis (2001) outlined a simple single-factor econometric model to determine whether hedge funds exhibit time-varying correlation. They tested both a linear as well as a quadratic regression model to estimate this change in correlation or market beta as a function of benchmark return and benchmark volatility. The changes in beta were found to be significant in all but long-short equity hedge funds (Evaluation Associates Capital Markets database) and discretion CTAs (MAR database). Spurgin *et al*. (2001) acknowledged, however, that a single factor model may not fully capture sources of return and the exposure to upside / downside volatility in hedge funds.

Schneeweis, Karavas and Georgiev (2002) estimated the risk-return relationship between a traditional stock/bond portfolio and alternative investments in periods of extreme return movements. They found wide variations in the return and correlation relationships between the alternative investment portfolios and traditional stock/bond portfolios. More specifically, in the extreme return periods of the stock/bond portfolio, previously uncorrelated alternative investments became highly correlated with traditional benchmark indices, such as the S&P 500.

One way to account for phase-locking behaviour and sudden changes in correlation is to allow for different states or regimes in pricing models. Billio, Getmansky and Pelizzon (2007) argued that beta switching regime models are able to forecast and capture the transition of the idiosyncratic risk factor in terms of changes from low volatility regimes to distressed states. Firstly, a regime with low S&P 500 expected returns and high volatility was defined as “down-market”, and a regime with high expected returns and low volatility as “up-market”. Next, the transition probabilities of moving from one market regime into another were defined. The regime switching process is described by a transition probability matrix. In Billio *et al.* (2007) the distribution of future returns depended only on the unconditional properties of the Markov chain. In a Markov switching regime, systematic and un-systematic events may change from the presence of discontinuous and random shifts in expected return and volatility.

Literature has shown that the frequency distribution of hedge fund returns is decidedly non-normal for the majority of strategies under observation. In addition, many hedge fund time series are found to be serially correlated, either as a result of investment asset illiquidity or deliberate performance management. There is some evidence that hedge fund returns are adequately described by univariate models and are linearly dependent on their own past observations and / or moving-average process ($AR(p)MA(q)$). Recent research focussed on non-linear models with hedge fund returns modelled on conditional volatility ($GARCH$). Multivariate models are the natural extensions of the existing research. VARs were used in the estimation of other alternative investment classes such as REITs. The review concluded by presenting regime-switching and conditional models that emphasize the benefits of such models in the presence of phase-locking behaviour. The following section looks at FoHFs, hedge fund indices and CTAs in comparison to single manager hedge funds and the degree to which some of the aforementioned undesirable data properties are mitigated.

3.4 Benchmarking, FoHFs and style drift

Some literature is available with respect to the various proxies and benchmarks that can be used in assessing the hedge fund universe. In the following, a differentiation was made between single manager funds, FoHFs and hedge fund indices and the merits and shortcomings of each were discussed. In addition, style drift and differences between hedge funds and managed futures were addressed as discussed in past research.

3.4.1 Style drift

Style drift can be thought of as a hedge fund's divergence from its stated investment objective or purpose.[8] Managers may decide to shift a fund's capital allocation to benefit from short-term arbitrage opportunities. Fung and Hsieh (2004b) stated that hedge funds adjust their portfolios in response to cataclysmic disruptions in the market such as the demise of the LTCM fund in 1998 and that appropriate models should allow for time-varying behaviour in regression alphas and betas. Fung and Hsieh (2004a) were able to

[8] The term 'style drift' rather than 'style shift' will be used throughout. Note that the terms are synonymous.

identify breakpoints that trigger changes in the regression regime by observing the behaviour of cumulative recursive residuals. When residuals violated set upper or lower bounds, a disruptive market event could be identified. Changes in asset allocation and in exposure to different asset classes cause hedge funds to gradually shift away from their stated strategy. Flexible regime-switching regression models increase the goodness of fit of the regression function substantially.

Baghai-Wadji and Klocker (2007) claimed that style shifts are a characteristic feature of hedge funds. They were also able to show that select single funds are able to improve upon their performance by abandoning their stated investment style. Baghai-Wadji and Klocker (2007) indicated that style shifts are beneficial to most poorly performing funds, whilst top performers rarely benefit from changing investment strategies. Frumkin and Vandegrift (2009) identified increasing fund age as a driver of style drift with fund managers seeking investment opportunities outside their area of expertise. This in turn caused hedge funds to perform poorly.

3.4.2 Difference between hedge funds and managed futures

A number of authors have explicitly differentiated between hedge funds and managed futures. Managed futures are overseen by national regulatory commissions that control the fund's adherence to local legislation. Some researchers have claimed that managed futures must be seen separately from hedge funds because of their higher liquidity and correlations with other assets (Liang, 2003a). Others have claimed that the return profiles of managed futures are increasingly correlated with the development of hedge funds (Cottier, 2000). Similar to hedge funds, managed futures have been weakly correlated with equities and treasury bonds throughout their history and thus increased the diversification of traditional portfolios (Chandler, 1994: 23–27).

Schneeweis and Spurgin (1998) are amongst the first to have used multi-factor models in the return estimation of managed futures and hedge funds. Their regression variables included the performance of broad equity indices, government and corporate bond indices, a commodity index, the return on Treasury Bills, as well as an US Dollar proxy. Additionally, the intra-month volatilities of the regression indices were used as proxies for the ability of hedge funds and managed futures to take both long and short positions. Thus, they explicitly allowed for dynamic trading strategies in alternative investment funds. Their findings show that the performance of managed futures is positively related to market trends, whereas the return of hedge funds is associated with the performance of the underlying markets and the volatility thereof (Schneeweis & Spurgin, 1998).

Edwards and Liew (1999) tried to differentiate between hedge funds and managed futures as components of a broadly diversified asset portfolio. They found that both types of alternative investments retain low correlations to other asset classes and confirmed that hedge funds as well as managed futures increase the portfolio Sharpe Ratio. Liang (2003a) on the other hand explicitly considered managed futures and hedge funds as distinct investment classes. Most notably, managed futures differ from hedge funds with respect to attrition rates, survivorship bias, liquidity and correlation structure: the empirical results suggested that whilst hedge funds outperform managed futures, the latter offer better downside risk protection in down markets with a liquidity squeeze. Due to their low correlations, hedge funds and managed futures are complementary additions to the standard asset portfolio.

Brown, Goetzmann and Park (2001) found three differences between the performance of hedge funds and managed futures. Firstly, hedge funds performing poorly on an annual basis are more likely to be terminated than their managed futures counterparts. Secondly, short-term performance and information is more relevant in the context of CTAs due to regulatory and transparency requirements. Lastly, due to high-watermarks, hedge fund managers are more likely to take on higher levels of systemic risks resulting in an overall increased default risk. Controlling for differences in relative performance as well as seasonality, it was revealed that excess volatility entails significantly higher default risks for hedge funds in comparison to managed futures.

3.4.3 FoHFs vs single manager funds

Research to date has painted a somewhat ambivalent picture of the diversification benefits of FoHF investments. For example, Brown, Goetzmann and Liang (2004) stated that most FoHFs do not earn the fees they charge. They attributed the underperformance of FoHFs in comparison to single manager funds to the fee arrangement: there is a significant probability that in highly diversified FoHFs, the incentive fees incurred at the constituent fund level absorb all of the excess return of the FoHF. Since investors do not understand the underlying hedge fund positions when investing in FoHFs, they cannot hedge against the incentive fee component. Consequently, incentive fees are associated with higher risk-adjusted performance at the individual hedge fund level only.

In the event of hurdle rates and high-water marks, the incentive fee works like a call option where the manager participates proportionately in the profits of the hedge funds but is not liable for losses (Agarwal *et al.,* 2009). With a hurdle rate, the fund manager earns an incentive fee only if the fund returns exceed the specified hurdle at the beginning of each year. With a high-water mark provision, the payment of the incentive fee is conditioned upon exceeding the previously achieved maximum. In the event that a hedge fund has incurred losses in the previous year, or has not fully recovered from past losses, the manager's options are effectively out of the money. In consequence, the incentive fee contract is a call option written by the investors on Assets-under-Management (AUM), where the strike price is determined by the net asset value at which investors enter the fund, the hurdle rate and the high-water mark provisions (Liang, 2003b).

Constructing equally weighted portfolios from single manager funds based on alpha, the Sharpe ratio and the Information ratio, Gregoriou, Hübner, Papageorgiou and Rouah (2007) showed that simple portfolios of hedge funds are able to outperform FoHFs, in particular when considering the double-fee structure. In contrast, Beckers, Curds and Weinberger (2007) acknowledged that historically FoHFs have delivered alpha, but that naïve selection of FoHFs may lead to a portfolio consisting of funds with significant factor exposure. They argued that a considerable portion of hedge fund alpha is in fact owed to directional exposure to easily replicated market factors. For the ten-year period prior to January 2005, alpha was still positive and significant when accounting for factor exposure. It can be argued whether the simplistic eight factor model employed fully encapsulated the risk exposures of FoHFs.

Ammann and Moerth (2008b) confirmed that the return difference between FoHFs and single manager funds is on a par with the additional fee load charged by FoHF managers. However, analysis of survivorship revealed a lower attrition rate for FoHFs. Using an eleven asset-class factor model to determine managerial skill in FoHFs, Ammann and Moerth (2008b) found no evidence of significant alphas. Analysing 646 FoHFs from the TASS database, Heidorn, Kaiser and Roder (2009) questioned their ability to deliver absolute returns and to provide downside protection in times of financial turmoil. However, Amo, Harasty and Hillion (2007) rejected the traditional time-series mean and standard deviation approach and stated that FoHFs do offer diversification benefits in a terminal wealth framework. Diversification benefits of FoHFs were expressed in the context of accumulated wealth at the end of the investment horizon. The accumulated wealth on baskets of randomly picked and equally weighted hedge funds simulated many times was compared to the performance of FoHFs.

By comparing the investor's utility function in two cases denoted 'no fund-of-funds' and 'with fund-of-funds', Ang, Rhodes-Kropf and Zhao (2008) argued that unskilled investors are willing to invest through FoHFs to enter a market where all other participants have superior skills to evaluate hedge funds. Thus, the correct FoHF benchmark for investors is the return the investor would achieve from direct hedge fund investment in a case without FoHFs. For the TASS database, they confirmed this notion empirically between 1992 and 2003: investors need only believe that they will earn slightly worse expected returns on their own direct investments compared to the median return of hedge funds in the TASS database (1% per annum lower) for FoHFs to add value.

Agarwal and Kale (2007) commented on the relative performance of multi-strategy funds and FoHFs. They found that multi-strategy funds of the TASS database continuously outperformed their FoHF peers by 2.6 - 4.8 percent annually between 1995 and 2004. Since FoHFs underperformed both on net- and gross-of-fee basis, the underperformance cannot be attributed to the double-fee structure, exclusively. Accounting for agency risk, liquidity and fee structure, Agarwal and Kale (2007) attributed higher performance to 'self-selection' by managers with superior ability. They argued that managers with better skill self-select into becoming multi-strategy fund managers, and that as a result, better investments accrue to investors in the form of superior returns.

3.4.4 Hedge fund indices

Amenc and Martellini (2001) opined that there is considerable heterogeneity with respect to the information conveyed in hedge fund indices from different providers. Differences between providers exist mainly in selection criteria such as minimum track record, the number of distinct style classifications, the weighting scheme (equal vs. asset-weighted) and rebalancing (e.g. monthly vs. annually). The contrasts in index construction are confirmed when observing low correlation coefficients between indices of the same style classification from different providers. Regardless of the index providers, all hedge indices suffer from the same shortcomings: firstly, style indices are never truly representative of all hedge funds following a given strategy; secondly, inherent hedge fund data biases are transferred to the index level (e.g. history-backfilling and database selection).

Brooks and Kat (2002) commented on the divergence between hedge fund indices as industry proxies and investable indices. They also emphasised that most hedge funds are calculated as straightforward population averages and are therefore non-investable (i.e. cannot be used in the portfolio allocation decision process). In addition, hedge fund indices suffer from the same undesirable statistical properties as single manager funds: non-normality of the return distribution and serial correlation. Finally, Brooks and Kat (2002) found evidence of significant survivorship bias in index returns for some providers (e.g. HFN).

Fung and Hsieh (2002b) found that hedge fund indices constructed from individual databases inherit their measurement biases: survivorship, instant-history and selection bias. Additionally, index construction may vary across providers depending on weighting schemes or the exclusion of sub-classifications of alternative investments funds (for example, the exclusion of CTAs/managed futures in HFR). FoHFs, on the other hand, were found to be less susceptible to measurement biases: The performance of single manager funds that stop reporting to a database vendor continued to be contained within the FoHF performance; similarly, instant-history bias and self-selection bias did not prevail in the returns of FoHFs.

Amenc, Martellini and Vaissié (2003b) emphasised the impact of style drift when managers shift their fund's investment focus away from the stated main strategy. Due to the lack of transparency in hedge fund allocation, it is not possible to accurately quantify the impact of style drift on hedge fund style indices. Equally-weighted indices correspond to a contrarian, value-weighted indices to a momentum strategy. For example, HFR indices are equally-weighted and re-balanced on a monthly basis. CSFB indices, on the other hand, are value-weighted. Value-weighted indices simulate a trading strategy of buying past winners, as their portfolio weight increases proportionately to their past performance, whereas equally-weighted indices correspond to contrarian trading strategies. On the re-balancing date, equally-weighted indices sell past outperformers and buy past underperformers to equalize the future performance contribution of each constituent. In the event where past sub-par performance indicates future earnings potential, re-balancing corresponds to a contrarian investment philosophy.

Depending on the weighting scheme, indices of the same classification from different providers exhibit different risk exposures in a factor model. Kat and Palaro (2006) pointed out that index replication of non-investable indices conveys further problems: a large proportion of hedge funds comprising an index are closed to investments. In addition, investors may be able to circumvent the FoHF fee structure, but are still subjected to the management fee at the individual fund level.

This dichotomy of representativeness and investability of hedge fund indices was discussed in Goltz, Martellini and Vaissié (2007). They pointed out that non-representative investable indices share common traits with FoHFs and they recommended the use of factor replicating portfolios as investable benchmarks. Kugler *et al.* (2010) considered factor analysis applied at the index level to create representative style indices.

Amenc and Goltz (2008) indicated that some shortcomings outlined for hedge funds are common to other widely accepted indices as well. To overcome the data bias in indices, Amenc and Goltz advocated the approach described in Amenc and Martellini (2001) and Goltz, Martellini and Vaissié (2007) whereby PCA is used in conjunction with factor replicating portfolios. Principal components are extracted from samples containing open as well as closed funds to maximise representativeness of the factor. Amenc and Goltz (2008) selected 10 hedge funds categorised by factor loadings and weighted so as to maximise correlation with the respective factor components. They justified the small number of funds used in index construction with the average number of index constituents used in stock benchmarks.

Schneeweis, Kazemi and Szado (2012) established significant differences in descriptive statistics comparing hedge fund indices from the CISDM, HFR and CSFB. They also differed substantially with respect to the tracking ability of an eight-factor asset model, which might have been partially attributable to a different weighting structure (asset-weighted versus equal-weighted) and re-balancing intervals. As a result, hedge fund indices are more representative of the particular database they are constructed from, than the entire universe of hedge funds. Deciding on a particular index provider carries database selection bias, which will be discussed in more detail in the following section.

It is a current focus of research to identify circumstances under which hedge fund managers depart from a fund's stated investment purpose and what the consequences of style drifting behaviour are. Studies on FoHFs and hedge fund indices showed how some of the risks of single manager hedge funds, including the risks associated with style drift, could be addressed by diversifying across many single hedge funds. However, some research suggested that risk mitigation comes with a hefty price tag due to the double-fee structure of FoHFs or investable indices. In this context, the impact from hurdle rates and incentive fees were discussed, as well as the inability of investors to hedge against such effects at the FoHF level. For hedge fund indices, the established index construction methodologies were presented and difficulties of reconciling investability and representativeness as discussed in related research were shown. The quality of the sampling process required to construct style indices and similar proxies is conditional on the avoidance of data bias effects. The research pertaining to data bias is discussed in the following section.

3.5 Hedge fund data bias effects

Bias effects are inherent to economic time series, where the pricing of assets rests with fund managers or database vendors, because assets are illiquid or there is no secondary market for trading them. In addition, hedge fund managers have no legal obligation to regularly publish performance reports on the funds they manage. Performance reporting to database vendors is voluntary, and often there is little incentive for fund managers to do so. Some of the most common bias effects in hedge fund performance data include: database selection bias, self-selection bias, survivorship bias, instant history bias, and look-ahead bias. A brief review and the available literature on hedge fund data bias is given in this section.

3.5.1 Database selection and self-selection bias

Since reporting to database vendors occurs on a voluntary basis, the sample of hedge funds observed does not constitute a true random sample of the entire population. Characteristics from reporting funds may differ widely from characteristics from non-reporting funds. Additionally, hedge fund managers may opt to report to one or two database vendors, but rarely report to all. Thus, selecting a database for statistical analysis resulted in a sample selection bias towards particular segments of hedge funds (some providers exclude certain investment strategies from their database). By comparing the data of multiple providers, the impact of selection bias could be mitigated.

Some examples from research into the TASS database and at least one complementary database: In Liang (2000) the joint sample of HFR (without managed futures) and TASS resulted in a common sample of 465 funds versus 2,324 unique funds; Fung and Hsieh (2004b) established that from 1,061 hedge funds in TASS, 1,151 in HFR and 909 in Zurich Capital, only 305 funds reported to all three databases (for all funds up to December 2000).

Some hedge funds choose not to publish their performance - either because the performance does not appear satisfactory or they have already reached their critical size (self-selection bias). It is therefore difficult to know whether this bias has a positive or negative impact on the average performances. Empirical research suggests, however, that funds that stop reporting do so because of underperformance rather than reaching the cap of their capital requirements (Grecu, Malkiel & Saha, 2007).

3.5.2 Survivorship

Survivorship bias results from the tendency of funds to be excluded from databases for the simple reason that they no longer exist. Thus, performance assessment based on surviving funds is likely to positively skew the expected performance of the average hedge fund. However, the reasons for exclusion from a database can be many: the fund has been liquidated due to financial losses, the fund has been closed (no more investors will be allowed into the fund), the fund has been merged with another fund, or the fund has simply stopped reporting for different reasons without being liquidated.

The potential survivorship bias could be addressed by looking at both reporting funds, as well as funds that stopped reporting to the database vendor. It was found that the monthly returns were overstated for most hedge fund strategies, when the graveyard funds were excluded from the analysis. Since TASS has included defunct funds in a graveyard database since 1994, their database is a popular starting point to quantify survivorship bias. However, considering funds that were dropped from the sample prior to 1994, a certain degree of remaining survivorship bias is to be expected for the following years.

Fung and Hsieh (2000: 294–297) quantified survivorship bias as the expected returns between an observable portfolio investing in all funds in a database and the surviving portfolio excluding defunct funds. For 1994 through 1998, the survivorship bias was 3.0 percent per annum (p.a.) and 1.3 percent p.a. for single manager funds and FoHFs respectively. An identical approach was employed for managed futures in the TASS database, setting survivorship bias at 3.5 percent p.a. for 1989 through 1995 (Fung & Hsieh, 1997b) and 3.6 percent for 1991 through 1997 (Fung & Hsieh, 2000). Liang

(2001) confirmed the results for single manager and FoHFs up to 1999, quantifying TASS survivorship bias at 2.4 percent annually.

Barry (2002) established that survivorship has increased in the years following 1998 due to higher attrition among managed futures and fixed income arbitrage strategies. The average performance difference between surviving and defunct hedge funds was 3.8 percent p.a. for the seven year period prior to June 2001. For the same time period, Amin and Kat (2003) found that annual survivorship bias amounted to 1.9 percent for single manager funds and 0.6 percent for FoHFs. However, Amin and Kat (2003) emphasised that survivorship bias may be much higher for small, young and leveraged funds. In addition, bias appeared persistent in the estimation of higher moments of the return distribution.

Getmansky *et al.* (2004: 75–76) expanded previous research by including a much larger observation period from November 1977 to January 2001. For hedge funds with a continuous five-year track record, the annual performance difference over the 24-year period between alive and dead funds was 4.1 percent. Considering the inherent bias of limiting the analysis to funds with a minimum return history and the 1994 TASS database cut-off for defunct funds, the results may not be directly comparable to previous research. Malkiel and Saha (2005) increased the survivorship bias to 4.4 percent p.a. for TASS from 1996 to 2003. For 1995 to 2006, Ibbotson and Chen (2006) set the return difference at 2.8 percent p.a. before accounting for backfill bias.

Whilst not quantifying the performance difference between survivors and defunct funds in the TASS database, Grecu *et al.* (2007) found that funds perform significantly worse shortly before they stop performing, suggesting that funds cease to report to database vendors due to inferior performance. More recently, Aggarwal and Jorion (2010) identified a previously unreported bias in TASS hedge fund returns due to the merger of Tremont with TASS. By subdividing the analysis in two sample periods (1994-2001 and 2002-2008), they found that the returns of the survived Tremont funds are on average 5.4 percent p.a. higher than those of the TASS pre-Tremont funds.

Survivorship estimates for other databases vary considerably, depending on compositional differences in the databases, the methodologies used in constructing survivor and defunct portfolios, the inclusion of FoHFs, and the specified timeframe. For the HFR database, Ackermann, McEnally and Ravenscraft (1999) estimated average survivorship bias at approximately 0.2 percent per month. Edwards and Caglayan (2001) approximated that the performance difference between dead and surviving funds of the MAR database was as high as 1.9 percent p.a.

A further component to survivorship bias is look-ahead bias, which stems directly from the methodology employed: An ex-post analysis of hedge fund time series may suffer from implicit survivorship bias if funds are selected on the basis of their past track record. While funds entering the sample may be derived from survivor and graveyard databases, in an *ex post* framework, the study may still be biased towards hedge funds with a minimum number of consecutive return observations. Baquero, Horst and Verbeek (2005) placed look-ahead bias at 3.8 percent at the one-year horizon for the TASS database by estimating persistence in hedge fund returns. Since attrition is higher in hedge funds than in mutual funds, look-ahead bias is of more severe consequence.

3.5.3 Attrition and survival time

Whilst not a data bias effect *per se*, attrition rates are good indicators of the average survival times and, consequently, the bias associated with selecting funds on the basis of a minimum continuous track record. According to an early study on TASS hedge funds, the annual attrition rate was 8.54 percent p.a. between 1990 and 1999 (Liang, 2001). It was also shown that attrition rates were higher in times of cataclysmic events such as the LTCM downfall of 1998 (13% of all reporting hedge funds disappeared from the TASS database).

Brown, Goetzmann and Ibbotson (1999) provided an early estimate of survival times of hedge funds in TASS using the available data and augmenting it to account for missing data on defunct funds prior to 1994. Their study put the half-life of hedge funds at 2.5 years. Differences did exist for various hedge fund styles, however, Gregoriou (2005) found mean survival time for Event Driven and Market Neutral funds to be as high as

7.25 and 6.08 years, respectively. This may be compared to the median survival time of 5.50 years for all hedge fund classifications (Gregoriou, 2002: all results for Zurich Capital Partners database between 1990 and 2001). The hazard function has been found to peak around six years for Event Driven and eight years for Market Neutral funds. In a study on micro hedge funds, Gregoriou (2006) conclusively showed that attrition rates are related to fund size as well as strategic classification. Amin and Kat (2003) showed that hedge fund attrition for FoHFs is lower than for single manager funds in TASS (1994-2001). For the 1995-2004 period, Gregoriou, Kooli and Rouah (2008) confirmed these earlier results for the HFR database.

Getmansky, Lo and Mei (2004) documented the attrition rate for hedge funds in the TASS database between 1994 and 2004. The attrition rates were found to differ significantly between style classifications: from as low as 5.2 percent for Convertible Arbitrage funds to 14.4 percent for managed futures. Getmansky *et al*. (2004) confirmed that poor performance is a main driver behind hedge funds that stopped reporting to the database. In addition, live funds exhibited higher illiquidity exposure as measured by serial correlation than dead funds. One explanation is that live funds are more successful at controlling risk and, as a result, display smoother returns.

The hazard model employed in Grecu *et al.* (2007) suggested that, in sharp contrast to the results for mutual funds, the estimated likelihood of hedge fund failure remains high even ten years after inception of the fund. The hazard rate decreases only gradually after reaching its maximum at 66 months. Thus, longer living funds do not have a radically lower probability of failing than young funds.

3.5.4 Instant history bias

Many database vendors allow newly joined funds to backfill their historical returns, even though they were not part of the database in previous months, resulting in instant history bias. This gives fund managers the incentive to delay database inclusion until the mean performance displayed by the fund during its incubation period exceeds that of funds that already belong to the database, consequently biasing the past performance upwards. The easiest way to quantify instant history bias is to compare the average returns since inception to the average returns since the fund's inclusion date.

By dropping the first twelve months of performance reporting (average incubation period) and comparing the adjusted to the acclaimed performance of hedge funds, Fung and Hsieh (2001) were able to estimate instant history bias at 1.4 percent. Considering additional data up to 2001, Barry (2002) confirmed these results for the TASS database. However, Barry emphasised that the instant history estimate reflects a bias resulting from the methodology itself: The truncation of return series removes a large proportion of returns of newly established funds, many of which are found to be long-bias equity funds outperforming other hedge fund styles.

Different starting dates for hedge funds across databases aggravate the determination of instant history bias. Liang (2000) found that only 33.1 percent of funds in both TASS and HFR start reporting their returns at the same month to both databases. For 1995 to 2004, Ibbotson and Chen (2006) attributed roughly half of the 5.7 percent performance difference between survivors and defunct funds to instant history bias. Aggarwal and Jorion (2010) pointed out that the TASS / Tremont merger may have resulted in additional instant history bias of 1.5 percent.

The results are comparable across databases (see for example Edwards & Caglayan, 2001: for the MAR database they estimated instant history bias of 1.2% per annum). By deleting the first 12, 24, 36, 48, and 60 months of the return series in a stepwise fashion, Eling (2009), based on earlier work conducted by Capocci, Corhay and Hübner (2005), identified instant history bias for MAR/CISDM of 0.18, 0.38, 0.38, 0.40, and 0.31 percent for single manager funds and 0.03, 0.02, 0.06, 0.08 and 0.08 percent for FoHFs, respectively.

The section on data bias effects emphasizes the need to address any potential short-comings of the sampling process. In particular, effects from survivorship, data selection and instant-history bias are likely to impair the reliability and robustness of the results. Research showed that survival and attrition times vary greatly between different hedge fund investment strategies. Moreover, the exclusion of a graveyard database would greatly skew the results and interpretation thereof in favour of riskier strategies with amplified total loss risks. The literature pertaining to hedge fund investments and the assessment thereof has been reproduced in the appendix in tabular form (see Table A.1). The research has been sorted chronologically as well as according to subject area.

CHAPTER 4: METHODOLOGY

4.1 Introduction

In this chapter the methodologies, models and statistical tests used throughout the remainder of the dissertation are briefly discussed. The chapter is structured as follows: firstly, some standard statistical tests for normality, autocorrelation, stationarity and co-integration are introduced. The general factor model and the derivation of a statistical factor model are subsequently discussed. Next, principal factor analysis, factor rotation and creation of factor portfolios are outlined. In conclusion, the VECM and a brief summary of the Granger causality, impulse response tests and variance decomposition are provided. There are some detailed derivations for the QR-method to be found in the appendix. For reasons of consistency, a short list of subscripts, variables and matrix notations used in this chapter is included. The most common notation throughout text books and previous research has been used. Some concessions have been made to avoid confusion.

Table 4.1: Notations and variables

Var	Description.	Constituents (vector/matrix only)
$t,(s)$	Time subscript, (alternate subscript).	
T	Number of individual return observations for each asset.	
n	The number of hedge funds in the sample.	
k	The lag subscript.	
m	The total number of factors used in the model. Note that for all common factor models $m \leq n$.	

α_i A constant representing the intercept.

f_{jt} The tth observation for the jth observed / extracted factor.

β_{ij} the loading of the ith variable (hedge fund series) on the jth factor.

r_{it} The estimated return of the ith variable at time t.

ϵ_{it} The specific error of the estimated return r_{it} at time t.

a The intercept matrix.

$$\underset{T\times n}{\mathbf{a}} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}_T \begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_n\end{bmatrix}$$

F The factor score matrix.

$$\underset{T\times m}{\mathbf{F}} = \begin{bmatrix}\boldsymbol{f}_1\\\boldsymbol{f}_2\\\vdots\\\boldsymbol{f}_T\end{bmatrix} = \begin{bmatrix}f_{11} & f_{12} & \cdots & f_{1m}\\ f_{21} & f_{22} & \cdots & f_{2m}\\ \vdots & \vdots & \ddots & \vdots\\ f_{T1} & f_{T2} & \cdots & f_{Tm}\end{bmatrix}$$

b The factor loading matrix.

$$\underset{m\times n}{\mathbf{b}} = \begin{bmatrix}\beta_{11} & \beta_{12} & \cdots & \beta_{1n}\\ \beta_{21} & \beta_{22} & \cdots & \beta_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mn}\end{bmatrix}.$$

R The return matrix for n assets.

$$\underset{T\times n}{\mathbf{R}} = \begin{bmatrix}r_{11} & r_{12} & \cdots & r_{1n}\\ r_{21} & r_{22} & \cdots & r_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ r_{T1} & r_{T2} & \cdots & r_{Tn}\end{bmatrix}$$

$\mathbf{e}$ The error estimate matrix.

$$\underset{T \times n}{\mathbf{e}} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_T \end{bmatrix}' = \begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1n} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2\mathrm{n}} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{T1} & \epsilon_{T2} & \cdots & \epsilon_{Tn} \end{bmatrix}$$

y_t The observation for a time series at time t.

x_{it} The tth observation for the ith regressor.

$\mathbf{Y}$ A vector representing a time series.

$$\underset{T \times 1}{\mathbf{Y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}$$

$\mathbf{X}$ A matrix of regressors.

$$\underset{T \times m}{\mathbf{X}} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T1} & X_{T2} & \cdots & X_{Tm} \end{bmatrix}$$

λ_j The jth eigenvalue of a nonsingular matrix.

c_j The eigenvector corresponding to the jth eigenvalue.

$\mathbf{D}$ A diagonal matrix containing k eigenvalues of a correlation matrix on the diagonals.

$$\underset{k \times k}{\mathbf{D}} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$$

$\mathbf{\Psi}$ The $n \times n$ error covariance matrix.

$$\mathbf{\Psi} = cov(e) = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_n \end{bmatrix}$$

P The $n \times n$ correlation matrix of the variables $\mathbf{P} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1\mathrm{n}} \\ r_{21} & 1 & \cdots & r_{2\mathrm{n}} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{\mathrm{n}2} & \cdots & 1 \end{bmatrix}$

Note that vectors and matrices are in bold font. Dimensions appear underneath the matrix/vector notations.

Other variables that are not listed here are explained in more detail in the context of the equations in which they appear. Note also that whenever reference is made to a single time series, the variable subscript n is dropped from the error term ϵ as well as from the coefficient estimates β, and y is used to represent a time series (in particular, section 4.2 on statistical tests).

4.2 Statistical tests

In order to account for the statistical properties of hedge funds, the original time series of single manager funds was subjected to tests for normality and autocorrelation (see section 6.2). A differentiation was made between parametric and non-parametric normality tests. The former was based on skewness and kurtosis of the return frequency distribution and has been defined according to Jarque and Bera (1987):

$$\chi_{JB} = \frac{T}{6}\left(S^2 + \frac{K^2}{4}\right) \qquad \text{...(4.1)}$$

where S is the third moment or skewness of the return distribution and K constitutes the kurtosis in excess of 3 (fourth moment). The test statistic χ_{JB} is asymptotically chi-squared distributed and computed from measuring the distance between the skewness and kurtosis with those from the hypothesized normal distribution. The probability associated with the test statistics represents the likelihood of observing an extreme sample test statistic in excess of the test statistic, whilst the chi-squared statistic derived from the population is equal to or less than the critical value (a probability of $< .05$

denotes sufficient evidence to reject the null hypothesis of normality). To account for smaller sample sizes, the observed frequencies of the return distribution could be compared to the expected frequencies of the normal distribution using the Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test (Lilliefors, 1967):

$$d_n = \max_{\mathbf{Y}} |F_{norm}(\mathbf{Y}) - F_0(\mathbf{Y})| \qquad \text{...(4.2)}$$

The test statistic d_n is a function of the maximum difference between the empirical return distribution function F_0 and the hypothesised normal probability distribution F_{norm}. The critical values for rejection of the null hypothesis are derived from Monte Carlo simulations. The null hypothesis states that there the distribution is approximately normal (i.e the maximum difference between the empirical distribution function and normal distribution function is marginal). The null hypothesis is rejected where the associated probability is $< .05$.[9]

Autocorrelation coefficients $\hat{p}_k$ describe the correlation between observations of the return series at different points in time t. Similarly, the partial autocorrelation coefficient measures the remaining correlation between current observations and observations at the mth lag after controlling for observations at intermediate lags. The significance of the cumulative autocorrelation coefficients for various lags (Portmanteau test) was established using the Ljung-Box statistic (Ljung & Box, 1978):

$$\chi_{LB} = T(T+2)\sum_{k=1}^{h}\frac{\hat{p}_k^2}{T-k} \qquad \text{...(4.3)}$$

[9] Most statistical packages report a number of non-parametric test statistics. In the case of EViews those include, besides the Lilliefors-test: Cramer von-Mise (for computational details see Sörgö & Faraway, 1996), Watson (for details see Durbin, 1970) and Anderson-Darling (Anderson & Darling: 1952, 1954). In rare circumstances, the test statistics may lead to conflicting results.

where h is the maximum number of lags being tested and the test statistic is asymptotically chi-squared distributed.

Unit roots in the long-run data generating process were tested for using the Augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1979) and Phillips-Perron (PP) test (Phillips & Perron, 1988). Because of the low power of the ADF test, this test was complemented, as part of a confirmatory data analysis, with the Kwiatkowski-Phillips-Schmidt-Shin ($KPSS$) test for stationarity (Kwiatkowski, Phillips, Schmidt & Shin, 1992). A process was defined as integrated of order one ($I(1)$) if the null hypothesis of the ADF test (H_0: $y_t \sim I(1)$ or unit root) is not rejected and the null hypothesis for the $KPSS$ test (H_0: $y_t \sim I(0)$) is rejected. The general ADF test statistic is defined as follows:

$$\Delta y_t = x_t'\delta + \psi y_{t-1} + \sum_{j=1}^{k} \beta_j \Delta y_{t-j} + \epsilon_t \qquad \text{...(4.4)}$$

The above equation was augmented to account for autocorrelation in the dependent variable, which was modelled using k lagged terms Δy_{t-j}. The optimal lag length was specified using the Schwarz Bayesian information criterion (henceforth $SBIC$). The relevant test statistic was the coefficient ψ subjected to the critical values of the Dickey-Fuller tables. The model is correctly specified if the error term ϵ_t is white noise. The ADF test can be conducted allowing for exogenous variables x_t (a constant or a constant and a linear time trend). Rather than introducing lags of the first differenced term, the PP estimates the non-augmented DF equation and adjusts the t ratio of the ψ coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. The $KPSS$ test differs from the ADF test in that the series is assumed to be stationary under the null hypothesis. The test statistic is based on the residuals of the regression of y_t on the exogenous variables x_t:

$$y_t = x_t'\delta + \epsilon_t \qquad \text{...(4.5)}$$

The residuals ϵ_t are obtained from ordinary least squares (OLS) regression. The KPSS test is a Lagrange Multiplier test (LM) of the hypothesis that the random walk has zero variance. The test statistic is defined as:

$$LM = \sum_t \frac{S(t)^2}{(T^2 f_0)} \qquad ...(4.6)$$

where $S(t)$ constitutes the partial sum $S(t) = \sum_{i=1}^{t} \epsilon_i$ based on the residuals $\hat{\epsilon}_t = y_t - x_t{}'\delta(0)$ and f_0 is an estimator of the residual spectrum at frequency zero. The kernel-based estimator of the frequency zero spectrum is based on a weighted sum of the autocovariances:

$$\hat{f}_0 = \sum_{j=-(t-1)}^{T-1} \hat{\gamma}(j) K\left(\frac{j}{l}\right) \qquad ...(4.7)$$

where l is a bandwidth parameter (truncation lag) according to West (1994), K is the Bartlett kernel function:

$$K(x) = \begin{cases} 1 - |x| \text{ if } |x| \leq 1 \\ 0 \text{ otherwise} \end{cases} \qquad ...(4.8)$$

and $\hat{\gamma}$ is defined as:

$$\hat{\gamma}(j) = \sum_{t=j+1}^{T} \frac{\left(\tilde{\epsilon}_t \tilde{\epsilon}_{t-j}\right)}{T} \qquad ...(4.9)$$

Robust simulation results for the critical values could be gathered from Kwiatkowski *et al.* (1992: 171–172). Cointegrating relationships are established in a VECM (see section 7.2).

4.3 Factor models

The factor model was developed in three steps: firstly, a general factor model for k assets was introduced. Secondly, from the basic factor model, a statistical factor model was derived to identify and group hedge funds according to a small number of unobservable factors derived from common factor analysis. Lastly, the benefits of factor rotation were outlined to facilitate the interpretation of the factor loadings. This is in contrast to earlier research using economic or asset-class factor models using a number of passive indices and option portfolios to identify the exposure of hedge funds to different sources of systematic risk (see section 3.2.2). The difference between the principal component analysis applied to hedge fund data in previous research (see section 3.2.3) and the common factor model introduced here is demonstrated.

4.3.1 General factor model

Assuming there are n assets and T observations for each asset, let r_{it} be the return of asset i at time t. It is then assumed that a limited number of m exogenously specified systematic factors describe a significant proportion of the variation in the return of all assets. The return of a specific asset can be expressed as a function of an intercept, the factor exposure and an error term:

$$r_{it} = \alpha_i + \sum_{j=1}^{m} \beta_{ij} f_{jt} + \epsilon_{it} \qquad \text{…(4.10)}$$

where β_j represents the sensitivities of r_t to the factors f_1 through f_m and α is a constant representing an intercept. Alternatively, β_j can be interpreted as the portfolio weights of the different factors, provided that the following constraints hold true: $\sum_{i=1}^{m} \beta_i = 1$ (sum-of-coefficients constraint) and $\beta_1, \beta_2, \dots, \beta_m \geq 0$ (short-sale constraint). The proportion of return variability not explained by asset allocation can be attributed to security selection within asset classes and is denoted as specific error ϵ_t. If the model in

4.7 is a general factor model (i.e. all assets under observation load on the same exogenous factors), it can be expressed in matrix notation:

$$\mathbf{R} = \mathbf{a} + \mathbf{bF} + \mathbf{e} \qquad \text{...(4.11)}$$

where $\mathbf{R}$, $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{F}$ are the matrices as defined in Table 4.1. In an asset-class-factor model, $\mathbf{F}$ represents the return series of a number of passive indices representing different asset classes. The factor $\mathbf{F} = (\boldsymbol{f}_1, \boldsymbol{f}_2, \dots, \boldsymbol{f}_T)'$ is assumed to be a $T \times m$-dimensional stationary process such that:

I. $E(\boldsymbol{f}_t) = \alpha_f$ (stationarity) ...(4.12)

II. $Cov(\boldsymbol{f}_t) = \Sigma_f$ ($m \times m$ matrix) ...(4.13)

and the asset specific factor $\mathbf{e} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_n]$ is a white noise process that is uncorrelated with the common factors, i.e.:

III. $E(\boldsymbol{\epsilon}_i) = 0$ ...(4.14)

IV. $Cov(\mathbf{F}, \mathbf{e}) = \underset{m \times m}{\mathbf{I}}$ (the $m \times m$ identity matrix). ...(4.15)

V. $Cov(\mathbf{e}) = \underset{m \times m}{\boldsymbol{\Psi}}$ (a $m \times m$ matrix with the variance ψ_i as diagonal and 0 as all other elements) ...(4.16)

Thus, the common factors are uncorrelated with the specific errors and the specific errors are uncorrelated amongst each other. Contrary to the orthogonal factor model derived in section 4.3.3, the common factors need not be uncorrelated with each other.

4.3.2 Statistical factor analysis

Statistical factor analysis can be thought of as a dimensionality reduction technique to reveal a common structure hidden in the data. A limited number of factors are extracted to account for most of the joint variation in the correlation matrix of the observed data. While statistical factor analysis foregoes the economic interpretation of the factors (refer to section 4.4.2 for further details), it is useful in the grouping or clustering of variables based on the factor loadings. Contrary to earlier research (see section 3.2.3), principal factors were used rather than principal components. In the following paragraphs, a clear distinction was made between factor analysis, principal component analysis, the (truncated) principal component method, the principal factor method and factor rotation. In existing literature, the methodologies and terminologies are often not clearly distinguishable.

At the outset, a distinction was made between factor analysis and PCA: in factor analysis, the variables are expressed as linear combinations of the factors, whereas in PCA the principal components are expressed as functions of the variables. Identifying the underlying communality of the data is the purpose of factor analysis. In factor analysis, a further differentiation is made between the component model and the common factor model. The former model implies that the variables can be directly calculated from the factors by applying the weights. The same factor scores produce all variables simply by altering the weights. The component model may be a truncated component model (i.e. the number of factors retained is strictly smaller than the number of variables). In component analysis, unit length vectors are not preserved. One way to extract the m relevant factors is via principal components.

Common factor models, on the other hand, divide factors into two groups: the common factors themselves consist of those factors which contribute to two or more of the variables (i.e. several variables have these factors in common), whereas the unique factors contain the individual scores necessary to complete the prediction of a variable. Figure 4.1 compares the principal factor method as the extraction of principal factors under the component model to the principal axis method or communality estimation under the common factor approach. It can be seen that the differences arise from the communality estimates in the diagonal of the correlation matrix. Other estimation methods exist that are not discussed here. ML estimation is a popular method but requires normality and the pre-specification of the number of common factors (Tsay, 2005: 428–429).

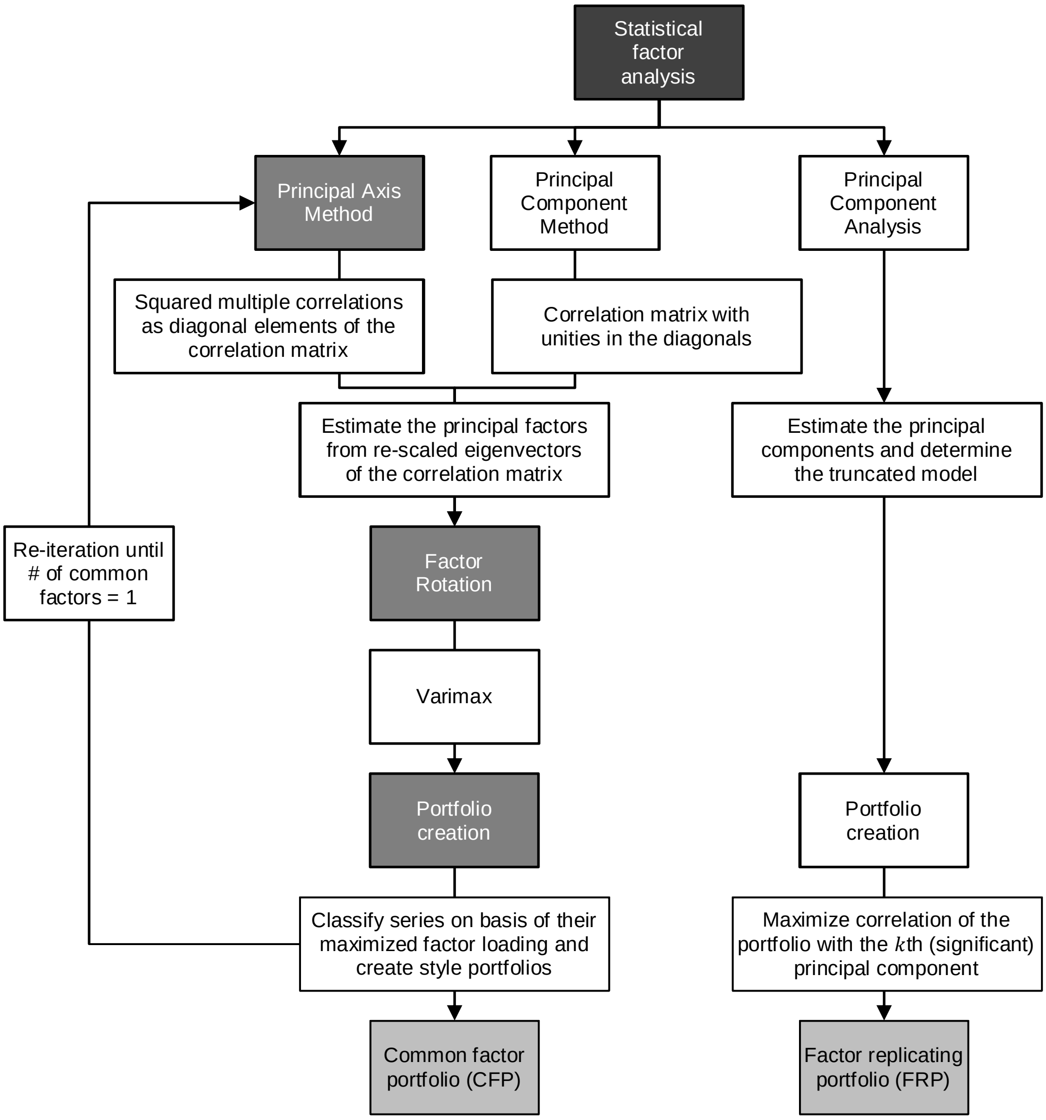

Figure 4.1: Dimensionality reduction techniques

The study continues by providing the rationale for choosing factor analysis over PCA in the categorisation of hedge funds. From the sample of hedge fund return series, an attempt was made to identify underlying common factors driving performance in hedge funds. It was postulated that a few central thrusts explain a significant proportion of the covariance between hedge funds. Here, the off-diagonal elements of the correlation matrix and the communality portions of the diagonal elements are of interest. This is referred to as the common factor model.

In a similar vein, the choice of the principal axis over the (truncated) principal component method is explained. If the principal factor procedure is applied to the correlation matrix with unities in the diagonal, principal components result. In the principal component method, the principal factors are extracted from the correlation matrix with unities as diagonal elements. The factor then gives the best least-square fit to the entire correlation matrix, and each succeeding factor accounts for the maximum amount of the total correlation matrix obtainable. The main diagonal remains unaltered. The procedure attempts to account for all the variance of each variable assuming that all variance is relevant. A truncated component solution may be applied and the inaccuracies in reproducing the correlation matrix are attributed to errors on the model in that sample.

If the principal factor procedure is applied to a correlation matrix where the diagonals have been adjusted to communality estimates, common factors result. Unities are appropriate estimates of initial communalities only if the component model is used. However, the actual communalities resulting from a truncated component solution are rarely unity. The procedure is often inaccurate and may produce communalities considerably higher than the reliability estimates (Gorsuch, 1974: 95). In particular, principal factor methods more accurately reflect small correlations amongst variables since they emphasise the off-diagonal elements. Jolliffe (2004: 159) provided an extensive list of research criticising the appropriateness of unities as initial estimates of the communalities. It should also be noted that PCA assumes that the number of assets n is strictly smaller than the number of time periods T.[10]

Lastly, the structure of the factor matrix was simplified to allow for a discrete classification of hedge funds. Orthogonal factor rotation maximises the variance of the squared loadings of a factor on all variables in a factor matrix so that each variable will have either large or small loadings on any particular factor.

4.3.3 Statistical factor model

The purpose of statistical factor analysis is to identify a limited number of unobservable factors that account for most of the common variability in the observed data. The common factor model postulates that $\mathbf{R}$ is linearly dependent on a few unobserved random variables $\mathbf{F}$ (common factors) and additional noises $\mathbf{e}$ (specific errors). The orthogonal factor model is of the form:

$$\mathbf{R} - \mu = \mathbf{bF} + \mathbf{e} \qquad \text{...(4.17)}$$

where $\mathbf{b}$ is the matrix of factor loadings, $T \times m$-dimensional matrix $\mathbf{F}$ represents the common factors, μ is the mean return, and $\mathbf{e}$ contains the specific factors (Tsay, 2005: 426–427). If the return series is normalised, the mean return equals zero ($\mu = 0$). The formulation above is similar to Equation 4.8. However, the intercept a is now replaced by the mean return μ and the specific factors $\mathbf{F}$ and the loadings $\mathbf{b}$ are unknown. In contrast to PCA, statistical factor analysis attempts a reduction in dimension by invoking a specific model of m factors. In full component models, the specific factor term $\mathbf{e}$ is dropped. Note also that the orthogonal factor representation in Equation 4.18 needs not have a unique solution. This allows for rotating the factors to improve the

[10] In order to conduct PCA where $T < n$, it is possible to use asymptotic PCA (ACPA). The methodology is outlined in Connor and Korajczyk (1986, 1988).

interpretations. The factor model is an orthogonal factor if the following assumptions are satisfied for $\boldsymbol{f}_t = (f_{1t}, f_{2t}, \dots, f_{mt})'$:

I. $$E(\boldsymbol{f}_t) = 0 \qquad \text{...(4.18)}$$

II. $$Cov(\boldsymbol{f}_t) = \underset{m \times m}{I} \quad \text{(the } m \times m \text{ identity matrix)} \qquad \text{...(4.19)}$$

And, as before, the asset specific factor ϵ_{it} is an independent and identically distributed (i.i.d) random variable (see assumptions III – V in Equations 4.11 through 4.13).

Several estimation techniques exist to estimate $\mathbf{F}$ and $\mathbf{b}$. The estimation of factors and loadings usually proceeds in two stages: firstly, restrictions are placed on $\mathbf{b}$ in order to find an initial optimal solution; secondly, other solutions can be found by rotating $\mathbf{b}$ (i.e. multiplying by an orthogonal matrix $\mathbf{T}$). From the number of possible solutions, one is selected that makes the structure of $\mathbf{b}$ as simple as possible, where most of the elements β_{ij} are either close to zero or far from zero. This is achieved using factor rotation.

From Equation 4.14, considering that μ does not affect variances and covariances and $\mathbf{I}$ is the identity matrix and $\mathbf{\Psi}$ is defined as in Table 4.1, it follows that:

$$\begin{aligned} \mathbf{\Sigma} = cov = cov(\mathbf{bF} + \mathbf{e}) \\ = \mathbf{b}cov(\mathbf{F})\mathbf{b}' + \mathbf{\Psi} \\ = \mathbf{bIb}' + \mathbf{\Psi} \\ = \mathbf{bb}' + \mathbf{\Psi} \end{aligned} \qquad \text{...(4.20)}$$

The principal factor method uses an initial estimate $\widehat{\mathbf{\Psi}}$ and factor $\mathbf{\Sigma} = \mathbf{P} - \widehat{\mathbf{\Psi}}$ to obtain:

$$\mathbf{P} - \widehat{\mathbf{\Psi}} \cong \hat{\mathbf{b}}\hat{\mathbf{b}}' \qquad \text{...(4.21)}$$

where $\widehat{\boldsymbol{\Psi}}$ is an estimate of the error covariance matrix, $\mathbf{P}$ is the correlation matrix (both defined in Table 4.1) and $\mathbf{P}-\widehat{\boldsymbol{\Psi}}$:

$$\mathbf{P}-\widehat{\boldsymbol{\Psi}}=\begin{bmatrix}\hat{h}_1^2 & r_{12} & \cdots & r_{1k}\\ r_{21} & \hat{h}_2^2 & \cdots & r_{2k}\\ \vdots & \vdots & \ddots & \vdots\\ r_{k1} & r_{k2} & \cdots & \hat{h}_k^2\end{bmatrix} \qquad ...(4.22)$$

Squared multiple correlations are used as a lower bound for the communality estimates in the diagonals. The squared multiple correlations $\hat{h}_i^2$ are easily calculated from the inverse of the correlation matrix with unites in the diagonal:

$$\hat{h}_i^2=1-\frac{1}{r^{ii}} \qquad ...(4.23)$$

where r^{ii} is the diagonal element from the inverse of the correlation matrix $\mathbf{P}^{-1}$. Once the initial estimates are obtained, factor rotation is used to maximise on the factor loadings. Using spectral decomposition, Equation 4.18 can be rewritten as:

$$\mathbf{P}_m-\widehat{\boldsymbol{\Psi}}_m=\mathbf{C}_m\mathbf{D}_m\mathbf{C}_m' \qquad ...(4.24)$$

where $\mathbf{C}$ is an orthogonal matrix constructed with normalised eigenvectors of $\mathbf{P}_m-\widehat{\boldsymbol{\Psi}}_m$ as columns and $\mathbf{D}$ is a diagonal matrix as defined in Table 4.1. Taking the square root of each eigenvalue to form two diagonal matrices:

$$\mathbf{D}_m=\sqrt{\mathbf{D}_m}\sqrt{\mathbf{D}_m} \qquad ...(4.25)$$

Substitution yields:

$$\mathbf{P}_m-\widehat{\boldsymbol{\Psi}}_m=\left(\mathbf{C}_m\sqrt{\mathbf{D}_m}\right)\left(\mathbf{C}_m\sqrt{\mathbf{D}_m}\right)' \qquad ...(4.26)$$

$\mathbf{b}$ is estimated by the first m columns of $\mathbf{C}\sqrt{\mathbf{D}}$ ($\underset{k \times m}{\mathbf{b}}$ is required with $m < k$):

$$\hat{\mathbf{b}} = \mathbf{C}_1\sqrt{\mathbf{D}_1} = \left(\sqrt{\lambda_1}c_1, \sqrt{\lambda_2}c_2, \ldots, \sqrt{\lambda_m}c_m\right) \quad \text{...(4.27)}$$

Thus, principal factors are rescaled eigenvectors. The eigenvalues are equal to the sum of the squared loadings on the principal factors and are indicative of the covariance accounted for by each factor. The sum of the characteristic roots is equal to the trace of the communality estimates used in the diagonals of the correlation matrix. Note that where the diagonal elements are less than unity (as with the common factor approach), there will be fewer factors than variables. To solve for eigenvalues and eigenvectors can be cumbersome. The QR-Method is a computationally inexpensive and accurate method. A brief introduction is given in Appendix 3. The proportion of variance explained by the ith factor is:

$$\frac{\lambda_i}{tr(\mathbf{P} - \mathbf{\Psi})} = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j} \quad \text{...(4.28)}$$

where tr is the trace of the matrix.

In order to determine the number of non-trivial factors to extract, a Monte Carlo approach was selected. One such Monte-Carlo based simulation method is Horn's parallel analysis (Horn, 1965). The analysis is conducted by generating multiple random data sets of independent random variables. The variables have the same number of observations and variances as the original data. From the simulated data, the Pearson correlation matrix and eigenvalue decomposition are computed. The number of factors retained is based on the number of eigenvalues that exceed their simulated counterparts. The roots from the random data and from the real data are plotted on the same graph. The point where the real data curve crosses the random data curve is designated as the point where real factors give way to random factors. Knuth (2009) provides the appropriate random number generator. The threshold can be the mean values of the simulated eigenvalues or some specified threshold (Glorfeld, 1995). In contrast to other determining methods, parallel analysis is the most robust method in limiting the number of factors. In contrast to other methods, parallel analysis is unbiased since it involves comparing eigenvalues of the dispersion matrix to the results obtained from using uncorrelated simulated data. The large number of iterations ($i = 100$) reduces the error margin.[11]

4.3.4 Correlation versus covariance matrix

The derivation of principal factors above is based on eigenvectors and eigenvalues of $\mathbf{P} - \widehat{\mathbf{\Psi}}$, where $\mathbf{P}$ is the correlation matrix of standardised variables, rather than the covariance matrix. Since the emphasis in factor analysis is on reproducing the covariances / correlations of the variables rather than the variance, the use of $\mathbf{P}$ often gives better results than the sample covariance matrix (Rencher, 2002: 418–419). Jolliffe (2004: 21–26) provided some additional arguments for choosing the correlation matrix over covariance:

- The results from different analyses can be more easily compared when the principal factor estimates are based on the correlation matrix
- Principal factors are used as an explorative / descriptive technique rather than for inferential purposes. Thus, the advantage of covariance in this context is irrelevant
- The standardisation of variables facilitates the interpretation of the principal factors.

There are other benefits to using correlation matrices. However, they were not relevant in the context of this research and shall not be discussed in detail.

[11] For a detailed discussion on other criteria (and their shortcomings in comparison to parallel analysis) such as Kaiser-Guttmann and Minimum Eigenvalue refer to Preacher and MacCullum (2003).

4.3.5 Factor rotation

The loadings of the model in Equation 4.14 can be multiplied by an orthogonal matrix without impairing their ability to reproduce the $\mathbf{P}-\widehat{\mathbf{\Psi}}$ matrix. Let $\mathbf{T}$ be an arbitrary orthogonal matrix. Then $\mathbf{T}'\mathbf{T}=\mathbf{I}$ (identity matrix) and when inserted into the factor model:

$$\mathbf{R}-\mu=\mathbf{b}\mathbf{T}\mathbf{T}'\mathbf{F}+\mathbf{e} \qquad \text{...(4.29)}$$

with $\mathbf{b}^*=\mathbf{bT}$ and $\mathbf{F}^*=\mathbf{T}'\mathbf{F}$ the following is obtained:

$$\mathbf{R}-\mu=\mathbf{b}^*\mathbf{F}^*+\mathbf{e} \qquad \text{...(4.30)}$$

Replacing $\mathbf{b}$ in equation 4.18 with $\mathbf{b}^*=\mathbf{bT}$**:**

$$\begin{aligned}\mathbf{P}-\mathbf{\Psi}&=\mathbf{b}^*\mathbf{b}^{*\prime}=\mathbf{bT}(\mathbf{bT})' \qquad \text{...(4.31)}\\ &=\mathbf{bTT}'\mathbf{b}'\\ &=\mathbf{bb}'\end{aligned}$$

Considering $\mathbf{TT}'=\mathbf{I}$. The new loadings $\mathbf{b}^*$ reproduce the covariance matrix just as $\mathbf{b}$ and the factors $\mathbf{F}$ fulfill the assumptions of 4.15 and 4.16. The communalities h_i^2 remain unaffected by the transformation. To see that this is so, let $\boldsymbol{\beta}_i{}'$ denote the ith row of $\mathbf{b}$**.** The sum of squares of the ith row of $\mathbf{b}$ is $h_i^2=\boldsymbol{\beta}_i{}'\boldsymbol{\beta}_i$. Similarly, $\boldsymbol{\beta}_i^{*\prime}=\boldsymbol{\beta}_i{}'\mathbf{T}$ for $\mathbf{b}^*=\mathbf{bT}$ and the corresponding communality:

$$h_i^{*2}=\boldsymbol{\beta}_i{}'\mathbf{TT}'\boldsymbol{\beta}_i=\ \boldsymbol{\beta}_i{}'\boldsymbol{\beta}_i=h_i^2 \qquad \text{...(4.32)}$$

Multiplication of a vector by an orthogonal matrix is equivalent to rotation of axes (Rencher, 2002: 414–415). If one can achieve a rotation where every point is close to an axis, then each variable loads highly on the axis-corresponding factor and has small loadings with respect to all other factors. A rotation that yields an interpretable loading pattern is required in order to identify natural groupings of variables (hedge funds). Varimax rotation seeks rotated factor loadings that maximise the squared loadings in each column of $\hat{\mathbf{F}}^*$ (Kaiser, 1958).

It is the objective of the rotation to maximize on factor loadings whilst keeping the extracted factors orthogonal (contrary to oblique rotation techniques). The Varimax criterion requires the least amount of iterations to maximize on a single common factor. As before let $\mathbf{b}^* = \mathbf{bT}$ and $\mathbf{b}^*$ consist of the elements β_{ij} with $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, m$, then for varimax rotation the orthogonal rotation matrix $\mathbf{T}$ is chosen in order to maximise:

$$Q = \sum_{j=1}^{m}\left[\sum_{i=1}^{k}\beta_{ij}^{4} - \frac{1}{k}\left(\sum_{i=1}^{k}\beta_{ij}^{2}\right)^{2}\right] \quad \text{...(4.33)}$$

The criterion Q tends to drive the factor loadings towards -1, 0 or 1.The rotation can be applied to the principal factors in order to simplify the interpretation of factors or rotated principal factors (Jolliffe, 2004: 252–254). Each variable will have either large (close to 1, -1) or small (close to 0) loadings of a particular factor. The varimax rotation cannot guarantee that all variables will load highly on only one factor. If the variables are not well clustered, it may be hard to distinguish between one factor loading and another.

One remedy is to conduct factor analysis in an iterative framework. The initial communalities and principal factors were estimated as described in section 4.3.3. Next, the axis was varimax-rotated so as to maximise the squared factor loadings. Variables (hedge funds) were grouped on the bases of the rotated factor loadings as well as the sign of the factor loading. Next, step one was repeated by extracting principal factors from the resulting strategic cluster of hedge funds. The procedure was stopped when all hedge funds within a strategic cluster loaded on one factor only.

4.3.6 Correcting for serial correlation in returns (ARMA process)

Hedge funds are often invested in highly illiquid securities that are not publicly traded or for which there is no secondary market. In consequence, hedge funds are thought to exhibit stale or managed prices (see section 3.3.2). To detect serially correlated price series, the significance of cumulative autocorrelation coefficients was estimated using the Ljung-Box statistic (section 4.2). Some researchers suggest that autoregressive models may be used to predict the proportion of hedge fund performance attributable to illiquidity or smoothed performance (e.g. Lòpez de Prado & Peijan, 2004).

Tsay (2005: 426–427) emphasised that statistical factor analysis assumes that the source data is not serially correlated. He recommended using parametric models to eliminate any linear dependence and to apply factor analysis to the residuals. If the returns are serially dependent, autoregressive models should remove any serial dependence (Tsay, 2005: 406). Let the current value of series y be a linear function of its own previous values and a combination of contemporaneous and lagged values of a white noise error term, the model is written as:

$$\begin{aligned} y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} \\ + \cdots + \theta_q u_{t-q} + u_t \end{aligned} \qquad \text{...(4.34)}$$

The characteristics of an $ARMA(p,q)$ model is a combination of the characteristics from an autoregressive process $AR(p)$ and a moving average process $MA(q)$. The (partial) autocorrelation coefficients are useful in distinguishing between pure AR or MA processes and $ARMA$ (compare sub-section 6.2). The appropriate lag length denoted by p and q is determined according to $SBIC$. For the purpose of determining the correct model orders, the $HQIC$ and AIC were produced as well. For the $SBIC$, the variation in specifications between models estimated from the population are larger compared to AIC (i.e. consistent but inefficient). However, asymptotically, the $SBIC$ will deliver the correct model in terms of lag specification, whilst the AIC will estimate too large a model. $SBIC$ delivers the most parsimonious of models reducing the uncertainty from estimating parameters associated with larger lags. The upshot is that some of the autoregressive and moving-average processes might not be removed from the residual series.

The models were estimated from the lowest to the maximum rank and the value of the information criteria recorded in each case. Upon observation of the observed autocorrelation and partial autocorrelation coefficients, the highest order model considered was an $ARMA(3,3)$ model. The residuals from OLS-regression rather than the original series were used to estimate the squared multiple correlations in the matrix of equation 4.19.

4.4 Factor portfolios

Due to performance differences in monthly returns across the various hedge funds that make up the factor portfolios, the weighting of the individual series gradually shifts away from its optimum. Here, optimum may be defined as the vector of portfolio weights that maximises on the correlation of the weighted series with the extracted factor from principal axis. In order to create time series for a portfolio of hedge funds, the portfolio constituents as well as the portfolio itself needed to be re-based in (regular) intervals. All portfolios were rebalanced on a 12-monthly basis. It was decided on that interval with regards to the maximum lock-up period found in hedge funds entering the sample

4.4.1 Portfolio rebalancing

Depending on the time index t, the input parameters to calculate the index level of the portfolio were calculated as follows:

$$t\ Mod\ 12 = 0 \begin{cases} P'_t = P_{t-1} \\ S'_{kt} = S_{kt} \\ S_{kt} = S'_{kt-1} * (1 + r_{kt-1}) \\ w'_{kt} = w_{kt-1} \end{cases} \quad \text{...(4.35)}$$

$$t\ Mod\ 12 > 0 \begin{cases} P'_t = P'_{t-1} \\ S'_{kt} = S'_{kt-1} \\ S_{kt} = S_{kt-1} * (1 + r_{kt-1}) \\ w'_{kt} = w'_{kt-1} \end{cases}$$

for $t > 0$ and where $S_{10}, S_{20}, \dots, S_{k0} = 1$, Mod denotes the remainder of the division $t/12$, P_t denotes the index level of the portfolio at time t, S_{kt} is the index level of the kth individual hedge fund, P'_t and S'_{kt} are the re-based index levels, and w_{kt} is the portfolio weight for the kth hedge fund at time t. Then, the portfolio index level may be calculated with:

$$P_t = P'_t * \sum_{i=1}^{k} \frac{S_{kt} * w'_{kt}}{S'_{kt}} \qquad \text{...(4.36)}$$

for $t > 0$ and where $P_0 = 1$. The portfolio return r_p at time t is then:

$$r_{Pt} = \frac{(P_t - P_{t-1})}{P_{t-1}} \qquad \text{...(4.37)}$$

The above algorithm for a regularly rebalanced portfolio can be easily implemented in VBA.[12] The constituents of the portfolios are the clusters of hedge funds defined using principal factors and varimax rotation.

4.4.2 Economic interpretability

In order to label the common factors extracted, the returns of the CFPs could be regressed against a number of representative indices in the spirit of asset-class-factor

---

[12] An example including the manual calculation as well as a macro version is available from the author on request.

models and additional asset-bases-style (ABS) factors as described in Fung and Hsieh (2003). The model is of the type described in section 4.3.1: asset returns are the returns of the factor portfolios and the factors are observable returns on asset indices, yield spreads or returns on dummy portfolios. The choice of regressors was in accordance with previous research on asset-class-factor models and their application to hedge funds (see section 3.2.2). In particular, a distinction was made between equity indices, yield curve proxies, the Fama-French and momentum factors, cash proxies and the primitive-trend-following factors derived from Fung and Hsieh (2002a). A brief explanation for each regressor can be found in Table 5.1, outlining the composition as well as the data source.

In addition to Sharpe's asset factors, the factors derived from Fama and French (1992, 1993, 1996) are considered as explanatory factors in pricing models. Empirical evidence shows that both firm size and book-to-market ratio are proxies for exposure to systematic risk in hedge funds not captured by the Sharpe's original factors. The two additional factors are denoted SMB (small minus big firm size) and HML (high book-to-market ratio minus low ratio):

$$SMB = \frac{1}{3}\left(\frac{S}{L} + \frac{S}{M} + \frac{S}{H}\right) - \frac{1}{3}\left(\frac{B}{L} + \frac{B}{M} + \frac{B}{H}\right) \qquad \text{...(4.38)}$$

$$HML = \frac{1}{2}\left(\frac{S}{H} + \frac{B}{H}\right) - \frac{1}{2}\left(\frac{S}{L} + \frac{B}{L}\right) \qquad \text{...(4.39)}$$

with S as the return of portfolios with low, medium, and high-market-capitalisation stocks and B as the return of portfolios sorted according to low, medium, and high book-to-market ratios. Here SMB is the difference in return between an equally weighted long position in the three small firm portfolios and a short position in three big firm portfolios. HML is defined as the difference in return between an equally weighted long position in high B/M portfolios and a short position in low B/M portfolios.

A natural extension of the Fama-French model is the momentum factor WML (winners minus losers) introduced by Carhart (1997). A momentum strategy could be described as 'buying past winners and shorting past losers' (Jegadeesh & Titman, 1993). Conversely, a contrarian strategy focuses on past underperformers. Carhart (1997: 73) finds no evidence that momentum funds outperform contrarian funds. WML is the return on a zero-investment portfolio that is long past winners and short past losers. Note here that Fama-French factors are thought to proxy for higher systematic moments (compare Chung *et al.*, 2006). One may differentiate between factors describing the exposure to passive indices (location choice) and factors serving as proxies for a particular trading strategy (style choice).

Fung and Hsieh (2001: 317–326) defined the PTFS as a long position in a look-back straddle. The straddle consists of a call and a put with the same exercise price and expiration, designed to capture the difference between the price of the underlying asset upon maturity and the exercise price. The PTFS attempts to capture the largest price movement of the underlying asset during a specified time interval, where the optimal payout of the PMTS is the maximum price less the minimum price of the asset. The payout profile of the PMTS can be simulated by dynamically rolling standard straddles over the life of the look-back straddle. Fung and Hsieh (2002a) showed empirically that trend-following hedge fund returns are strongly correlated with the returns of the PTFS. A complete list of the regressors including sources is depicted in Table 5.1.

Allowing for lagged factor exposure for up to three lags, the general asset-class factor model can now be written:

$$r_t = \alpha + \beta_{1t}F_{1t} + \beta_{1t-1}F_{1t-1} + \cdots + \beta_{1t-3}F_{1t-3} + \beta_{2t}F_{2t} + \beta_{2t-1}F_{2t-1} + \cdots + \beta_{2t-3}F_{kt-3} + \beta_{kt}F_{kt} + \beta_{kt-1}F_{kt-1} + \cdots + \beta_{kt-3}F_{kt-3} = \alpha + \sum_{j=1}^{k}\sum_{i=0}^{3}\beta_{kt-i}F_{kt-i} \quad \ldots(4.40)$$

The coefficients are estimated using Ordinary-Least-Squares (OLS). A HAC-consistent standard error estimate of the coefficients is given by the diagonal elements of the matrix below (Newey & West, 1987):

$$\hat{\Sigma}_{NW} = \frac{T}{T-m}(\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{\Omega}}(\mathbf{X}'\mathbf{X})^{-1} \quad \ldots(4.41)$$

and $\hat{\Omega}$ is defined as follows:

$$\hat{\mathbf{\Omega}} = \frac{T}{T-m}\left\{\sum_{t=1}^{T}\epsilon_t^2 x_t x_t{}' + \sum_{v=1}^{l}\left(\left(1-\frac{v}{l+1}\right)\sum_{t=v+1}^{T}\left(x_t\epsilon_t\epsilon_{t-v}x_{t-v}' + x_{t-v}\epsilon_{t-v}\epsilon_t x_t'\right)\right)\right\} \quad \ldots(4.42)$$

where the weighting function is the Bartlett kernel and l is the truncation parameter. The truncation lag is $l = floor\left(4(T/100)^{2/9}\right)$ and m is the number of regressors (the floor and function maps a real number to the largest previous or the smallest following integer). In order to avoid multicollinearity between the regressors, a stepwise regression algorithm was employed.

The forward stepwise regression algorithm is described in detail in Neter, Kutner, Nachtsheim and Wasserman (1996: 348–352). At every step, a regressor was entered into the model if the increase in explanatory power outweighed the ‘cost’ associated with the decrease in degrees of freedom. Entering variables were tested for collinearity with existing regressors and would not enter the model if any $\hat{R}^2$ (adjusted R^2) from regressing all regressors on all other independent variables exceeded a specified threshold. Conversely, existing regressors were removed if their removal did not cause a significant decrease in explanatory power in the movements of the regressand. The partial F-test allowed for a formal comparison of the reduced model to the full model at every step. In the two variable case where k and n are indicators of regressors from X:

$$F_k^* = \frac{MSR(X_k|X_n)}{MSE(X_n, X_k)}, \qquad \text{...(4.43)}$$

$$= \frac{SSE(X_n) - SSE(X_k, X_n)}{df_{X_n} - df_{X_k X_n}} \div \frac{SSE(X_k, X_n)}{df_{X_k X_n}}$$

$$= \left(\frac{b_k}{s\{b_k\}}\right)^2$$

where MSR is mean square due to regression, MSE is the mean square error, SSE is the sum of squares for error, and df denotes the degrees of freedom. The coefficient estimates and standard error thereof are denoted as b_k and $s\{b_k\}$ respectively. The equality above holds in the absence of heteroscedasticity and autocorrelation in the error terms. The partial F-statistic can be calculated from $(b_k/s\{b_k\})^2$ to account for HAC-adjusted standard errors as in Equation 4.36.

It should be noted that the stepwise regression procedure does not always reliably remove multicollinearity from the regressors. The reasons for this are two-fold: firstly, in a stepwise process variables are either added or removed from the list of regressors. This accounts for co-linearity at the individual series level but ignores any effects from entering variables being jointly co-linearly related to existing regressors or combinations thereof. Secondly, the calculated F-statistics do not take into consideration the iterative nature of the estimation procedure. The p-values associated with the inclusion/exclusion of parameters from the iterative model should be adjusted to account for the sequential parameter selection process (compare Brooks, 2003: pp. 104, 105). Whilst these limitations were acknowledged, it was the purpose of the stepwise procedure to reduce the number of regressors to facilitate interpretation of the factor exposures rather than to minimize the impact from multi-collinearity. This approach is deemed appropriate since the results from the procedure were used in the *nomenclature* of hedge fund portfolios/indices only.

4.5 Vector error correction model

A portfolio of traditional investments in stocks, bonds and commodities in combination with an investment into hedge funds was tested. Each investment is represented by an index. For the hedge fund investments, those are the indices created from CFPs established in section 6. The proxies for the stock market, bond market and commodities are the MSCI World US Index ($MSCIW$), the Barclays Aggregate Bond Index ($BABDIDX$), and the Goldman Sachs Commodity Index ($GSCI$), respectively.

The error correction model was first established for the bivariate case (compare Engle & Granger, 1987). Consider two series y_t and x_t that are both $I(1)$. In the context of univariate modeling, accounting for non-stationarity of the series usually involves using the first-differenced terms Δy_t and Δx_t in the further modeling process. However, in the event where a co-integrating relationship exists, the first difference model has no long-run solution and an equilibrium relationship between y and x cannot be established (Brooks, 2008: 338). The first difference model is given by:

$$\Delta y_t = \beta \Delta x_t + \epsilon_t \qquad \text{...(4.44)}$$

The proposed model includes an error correction term $y_{t-1} - \gamma x_{t-1}$:

$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 (y_{t-1} - \gamma x_{t-1}) + \epsilon_t \qquad \text{...(4.45)}$$

Provided that y_t and x_t are co-integrated with the co-integrating coefficient γ, then the error correction term will be $I(0)$ even though the constituents are $I(1)$. In consequence, OLS can be used in estimating the above equation. Here, the change in y between t and $t-1$ corresponds to changes in the explanatory variable x between t and $t-1$ after accounting for the disequilibrium between y and x in the previous period (the error correction term). The interpretation of the coefficients is as follows:

β_1 = short-run relationship between changes in x and y ...(4.46)

β_2 = coefficient for the speed of adjustment back to equilibrium

γ = long-run relationship between x and y

4.5.1 Simultaneous equations

The above bivariate model may also be expressed in the context of simultaneous equations (i.e. changes in y may be inducing changes in x). In addition, the error correction model can be extended to the multivariate case (compare Johansen, 1991). The cointegrating relationship between $k \geq 2$ variables may be estimated in a VECM. First, the notation for variables in a multivariate model is established:

Table 4.2: Notations and variables in the VECM

Var	Description.	Constituents (vector/matrix only)
t	Time subscript.	
m	The number of variables ($m = 4$).	
k	The number of lags considered.	

$\mathbf{y}_t$ The dependent variable. $\underset{k\times 1}{\mathbf{y}_t} = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{bmatrix}$

$\mathbf{y}_{t-k}$ The kth lagged parameter of y_t. $\underset{k\times 1}{\mathbf{y}_{t-k}} = \begin{bmatrix} y_{1t-\mathrm{k}} \\ y_{2t-k} \\ \vdots \\ y_{kt-\mathrm{k}} \end{bmatrix}$

$\mathbf{\Delta y}_{t-k}$ The kth lagged differenced form of y_t. $\underset{k\times 1}{\Delta\mathbf{y}_{t-m}} = \begin{bmatrix} \Delta y_{1t-\mathrm{k}} \\ \Delta y_{2t-\mathrm{k}} \\ \vdots \\ \Delta y_{kt-\mathrm{k}} \end{bmatrix}$

$\mathbf{I}_g$ A $m \times m$ identity matrix.

$\mathbf{\Pi}$ A matrix of the coefficients for the lagged parameter. $\underset{m\times m}{\mathbf{\Pi}} = \left(\sum_{i=1}^{k} \beta_i\right) - \mathbf{I}_m$

$\mathbf{\Gamma}_i$ The ith coefficient matrix of the differenced parameter. $\underset{m\times m}{\mathbf{\Gamma}_\mathrm{i}} = \left(\sum_{j=1}^{i} \beta_j\right) - \mathbf{I}_m$

Note that vectors and matrices are in bold font. Dimensions appear underneath the matrix/vector notations.

From the general Vector Autoregressive Model (VAR) with $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})'$ and $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \epsilon_{2t}, \dots, \epsilon_{kt})'$

$$\mathbf{y}_t = \boldsymbol{\beta}_1\mathbf{y}_{t-1} + \boldsymbol{\beta}_2\mathbf{y}_{t-2} + \cdots + \boldsymbol{\beta}_k\mathbf{y}_{\boldsymbol{t-k}} + \boldsymbol{\epsilon}_t \qquad \text{...(4.47)}$$

the VECM of the form

$$\Delta\mathbf{y}_t = \mathbf{\Pi}\mathbf{y}_{t-k} + \mathbf{\Gamma}_1\mathbf{\Delta y}_{t-1} + +\mathbf{\Gamma}_2\mathbf{\Delta y}_{t-2} + \cdots + \mathbf{\Gamma}_{k-1}\mathbf{y}_{t-(k-1)} + \boldsymbol{\epsilon}_t \qquad \text{...(4.48)}$$

follows, where the coefficients $\mathbf{\Pi}$ and $\mathbf{\Gamma}_i$ are defined as in Table 4.2. Models with different cointegration test specifications for deterministic trend assumptions were compared and selected according to modified information criteria (Brooks, 2008: 294), which are given below:

$$MAIC = \log\left|\hat{\mathbf{\Sigma}}\right| + 2\upsilon'/T \qquad \text{Akaike (1974)} \qquad ...(4.49)$$

$$MSBIC = \log\left|\hat{\mathbf{\Sigma}}\right| + \upsilon'/T\log(T) \qquad \text{Schwartz (1978)} \qquad ...(4.50)$$

$$MHQIC = \log\left|\hat{\mathbf{\Sigma}}\right| + \upsilon'/T\log(\log(T)) \qquad \text{Hannan and Quinn (1979)} \qquad ...(4.51)$$

with $\hat{\mathbf{\Sigma}}$ as the covariance matrix of residuals and υ as the total numbers of regressors in all equations ($\upsilon = k^2m + k$ where k is the number of equations and m the number of lags considered). The information criteria were constructed for various lags, where the appropriate lag number was that minimising the value of the information criterion. Note that the criteria need not agree on a lag length. The appropriate lag length for the first-differenced terms was confirmed with Wald lag exclusion tests. The diagnostics of the residual series for the VEC model included multivariate extensions of the Ljung-Box statistics for autocorrelation, a multivariate LM test statistic for autocorrelation (Johansen, 1995) and extensions of the Jarque-Bera normality tests with different choices for the factorisation matrix of the residuals: the inverse of the lower triangular Cholesky factor of the residual covariance, the inverse square root of the residual correlation matrix (Doornik & Hansen, 2008), and the inverse square root of the residual covariance matrix (Urzua, 1997). The multivariate extensions of the White heteroskedasticity test with no cross terms were used (see Kelejian, 1982).

4.5.2 Granger causality, variance decomposition and impulse response tests

Since all variables in the error correction model are stationary, the F-test could be used to determine the significance of the various coefficients. The F-statistic may be calculated from the Sum-of-Squares-for-Error (SSE) of the unrestricted OLS model versus the SSE of the restricted model. One may also test for one-directional causality or Granger-causality in variables, defined as the correlation between the current value of one variable and past observations of another (Granger, 1969).

The impulse response is a measurement of the responsiveness of the dependent variable to shocks to each of the variables. For each of the k variables in the equation, a unit shock is applied to the error, and the effects upon the VECM system over time are recorded. Consider Equation 4.43 for the simplified case where lag length $m = 1$ and variables $k = 2$ and the elements of the vectors and matrices are written out:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{matrix} \Delta y_{1t-1} \\ \Delta y_{2t-1} \end{matrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \quad \text{...(4.52)}$$

It is easy to see that one may consider the effect of a unit shock to Δy_{1t} and Δy_{2t} at time $t = 0$ by setting $\Delta y_0 = \begin{bmatrix} \epsilon_{10} \\ \epsilon_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\Delta y_0 = \begin{bmatrix} \epsilon_{10} \\ \epsilon_{20} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, respectively. It should be evident that the required number of tests increases exponentially with the number of variables (a total of k^2 impulse responses can be generated). To generalise, impulse tests allow for observing the impact of a unit shock in $t = 0$ (one standard deviation shock) to the entire system of equations. Brooks (2008: 301) emphasised the importance of the ordering of the variables. In order to account for the common component in the error term of the variables (assuming the error terms are correlated across equations), it is possible to generate orthogonalised error responses. For the bivariate example above, the whole of the common error component was attributed somewhat arbitrarily to the first variable, implying an ordering of the variables. If financial theory fails to suggest an ordering, the sensitivities of the results to changes in the ordering may be established by re-computing the results for the impulse tests for different orders.

Variance decomposition, on the other hand, yields the proportion of movements in the dependent variables due to their own shock in relation to shocks to the other variables. Put differently, variance decomposition estimates how much of the s-step-ahead forecast error variance of a given variable is explained by changes to each explanatory variable. The same principles of ordering outlined in the previous paragraph apply to variance decomposition.

Tests for normality and autocorrelation were applied at the single manager fund level. The residual series for all sample hedge funds were derived from $ARMA(p,q)$ models selected on the basis of information criteria. The iterative estimation of communalities using principal factor axis methodology was conducted for all residual series. Factor extraction was conducted in a rolling-window framework: the communalities of single manager hedge funds were derived from a 120-month estimation period. The results from the estimation period were used to classify single manager hedge funds on the basis of their past performance. Hedge funds belonging to the same classification were sorted into CFPs and portfolio weights were estimated using a non-linear optimization technique. The results were re-iterated for 120 months and the weighted series, re-balanced annually, could be seen as a pure style index representation of hedge fund strategies. The style indices were entered into a multivariate framework to estimate their short-term and long-run exposure to other asset classes. Modified information criteria and the Wald-test determined the appropriateness of the estimated VECM. The tests applied to determine causality and interdependence of the model variables were variance decomposition, impulse response and Granger causality.

CHAPTER 5: DATA SOURCING

5.1 Introduction

The source data used throughout this thesis was provided in a number of different file formats, including Excel spreadsheets, SQL tables and text files. All data was normalized and synchronized to appear in a centralized SQL database and to be used in queries when sourcing for particular information. All data for this research was financial time series data and series metadata and collected in a central database. The resulting database contains all single manager hedge funds and FoHFs reporting to HFR and HFN as well as a number of additional time series to be used in analysis (including equity and bond indices, commodity and currency proxies, Treasury rates and interest spreads as well as the return on trading strategies such as primitive trend-followers and the Fama-French portfolios). All hedge fund indices, investable or not, were included.

An Access-SQL database was used as a central back-end repository for movement data (i.e. time series observations) as well as metadata (i.e. additional descriptive information on time series). Metadata was the additional information associated with a time series such as fund manager, nomination currency etc. The metadata and movement data were stored in different tables and linked together to improve upon the performance of data queries. Excel was used as front-end to retrieve relevant data using VBA/SQL and to display and format results. The data query was custom-built in Excel-VBA based on selection criteria entered in a userform and the relevant information extracted to a pivottable to retain the multidimensionality of the source data, e.g. Return-on-Investment (ROI) and Assets-under-Management (AUM) report filters. The upshot was an integrated data storage and retrieval solution based on MS Access and MS Excel. The database and front-end data sourcing tools are retained to be used in future research. The database itself can be easily complemented with additional series and/or time periods.

Most statistical tests and modelling were conducted from within Eviews. Series were imported from Excel and results exported as .csv files to be viewed in Excel. Data sources were hedge fund databases provided in Excel or Access format, publicly available pricing data published online and pricing data from the Bloomberg terminal. The data from HFR and HFN was not altered but was checked for consistency / accuracy.

Existing research into statistical factor models for hedge funds has been focused primarily on the TASS, HFR and CISDM/MAR databases (compare Table A.1). For this paper, all samples were created from hedge funds reporting to one of two databases: HFR and HFN. The HFR database was complemented by the HFN for two reasons: firstly, only some hedge funds report to both databases. As a consequence, the sample size was increased for the combined samples of the HFR and HFN databases. Secondly, the HFN database includes managed futures/commodity trading advisors (CTAs), whereas HFR does not. To account for attrition rates and survivorship bias, defunct or derelict funds formerly reporting to HFR were included in the analysis in the form of the HFR graveyard database.

In this context, survivorship bias may be defined as the difference in average performance between the entire HFR sample including surviving as well as graveyard funds and the HFR sample including live funds only. Note that HFN does not provide a graveyard database. It is acknowledged that inclusion of the HFN database biased the analysis towards survivors. In the opinion of the author, the benefits from a larger sample size outweighed the estimation error due to survivorship bias. All results have been reproduced for the standalone HFR database and included in the appendix (cross-references are provided where appropriate).

This study covers the period from July 1990 to June 2010. The timeframe was selected so as to include the demise of the LTCM fund in 1998, the subsequent period of economic recovery, as well as the subprime lending and banking crisis of 2007. This allowed for testing of the results throughout different economic cycles. Firstly, it was postulated that the classification according to common factor loadings yields meaningful strategic clusters of hedge funds. Secondly, this classification is robust with respect to macroeconomic impact factors (i.e. all hedge funds within a style classification are expected to react in a similar fashion to external shocks). To test for the persistency of the results, it was decided to conduct the analysis for 120 rolling-window estimation periods commencing with the July 1990 to June 2000 timeframe. The results from the estimation periods were then used as forecasts for the composition of common factor portfolios. The estimates are unbiased since each prediction relies upon the information that is available at time t, thus avoiding look-ahead bias.

All classifications and sub-strategies for HFR and HFN are displayed in Figure 5.1 and Figure 5.2. They show the reported main strategy and sub-strategy that can be expected to deviate from the classification as a result of factor analysis. It is evident that the self-reported classifications are diverse and inconsistent across database providers. For the HFN database, the differentiation between market neutral and directional hedge funds, as well as the broader strategic themes, were included to facilitate comparison between the databases. The following section briefly describes the two databases used. This section concludes by outlining the selection criteria/minimum requirements and the data sampling process.

5.2 HFR and HFN database

Single manager hedge funds reporting to professional database vendors such as HFR and HFN are classified according to their self-reported investment strategy. However, strategic classifications of hedge funds are not consistent across database providers and style attribution of individual funds within databases may depend on a number of factors such as the predominant trading strategy, geographic or sector focus, financial gearing and hedge overlay, or type of investment instruments employed. It is questionable whether self-classification yields homogenous groups of funds representative of a distinct investment style and focus. In times of financial turmoil such as the subprime lending crisis in 2007, hedge fund managers were often found to gradually shift a fund’s investment focus away from its stated investment objective, either due to tightening liquidity or due to the devaluation of asset-based securities such as collateralised debt obligations (see section 3.4.1 on style shifts).

To give an idea of the complexity surrounding hedge fund strategies and style classification, two figures have been included depicting the classification of hedge funds in the HFR and HFN database (Figure 5.1 and Figure 5.2). For visual purposes, a distinction is made between market directional and market neutral funds. It should be noted, however, that not all hedge funds belonging to the various sub-categories are easily defined as strictly directional or market neutral. HFR includes 31 different sub-strategies and differentiates between four major classifications (macro, equity hedge, event driven and relative value). HFN, on the other hand, makes provisions for 28 single manager strategies. In addition, HFR has four FoHFs and HFN three FoHFs strategies. Within HFN, the 28 sub-strategies are referred to as main strategies. No further information is given with respect to a broader classification. These classifications for HFN are provided by the author to facilitate the comparison of the two databases.

It should be apparent that, whilst impossible to cater for all sub-strategies depicted, hedge funds may be classified according to four broad strategic themes as defined in the HFR database: Equity Hedge, Event Driven, Macro and Relative Value. The sub-strategies of the HFN database were sub-classified according to the same HFR themes: Equity and Equity Market Neutral were joined to form the equity hedge category, Fixed Income strategies were subsumed under relative value, CTA/managed futures remained as a separate classification and all multi-strategy, option strategies and short-term trading funds were attributed to the newly formed classification ‘other’. The remaining classifications event driven and (Global) macro remained as they are. The appendix gives brief explanations for the more common investment strategies in hedge funds (Table A.2).

Results for the VEC models in section 7.6 were reproduced using commercial indices from the HFR and HFN database. The HFN database provides several classification indices based on investment philosophy as well as sector focus. Thirteen index series were selected that best correspond to the distinct classifications as identified in Chapter 6. For HFR, the HFRX index series was chosen (all series are included with the exception of the composite equal weighted index). Their composition is based on correlation and cluster analysis, and hence, the index construction philosophy is similar to index composition outlined in sub-section 6.3. There are also similarities in terms of number of constituents included and the weighting of constituents to maximise the correlation within strategic groups.[13]

[13] For more information on HFRX indices: https://www.hedgefundresearch.com/?fuse=hfrx-faq&1254841202.

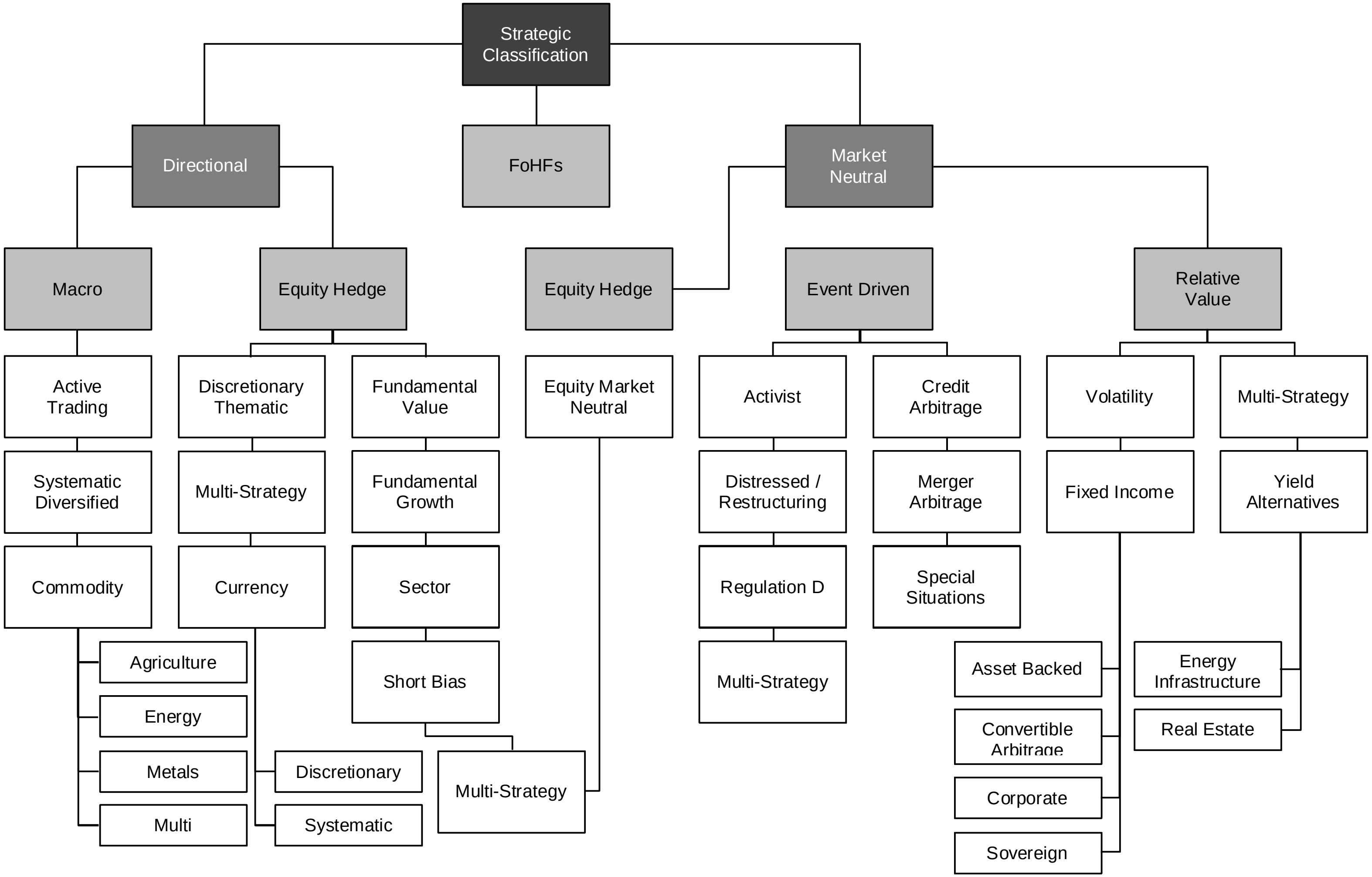

Figure 5.1: Hedge fund classifications in the HFR database

Source: https://www.hedgefundresearch.com/index.php?fuse=indices-str&1291127795

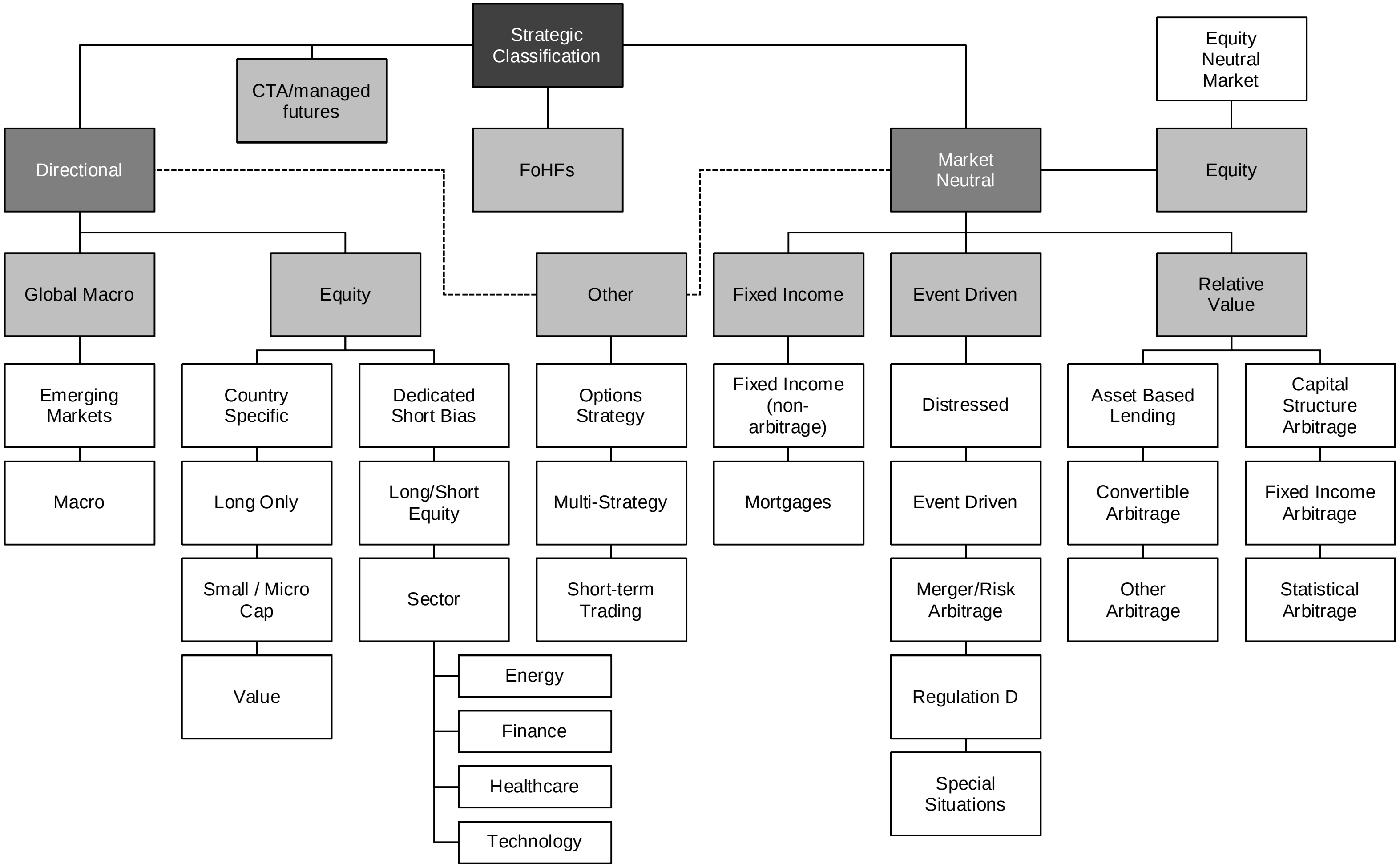

Figure 5.2: Hedge fund classifications in the HFN database

Source: https://www.evestment.com/resources/indices/fund-classification-guide

5.3 Sampling

The information extracted from the two databases included the monthly return on investment (ROI), main investment strategy and sub-strategy. For hedge funds reporting to the HFR graveyard database, the date that the fund stopped reporting to the database was included. For both databases, ROI has been defined as change in net asset value during the month, assuming the reinvestment of any inflows on the fund's reinvestment date, divided by net asset value at the beginning of the month. In general, returns were reported net of management fees, incentive fees or other expenditures. Net-of-fee performance was calculated and provided by the fund managers. Reported returns were assumed to be an accurate representation of investors' realised returns. The sampling criteria comprised the following:

1. Single-manager funds only
2. Continuous track records of 123 return observations; no inconsistencies or performance gaps
3. USD as returns currency
4. At least monthly reporting frequency
5. Reporting style: net-of-all-fees.

Double-reporting funds within as well as between databases were accounted for. For example, hedge funds reporting to a database may elect to include time series for onshore and offshore investment vehicles separately. While the after-tax return may differ with respect to investor residency, the ROI reported to the database is identical for both onshore and offshore funds. Additionally, some managers offer several classes of the same basic investment strategy that differ with respect to the underlying currency (e.g. USD, EUR or GBP). To avoid accounting twice for the same investment fund, the analysis was limited to funds reporting in USD. This is on par with Amenc, Martellini and Vaissié (2003a): in order to avoid currency fluctuations, they included funds reporting in USD only. Since most hedge funds have a USD version, focusing on USD-denominated funds helps to avoid errors from duplicated funds.

Lastly, a fund manager may offer several classes of the same basic investment strategy (e.g. market neutral) but that differ with respect to hedge overlay and leverage. Alternatively, similar series of the same funds may be offered as different share classes for regulatory and accounting reasons. Where these funds produced identical time series, one of the two series was eliminated. In the source SQL database, hedge funds reporting to HFR and HFN were compared on the basis of their coefficients of cross-correlations as well as the degree of similarity of the text strings for fund and manager name. If there were sufficient indications that two series are representative of the same underlying fund, one of the two series was removed. In the event where funds were found to report both to HFR and HFN, the HFN series was removed.

In the combined sample, if funds were found to report to both databases, only one of two records was retained. The HFR database was used as the primary database, and the samples were complemented with hedge funds reporting only to the HFN database. In the event where duplicate hedge funds belong to different strategy classifications in the HFR and HFN databases, the record was removed from the HFN classification. This could have led to fewer funds of that particular strategy entering the joint sample. The combined database of HFR and HFN includes 26 300 funds, of which 18 471 are single manager funds. Between April 1995 and June 2010, after accounting for double reporters, 1 055 funds provided a continuous performance track record of $T = 123$ whilst fulfilling the sampling criteria outlined above (three funds were removed due to inconsistencies in their track record). An extensive list of the asset-based factors used in found below.

Table 5.1: Regressors and data sources

Regressor	Source
[$GOLDUSD$] Price change in Gold quoted in USD (London PM Fix)	http://www.gold.org/assets/file/value/stats/statistics/xls/monthly_prices.xls
[$GSCI$] S&P Goldman Sachs Commodity Total Return Index	Bloomberg (Ticker: SPGSCITR Index)
[$BABDIDX$] Barclays Aggregate Bond Index	Bloomberg (Ticker: BABIDX Index)
[$MSCIEM$], [$MSCIUS$], [$MSCIEXUS$] USD returns on Morgan Stanley Capital International Emerging Markets Index, MSCI World Index and MSCI World Index excluding USA	Bloomberg (Ticker: MXEF, MXWO and MXWOU Index)
[$USDIDX$] Federal Reserve Traded Weighted Index of the US Dollar against Major Currencies	http://www.federalreserve.gov/releases/h10/Summary/indexb_m.txt.
[$CBOEVIX$] Chicago Board Options Exchange Volatility Index	www.cboe.com/VIX
[SMB], [HML], [WML] Returns on the SMB, HML, WML portfolios	SMB and HML:

constructed from US equities	http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/6_Portfolios_2x3.zip WML: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/ftp/F-F_Momentum_Factor.zip
[$US10Y3M$], [$USBAA10y$], [$USMO10Y$] Yield curve spreads: Between 10-year Treasury Bill and 3-month Treasury, between Moody's yield on seasoned corporate bonds – all industries – Baa and 10-year Treasury Bill , and between the contract rate on 30-year, fixed-rate conventional home mortgage commitments (US) and 10-year Treasury Bill	http://www.federalreserve.gov/releases/h15/data/Monthly/H15_TCMNOM_Y10.txt http://www.federalreserve.gov/releases/h15/data/Monthly/H15_TCMNOM_M3.txt http://www.federalreserve.gov/releases/h15/data/Monthly/H15_BAA_NA.txt http://www.federalreserve.gov/releases/h15/data/Monthly/H15_MORTG_NA.txt
[$PTFSBD$], [$PTFSFX$], [$PTFSCOM$]. [$PTFSIR$], [$PTFSSTK$] PTFS: Bond look-back straddle, currency look-back straddle, commodity look-back straddle, short-term interest look-back straddle, and stock index look-back straddle	http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-Fac.xls

All regressors are considered up to the third lag. Regressors are entered into the model if they improve upon the explanatory power and are not highly correlated to any other regressors or combinations thereof.

CHAPTER 6: FACTOR AXIS AND ROTATION

6.1 Introduction

The factor model applied to the sample hedge funds served two purposes: firstly, to classify and group hedge funds based on their past performance and, secondly, to create weighted indices from the return series of hedge funds within a classification. The indices could be thought of as pure style indices representing the common return component or broad strategic themes of hedge funds grouped together. Hedge funds were classified according to their past ten-year performance. As an example, the covariance of hedge funds with continuous track records between July 1990 and June 2000 provided the initial communality estimates and, ultimately, classification of hedge funds for July 2000.

Throughout the ten-year rolling timeframes under observation, new hedge funds entered the sample whilst others stopped reporting or suspended operations. Other hedge funds gradually shifted their market exposure or adjusted investment strategies to benefit from short-term arbitrage opportunities. It was thus decided to re-evaluate the classification of hedge funds every month using the same methodology as for the initial communality estimates. This allowed for newly entered hedge funds to be attributed to existing classifications and existing sample hedge funds to be re-classified where they had departed from their initial strategy (as evidenced by a deceasing degree of communality with other hedge funds from the same classification). In addition, emerging sub-clusters could be identified that warranted a separate classification. The weights of style indices were adjusted on an annual basis to re-balance constituents and to include additional sample hedge funds.

This chapter is subdivided into four parts. Section 6.2 describes the properties of the hedge fund time series in the sample as well as the properties of the adjusted series. Section 6.3 displays the results of the factor model and discusses possible implications. Section 6.4 provides economic interpretations of the factor portfolios. For the purpose of the initial analysis, a differentiation was made between the following five major style classifications: equity hedge, event-driven, macro, relative value and other. The 'other' classification was added for hedge funds that did not belong to one of the previous categories or for which there was not enough information on the main strategy to classify. For the HFN database, CTA/managed futures were included as a separate category. The classification was preliminary and differed from the style classification derived from factor analysis. The chapter concludes with the statistical properties of the index series outlined in section 6.5.

6.2 Properties of hedge fund return series

Classical linear regression models require that the regression error term is i.i.d. and approximately normal, or $\epsilon_{it} \sim N(0, \sigma^2)$. Similarly, some estimation techniques require explicit assumptions about the frequency distribution (e.g. maximum likelihood). It is easy to see from Table 6.1 that these assumptions are unlikely to hold in the context of the distribution of hedge fund returns. Table 6.1 gives the first four moments of the frequency distribution of returns as well as a parametric and nonparametric test statistic to estimate the deviation from the normal distribution function. It was assumed that the first four moments of the distribution describe the frequency distribution of returns in sufficient detail. This implies that investors consider the mean, standard deviation, skewness and kurtosis of a time series when comparing investments. The results are for the reported strategic classifications from the HFR and HFN databases. One upshot of using the factor axis methodology is that the unique component of the model subsumes the abnormal return component of hedge funds and, hence, non-normality of the original series is less of a concern.

The results show that the sample frequency distribution is decidedly different from normal. This is confirmed for both the parametric Jarque-Bera test relying on skewness and kurtosis as well as for the non-parametric Kolmogorov-Smirnov-Lilliefors statistic. It is well established in literature that hedge fund returns exhibit negative skewness and excess kurtosis (see for example Füss *et al.*, 2007).[14] In consequence, the frequency distributions tend to be more sensitive to the parametric Jarque-Bera test statistic. Overall, the results are of similar scope and lead to the conclusion that hedge fund returns are non-normal.

Arbitrage-oriented absolute return strategies in particular exhibit returns that are clustered around the mean and deviate little from their expected value. On rare occasions, disruptive events lead to extreme negative returns and ensuing unexpected losses for the investor. As in the case of the downfall of the Bear Stearns structured credit funds in 2007, excessive leverage and insufficient credit insurance in an environment of tightening liquidity can lead to greatly amplified losses. The resulting frequency distribution of returns exhibits the characteristic heavy left-sided tails and excess kurtosis.

[14] Financial literature suggests that investors prefer positive skewness (i.e. decreased probability for extreme losses). Agarwal and Naik (1999), however, found that negative skewness warrants higher returns. The underlying assumption is that investors require risk premiums for negative skewness. It is questionable whether investors are able to correctly assess the required payment for higher moments of the distribution.

Table 6.1: Statistical properties of the frequency distribution of returns

	m	$\bar{\mu} \times 10^{-2}$	$\bar{\sigma} \times 10^{-2}$	$\bar{m}_3$	$\bar{m}_4$	$\bar{\chi}_{JB}$	$\%\chi_{JB}$	$\bar{d}_n$	$\%d_n$
CTA	38	1.281	7.180	1.026	9.222	2572.5	89.5	0.106	86.8
EH	121	0.879	3.107	-0.564	9.182	457.0	95.9	0.122	94.2
ED	524	1.046	5.446	0.231	7.496	370.9	91.0	0.096	82.3
M	226	1.053	5.134	0.607	6.635	341.0	77.9	0.089	72.6
O	11	0.741	3.183	-1.153	15.421	5401.3	90.9	0.122	90.9
RV	132	0.784	2.016	-1.419	17.438	3929.6	97.0	0.157	95.5
Total	1052	1.001	4.777	0.028	8.898	953.0	89.4	0.106	83.5

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is April 1990 to June 2010. Here n denotes the number of funds in the subsample as per main strategy, $\bar{\mu}$ and $\bar{\sigma}$ give the sample mean and standard deviation, $\bar{m}_3$ and $\bar{m}_4$ represent the third and fourth moment used in calculating the Jarque-Bera test statistic ($\bar{\chi}_{JB}$) and $\%\chi_{JB}$ is the proportion of sample hedge funds with significant p-values (95% confidence level) for the Jarque-Bera test, , $\bar{d}_n$ is the average Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test statistic for all funds within the sample, $\%d_n$ is the proportion of funds with significant p-values for the Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test (i.e. the proportion of funds for which the null assumption of a normal distribution is violated). Acronyms for strategies as reported in the databases are as follows: CTA = CTA/managed futures, EH = equity hedge, ED = event driven, M = macro, O = other, RV = relative value.

Table 6.2 provides the average autocorrelation coefficients for the original series, Table 6.3 displays the results after correcting for autoregressive and moving-average processes described in the method section. It is evident that most of the autocorrelation at the first three lags is removed. In addition, the number of sample hedge fund return series exhibiting cumulative significance of serial correlation at lags 4 through 12 is substantially reduced. It should be evident that significant autocorrelation may skew the results from factor analysis and may bias the initial covariance estimates between hedge fund return series.

Table 6.2: Serial correlation in consecutive returns

	$k=$											
CTA	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{p}_k$	0.025	-0.060	0.009	-0.027	0.000	0.004	-0.009	0.009	0.040	0.000	0.006	-0.010
$\hat{p}'_k$	0.025	-0.076	0.007	-0.045	-0.008	-0.019	-0.011	-0.014	0.033	-0.011	0.004	-0.027
$\%\chi_{LB}$	23.7	36.8	31.6	36.8	36.8	42.1	42.1	36.8	44.7	44.7	39.5	36.8
ED												
$\hat{p}_m$	0.276	0.123	0.097	0.073	0.050	0.028	0.055	0.039	0.010	0.017	0.008	0.009
$\hat{p}'_m$	0.276	0.036	0.046	0.016	0.011	-0.003	0.042	-0.001	-0.005	-0.001	-0.007	-0.008
$\%\chi_{LB}$	81.8	77.7	71.1	71.9	74.4	71.9	71.1	71.9	69.4	68.6	67.8	67.8
EH												
$\hat{p}_m$	0.156	0.050	0.039	0.028	-0.008	0.018	0.043	0.034	0.007	-0.006	0.005	-0.017
$\hat{p}'_m$	0.156	0.009	0.024	0.003	-0.019	0.012	0.041	0.006	-0.003	-0.017	0.007	-0.030
$\%\chi_{LB}$	50.2	50.4	48.5	47.1	46.0	47.3	48.1	46.2	46.0	44.8	44.7	45.6
M												
$\hat{p}_m$	0.063	-0.011	0.008	-0.022	0.013	-0.016	-0.003	0.024	0.035	0.017	0.010	-0.016
$\hat{p}'_m$	0.063	-0.029	0.004	-0.040	0.003	-0.033	-0.005	0.005	0.028	0.005	0.010	-0.025
$\%\chi_{LB}$	21.7	24.8	23.0	25.2	26.1	27.0	26.5	25.7	27.4	27.4	28.3	28.3
O												
$\hat{p}_m$	0.168	0.046	0.065	0.103	0.018	-0.009	-0.010	0.025	0.006	-0.031	0.002	-0.025
$\hat{p}'_m$	0.168	-0.003	0.039	0.062	-0.011	-0.010	-0.003	0.020	0.000	-0.034	0.017	-0.050
$\%\chi_{LB}$	54.5	63.6	54.5	63.6	63.6	54.5	54.5	54.5	54.5	54.5	54.5	54.5
RV												
$\hat{p}_m$	0.287	0.148	0.113	0.079	0.039	0.030	0.052	0.028	0.026	0.031	0.019	0.010
$\hat{p}'_m$	0.287	0.016	0.048	0.007	0.000	-0.010	0.028	-0.007	0.004	0.018	-0.009	-0.010
$\%\chi_{LB}$	81.8	81.1	78.8	78.0	78.0	78.0	79.5	78.8	78.8	77.3	75.8	75.8

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is April 1990 to June 2010. Here $\hat{p}_k$ denotes the average autocorrelation coefficient across the subsample at lag k, $\hat{p}'_k$ is the average partial autocorrelation coefficient at distinct lag k and $\%\chi_{LB}$is the proportion of funds with significant p-values for the Ljung-Box statistic (95% confidence).

Table 6.3: Serial correlation in adjusted consecutive returns

	$k =$											
CTA	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{p}_k$	-0.014	-0.028	-0.012	-0.010	-0.011	-0.011	-0.011	0.003	0.022	-0.004	0.001	-0.030
$\hat{p}'_k$	-0.014	-0.031	-0.013	-0.017	-0.015	-0.015	-0.009	-0.006	0.025	-0.004	0.000	-0.038
$\%\chi_{LB}$	0.0	5.3	2.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
EH												
$\hat{p}_m$	0.018	0.014	0.018	0.020	0.015	-0.005	0.031	0.014	-0.018	0.003	-0.007	-0.002
$\hat{p}'_m$	0.018	0.011	0.018	0.019	0.015	-0.010	0.029	0.002	-0.016	0.001	-0.008	-0.013
$\%\chi_{LB}$	3.3	0.8	2.5	4.1	4.1	5.0	5.0	7.4	6.6	7.4	5.0	8.3
ED												
$\hat{p}_m$	0.018	0.006	0.014	0.011	-0.016	0.005	0.032	0.019	-0.002	-0.012	0.002	-0.020
$\hat{p}'_m$	0.018	0.002	0.014	0.006	-0.016	0.001	0.030	0.010	-0.002	-0.017	0.003	-0.027
$\%\chi_{LB}$	1.5	2.5	2.3	3.1	4.0	5.5	6.7	6.7	6.7	6.3	7.1	8.0
M												
$\hat{p}_m$	0.008	-0.014	-0.003	-0.017	0.004	-0.023	-0.007	0.015	0.028	0.013	-0.001	-0.024
$\hat{p}'_m$	0.008	-0.017	-0.003	-0.024	-0.001	-0.031	-0.007	0.006	0.025	0.008	0.004	-0.029
$\%\chi_{LB}$	2.2	2.7	1.8	4.0	4.0	6.6	7.5	7.1	8.8	8.0	7.5	7.1
O												
$\hat{p}_m$	0.014	0.016	0.052	0.074	0.001	-0.021	-0.016	0.008	0.006	-0.041	0.012	-0.034
$\hat{p}'_m$	0.014	0.010	0.044	0.069	-0.002	-0.030	-0.026	0.006	0.008	-0.031	0.019	-0.049
$\%\chi_{LB}$	9.1	9.1	0.0	9.1	9.1	9.1	9.1	9.1	9.1	9.1	9.1	0.0
RV												
$\hat{p}_m$	0.002	0.004	0.018	0.019	-0.006	-0.016	0.023	0.001	-0.009	0.014	0.002	-0.003
$\hat{p}'_m$	0.002	0.002	0.019	0.016	-0.006	-0.022	0.023	-0.005	-0.007	0.016	-0.003	-0.009
$\%\chi_{LB}$	0.8	1.5	3.0	5.3	3.8	3.0	5.3	6.1	6.1	6.8	6.1	5.3

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is April 1990 to June 2010. Here $\hat{p}_k$ denotes the average autocorrelation coefficient across the subsample at lag k, $\hat{p}'_k$ is the average partial autocorrelation coefficient at distinct lag k and $\%\chi_{LB}$ is the proportion of funds with significant p-values for the Ljung-Box statistic (95% confidence).

The results for autocorrelation vary greatly depending on the style classification. Return-smoothing appears to be most prominent in event-driven and relative value funds. It is possible, however, that the observed serial correlation comes about as a result of trading in illiquid securities for which there is no secondary market. Many event-driven funds engage in private equity and venture capital investments that are inherently illiquid and require a long-term commitment. In addition, event-driven hedge funds place bets on the outcome of events that have an impact on the performance of a single asset or a specific sector. Fund managers either respond successfully to events that induce changes in the price of an asset or actively influence the outcome of such an event. It is often difficult to predict when such an event will come about. Hostile or friendly take-over attempts, acquisitions, spin-offs, mergers, insolvency of a company, or the restructuring of companies / part of a company in distress require a long-term commitment on behalf of the fund managers and, in consequence, aggravate the accurate pricing of the share-holding. Due to the uncertainty of the outcome of the event, it is at the fund manager's discretion to value such investments.

Relative value or absolute return funds, on the other hand, have a vested interest to make their return profile appear as smooth as possible. Fund managers are judged by their ability to hedge their exposure to economic factors and their ability to create factor neutral portfolios. The investment thesis is predicated on realisation of pricing discrepancies between related securities whilst hedging the market-related risks using derivative instruments and short-selling. Investors expect relative value funds to have low correlations with standard asset indices and to exhibit low historic volatility. Managing prices and smoothing returns is one way to increase the attractiveness of a hedge fund.

Since CTAs deal in standardised financial futures contracts and are overseen by national regulatory commissions, there is no room for stale or managed prices. Similarly, systematic commodity and currency macro funds focus on highly liquid futures markets. Systematic traders rely upon computer-generated trading signals and algorithmic models to identify markets with trending or momentum behaviour. Typically, this requires systematic hedge funds to focus on highly liquid instruments. Other macro funds engage in active trading methods with high frequency position turnover and short holding periods that involve trading in volatile but liquid equity markets. For the majority of CTA and Macro funds, serial correlation at the first lag is insignificant. Lastly, it can be observed from partial autocorrelation coefficients across style classifications that most serial correlation occurs at the first lag.

Autocorrelation is a good proxy of return-smoothing behaviour (compare Lo, 2001). An autoregressive moving-average $ARMA[p, q]$ model is estimated for all hedge fund series used in further analysis. The appropriate model is selected on the basis of the Schwarz Information Criterion (Schwarz, 1978). Upon observation of the observed autocorrelation and partial autocorrelation coefficients, the highest order model considered is an $ARMA[3, 3]$ model. In order to decide on the correctly specified model, all possible $ARMA[p, q]$ models for each hedge fund series of the sample were considered and the results of the information criteria were recorded in a matrix. The specific model minimizing the information criterion was selected. As outlined in in the meothodology section, the $SBIC$, $HQIC$ and AIC were recorded, Where the criteria suggested different models, the $SBIC$ would be used to decide on the model order. No further tests were applied to the residuals to test for the validity of the model. The residuals from OLS-regression, rather than the original series are used to estimate the squared multiple correlations in the matrix of Equation 4.19.

It is the purpose of the model to yield a residual series that is purged of effects from serial correlation and moving-average processes. It is not argued that the univariate model is fully specified. It is the underlying assumption that current return observations are a function of previous return observations, a combination of current and previous values of a white noise error term and an unspecified residual component. The residual series of OLS provides the appropriate estimator for cross-correlations between ‘unsmoothed’ hedge fund return series.

Correlations with other assets tend to be understated when the presence of autocorrelation in time series is ignored (see for example Kat & Lu, 2002). Similarly, autocorrelation may obscure the initial estimates for multiple squared correlations in factor models. The factor axis methodology relies on the covariance between hedge funds rather than the variance (as in PCA) to generate the initial factor estimates. Squared multiple correlations ($\widehat{h}_i^2$) are calculated from the inverse of the correlation matrix with unities in the diagonal. In the presence of serially correlated returns, $\widehat{h}_i^2$ will be a biased initial estimator of the communalities. In consequence, it is necessary to remove serial correlations from the variables entering the factor model.

Table 6.3 reveals (partial) autocorrelation coefficients and Ljung-Box statistics for various lags for the adjusted series. Correcting for the first three orders of autocorrelation and moving average processes not only removed any significant serial correlation at the respective lags, but also eliminated most of the residual autocorrelation at higher lags. The improvements were substantial across all six style classifications with only a small proportion of hedge funds exhibiting cumulatively significant autocorrelation coefficients at lags one through twelve. At higher lags, accounting for autocorrelation poses practical problems. For every additional lag considered, the residual series in factor analysis reduces by one observation. Fortunately, selecting the appropriate model according to $SBIC$ does not support models where the maximum lag length $p > 3$ for all but a few hedge funds (the initial model estimates considered a maximum lag length of $p = 12$).

All results were reproduced for the stand-alone HFR database. Results for the frequency distribution of returns are contained in Table A.4 of the appendix and autocorrelation coefficients are displayed in Table A.5. The results confirm the presence of serially correlated returns and non-normality.

6.3 Creating common factor portfolios

The classification according to principal factor extraction was expected to be an improvement over the self-selected classification of hedge fund managers in databases and over the results from PCA and classification according to the first k principal components. The common factor loadings were expected to be statistically significant:

the specific return component, whilst significant, does not fully explain the variation in the performance of the individual hedge fund.

The extracted factors explaining the communalities across hedge funds vary according to the number of observations T included. It is a reasonable assumption that longer time series yield classifications that are more robust and persistent. Similarly, larger samples of hedge funds were expected to require an increasing number of extracted factors explaining the covariance between the sample hedge funds. For example, the initial sample for the July 1990 to June 2000 included 129 single manager hedge funds meeting the selection criteria outlined below. For the final July 2000 to June 2010 window, that number increased to 612 funds. It is acknowledged that the hedge fund industry is growing and that an increasing number of hedge funds warrant additional style classifications as represented by a larger number of extracted common factors.

Factor axis and rotation were assumed to reveal style classifications that prevail over time. Assuming that style classifications are persistent, existing hedge funds could be expected to fall within the same classification as in previous periods. This was expected to be true in the long run despite style drift and phase-locking behaviour. Although hedge fund returns may become synchronised due to market disruptions or may dynamically shift their exposure to benefit from short-term arbitrage opportunities, it is expected that they revert back to a broad strategic theme after a short while. Conversely, comparing covariances between hedge funds over extended periods of time helps to identify those funds that have permanently departed from their initial main strategic focus.

For the timeframes under observation, some evolutionary development of the prevailing style classifications was expected as the number of hedge funds in the sample increased (i.e. as new funds entered the sample, they warranted their own classification). The decision to include an additional classification was unbiased since the selected procedure relied upon parallel analysis to decide on the number of common factors extracted. It was not possible to discern or influence the number of iterations or distinct categories required to account for all hedge funds in the sample *a* *priori*. The complexity of the factor model is a function of the existing communalities between hedge funds and the number of hedge funds in the initial sample. There were numerous criteria to determine the number of factors required. Horn's parallel analysis was chosen here (Horn, 1965). It is an unbiased estimator since the number of required factors is confirmed in Monte Carlo simulations. Other standard criteria such as Kaiser-Guttman or Broken Stick could have produced different results. Principal axis was deemed appropriate if a) some persistence was observed with respect to hedge funds belonging to one classification or another throughout their reporting history; and b) the main style classifications prevailed throughout different estimation windows.

After correcting for autocorrelation, the remaining sample of hedge funds was entered into the factor model. Upon manual inspection, three funds were removed from factor analysis due to inconsistencies in performance reporting. The historic time series suggest that the funds initially started out as a quarterly reporting fund and changed the reporting standards at a later stage. In consequence, the subtotal changed from $m = 1055$ to $m = 1052$ hedge funds in the sample.

The factor axes were rotated so that the majority of hedge funds maximised upon one factor loading only. All hedge funds in the sample were initially classified according to their maximum factor loading. Principal factor estimation was repeated for every subsample of hedge funds within a classification. In an iterative process, hedge funds within one classification were further categorised by their newly acquired rotated absolute factor loadings. This process was repeated until hedge funds within one category loaded on one factor only (and all factor loadings were positive). Within each classification, hedge funds loaded on one common factor f_t only. The unique return contribution ϵ_{it} could be large and significant. However, it represented a unique factor that had no significant impact on other hedge funds within the same classification. Put differently, all hedge funds within a classification were now described by exactly one common and one specific factor.

An example is given for the July 1990 to June 2000 estimation period: the 129 hedge fund series of the initial sample were explained by seven common factors and one factor unique to each hedge fund of the sample (the error term). In factor axis, this error term at t can be large and significant. However, it is unrelated to the specific factor component of other hedge funds in the same classification. Since funds can display positive as well as negative factor loadings (compare to equation 4.30 in section 4.3.5: the Varimax criterion tends to drive the factor loadings towards -1, 0 or 1), the maximum number of factor portfolios is 14. From the original sample, 43 hedge funds were found to maximise on the first factor loading. Second-level factor analysis yielded two common factors driving the performance of the 43 funds in the sub-sample (again, all hedge funds were sorted into two samples of 38 and 5). Third-level factor extraction yielded one common factor within each sample and thus the final statistical cluster for all hedge funds.

The same principle of iterative factor extraction was applied to all subsamples of the initial factor classification. A total of eight statistical clusters were required to identify homogenous classifications across the initial sample of 129 hedge funds. It stands to reason that some of these clusters contain very few funds that are not representative of a dominant strategic theme. Hence, only the largest of these clusters were retained, the remaining funds from smaller clusters were attributed to the retained clusters. The remaining funds were attributed to retained clusters on the basis of the significance of the coefficient when regressing the return series against the factor score of the existing portfolio. Funds were entered only if their inclusion did not require an additional factor to explain the communalities across the hedge funds within the portfolio. The statistical clusters are henceforth referred to as common factor portfolios (CFPs).

In the following estimation period, after allowing for hedge funds to exit the sample (derelict funds or funds that stopped reporting), factor extraction was repeated for the existing CFPs. As before, factor extraction was repeated until all hedge funds within one portfolio had loaded on one common factor only. If all hedge funds within a CFP required more than one common factor to explain the communalities between them, new CFPs were created for each such factor (as before, the CFPs were retained only if they represented a significant number of hedge funds from the sample). New funds entering the sample were attributed to the existing CFPs on the basis of their correlation with the existing factor scores. New factor portfolios were created if the resulting common factor was orthogonal to the existing factors and the CFPs were representative of a significant number of hedge funds.

Rolling-window estimation periods were used to create unbiased estimators of the CFP constituents. The results from factor extraction of the last 120 observations were used in predicting the CFP composition one month into the future. As an example, the July 1990 to June 2000 estimation period provided the CFP constituents for July 2000. To account for lockdowns and other trading restrictions, the results from the estimation periods were used as forecasts for the particular month following one month after the last observed month of the estimation period. For 120 windows, there were 120 estimates for the composition of each CFP. The initial estimate for July 2000 yielded five CFPs as strategic representations of the 129 hedge funds in the sample. As time progressed and more hedge funds entered, additional portfolios were required to incorporate the increasing strategic diversity of the funds in the sample. For the last July 2010 estimate, the initial CFPs were split several times to yield the final 10 CFPs.

All returns within a CFP were weighted and the components were rebalanced annually to create style indices (i.e. indices that are comprised of hedge fund return series that load on the same common factor). The portfolio constituent weights were selected so as to maximise the correlation of the weighted index with the extracted factor score for each estimation window. The common factors used to explain the covariance structure of the observed data were unobserved. However, the factor could be estimated from the observed data and the loadings of each variable using regression as the coefficient estimation method (for details see Gorsuch, 1974).

For 120 estimation periods, the resulting style index return series was comprised of up to $n = 120$ observations (July 2000 to June 2010). Since the first five CFPs split into 10 CFPs over the course of 120 months, some indices had the same initial performance history. It was decided to use some of the existing classification terminology to label the style indices. Indices were named in accordance with the prevailing reported hedge fund strategy within the CFPs, as well as the correlation of said hedge funds with strategy indices from the HFN and HFR database and results from the following regression in Table 6.4 and Table 6.5. Figure 6.1 below provides an overview of the 10 distinct classifications.

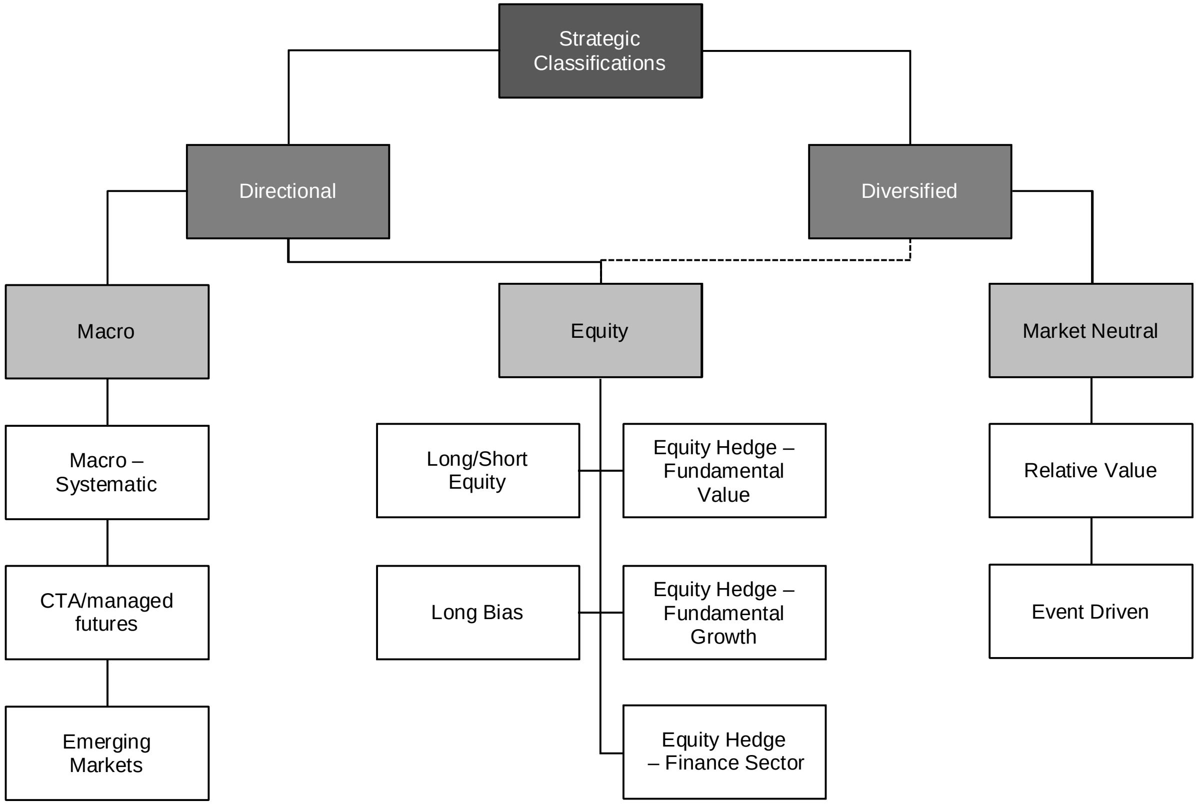

Figure 6.1: Style classifications as determined by factor analysis

The figure above is for the joint sample of the HFR and HFN databases. Classification as of July 2010.

The common factors from the principal factor axis methodology lacked any economic meaning or interpretability. However, in a reiterative framework of consecutive factor extraction and rotation, most single manager hedge funds from the sample would load on of a handful of unobsevable common factors explaining the co-variance between hedge funds loading on the same common factor. Clustering single manager hedge funds on the basis of their common factor loadings allowed for some initial estimation of the strategic theme represented by each common factor.

The reported strategies of hedge funds provided an estimate of the pre-dominant style whithin each group. (compare Table 6.4). In addition, it allowed for the creation of a weighted series, provided factor extraction is repeated for consecutive months, by viewing hedge funds from the same cluster / group as portfolio constituents and assigning weightings that sum up to one. This coined the term of CFP: the hedge funds belonging to a strategic cluster could be viewed as making up a portfolio, the constituents of which loaded on the same common factor. Once the weighted index series had been created, the index series could be regressed against asset-based factors to improve upon economic interpretability (Table 6.5).

Figure 6.1 uses some of the same terminology and labels used to describe hedge fund strategies in HFR and HFN. Whilst it was argued here that self-classification is arbitrary and inconsistent (refer to section 3.4.1 on style drift and section 5.2 on the differences between HFR and HFN), a certain degree of style convergence was expected within the CFPs. It may also have helped to identify reported strategies that are empirically consistent as well as persistent throughout time. Table 6.4 displays the two most common main strategies as well as sub-strategies within each CFP. It is important to note that Table 6.4 represents a snapshot of the latest estimation period July 2000 to June 2010. The composition of the CFPs changed on every re-balancing date. This is due to funds exiting the dataset (derelict funds or funds that stop reporting), style drift or increasing diversity of trading strategies and an overall larger dataset.

Table 6.4: Strategy classifications within CFPs

CFP	Main Strategy 1	Main Strategy 2	$n1$	$n2$	Total
LS	Equity Hedge	Event-Driven	61	3	68
EHv	Equity Hedge	Event-Driven	46	10	63
EHg	Equity Hedge	Macro	41	6	54
CTA	Macro	CTA/managed futures	34	10	50
RV	Equity Hedge	Event-Driven	16	16	50
L	Equity Hedge	Event-Driven	40	6	49
M	Macro	CTA/managed futures	28	5	48
ED	Event-Driven	Relative Value	20	11	37
EM	Equity Hedge	Macro	19	5	36
EHf	Equity Hedge	Event-Driven	26	1	31

CFP	Sub Strategy 1	Sub Strategy 2	$n1$	$n2$	Total
LS	Fundamental Growth	Long/Short Equity	22	12	68
EHv	Fundamental Value	Special Situations	22	8	63
EHg	Fundamental Value	Fundamental Growth	19	8	54
CTA	Systematic Diversified	CTA/managed futures	24	10	50
RV	Equity Market Neutral	Convertible Arbitrage	7	7	50
L	Fundamental Value	Long/Short Equity	12	7	49
M	Systematic Diversified	CTA/managed futures	17	5	48
ED	Distressed/Restructuring	Multi-Strategy	7	6	37
EM	Fundamental Growth	Multi-Strategy	13	3	36
EH	Fundamental Value	Finance Sector	17	3	31

The table contains the self-reported strategic classifications as they appear in HFR and HFN. The number of funds within a CFP belonging to a particular reported classification is a snapshot of the latest estimation period (July 2000 to June 2010). With every re-balancing date, some changes in the proportions of reported strategies within each CFP are expected. CTA = CTA/managed futures, ED = event driven/distressed securities, EH – Finance = equity hedge - FV finance sector, EH – Growth = equity hedge – fundamental growth, EH – Value = equity hedge – fundamental value, EM = emerging markets, LS = long/short equity – quantitative directional, L = long bias, M = macro system/trend, RV = relative value.

As expected, the CFPs did not match the self-reported style classifications perfectly. The reasons are threefold: firstly, hedge funds may deviate from their reported style due to style drift. Secondly, hedge funds often employ more than one sub-strategy but only report under one. Lastly, there are some overlaps between strategies: for example, a

CTA fund invested in currency futures may report as CTA/managed futures or, alternatively, macro – currency – discretionary. Some style classifications, however, are more likely to appear within particular CFPs than others. Thus, the strategy terminology from HFR and HFN were used as distinct classifications. However, hedge funds belonging to the macro CFP do not necessarily adhere to the definitions of a macro – trend-following fund in the strictest sense, but they share some common traits with macro funds (for example some CTAs are more closely related to macro hedge funds than other CTAs). It was found that the CTA classification of the joint sample is closely related to the Macro –Currency/Commodity classification of the HFR sample. The terminology was used as a means of convenience rather than precise identification of the underlying common strategy. The following regression analysis sheds some light onto the factual exposure of each CFP and associated index series.

6.4 Index regression results

The style index returns from July 2000 to June 2010 period were regressed against a number of asset-based factors and simulated portfolios representative of particular trading strategies. Lagged factors were included and the error estimates were heteroskedasticity-autocorrelation-consistent (HAC). Due to the large number of factors considered (see Table 5.1), and to avoid multicollinearity, a stepwise forward regression algorithm was employed. The aggregate results for the 10 CFPs are displayed in Table 6.5.

Table 6.5: Regression results for CFPs

CFPs	m	R^2	$\hat{R}^2$	F	k
CTA	50	57.2	51.8	$10.604^{\cdots}$	14
ED	36	85.5	84.7	$107.810^{\cdots}$	7
EHf	29	76.0	74.9	$70.295^{\cdots}$	6
EHg	54	79.3	77.6	$45.584^{\cdots}$	10
EHv	112	89.3	88.5	$112.360^{\cdots}$	9
EM	31	80.1	78.3	$42.784^{\cdots}$	11
L	48	90.1	89.2	$107.722^{\cdots}$	10
LS	68	85.4	83.9	$55.837^{\cdots}$	12
M	37	52.3	47.3	$10.479^{\cdots}$	12
RV	50	76.0	72.9	$25.054^{\cdots}$	14

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is July 2000 to June 2010. All error estimates are HAC at every step of the regression algorithm. Significance for overall fit as follows: $^{\cdot}$ denotes significance at 10% level, $^{\cdot\cdot}$ denotes significance at 5% level, and $^{\cdots}$ denotes significance at 1% level. The initial critical value of the F distribution for F-to-enter = 3.9 and F-to-remove = 3.8 (this corresponds roughly to a significance level of 5% for 120 observations). The intercept is always entered. The timeframe under consideration is July 2000 to June 2010 (n = 120 observations). Here k = number of coefficients including the intercept and $\hat{R}^2 = R^2$ adjusted for k. F denotes the joint significance for the regressors entered. CTA = CTA/managed futures, ED = event driven/distressed securities, EH – Finance = equity hedge - FV finance sector, EH – Growth = equity hedge – fundamental growth, EH – Value = equity hedge – fundamental value, EM = emerging markets, LS = long/short equity – quantitative directional, L = long bias, M = macro system/trend, RV = relative value.

Table 6.6: Regression coefficients

CTA_t	$= -0.002 + 0.\dddot{0}54PTFSFX_t - 0.\dddot{2}67SMB_{t-1} + 0.\dddot{0}50PTFSCOM_t + 0.\ddot{8}11BABDIDX_{t-2} + 0.\dddot{2}13GOLDUSD_t + 0.\dddot{2}42HML_t - 0.\dddot{1}17GSCI_{t-1} + 0.\ddot{2}47HML_{t-1} + 0.\ddot{0}24PTFSIR_{t-1} - 0.\ddot{4}08USDIDX_t - 0.\ddot{0}39PTFSCOM_{t-1} + 0.\dot{1}28WML_t + 0.\ddot{0}84MSCIEM_t$
ED_t	$= 0.\ddot{0}03 + 0.\dddot{1}46MSCIEM_t + 0.\dddot{1}94MSCIUS_t - 0.\ddot{0}07PTFSIR_{t-2} - 0.\dddot{4}50USDIDX_t + 0.\dddot{1}35SMB_t - 0.\dddot{0}29CBOEVIX_t$
EHf_t	$= 0.\dddot{0}07 + 0.\dddot{1}16MSCIEM_t - 0.\dddot{0}57CBOEVIX_t - 0.\ddot{0}20CBOEVIX_{t-1} + 0.\dddot{0}34PTFSSTK_t - 0.\dddot{0}07PTFSIR_{t-2}$
EHg_t	$= 0.003 + 0.\dddot{1}25MSCIEM_t - 0.\dddot{0}26CBOEVIX_t + 0.\dddot{3}48BABDIDX_{t-2} + 0.\dddot{1}73MSCIEXUS_t + 0.\dot{0}12PTFSFX_t + 0.\ddot{0}16PTFSCOM_{t-3} + 0.\dddot{1}45SMB_t + 0.\ddot{0}55GOLDUSD_t - 0.\ddot{1}86US10Y3M_t$
EHv_t	$= 0.\dddot{0}05 + 0.\dddot{1}32MSCIEM_t + 0.\ddot{2}44MSCIUS_t - 0.\ddot{0}08PTFSIR_{t-1} + 0.\dddot{1}55SMB_t - 0.\dddot{2}75USDIDX_t - 0.\ddot{0}35CBOEVIX_t + 0.\dddot{0}34PTFSSTK_t - 0.\ddot{0}10PTFSIR_t$
EM_t	$= 0.\ddot{0}07 + 0.\dddot{2}32MSCIEM_t + 0.\dddot{2}88BABDIDX_{t-2} - 1.\ddot{0}26USBAA10Y_t + 0.\ddot{8}78USBAA10Y_{t-3} + 0.\dddot{0}50WML_{t-2} + 0.\dddot{0}65GOLDUSD_t - 0.\ddot{0}31CBOEVIX_t - 0.\dot{0}82SMB_t + 0.\ddot{0}32PTFSSTK_t - 0.\ddot{0}11PTFSIR_t$
L_t	$= -0.\ddot{0}11 + 0.\dddot{1}60MSCIEXUS_t + 0.\dddot{3}39MSCIUS_t + 0.\dddot{1}72HML_t - 0.\ddot{0}22PTFSCOM_{t-2} - 0.\ddot{0}20CBOEVIX_t - 0.\ddot{0}17PTFSCOM_t + 1.\ddot{8}66USMO10Y_t + 0.\ddot{0}37MSCIEM_{t-2} - 1.\ddot{0}58USMO10Y_{t-3}$
LS_t	$= 0.\dddot{1}30MSCIEM_t + 0.\dddot{3}20MSCIUS_t + 0.\dddot{1}01WML_t + 0.\dddot{2}33SMB_t + 0.\ddot{0}47WML_{t-2} + 0.\dddot{3}65BABDIDX_{t-2} - 0.\ddot{0}64GOLDUSD_{t-2} + 0.\ddot{0}52GOLDUSD_{t-3} + 0.\dot{0}17PTFSSTK_{t-2} + 0.\dddot{0}52MSCIUS_{t-2} + 0.\ddot{0}05PTFSIR_{t-2}$
M_t	$= -0.001 + 0.\dddot{0}59PTFSFX_t - 0.\ddot{1}28SMB_{t-1} + 0.\ddot{1}88GOLDUSD_t + 0.\dddot{1}73HML_t - 0.\dddot{2}00MSCIEM_{t-3} + 0.\ddot{2}31MSCIEXUS_{t-3} + 0.\dddot{4}05BABDIDX_{t-2} - 0.\dddot{0}29CBOEVIX_t + 0.\ddot{1}44HML_{t-1} + 0.\ddot{0}17PTFSIR_{t-1} - 0.\ddot{0}32PTFSCOM_{t-1}$
RV_t	$= 0.\ddot{0}06 + 0.\dddot{0}55MSCIEM_t + 0.\ddot{0}61MSCIEXUS_{t-1} - 0.\dddot{0}19CBOEVIX_t - 0.\dddot{0}06PTFSIR_{t-3} + 0.\dddot{2}02BABDIDX_{t-3} + 0.\dddot{1}72USDIDX_{t-2} - 0.\ddot{0}06PTFSIR_{t-2} - 0.\ddot{0}06PTFSIR_t + 0.\ddot{0}09PTFSFX_{t-1} - 0.\dddot{1}29US10Y3M_{t-3} - 0.\ddot{0}30HML_{t-2} + 0.\ddot{1}48BABDIDX_t + 0.\ddot{0}49SMB_t$

Results are for the combined sample of the HFR and HFN databases for July 2000 to June 2010. Statistical significance is denoted by accents: $\dot{}$ denotes significance at 10% level, $\ddot{}$ denotes significance at 5% level, and $\dddot{}$ denotes significance at 1% level. Time indices reflect lagged coefficients (e.g. $t-3$ is the 3-month lagged exposure). All regressors are in the same order as they enter the equation. All acronyms of dependent variables and regressors according to Table 6.5 and Table 5.1.

Table 6.6 gives the regression function and statistical significance of the coefficient estimates. From $\hat{R}^2$ in Table 6.5, it is evident that asset-based factor models explain a significant proportion of the common variation in hedge fund index returns. In large portfolios, the specific return component is expected to be diversified away, leaving the common factor to impact on the performance of the style index. The results from Table 6.5 and Table 6.6 may suggest that the performance of some hedge fund portfolios is easily replicated.

There are, however, some important limitations: firstly, $\hat{R}^2$ describes the in-sample fit rather than the tracking or predictive ability of the model. Secondly, the model should be considered as a long-term equilibrium. The goodness-of-fit statistics differ significantly between different strategies. For CTAs and Macro funds in particular, the proposed models explain just half of the in-sample variation in returns. For Macro hedge funds, this is indicative of effective hedging against linear market exposures. The low explanatory power of the model for commodity traders indicates that they are exposed to different risk factors not encompassed by the passive indices and simplistic trading strategies of Table 6.6. Phase-locking behaviour and style drift are likely to necessitate conditional or regime-switching models to match the performance over shorter intervals. Thirdly, the proposed model is not a passive index or asset-class model. Some coefficients are difficult to interpret, in particular where lagged coefficients have different signs from contemporaneous coefficients. In addition, it is assumed that short selling is allowed and that all roll-over dates/maturities match the monthly reporting frequency of the style indices. Despite these simplifying assumptions, the regression results allow for an objective assessment of the prevailing investment philosophy in each classification.

Despite the limitations to track the performance of style indices, the results from the regression model may be used to more closely define the prevailing strategy / investment philosophy within CFPs. Additionally, the regression model allows for some general comments on the performance of hedge funds: the exposure of style indices to the five Fung – Hsieh PTFS dummies suggests that a large proportion of hedge funds are trend followers or employ a momentum strategy and that their performance can be matched using the PTFSs in conjunction with asset-class factors. Furthermore, most equity hedge funds are long- or short-biased and display significant directional exposure to equity and emerging market indices. Currency and commodity PTFS dummies are found to be good estimators of the performance of trend following CTAs. Some exposures are surprising in the context of the self-acclaimed strategies of the funds within the CFPs and are indicative of style drift.

The *equity hedge – fundamental value* (EHv) portfolio exhibits significant directional exposure to broad equity indices such as the MSCI US and the emerging markets Index. EHv strategies maintain long and short positions in equities and equity derivatives and use valuation matrices to identify companies which are inexpensive and undervalued when compared to relevant benchmarks. The Fama-French smb portfolio proxies for higher moments and nonlinear risk exposure (compare Chung *et al.*, 2006). The $cboevix$ coefficient (volatility proxy) suggests that these funds have a long bias in United States equity. Similarly, the *equity hedge - fundamental growth* (EHg) portfolio correlates with non-US equities and with the MSCI emerging markets index. The EHg strategy primarily focuses on exchange-traded companies with prospects for earnings growth and capital appreciation in excess of a proposed benchmark index.

The performance of the *long/short equity* (LS) portfolio correlates with equity indices, but the explanatory power of the model is lower compared to the results from *long only* (L). Exposure to the wml as well as $ptfsstk$ portfolios suggests that a number of equity hedge growth managers employ a momentum strategy of buying past winners and shorting past losers. Fund returns result from directional exposure to equity markets as well as spread trades on the stock market. Fung and Hsieh (2011) confirmed that price momentum and the spread between small and large capitalisation stocks (smb) explain most of the return variation and that LS funds are unlikely to deliver returns not related to asset factors. The *equity hedge – finance sector* (EHf) index exhibits some exposure to the same asset-based factors as the other equity hedge portfolios, however, with significantly lower explanatory power in regression. The volatility proxy explains a large proportion of the change in returns, implying that these fund managers trade in options markets to benefit from changes in intra-month volatility (e.g. using option straddles). This CFP represents the smallest portfolio of hedge funds and is likely representative of hedge funds with particular exposure to the financial services industry in the wake of the sub-prime lending crisis.

Macro system/trend hedge funds are characterised by trend-following behaviour as evidenced by their linear exposure to the Fung-Hsieh factors. Systems traders employ technical trading rules or rely upon mathematical and algorithmic models to identify investment opportunities in markets exhibiting trending or momentum characteristics. Similar to systematic macro funds, CTAs invest mainly in listed options and futures on commodities or currencies, often using a long-term trend-following strategy (for details on trend followers refer to Collins, 2003). There is some evidence of positive correlation with the gold price, suggesting that CTAs were able to capitalise on soaring precious metal prices in recent years. The exposure to fixed income markets can be explained by the required cash margins to trade on the futures exchange, which are invested in riskless bonds, as well as spread trades between the short and long end of the yield curve.

In contrast to trend-following funds, *emerging markets* portfolio returns correlate with movements of broad equity indices (changes in the MSCI Emerging Market index alone explain 68% of the in-sample variation in prices of the emerging markets portfolio). *Relative Value* funds show lower correlations with most asset-based factors and correspond most closely to the description of market-neutral funds. RV hedge funds use mathematical, fundamental or technical analysis to identify pricing discrepancies between related instruments including equity, fixed income, derivatives and other security types. As a sub-classification of relative value funds, fixed income corporate funds focus primarily on high-yield corporate bonds with low or no credit rating. This explains the strategy's correlation with movements of the yield curve and the fixed income markets. Both the $ptfsir$ and USD index are found to be highly significant in explaining changes in portfolio performance.

Like all market neutral strategies, *event driven* funds are less susceptible to price movements in equities. However, the portfolio reveals some exposure to emerging markets, which may be attributable to style drift and the recent higher-than-average performance of emerging markets funds. The inverse relationship between the USD index and portfolio performance confirms an investment strategy biased towards offshore equity investments. As with macro funds, the five Fung-Hsieh factors prove to be good estimators of the underlying investment strategy, albeit less significant in explaining overall portfolio performance.

6.5 Statistical properties of CFPs

All regression results were replicated for the sample consisting of HFR hedge funds only (see Table A.5 and Table A.6). The goodness-of-fit statistics as well as coefficient estimates are comparable to the results discussed above. The following Table 6.7 and Table 6.8 give the asymptotic properties and (partial) autocorrelations of the index series (Results for HFR in Table A.7 and Table A.8 of the appendix).

Table 6.7: Statistical properties of the frequency distribution of style index returns

	n	$\mu \times 10^{-2}$	$\sigma \times 10^{-2}$	m_3	m_4	χ_{LB}	$p_{\chi_{JB}}$	d_n	p_{d_n}
CTA	50	0.828	3.933	0.089	3.314	0.7	0.720	0.038	#N/A
ED	36	0.429	2.969	-1.315	6.922	112.5	0.000	0.086	0.027
EHf	29	0.467	2.114	-0.348	5.546	35.1	0.000	0.081	0.052
EHg	54	0.353	2.530	-0.433	3.373	4.5	0.106	0.069	#N/A
EHv	112	0.444	3.142	-1.093	5.855	65.2	0.000	0.077	0.077
EM	31	0.487	2.647	-0.894	7.046	98.7	0.000	0.115	0.001
L	48	0.454	2.882	-0.993	5.466	50.6	0.000	0.084	0.035
LS	68	0.328	2.632	-0.341	2.691	2.8	0.243	0.070	#N/A
M	37	0.394	2.568	0.252	3.582	3.0	0.224	0.061	#N/A
RV	50	0.516	1.159	-1.171	5.716	64.8	0.000	0.102	0.003

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is July 2000 to June 2010. Statistical significance is denoted by accents: $\dot{}$ denotes significance at 10% level, $\ddot{}$ denotes significance at 5% level, and $\dddot{}$ denotes significance at 1% level. The mean return is denoted by μ, standard deviation is σ, the third and fourth moment of the distribution are m_3 and m_4, χ_{JB} is the Jarque-Bera test statistic and $p_{\chi_{JB}}$ is the associated p-value, d_n is the Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test statistic and p_{d_n} is the associated p-value.

The expected performance varies depending on the index re-balancing intervals as well as the performance impact from defunct hedge funds. It was assumed that the full loss potential is realised in the periods leading up to the liquidation date and that investors receive the full residual value of their investment on the day that a fund stops reporting to the database vendor. The average performance across single manager hedge funds for the July 2005 to June 2010 interval was lower than for any of the previous reference periods. As a result, the weighted index performance across all style classifications was found to be comparatively low. All index series except event driven, macro systems/trends, cta and long/short equity exhibit statistically significant deviations from a normal distribution.

Additionally, testing for autocorrelation at cumulative lags yielded significant test statistics. Table 6.7 depicts the descriptive statistics for the ten CPPs of the HFR –HFN sample. The results of the mean-variance analysis are in line with expectations: relative Value funds show the lower expected returns but offer more stable returns over the 120 months under observation. Conversely, CTAs demand the highest risk premium. Confirming the results from single fund analysis, the index series for CTAs and macro funds exhibit no evidence for serially correlated returns, which is attributable to the highly liquid markets those funds trade in. However, effects from both non-normality as well as autocorrelation are partially reduced for all index series.

In summary, the statistical clusters of hedge funds are both meaningful as well as an improvement over the existing classifications. Principal axis as a dimensionality reduction technique greatly limits the number of statistical clusters representative of particular investment strategies in hedge funds. For the July 2000 to June 2010 sample, 84.2 percent of all sample hedge funds belonged to one of only ten CFPs. Conversely, roughly one in six hedge funds could not be attributed to any of the clusters. Recall that smaller clusters were omitted when creating the CFPs. Hedge funds belonging to those portfolios were attributed to larger clusters to the degree possible (i.e. as long as their inclusion did not cause the number of extracted common factors to increase). This left some funds unclassified that are not representative of a major strategic theme.[15] Hedge fund classification is an on-going process and an increasing number of hedge funds may warrant additional style classifications.

[15] Sun, Ashley and Zheng (2012) argued that 'strategic distinctiveness' is a performance criterion for hedge funds and indicative of superior investment skill. In the opinion of the authors, it is questionable whether the regressions used controlled for all fund characteristics that may have impacted on performance and that were related to distinctiveness (e.g. fund age, size, leverage etc.).

Table 6.8: Serial correlation in style indices

.	$k=$											
	1	2	3	4	5	6	7	8	9	10	11	12
CTA												
$\hat{p}_k$	-0.034	-0.174	-0.028	-0.053	-0.071	-0.199	0.015	0.123	0.158	-0.014	-0.013	-0.047
$\hat{p}_k'$	-0.034	-0.176	-0.042	-0.09	-0.095	-0.247	-0.059	0.017	0.138	-0.008	0.016	-0.079
χ_{LB}	0.143	3.952	4.051	4.414	5.067	10.170	10.201	12.185	15.511	15.538	15.561	15.863
ED												
$\hat{p}_m$	0.292	-0.013	0.073	0.17	0.006	-0.166	-0.061	-0.025	0.023	0.002	-0.007	-0.057
$\hat{p}_m'$	0.292	-0.107	0.12	0.122	-0.084	-0.144	0.013	-0.055	0.081	0.023	-0.007	-0.085
χ_{LB}	10.556	10.576	11.257	14.946	14.950	18.505	18.992	19.072	19.143	19.143	19.150	19.589
EHf												
$\hat{p}_m$	0.124	-0.158	-0.018	0.189	-0.024	-0.118	0.059	0.063	-0.094	-0.014	0.033	-0.011
$\hat{p}_m'$	0.124	-0.176	0.029	0.168	-0.08	-0.051	0.08	-0.018	-0.075	0.057	-0.03	-0.028
χ_{LB}	1.892	5.031	5.070	9.627	9.701	11.489	11.948	12.465	13.640	13.668	13.814	13.829
EHg												
$\hat{p}_m$	0.213	0.000	0.018	0.049	-0.034	-0.071	-0.013	-0.096	-0.107	-0.089	-0.047	-0.11
$\hat{p}_m'$	0.213	-0.048	0.03	0.04	-0.056	-0.052	0.011	-0.106	-0.061	-0.057	-0.027	-0.099
χ_{LB}	5.645	5.645	5.687	5.991	6.141	6.785	6.806	8.020	9.530	10.592	10.894	12.558
EHv												
$\hat{p}_m$	0.254	-0.05	0.074	0.193	0.008	-0.192	-0.029	-0.019	-0.043	-0.041	0.009	-0.029
$\hat{p}_m'$	0.254	-0.122	0.129	0.144	-0.077	-0.166	0.045	-0.088	0.019	0.027	0.008	-0.061
χ_{LB}	7.981	8.290	8.986	13.746	13.755	18.509	18.621	18.667	18.917	19.143	19.154	19.266
EM												
$\hat{p}_m$	0.235	0.154	0.055	0.214	-0.021	-0.061	0.002	-0.035	-0.154	-0.1	-0.123	-0.136
$\hat{p}_m'$	0.235	0.105	-0.002	0.199	-0.123	-0.087	0.056	-0.083	-0.124	0.006	-0.108	-0.078
χ_{LB}	6.876	9.856	10.241	16.050	16.104	16.589	16.590	16.753	19.910	21.254	23.317	25.86

.	$k =$											
	1	2	3	4	5	6	7	8	9	10	11	12
L												
$\widehat{p}_m$	0.201	-0.069	0.079	0.16	-0.039	-0.257	-0.027	0.074	-0.08	-0.082	0.065	0.102
$\widehat{p}_m^{'}$	0.201	-0.114	0.124	0.114	-0.088	-0.229	0.05	0.025	-0.054	0.019	0.048	0.014
χ_{LB}	$5.\ddot{0}20$	$5.\dot{6}13$	$6.\dot{4}03$	$9.\ddot{6}58$	$9.\dot{8}51$	$18.\dddot{3}95$	$18.\ddot{4}89$	$19.\ddot{2}12$	$20.\ddot{0}61$	$20.\ddot{9}53$	$21.\ddot{5}32$	$22.\ddot{9}58$
LS												
$\widehat{p}_m$	0.157	0.011	-0.009	0.039	-0.036	-0.076	0.076	0.015	-0.083	-0.043	-0.016	-0.068
$\widehat{p}_m^{'}$	0.157	-0.014	-0.009	0.043	-0.05	-0.064	0.102	-0.016	-0.086	-0.008	-0.02	-0.069
χ_{LB}	$3.\dot{0}51$	3.066	3.076	3.268	3.430	4.175	4.919	4.951	5.857	6.107	6.141	6.764
M												
$\widehat{p}_m$	-0.028	-0.193	-0.04	-0.052	0.016	-0.201	-0.01	0.128	0.091	-0.017	0.017	0.061
$\widehat{p}_m^{'}$	-0.028	-0.194	-0.053	-0.097	-0.009	-0.245	-0.042	0.025	0.073	-0.019	0.064	0.045
χ_{LB}	0.096	$4.\dot{7}42$	4.941	5.286	5.320	10.571	10.583	12.741	13.849	13.886	13.925	14.432
RV												
$\widehat{p}_m$	0.326	0.113	0.09	0.094	0.054	-0.105	-0.069	-0.039	0.083	0.036	-0.039	-0.121
$\widehat{p}_m^{'}$	0.326	0.008	0.057	0.053	0.004	-0.149	-0.001	-0.014	0.126	-0.006	-0.046	-0.138
χ_{LB}	$13.\dddot{1}54$	$14.\dddot{7}50$	$15.\dddot{7}62$	$16.\dddot{8}75$	$17.\dddot{2}56$	$18.\dddot{6}86$	$19.\dddot{3}01$	$19.\ddot{5}06$	$20.\ddot{4}17$	$20.\ddot{5}90$	$20.\ddot{7}97$	$22.\ddot{7}93$

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is July 2000 to June 2010. Statistical significance is denoted by accents: $\dot{}$ denotes significance at 10% level, $\ddot{}$ denotes significance at 5% level, and $\dddot{}$ denotes significance at 1% level. Autocorrelation coefficients at lag k are $\hat{p}_k$ and partial autocorrelation coefficients $\hat{p}_k^{'}$, χ_{LB} denote the significance of cumulative lags according to Ljung-Box.

Despite the aforementioned limitations, principal axis is a significant improvement over self-classification. Firstly, it identifies hedge funds that behave differently from their self-acclaimed strategy either due to trading restrictions or due to style drift. Secondly, it does not depend on managers reporting fund strategy to database vendors. Lastly, the common factor shared across hedge funds in one classification is a representation of a unique trading strategy that shares no common trait with other trading strategies (this is due to the orthogonality of the extracted factors). This is of particular interest for practitioners seeking to complement an existing portfolio with hedge fund investments and to compare diversification benefits across hedge fund classifications.

No evidence has been found to conclude that the distinct style classifications or their composition changed for estimation periods following the subprime lending crisis. This suggests that hedge funds within a particular classification reacted similarly to tightening liquidity following the demise of the Bear Sterns funds in 2007.[16] This is to be expected since hedge funds of a particular classification share some common traits with respect to financial gearing, leverage and hedge overlay. A high probability was observed for hedge funds belonging to one classification and then belonging to the same classification in the following period, irrespective of external shocks.

Principal factor axis yield better results compared to PCA, in particular where the number of principal components is larger than the number of observations ($k > n$) and asymptotic properties of the estimators do not hold. Furthermore, the explanatory power of the truncated component model is limited. In the example of Fung and Hsieh (1997), the five extracted style factors account for only 43 percent of the return variance across 409 hedge funds, despite a relatively short observation window of 36 months. This stems from small yet significant eigenvalues associated with extracted eigenvectors not included in the truncated model. Principal factors, on the other hand, acknowledge that

[16] It is important to understand that this does not indicate that no structural breaks were observed. This will be discussed in more detail on section 7.4.

part of a hedge fund's return variation is attributable to a unique component and seek to extract the communalities as defined by the covariance instead.

Considering the marginal differences between the eigenvalues, the truncated component model is arbitrary and of little statistical significance. Using the broken-stick methodology in PCA, additional extracted factors are discarded if their inclusion will not significantly improve the explanatory power of the model. One shortcoming of this approach results from discarded factors being jointly significant. Consequently, the dimensionality reduction comes at the cost of the lack of representativeness. Using parallel analysis in order to determine the number of nontrivial factors in principal axis prevents such selection bias. The methodology is also unbiased in contrast to earlier research due to the number of observations included (including 120 observations increased the confidence in establishing the long-term communalities between single-manager hedge funds across periods of financial distress as well as recovery).

A side-effect of hedge fund classification was the creation of pure style indices which could be used in further analysis. In order to assess the portfolio diversification benefits of different hedge fund strategies, it became necessary to reduce the dimensionality of the initial sample of single manager hedge funds by creating weighted indices that are both unbiased with respect to the constituent selection process as well as robust representations of distinct hedge fund investment styles. An arbitrary selection of single manager hedge funds or FoHFs to be used in further analysis would have raised concerns with regards to representativeness and generalized inferences from the results. Index creation presented the logical conclusion to retain the maximum amount of information whilst reducing the number of distinct datasets to a manageable number. The following chapter differentiates between commercial indices from the HFR and HFN database (see section 7.6) and the style indices from section 6.3.

CHAPTER 7: VECTOR MODEL AND COINTEGRATION

7.1 Introduction

In Chapter 6, factor axis methodology allowed for the identification of distinct style classifications and the construction of associated pure style indices. It is of interest to investors to identify hedge fund investment strategies that offer diversification benefits in the context of existing investment portfolios. The stepwise regression results from section 6.4 deliver an initial estimate of the relationships between hedge fund strategies and asset-based factors representative of various markets. The results indicate a lead-lag relationship between returns of broad asset indices and those of hedge fund indices.

The interaction between stock, bond and commodity markets is well established. For example, a rise in commodity prices results in increases in cost of goods. The increases in price are inflationary and interest rates rise to reflect the inflation. At higher interest rates the present value of bond payments to be received is lowered and bond prices decrease. With increasing costs of borrowing and higher risks due to inflation, stock performance is expected to deteriorate and investments become less attractive. Due to market imperfections and illiquidity of certain investments, there are response lags between each of the markets' reactions. It is reasonable to assume that changes in stock, bond and commodity markets also induce responses in alternative investments.

In accordance with EMH, it is assumed that the level of prices for different asset classes at time t reflects all available information and that past prices are poor estimators of future performance. Many economic time series, and asset index series in particular, are non-stationary but move in tandem over time, i.e. the series are bound by some long-term relationship. This equilibrium is expressed in the form of the EMH. The benefit of vector autoregressive models does not come about as a result of their ability to predict future outcomes, but rather as a result of describing the relationship between variables and their co-dependence. In the context of a well-diversified portfolio it is expected that each portfolio constituent is negatively correlated to or independent of contemporaneous as well as lagged effects from all other constituents.

One way to test this hypothesis is to construct a portfolio comprised of stock, bond and commodity investments and to complement such a portfolio with hedge funds. The portfolio weights are then chosen so as to minimise the variance of the weighted series for a particular target return, or alternatively, to maximise the return associated with a particular level of risk (as expressed through portfolio variance). However, this approach ignores the belated response times of markets due to frictions, market imperfections and illiquidity of certain assets. Additionally, it does not give any indication of the lead-lag relationships between assets. Lastly, it is not possible to differentiate between the susceptibility of portfolio constituents to their own past performance in comparison to their reactions to other markets.

Table 7.1 displays the results for portfolios comprised of a hedge fund index and the BABDIDX, GSCI and MSCIW indices. The objective function is the function that minimises the percentage deviation from an optimum for both a minimum-variance as well as maximum-performance portfolio. It is assumed that the investor assigns equal weighting to both objectives. The optimal portfolio weights are the variables and subjected to the sum-of-coefficient as well as non-negativity constraint (in the spirit of a strong style analysis, compare ter Horst *et al.,* 2004). The optimising algorithm was the Generalized Reduced Gradient (GRG) non-linear method and was applied to the return rather than the level index series. The selected model is appropriate since the objective function is non-linear and cannot be re-specified as a Simplex Linear Programming problem. The optimizations are run until the Karush-Kuhn-Tucker (KKT) conditions for a local optimum are satisfied (Karush, 1939; Kuhn & Tucker, 1951). The optimisation followed in three steps: maximising portfolio performance, minimising portfolio variance and minimising percentage deviation of both objectives (alternatively, a risk-return ratio such as Sharpe might have been used).

Portfolio composition is heavily impacted on by the sub-par performance across equities and selected commodities in the years following the financial crisis. As a result, the optimisation algorithm selects the global bond index as the major contributor to portfolio performance (both in terms of performance as well as risk diversification). With the exception of the CTA series, none of the other hedge fund indices outperform the returns on the BABDIDX (including GSCI and MSCIW). The performance-driven portfolio always selects BABDIDX with the exception of the portfolio including CTA. However, all portfolios include at least a small proportion of hedge fund investments to optimise investors' objectives from a risk-return perspective (i.e. inclusion of the hedge fund series decreases portfolio variance with minimal impact on overall performance).

Table 7.1: Exemplary portfolio composition (risk-return optimised)

	HF	BABDIDX	GSCI	MSCIW	$\mu \times 10^{-2}$	$\sigma \times 10^{-2}$	m_3	m_4
CTA	16.1%	83.9%	0.0%	0.0%	0.581	1.199	−0.209	0.741
ED	7.2%	92.8%	0.0%	0.0%	0.527	1.041	−0.792	2.312
EHF	14.8%	85.2%	0.0%	0.0%	0.524	0.987	−0.765	2.396
EHG	9.7%	90.3%	0.0%	0.0%	0.517	1.009	−0.678	2.133
EHV	6.8%	93.2%	0.0%	0.0%	0.528	1.044	−0.766	2.227
EM	9.4%	90.6%	0.0%	0.0%	0.530	1.034	−0.831	2.184
L	8.0%	92.0%	0.0%	0.0%	0.528	1.036	−0.795	2.194
LS	12.4%	88.5%	0.0%	0.0%	0.514	0.990	−0.758	2.186
M	7.7%	90.1%	1.0%	1.2%	0.515	1.037	−0.561	1.280
RV	40.5%	59.5%	0.0%	0.0%	0.527	0.830	−1.240	3.455

Results are for the combined sample of the HFR and HFN databases. The timeframe under consideration is July 2000 to June 2010. The mean portfolio return is denoted by μ, portfolio standard deviation is σ, the third and fourth moment of the distribution are m_3 and m_4.The objective function assigns equal weighting to the objectives of maximising return and minimising variance. Other acronyms are as in Table 6.5 and Table 6.6.

As discussed above, the optimisation model considers contemporaneous terms only. Implicitly, it was assumed that all information is reflected immediately in current return observations (i.e. there is no exposure of a series to its own lagged terms as well as lagged terms of the other portfolio constituents). It is shown in the following that such assumptions are not supported by the data.

In order to gauge the co-dependence of hedge funds in such a portfolio, a univariate model including exogenous variables could be used (compare section 6.4). Such a model assumes that there is no co-dependence between the exogenous variables. Put differently, all explanatory variables considered are *strictly* exogenous and their values are determined outside the regression equation. However, in the event where financial theory suggests that a relationship (or causality) exists between exogenous variables, the model is underspecified and the coefficient estimates are likely biased. Preferably, a set of structural equations differentiating between endogenous and exogenous variables is required that determines the coefficient estimates simultaneously.

As a generalisation of the simultaneous equation model, vector-autoregressive models are well suited to describe the relationship between portfolio constituents, since all variables entering the model are assumed to be endogenous. For stock, bond and commodity prices such a relationship is well established in economic theory. It is of interest how the 10 CFPs identified in Chapter 6 perform in the context of such a model and whether the assumption of endogeneity is justified. One indicator of the independence of hedge fund performance from other asset classes may be that the underlying assumptions of such a model are violated and that no relationship can be discerned between hedge fund investments and the performance of stock, bond and commodity proxies.

The drawback of multivariate models is the restrictions imposed in terms of the number of variables to be included in such a model. For every additional variable, the number of coefficients to be estimated increases exponentially. Considering the relatively few observations of $n = 120$ datapoints, additional variables quickly use up the available degrees of freedom in estimating model coefficients. This problem is aggravated when estimating the coefficients for the cointegrating relationships of the level series. Consequently, the multivariate analysis is limited to a stock, bond and commodity proxy in addition to a hypothesized hedge fund investment ($k = 4$ variables for all 10 models estimated).

In this chapter, the time series under consideration were the index level series rather than the period returns. Such economic series were expected to be non-stationary (a formal test for stationarity follows below). Upon determining the integration order of each series, the first question to answer was whether or not a cointegration relationship existed. In the case where the series were found to be non-stationary but no cointegration relationship existed, the appropriate model would have to employ specifications in first differences only. However, if portfolio constituents were cointegrated, it meant that the index levels would have a long-term relationship, which prevented them from wandering apart without bound. An error correction component was introduced to the vector autoregressive model to differentiate between a short-term lead-lag relationship and the long-term equilibrium.

The remainder of this chapter is structured as follows: Firstly, the unit root tests that were applied to the stock, bond and commodity level index series are described as well as each of the ten hedge fund indices. The series in each group were tested for cointegration relationships using several specifications and different lag structures in a Johansen framework. Upon determining the number of cointegrating equations, the appropriate VECMs, as selected by information criteria, were subjected to the testing of coefficient restrictions on adjustment coefficients and vector parameters. The model statistics and (multivariate) diagnostics for the individual equations as well as the joint model are given in a separate table to assess the appropriateness of the models selected. Lastly, Granger causality block significance tests, impulse response tests and variance decomposition reveal the nature of the relationship between the variables of the models.

7.2 Stationarity and cointegration

Table 7.2 gives the results of various unit root tests. The three test statistics of the confirmatory analysis are the Augmented Dickey-Fuller, Philips-Perron and Kwiatkowski–Phillips–Schmidt–Shin test statistic. The top half of the panel provides the results for including an intercept and the bottom half the results for including an intercept and a time trend in the test regression equation. The left quadrants show the outcome for first-differenced terms or evidence for a $\sim I(1)$ process, the right quadrants confirm that the series contain a single unit root by testing for higher orders of integration. The tests were conducted for the stock (MSCIW), bond (BABDIDX) and commodity (GSCI) index series as well as each of the hedge fund indices.

The left half of the panel shows that, for any index level series entering the model, the null hypothesis of a unit root cannot be rejected. With the exception of the macro index, this is true whether a time trend is included in the regression or not. The results from testing for stationarity in $KPSS$ confirm the results from unit root testing for all models without time trends. The results are less pronounced for models including a deterministic trend. However, for most series there is at least weak evidence to reject the stationarity hypothesis with the exception of the EHf and EM series. As expected, first-differencing removed the unit root from all series. There was no evidence to infer that the series are $\sim I(2)$. Where tests were inconclusive, it was likely attributable to the relatively small sample of continuous observations. Using the return series rather than the index levels will remove any unit root from the original series. In that sense, computing period returns is similar to first-differencing of the series. However, in doing so one foregoes the information contained within the cointegrating relationship of non-stationary processes.

For the VECM, it was assumed that all series entering the model are non-stationary. Firstly, there was no conclusive evidence to reject a unit root *and* no evidence in favour of non-stationarity for any of the index series. Secondly, of the four variables entered into the multivariate models, at least three presented with overwhelming evidence of non-stationarity. Before linear combination of the non-stationary series in a VECM, it was necessary to test for the order of integration of such a combination. The tests were conducted in a Johansen framework.

Table 7.2: Results for unit root tests

	Δy_t			$\Delta^2 y_t$		
	$H_0: y_t \sim I(1)$		$H_0: y_t \sim I(0)$	$H_0: y_t \sim I(2)$		$H_0: y_t \sim I(1)$
μ	ADF_τ	PP_τ	$KPSS_{LM}$	ADF_τ	PP_τ	$KPSS_{LM}$
BABDIDX	0.058	0.044	1.301	$-9\dddot{.}912$	$-10\dddot{.}623$	0.133
GSCI	-2.064	-2.060	0.530	$-8\dddot{.}216$	$-8\dddot{.}315$	0.099
MSCIW	-1.738	-1.807	0.327	$-8\dddot{.}230$	$-8\dddot{.}359$	0.115
CTA	-0.736	-0.469	1.254	$-11\dddot{.}046$	$-12\dddot{.}077$	0.077
ED	-1.034	-0.978	1.047	$-7\dddot{.}846$	$-7\dddot{.}910$	0.075
EHf	0.797	0.331	1.246	$-8\dddot{.}566$	$-8\dddot{.}322$	0.148
EHg	-0.642	-0.596	1.187	$-8\dddot{.}500$	$-8\dddot{.}491$	0.087
EHv	-0.979	-0.972	1.060	$-8\dddot{.}035$	$-8\dddot{.}077$	0.071
EM	-0.666	-0.783	1.151	$-8\dddot{.}167$	$-8\dddot{.}424$	0.061
L	-1.181	-1.151	1.078	$-8\dddot{.}449$	$-8\dddot{.}473$	0.063
LS	-0.499	-0.674	1.123	$-9\dddot{.}099$	$-9\dddot{.}113$	0.089
M	-2.102	-2.192	1.273	$-9\dddot{.}582$	$-12\dddot{.}734$	0.268
RV	-0.537	-0.626	1.288	$-7\dddot{.}603$	$-7\dddot{.}666$	0.060
	Δy_t			$\Delta^2 y_t$		
	$H_0: y_t \sim I(1)$		$H_0: y_t \sim I(0)$	$H_0: y_t \sim I(2)$		$H_0: y_t \sim I(1)$
$\mu + \lambda t$	ADF_τ	PP_τ	$KPSS_{LM}$	ADF_τ	PP_τ	$KPSS_{LM}$
BABDIDX	-1.885	-1.779	0.167	$-9\dddot{.}886$	$-10\dddot{.}591$	0.120
GSCI	-2.011	-1.965	0.191	$-8\dddot{.}216$	$-8\dddot{.}314$	0.046
MSCIW	-1.940	-1.984	0.145	$-8\dddot{.}208$	$-8\dddot{.}333$	0.111
CTA	-2.955	-3.017	0.251	$-10\dddot{.}996$	$-11\dddot{.}988$	0.074
ED	-2.080	-2.033	0.147	$-7\dddot{.}815$	$-7\dddot{.}879$	0.077
EHf	-2.084	-2.384	0.080	$-8\dddot{.}718$	$-8\dddot{.}294$	0.063
EHg	-2.774	-2.723	0.124	$-8\dddot{.}492$	$-8\dddot{.}479$	0.054
EHv	-2.190	-2.127	0.142	$-8\dddot{.}000$	$-8\dddot{.}041$	0.072
EM	-2.379	-2.568	0.086	$-8\dddot{.}148$	$-8\dddot{.}399$	0.048
L	-2.165	-2.080	0.153	$-8\dddot{.}407$	$-8\dddot{.}432$	0.061
LS	-2.460	-2.457	0.110	$-9\dddot{.}084$	$-9\dddot{.}096$	0.065
M	$-4\dddot{.}287$	$-3\dddot{.}875$	0.109	$-9\dddot{.}662$	$-13\dddot{.}559$	0.092
RV	-2.135	-2.093	0.130	$-7\dddot{.}567$	$-7\dddot{.}630$	0.057

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The table includes three test statistics from confirmatory analysis: Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin ($KPSS$). The intercept is denoted by μ, the time trend is λt, Δ^k is the k-differenced term of y_t, H_o the null hypothesis for the test statistic and $I(k)$ an integrated process of order k. $KPSS$ critical values at the 5% level are $LM_{.05} = 0.463$ (μ) and $LM_{.05} = 0.146$ ($\mu + \lambda t$). Other acronyms and designation of statistical significance are according to Table 6.5 and Table 6.6.

Table 7.3: Results for the Johansen cointegration test

CTA								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	37.402	1	0	2	yes	yes	yes
	SBIC*	37.620	1	0	0	yes	yes	yes
Δy_{t-1}	AIC	37.435	0	0	1	yes	yes	yes
	SBIC	37.921	0	0	0	yes	yes	yes
Δy_{t-2}	AIC	37.331	0	1	1	yes	yes	yes
	SBIC	38.293	0	1	0	yes	yes	yes
Δy_{t-3}	AIC*	37.257	1	1	1	yes	yes	yes
	SBIC	38.667	1	1	0	yes	yes	yes
ED								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.899	1	1	2	yes	yes	yes
	SBIC*	35.121	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	34.917	0	0	1	no	yes	yes
	SBIC	35.474	0	0	0	yes	yes	yes
Δy_{t-2}	AIC*	34.803	1	0	2	no	yes	yes
	SBIC	35.776	0	0	0	yes	yes	yes
Δy_{t-3}	AIC	34.836	1	1	2	yes	yes	yes
	SBIC	36.202	1	1	1	no	no	yes
EHf								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.688	1	1	2	yes	yes	yes
	SBIC*	34.946	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	34.673	1	0	2	yes	yes	yes
	SBIC	35.285	1	0	0	yes	yes	yes
Δy_{t-2}	AIC*	34.633	1	1	2	yes	yes	yes
	SBIC	35.659	1	1	1	no	no	yes
Δy_{t-3}	AIC	34.676	1	1	2	yes	yes	yes
	SBIC	36.117	1	1	1	yes	yes	yes
EHg								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.628	1	0	2	yes	yes	yes
	SBIC*	34.853	1	0	0	yes	yes	yes
Δy_{t-1}	AIC	34.735	0	0	1	yes	yes	yes
	SBIC	35.212	0	0	0	yes	yes	yes
Δy_{t-2}	AIC	34.617	0	1	1	yes	yes	yes
	SBIC	35.594	0	1	0	yes	yes	yes
Δy_{t-3}	AIC*	34.609	1	1	2	yes	yes	yes
	SBIC	36.011	1	1	0	yes	yes	yes

EHv								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC*	34.753	1	1	2	yes	yes	yes
	SBIC*	35.013	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	34.796	0	1	1	no	yes	yes
	SBIC	35.362	0	0	0	yes	yes	yes
Δy_{t-2}	AIC	34.784	1	1	2	no	yes	yes
	SBIC	35.758	0	0	0	yes	yes	yes
Δy_{t-3}	AIC	34.780	1	1	2	yes	yes	yes
	SBIC	36.158	1	1	1	no	no	yes
EM								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC*	34.988	2	1	2	yes	yes	yes
	SBIC*	35.331	2	1	0	yes	yes	yes
Δy_{t-1}	AIC	35.099	1	0	2	yes	yes	yes
	SBIC	35.718	1	0	0	yes	yes	yes
Δy_{t-2}	AIC	34.992	1	1	2	yes	yes	yes
	SBIC	35.988	1	1	0	yes	yes	yes
Δy_{t-3}	AIC	35.067	1	1	2	yes	yes	yes
	SBIC	36.491	1	1	0	yes	yes	yes
L								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.717	1	1	1	yes	yes	yes
	SBIC*	35.018	1	1	1	yes	yes	yes
Δy_{t-1}	AIC	34.756	0	1	1	yes	yes	yes
	SBIC	35.345	0	1	0	yes	yes	yes
Δy_{t-2}	AIC	34.753	1	1	2	yes	yes	yes
	SBIC	35.785	1	1	0	yes	yes	yes
Δy_{t-3}	AIC*	34.644	1	2	2	yes	yes	yes
	SBIC	36.076	1	2	1	no	yes	yes
LS								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC*	34.474	1	1	2	yes	yes	yes
	SBIC*	34.786	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	34.605	0	0	2	no	yes	yes
	SBIC	35.116	0	0	0	yes	yes	yes
Δy_{t-2}	AIC	34.543	1	1	2	no	yes	yes
	SBIC	35.537	1	1	0	yes	yes	yes
Δy_{t-3}	AIC	34.570	1	1	2	yes	yes	yes
	SBIC	35.986	1	1	0	yes	yes	yes

M								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	36.108	1	1	1	yes	yes	yes
	SBIC*	36.320	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	36.077	1	1	2	yes	yes	yes
	SBIC	36.675	1	1	0	yes	yes	yes
Δy_{t-2}	AIC*	36.024	1	2	2	yes	yes	yes
	SBIC	37.061	1	2	0	yes	yes	yes
Δy_{t-3}	AIC	36.056	1	2	2	yes	yes	yes
	SBIC	37.501	1	2	0	yes	yes	yes
RV								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	33.936	1	0	2	yes	yes	yes
	SBIC*	34.143	1	0	0	yes	yes	yes
Δy_{t-1}	AIC	33.936	0	0	2	yes	yes	yes
	SBIC	34.484	0	0	0	yes	yes	yes
Δy_{t-2}	AIC	33.855	1	1	2	yes	yes	yes
	SBIC	34.883	1	1	0	yes	yes	yes
Δy_{t-3}	AIC*	33.803	1	1	2	yes	yes	yes
	SBIC	35.253	1	1	1	no	no	yes

Results are for HFR and HFN database for the July 2000 to June 2010 timeframe. All cointegration tests for groups consisting of hedge fund index and $BABDIDX$, $GSCI$ and $MSCIW$. The table shows the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBIC) for several test specifications and different number of lags considered for the differenced series. Here $*$ denotes the selected model according to AIC and SBIC, λ_{trace} and λ_{max} denote the number of cointegrating relations according to trace and maximum eigenvalue respectively, Δy_{t-k} is the k-differenced series for all variables. Trend and Intercept are for CE. Other acronyms are according to Table 6.5 and Table 6.6.

Table 7.3 gives the aggregate results for various specifications within the Johansen framework. The appropriateness of the model selected was determined using either the Akaike or Schwarz Bayesian Information criterion. The five models under consideration are:

- no intercept or trend in cointegration equation (CE) or VAR
- intercept in CE and no intercept in VAR
- intercept and no trend in CE and VAR
- intercept and trend in CE, no intercept in VAR

- intercept and trend in CE, intercept in VAR.

None of the cointegrating tests suggest an intercept for the test VAR (the information is omitted from Table 7.4 for spatial reasons). In addition to the intercept and trend specifications, it was necessary to determine the appropriate lag length for the VAR using the same information criteria. The tests were conducted up to a lag length of 12. However, since no model specification suggests a lag length of $k > 3$, the results displayed in Table 7.3 are limited to the first three lags. Note that in a VECM the lagged series of the VAR are the first-differenced terms rather than the level series (Δy_{t-k}).The row heading provides the highest lag-order term of the VAR. Thus, Δy_{t-3} implies that Δy_{t-2} and Δy_{t-1} are included in the model. The panel headings denote the fourth variable (or hedge fund index) added to BABDIDX, GSCI and MSCIW series (as an example: the CTA-panel provides the cointegration results for a model consisting of the CTA, BABDIDX, GSCI and MSCIW series).

It is apparent from Table 7.3 that the two information criteria select the same model and underlying deterministic trend assumption irrespective of the number of lags in the VAR. The test results also confirm that a cointegration relation exists between the variables of the model (as evidenced by trace and maximum eigenvalue), despite different results for the rank of the cointegration equation. If all variables are cointegrated, then u_t (error term) will be a stationary process $I(0)$. The unit root rest applied to the residuals has the following null and alternative hypotheses: $H_0{:}\,\hat{u}_t{\sim}I(1)$ and $H_1{:}\,\hat{u}_t{\sim}I(0)$. If the null hypothesis is not rejected, there is no co-integration. The appropriate strategy for modeling in this case would be to employ specifications in first differences only. Such models would have no long-term equilibrium solution. This would not matter since no cointegration implies that there is no long-run relationship anyway. The lag specification $1\ to\ 3$ is confirmed by lag exclusion Wald tests (the test results are not displayed in Table 7.3, however, the results are consistent throughout the ten models under consideration).

7.3 Model restrictions

It was decided to include the first three lags of the differenced term in each equation as well as an equation describing the cointegrating relationship (or the error correction term). The rank of the cointegration equation was two for all models except for CTA, for which it was one. Each model included an intercept and trend in the CE. Model restrictions for the cointegrating relationship were tested prior to assessing the quality and appropriateness of the selected model. Restrictions were imposed on the adjustment coefficients of the cointegration relationship, the parameters of the cointegrating vector, as well as both (the adjustment coefficient measures the proportion of the last period’s equilibrium error that is corrected for). The results are displayed in Table 7.4, Table 7.5 and Table 7.6.

The test is a likelihood-ratio test for coefficient restrictions; the test statistic is a chi-squared distributed. The null hypothesis states that the imposed restrictions are binding (i.e. a small enough p-value indicates that the model improves significantly when relaxing the constraints). In the following three tables, a bold font indicates that restrictions may be imposed without significantly impairing the explanatory power of the model. Conversely, statistical significance as denoted by accents indicates that the restrictions are not supported by the data. More than one restriction could be imposed at the same time, which may or may not be jointly significant.

The left column in each table indicates the restrictions imposed, the bottom panel describes the associated coefficient restrictions. Note that for models with only one CE, the number of estimated coefficients halves (and so does the number of imposed restrictions). Because of the number of possible restriction combinations, the results in Table 7.6 are limited to those with the largest associated p-value of the test statistic (and ranked accordingly). The top panel denotes the restrictions or combinations thereof that are supported from the results of all three tables (bold face).

Note that not rejecting the null hypothesis of binding restrictions does not mean a cointegration relationship does not exist. It implies that individual equations for a variable do not include a particular cointegrating vector (restrictions on α_{ij}), or that some variables do not appear in the cointegration equation (restrictions on β_{ij}). However, the restrictions imposed should allow for some initial analysis of the long-term relationship between the variables. The coefficient restrictions in Table 7.4 through Table 7.6 are as follows:

- $\alpha_{11} = 0(, \alpha_{12} = 0)$: Adjustment coefficient(s) restrictions for the cointegrating component (s) of the hedge fund index equation
- $\alpha_{21} = 0(, \alpha_{22} = 0)$: Adjustment coefficient(s) restrictions for the cointegrating component(s) of the BABDIDX index equation
- $\alpha_{31} = 0(, \alpha_{32} = 0)$: Adjustment coefficient(s) restrictions for the cointegrating component(s) of the GSCI index equation
- $\alpha_{41} = 0(, \alpha_{42} = 0)$: Adjustment coefficient(s) restrictions for the cointegrating component(s) of the MSCIW index equation
- $\beta_{11} = 0, \beta_{12} = 0, \beta_{13} = 0, \beta_{14} = 0(, \beta_{21} = 0, \beta_{22} = 0, \beta_{23} = 0, \beta_{24} = 0)$: Restrictions imposed on the parameters within the cointegrating vector(s) where the second subscript 1 through 4 denote the hedge fund BABDIDX, GSCI and MSCIW index series.

Table 7.4: Testing restrictions for the adjustment coefficients of the cointegrating relationship

χ^2	CTA	ED	EHf	EHg	EHv	EM	L	LS	M	RV
(4)	6.822	6.768	8.740	18.962	4.801	15.387	**3.184**	18.551	**2.363**	4.863
(3)	**0.831**	**2.282**	6.476	**2.336**	**3.315**	**0.993**	5.027	**0.517**	**0.142**	**2.293**
(3)(4)	6.849	8.934	16.367	23.826	**7.700**	20.471	7.955	24.737	**2.540**	**6.563**
(2)	22.690	23.223	23.641	19.302	23.139	19.086	15.698	17.738	21.837	24.197
(2)(4)	26.674	27.017	35.406	43.081	25.940	38.263	16.840	43.100	24.036	27.692
(2)(3)	22.702	25.949	37.182	26.431	27.063	22.120	24.961	23.443	27.745	30.759
(2)(3)(4)	29.546	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
(1)	**0.008**	**4.420**	13.493	17.549	**4.031**	20.152	8.153	13.141	18.181	4.690
(1)(4)	7.087	14.029	14.025	20.394	13.499	31.871	19.134	21.215	20.111	15.524
(1)(3)	**1.117**	**6.198**	26.309	18.748	**6.938**	25.145	12.820	18.863	24.324	**6.687**
(1)(3)(4)	7.271	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
(1)(2)	22.750	24.955	46.424	37.815	25.906	39.971	22.782	38.897	46.329	26.445
(1)(2)(4)	30.970	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
(1)(2)(3)	22.861	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
	1 CE	2 CE								
(1)	$\alpha_{11} = 0$	$\alpha_{11} = 0, \alpha_{12} = 0$								
(2)	$\alpha_{21} = 0$	$\alpha_{21} = 0, \alpha_{22} = 0$								
(3)	$\alpha_{31} = 0$	$\alpha_{31} = 0, \alpha_{32} = 0$								
(4)	$\alpha_{41} = 0$	$\alpha_{41} = 0, \alpha_{42} = 0$								

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. All exclusion tests for models including hedge fund index and BABDIDX, GSCI and MSCIW . CE is the number of cointegrating equations selected according to Table 7.3, χ^2 is the relevant chi-squared distributed test statistic and α_{ij} is the coefficient of the cointegrating vector for the ith variable in the jth cointegrating relation. Other acronyms and designation of statistical significance are according to Table 6.5 and Table 6.6.

Table 7.5: Testing restrictions for the parameters in the cointegrating vector

χ^2	CTA	ED	EHF	EHG	EHV	EM	L	LS	M	RV
(8)	2.721	9.700	10.310	7.384	10.181	6.990	20.795	8.824	5.831	11.769
(7)	17.889	14.956	13.725	17.173	15.558	22.273	8.392	14.721	11.540	20.093
(7)(8)	26.611	31.948	46.505	28.942	33.983	33.614	38.649	32.373	33.682	44.438
(6)	7.884	14.640	17.302	8.096	13.404	5.367	12.514	6.289	13.465	12.924
(6)(8)	7.919	19.827	23.748	23.099	21.253	26.284	34.698	23.882	17.255	27.951
(6)(7)	24.111	27.093	46.130	30.367	29.013	33.492	24.197	29.569	26.592	35.378
(6)(7)(8)	28.677	41.440	48.867	33.963	42.773	39.349	49.735	40.664	34.857	51.541
(5)	4.242	7.307	7.608	15.581	8.535	10.610	18.255	14.109	18.705	11.003
(5)(8)	5.923	16.813	19.145	20.611	17.968	19.387	28.983	23.877	26.730	22.392
(5)(7)	18.365	32.098	34.463	33.674	34.404	37.902	41.249	36.163	37.914	42.787
(5)(7)(8)	35.359	48.829	57.976	49.890	51.363	53.051	57.143	54.028	54.667	57.867
(5)(6)	15.452	18.689	23.815	29.372	20.392	24.256	32.512	28.246	31.102	25.957
(5)(6)(8)	16.302	29.094	32.231	35.157	30.536	39.835	42.197	37.741	38.386	40.180
(5)(6)(7)	28.196	43.641	48.457	44.765	45.468	48.911	53.584	48.227	48.808	53.698
	1 CE	2 CE								
(5)	$\beta_{11}=0$	$\beta_{11}=0, \beta_{21}=0$								
(6)	$\beta_{12}=0$	$\beta_{12}=0, \beta_{22}=0$								
(7)	$\beta_{13}=0$	$\beta_{13}=0, \beta_{23}=0$								
(8)	$\beta_{14}=0$	$\beta_{14}=0, \beta_{24}=0$								

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. All exclusion tests for models including hedge fund index and BABDIDX, GSCI and MSCIW. CE is the number of cointegrating equations selected according to Table 7.3, χ^2 is the relevant chi-squared distributed test statistic and β_{ij} represents the jth coefficient in the ith cointegrating relationship. Other acronyms and designation of statistical significance according are to Table 6.5 and Table 6.6.

Table 7.6: Testing for combined restrictions on adjustment coefficient and vector parameters

p_{χ^2} rank	CTA	ED	EHF	EHG	EHV	EM	L	LS	M	RV
1	(1)	(3)	(3)	(3)	(3)	(**3**)	(**4**)	(**3**)	(3)	(3)
2	(1)(3)	(**1**)(**3**)	(5)	(**3**)(**8**)	(**1**)(**3**)	(6)	(3)(4)	(6)	(3)(4)	(**3**)(**4**)
3	(3)	(1)	(4)	(8)	(1)	(3)(8)	(3)	(3)(6)	(4)	(1)(3)
4	(1)(8)	(3)(4)	(3)(8)	(3)(6)	(3)(4)	(3)(6)	(4)(7)	(8)	(**3**)(**4**)(**8**)	(1)
5	(**1**)(**3**)(**8**)	(4)	(3)(5)	(6)	(4)	(8)	(1)	(3)(8)	(3)(8)	(4)
6	(1)(3)(5)	(5)	(1)(4)	(1)(3)(8)	(1)(4)(5)	(3)(5)	(7)	(1)(8)	(4)(8)	(1)(4)(8)
7	(3)(5)	(1)(4)(5)	(8)	(1)(4)(5)	(1)(4)(8)	(5)	(1)(3)	(1)(3)(8)	(8)	(1)(4)(5)
p_{χ^2} rank	CTA	ED	EHF	EHG	EHV	EM	L	LS	M	RV
1	**0.008**	**2.282**	6.476	**2.336**	**3.315**	**0.993**	**3.184**	**0.517**	**0.142**	**2.293**
2	**1.117**	**6.198**	7.608	**7.710**	**6.938**	5.367	7.955	6.289	**2.540**	**6.563**
3	**0.831**	**4.420**	8.740	7.384	**4.031**	8.735	5.027	11.329	**2.363**	**6.687**
4	**2.740**	8.934	13.912	11.984	**7.700**	9.742	10.848	8.824	**9.200**	4.690
5	**5.323**	6.768	13.944	8.096	4.801	6.990	8.153	13.136	**6.681**	4.863
6	**5.878**	7.307	14.025	18.894	14.277	14.448	8.392	13.276	8.144	16.869
7	**4.323**	14.381	10.310	20.651	15.769	10.610	12.820	19.415	5.831	17.094
	1 CE	2 CE								
(1)	$\alpha_{11}=0$	$\alpha_{11}=0, \alpha_{12}=0$								
(2)	$\alpha_{21}=0$	$\alpha_{21}=0, \alpha_{22}=0$								
(3)	$\alpha_{31}=0$	$\alpha_{31}=0, \alpha_{32}=0$								
(4)	$\alpha_{41}=0$	$\alpha_{41}=0, \alpha_{42}=0$								
(5)	$\beta_{11}=0$	$\beta_{11}=0, \beta_{21}=0$								
(6)	$\beta_{12}=0$	$\beta_{12}=0, \beta_{22}=0$								
(7)	$\beta_{13}=0$	$\beta_{13}=0, \beta_{23}=0$								
(8)	$\beta_{14}=0$	$\beta_{14}=0, \beta_{24}=0$								

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. All exclusion tests for groups consisting of hedge fund index and BABDIDX, GSCI and MSCIW . CE is the number of cointegrating equations and the results are ranked by the associated p-value of χ^2. Other acronyms and designation of statistical significance areaccording to Table 6.5 and Table 6.6.

From Table 7.4 and Table 7.6 it can be inferred that there is no long-term cointegrating relationship binding the index levels of GSCI to any of the other series. With the exception of the model including the EHf and the L series, all other models suggest that the adjustment coefficient(s) of the cointegration equation for GSCI can be restricted to zero without impairing the explanatory quality of the overall VEC model. For the model including L, there is at least some indication to support the restrictions on the adjustment coefficients. Additional adjustment parameter restrictions are supported for the MSCIW series.

When testing for joint restrictions in Table 7.6, however, the chi-squared statistic does not support the retention of the restriction except for the L, M and RV models. Interestingly, both models for CTA and ED support adjustment coefficient restrictions for the hedge fund index series. This may be some indication that no long-term equilibrium state can be discerned between the CTA / ED indices and the asset indices. This is not to say that the index levels are independent of one another. The VAR component of the VEC model may still reveal a susceptibility to system shocks as expressed in the coefficients of the lagged first-differenced series. In addition, the hedge fund index series may still be relevant in explaining changes in any of the other index series.

Most coefficient estimates for the parameter vectors are found to be significant in the VEC model. However, Table 7.6 reveals that MSCIW may be removed from the cointegrating equations for the CTA, EHg and M variables. This underlines the long-run independence of these strategies from equity markets. While it can be assumed that a relationship exists between the markets in which CTAs and macro funds trade and broad equity indices (i.e. they respond to the same impact factors such as the interest rate), a lead-lag relationship is not easily established. In part this can be attributed to the complexity of the trades conducted (spread trades, options on options).

7.4 Model statistics and diagnostics

The restrictions as depicted in Table 7.6 are imposed when deciding on the appropriate VEC model. Aggregated results of the goodness-of-fit statistics for the individual equations as well as the system are provided in Table 7.7. Note that the Johansen cointegration tests for the VEC including the CTA series suggest a cointegrating relationship of rank one. The upshot is that the number of coefficients estimated reduces by four (thus increasing the degrees of freedom). Multivariate adaptations of the $SBIC$ and AIC statistics, whilst not meaningful by themselves, allow for comparison between models. Extensive diagnostics are included in the appendix, both for the individual series as well as the model (Table A.14 – Table A.23). The interpretations of the results in Table 7.6 are subject to the implications from estimating simultaneous equations: since inclusion of a particular variable impacts on all endogenous variables simultaneously, it is impossible to make *ceteris paribus* inferences.

Table 7.7: Aggregate results for selected VECM

CTA	eq_1	eq_2	eq_3	eq_4	System
F	1.326	4.653	2.670	2.752	
R^2	0.142	0.368	0.250	0.256	
$\bar{R}^2$	0.035	0.289	0.157	0.163	
AIC	6.727	5.092	14.932	10.730	37.303
SBIC	7.056	5.420	15.261	11.059	38.735
k	14	14	14	14	56
ED	eq_1	eq_2	eq_3	eq_4	System
F	2.621	5.916	1.759	2.472	
R^2	0.263	0.446	0.193	0.252	
$\bar{R}^2$	0.162	0.370	0.083	0.150	
AIC	5.530	4.977	15.023	10.753	34.888
SBIC	5.883	5.329	15.375	11.105	36.532
k	15	15	15	15	60

EHf	eq_1	eq_2	eq_3	eq_4	System
F	4.038	3.938	3.095	2.914	
R^2	0.354	0.349	0.296	0.284	
$\bar{R}^2$	0.267	0.260	0.200	0.186	
AIC	4.787	5.138	14.886	10.709	34.676
SBIC	5.139	5.491	15.238	11.061	36.320
k	15	15	15	15	60
EHg	eq_1	eq_2	eq_3	eq_4	System
F	2.392	4.384	2.116	3.649	
R^2	0.245	0.373	0.223	0.332	
$\bar{R}^2$	0.143	0.288	0.118	0.241	
AIC	4.992	5.100	14.985	10.640	34.674
SBIC	5.345	5.452	15.337	10.992	36.318
k	15	15	15	15	60
EHv	eq_1	eq_2	eq_3	eq_4	System
F	2.862	5.255	1.877	2.327	
R^2	0.280	0.417	0.203	0.240	
$\bar{R}^2$	0.182	0.337	0.095	0.137	
AIC	5.658	5.028	15.010	10.768	34.838
SBIC	6.010	5.380	15.362	11.120	36.482
k	15	15	15	15	60
EM	eq_1	eq_2	eq_3	eq_4	System
F	4.104	4.011	1.682	4.084	
R^2	0.358	0.353	0.186	0.357	
$\bar{R}^2$	0.271	0.265	0.075	0.270	
AIC	5.397	5.132	15.031	10.601	35.076
SBIC	5.750	5.484	15.384	10.953	36.720
k	15	15	15	15	60
L	eq_1	eq_2	eq_3	eq_4	System
F	3.123	3.773	3.286	2.577	
R^2	0.298	0.339	0.309	0.259	
$\bar{R}^2$	0.203	0.249	0.215	0.159	
AIC	5.563	5.153	14.868	10.742	34.671
SBIC	5.915	5.505	15.220	11.094	36.315
k	15	15	15	15	60

LS	eq_1	eq_2	eq_3	eq_4	System
F	4.050	4.153	1.961	4.855	
R^2	0.355	0.361	0.210	0.398	
$\bar{R}^2$	0.267	0.274	0.103	0.316	
AIC	4.927	5.119	15.001	10.536	34.574
SBIC	5.279	5.472	15.353	10.888	36.218
k	15	15	15	15	60
M	eq_1	eq_2	eq_3	eq_4	System
F	3.103	3.866	2.086	2.628	
R^2	0.297	0.344	0.221	0.263	
$\bar{R}^2$	0.201	0.255	0.115	0.163	
AIC	5.205	5.145	14.988	10.737	36.134
SBIC	5.557	5.497	15.340	11.089	37.778
k	15	15	15	15	60
RV	eq_1	eq_2	eq_3	eq_4	System
F	2.861	5.420	2.127	2.260	
R^2	0.280	0.424	0.224	0.235	
$\bar{R}^2$	0.182	0.346	0.119	0.131	
AIC	3.922	5.015	14.983	10.775	33.859
SBIC	4.274	5.367	15.336	11.127	35.503
k	15	15	15	15	60

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. Here eq_1 represents the equation for the hedge funds index and eq_2, eq_3 and eq_4 are the equations for BABDIDX, GSCI and MSCIW, respectively. $\bar{R}^2 = R^2$ adjusted for k (estimated coefficients), F denotes the joint significance for the regressors entered. Other acronyms are according to Table 6.5 and Table 6.6.

Despite the difficulties of interpreting the results for one particular equation of the VECs in Table 7.7, it is possible to provide some general statements about the quality of the model. It is evident that the models including the CTA and M series produce the poorest results as evidenced by the multivariate Baysian estimators. There are, however, some substantial differences for the goodness-of-fit statistics of the single equations. The lagged first-differenced coefficients and cointegrating equations do not explain any of the changes in index level for CTA. Conversely, changes in the index series LS, EM and EHf are at least partially explained in the context of the VEC model. Note that the purpose of the VEC was not to improve upon the predictive power of established models but to identify the dependencies and lead-lag relationship between variables.

The residual series for all the models exhibit significant evidence of heteroskedasticity and deviations from the normal distribution functions (see Table A.14 through Table A.43), whilst serial correlation of consecutive residuals is less of a concern. This is the upshot of using index levels rather than period returns to estimate the VEC. For bigger values of the dependent variable, the residuals are expected to get bigger. For the log-transformed series the multiplicative error is transformed into an additive error, regardless of the values of the dependent series. Univariate normality is not needed for the least-square estimates of the regression parameter (see Gauss-Markov theorem), i.e. the coefficient estimates are still consistent albeit not unbiased: confidence interval estimates and hypothesis testing have better statistical properties if the variables exhibit multivariate normality. Another assumption of the linear model states that the variance must be the same for each possible expected value (homoscedasticity).

Assuming the VEC models in Table 7.7 are correctly specified, heteroskedasticity in the residuals as a result of non-stationarity is not a concern (the coefficient estimates are still unbiased in the presence of heteroskedasticity). However, when observing the residual clusters, there appears to be some evidence of structural breaks as a result of the global financial crisis. Recursive regression reveals outliers for the two months following the Lehman Brothers bankruptcy in September 2008. The VEC models were reproduced introducing a dummy variable that is equal to one for September and October 2008 and equal to zero for all other months, effectively removing the outliers. The results are presented in Table A.24 through Table A.33 of the appendix. They confirm that removal of the outliers invoked homoskedasticity and approximate normality in the residual series. However, the dummy variable artificially increases the goodness-of-fit for the individual equations as well as the overall model. The following analysis of causality and impulse response / variance decomposition is limited to the models without dummy.

7.5 Causality and impulse response/variance decomposition

One problem in determining causality and impulse responses results from simultaneous estimation of model coefficients: If one variable induced changes in another variable of the multivariate model, such change resulted in an immediate feedback. Thus, it was required to make explicit assumptions about the ordering of model variables or, put differently, to make some *a priori* estimates about the lead-lag relationship between variables. For the remainder of this section, it was assumed that changes in the hedge fund index series followed from changes in the BABDIDX, GSCI and MSCIW index series rather than *vice versa*.

Granger causality tests for the significance of each lagged variable in the equations for every endogenous variable of the simultaneous equation system. The initial specifications of the VECMs stated that all variables are endogenous. The joint significance of all other variables was used to determine strict exogeneity for any particular variable (i.e. from the Wald statistic there is no evidence to infer causality). It is noteworthy that Granger causality tests for the correlations between variables and other lagged endogenous terms but does not necessarily mean that one variable induces changes in another. The test indicates a chronological ordering of the model variables and is useful in identifying *potential* lead-lag relationships between them. Since the models estimated are VECs, the causality tests are for the first-differenced terms only, not for the level terms of the cointegrating equation.

In Table 7.8, variables in rows denote the predictor variables whereas differenced variables arranged in columns are the dependents. From Table 7.8, there is little evidence for strict exogeneity for any variables in the models (i.e. at least one other endogenous variable induces changes in the dependent variable) except for the ones including CTA and EHg. It was found, however, that the first-differenced series of the CTA index series was significant in explaining the changes in GSCI. This was not surprising since the bulk of commodity trades are not physical trades but are conducted in the futures market. Similarly, the EHg series explained a significant proportion in the changes of the BABDIDX series. It is unlikely to infer a lead-lag relationship between the hedge and index series and the global bond market. More likely, the index series reacted to the same external factor albeit at different intervals.

Assuming that the causality tests are indicative of a lead-lag relationship, it seems as if managed futures (as designated by the CTA series) are independent of the changes in the level series for the three other asset indices. The same could be said for growth-oriented equity hedge (EHG) and, to some degree, for macro funds (M). Since there are no explicit exogeneity / endogeneity assumptions in the VEC model, the coefficient estimates and significance thereof are unbiased.

Table 7.8: Granger causality and block significance

	D(CTA)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(EM)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(CTA)		11.924	8.687	1.626	D(EM)		7.813	0.762	0.595
D(BABDIDX)	4.834		1.457	6.336	D(BABDIDX)	2.342		1.641	5.595
D(GSCI)	3.897	16.942		3.517	D(GSCI)	11.795	13.977		7.595
D(MSCIW)	0.785	8.956	4.537		D(MSCIW)	3.931	5.048	2.779	
all	11.037	30.072	16.126	12.752	all	18.693	21.233	6.660	18.280
	D(ED)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(L)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(ED)		25.379	1.210	1.308	D(L)		4.521	15.054	5.623
D(BABDIDX)	7.755		1.202	9.294	D(BABDIDX)	6.445		2.159	5.763
D(GSCI)	4.596	18.624		4.983	D(GSCI)	5.476	11.474		4.543
D(MSCIW)	3.836	12.490	2.256		D(MSCIW)	9.933	1.348	18.744	
all	13.996	44.144	9.109	13.738	all	23.324	19.969	26.820	14.942
	D(EHF)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(LS)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(EHF)		9.351	4.689	6.132	D(LS)		7.872	4.216	4.728
D(BABDIDX)	5.229		1.944	5.213	D(BABDIDX)	1.491		3.326	3.060
D(GSCI)	7.912	12.648		7.350	D(GSCI)	3.414	13.360		9.263
D(MSCIW)	11.025	3.098	5.022		D(MSCIW)	11.585	6.561	7.761	
all	23.419	24.373	11.691	16.152	all	17.973	23.258	11.332	19.715
	D(EHG)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(M)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(EHG)		10.399	5.967	4.691	D(M)		5.482	5.324	6.183
D(BABDIDX)	5.856		3.097	6.162	D(BABDIDX)	2.521		2.755	8.280
D(GSCI)	2.181	16.251		5.276	D(GSCI)	6.635	10.442		2.123
D(MSCIW)	5.790	8.958	6.230		D(MSCIW)	1.810	6.046	6.845	
all	14.226	27.878	15.526	19.795	all	12.504	20.504	14.668	15.981
	D(EHV)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(RV)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(EHV)		18.936	1.072	2.262	D(RV)		20.747	5.927	0.173
D(BABDIDX)	8.528		1.265	7.876	D(BABDIDX)	6.886		2.676	6.690
D(GSCI)	4.538	18.749		5.438	D(GSCI)	2.773	17.349		3.448
D(MSCIW)	5.960	10.363	3.645		D(MSCIW)	8.767	6.961	6.573	
all	16.420	36.201	9.603	12.469	all	15.675	39.511	14.951	10.732

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. Acronyms and significance are according to Table 6.5 and Table 6.6.

For the hypothesised portfolios of hedge funds and other investments, the causality tests delivered an initial estimate for the hedge fund index series providing the best diversification benefits: managed futures/CTA showed no exposure to the lagged series of equity, bond or commodity proxies. Conversely, long-short equity (LS and L) as well as emerging markets (EM) exhibited significant lagged exposure to these markets.

Impulse response tests are dependent on the implicit ordering of the variables in the model. Impulse responses refer to a unit shock to the errors of one of the four variables (in this case that of the hedge fund index level series): Since the errors of the four variables could not be assumed to be independent, the common component of the errors was attributed to the first Cholesky-ordered variable (orthogonalised impulse response). Cholesky ordering in Eviews uses the inverse of the Cholesky factor of the residual covariance matrix to orthogonalise the impulses. The hedge fund index series was designated as the first variable in testing whether movements in the hedge fund variable were likely to follow movements in the other asset indices. Note that EViews does not produce confidence bands for impulse reponse tests of VECMs.[17]

Likewise, variance decomposition differentiates between the proportion of movements in the dependent variable attributable to their own shocks versus shocks to the other variables of the model, providing information on the relative importance of each random innovation in affecting the VEC variables. Each variable would be affected by shocks to its own series but the shock would also be transmitted to all other variables of the system for which the series was an explanatory variable. The Cholesky ordering had a significant impact on the outcome of the tests since the first period decomposition for the first VEC variable was due entirely to its own innovation (it is expected that own series shocks explain the most of the error variance of each series in the VEC). All results for impulse response and variance decomposition implying different Cholesky ordering have been omitted here for spatial reasons but are available on request.

[17] For a detailed description on the determination of confidence bands refer to Lütkepohl (2013).

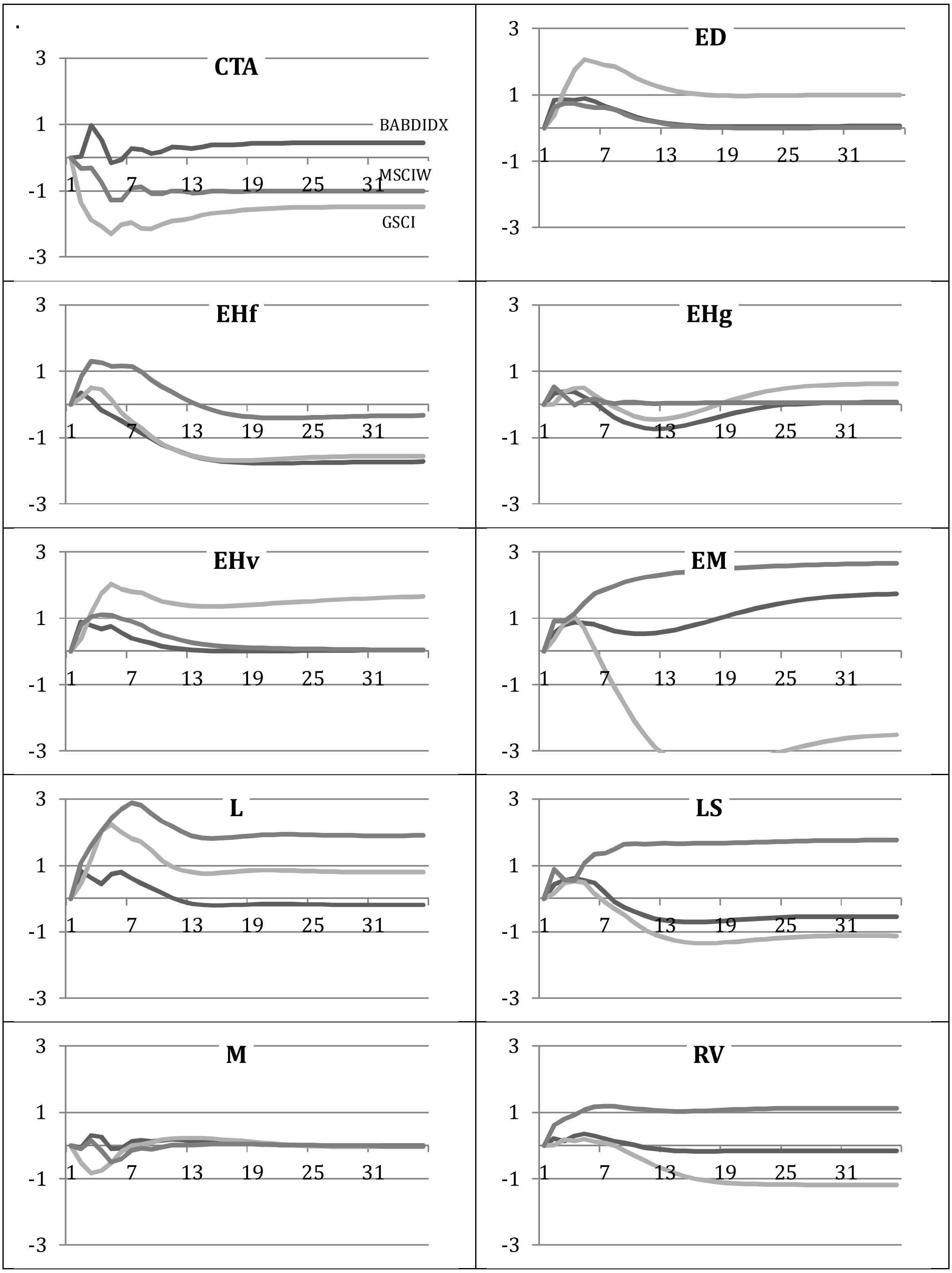

Figure 7.1: Impulse response test

Cholesky ordering: Hedge fund index, BABDIDX (dark grey), GSCI (light grey) and MSCIW (grey). Acronyms are according to Table 6.5 and Table 6.6.

The interpretation of the impulse response is somewhat difficult in the context of VAR models including an error term. For stationary VARs, the shocks are expected to die out to zero and the accumulated responses should asymptote to some non-zero constant. Since the variables in the models are jointly cointegrated, the model appears unstable (i.e. the shock is persistent). However, it is still possible to see at what point the variables return to their equilibrium. The charts in Figure 7.1 express the impact from shocks to any of the three index series BABDIDX, GSCI and MSCIW to the ten hedge fund series. The x-axis gives the time periods until equilibrium is achieved. Figure 7.1 confirms the results from Granger causality: Shocks to any of the other variables has little impact on the index series for CTA and M and the dependent series settle down to a steady state. For all other hedge fund strategies, the one standard deviation shock takes longer to work through the system. Generally speaking, the markets cannot be seen as acting independently from one another and, in contrast to EMH, information is passed through the system at a slow rate. The significance of the amplitudes is difficult to estimate since EViews does not compute confidence bands for impulse responses in VEC models.

Similar to the impulse response charts from Figure 7.1, the variance decompositions in Figure 7.2 are combined response charts for each of the hedge fund series. The charts are cumulative, i.e. the percentage of the errors that is attributable to own shocks plus the percentage attributable to shocks to other variables adds up to 100%. From Figure 7.2 it is deduced that most of the errors for the CTA, ED, EHg, EHv and M series are attributed to own system shocks. For the other hedge fund indices (EM, LS, and RV in particular), the asset indices explain a significant proportion of the variation in first-differenced terms. As before, the interpretation is difficult due to the non-stationarity of the model variables.

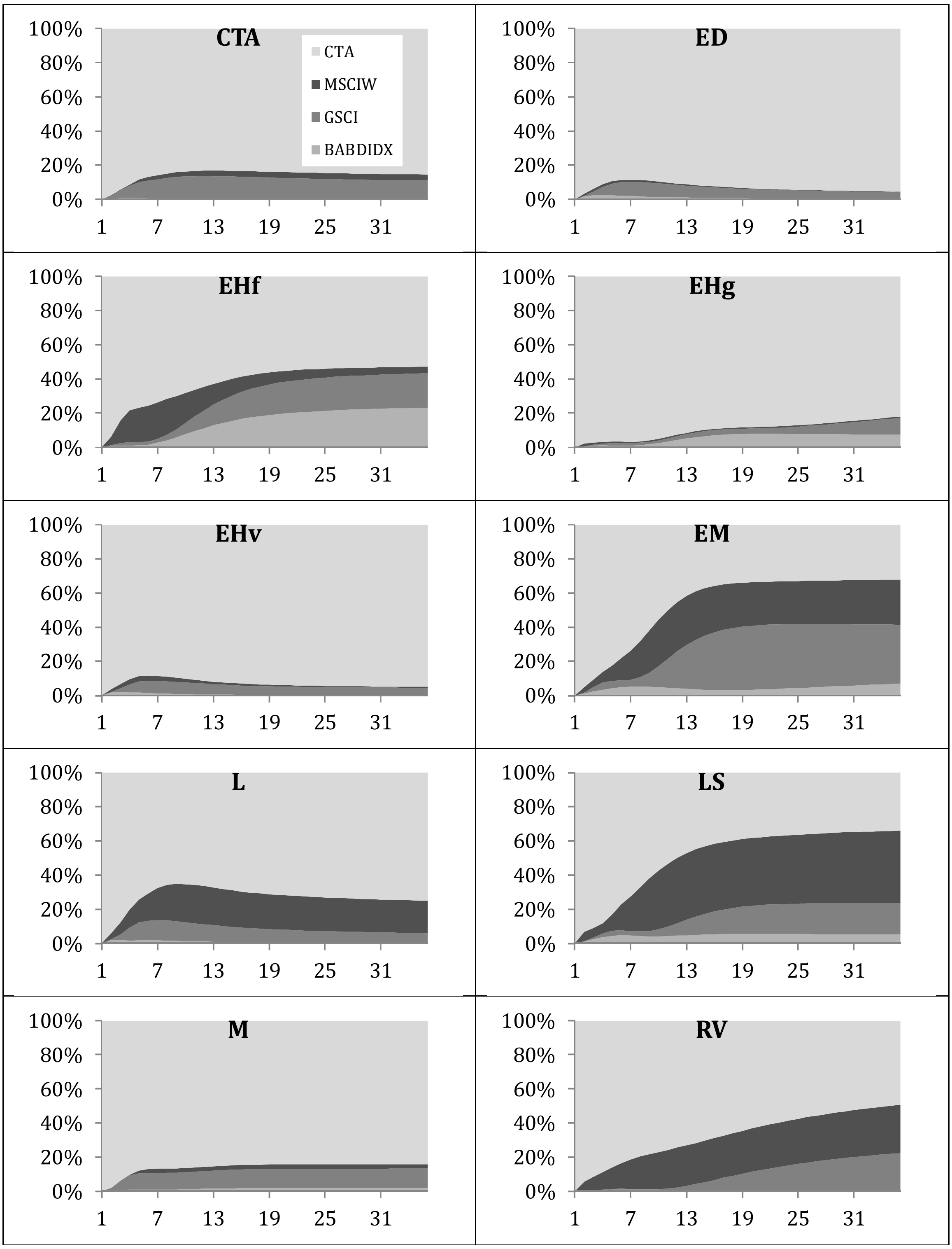

Figure 7.2: Variance decomposition

Cholesky ordering: Hedge fund index, BABDIDX, GSCI and MSCIW. Acronyms according to Table 6.5 and Table 6.6.

The causality, impulse response and variance decomposition tests give some initial idea of the lead-lag relationship between hedge fund and asset indices. Combining these with the results from Table 7.7 leads to the conclusion that, overall, managed futures are independent of equity, fixed income and commodity markets whereas many equity-oriented hedge fund strategies are not. It is of particular interest that this is confirmed for the cointegrating as well as the short-run relationship between CTA and other markets. At the same time, the CTA index exhibits the highest period returns of any of the ten indices under observation. Despite the increased volatility in the return series, managed futures, in the context of a well-diversified portfolio, may be the optimal complementary choice from a risk-return perspective. Similar results can be found for the M index series albeit displaying lower overall performance. The ED and EHg indices offer comparable diversification benefits. Conversely, directional equity funds such as EM and LS are highly correlated with the performance of the asset indices.

7.6 Results for HFRX and HFNI

The results in section 7 were replicated for the benchmark indices of the HFR and HFN database (HFRX and HFNI). For reasons of brevity, the results are not displayed in detail.[18] For each index considered, a separate VEC model was estimated (for 10 indices from HFR and 13 from HFN). Model specification, estimation and underlying assumptions are the same as for the style indices from CFPs. The indices were included for two reasons. Firstly, to test the robustness of VEC models in the context of commercial indices from database providers (i.e. the models can be applied to any hedge fund index to confirm their exposure to the other three asset indices), and secondly, to allow for comparisons between existing strategy indices and the indices constructed in section 6.

[18] In particular, details with respect to model specifications, cointegrating relationships, coefficient restrictions etc. are omitted. All calculations and results are available from the author upon request.

Table 7.9: Aggregated regression statistics from VECMs for HFRX and HFNI

	F	R^2	$\hat{R}^2$	AIC	SBIC	k	$MSBIC$
HFRX_AR	7.787	0.534	0.465	3.147	3.523	16	35.670
HFRX_CA	16.305	0.689	0.647	4.991	5.343	15	37.312
HFRX_DS	11.006	0.618	0.562	4.649	5.025	16	37.264
HFRX_ED	23.110	0.759	0.726	3.649	4.001	15	36.099
HFRX_EH	23.084	0.758	0.725	3.837	4.189	15	36.334
HFRX_EMN	2.438	0.249	0.147	3.166	3.518	15	35.799
HFRX_G	22.103	0.750	0.716	3.604	3.957	15	36.150
HFRX_M	4.089	0.338	0.256	5.479	5.808	14	37.672
HFRX_MA	4.236	0.365	0.279	3.140	3.492	15	35.526
HFRX_MD	32.103	0.814	0.788	4.202	4.554	15	36.709
HFRX_RV	28.428	0.794	0.766	3.710	4.062	15	36.203
	F	R^2	$\hat{R}^2$	AIC	$SBIC$	k	$MSBIC$
HFNI_CTA	3.938	0.330	0.246	5.437	5.766	14	37.731
HFNI_DS	38.698	0.840	0.819	4.165	4.517	15	36.699
HFNI_ED	39.766	0.844	0.823	3.605	3.957	15	36.092
HFNI_EH	62.236	0.894	0.880	3.554	3.906	15	36.137
HFNI_EM	36.298	0.819	0.797	6.031	6.360	14	38.218
HFNI_EMN	8.282	0.530	0.466	2.322	2.674	15	34.598
HFNI_FIA	12.607	0.631	0.581	3.281	3.634	15	35.691
HFNI_FS	11.415	0.627	0.572	4.687	5.062	16	37.347
HFNI_G	22.284	0.752	0.718	3.593	3.946	15	35.919
HFNI_LO	88.082	0.923	0.912	4.102	4.454	15	36.666
HFNI_LS	55.038	0.882	0.866	3.541	3.894	15	36.125
HFNI_M	9.300	0.538	0.480	3.860	4.189	14	36.070
HFNI_RV	41.938	0.851	0.830	2.611	2.963	15	35.130

Results are for HFRX (Hedge Fund Research indices) and HFNI (HedgeFund.Net indices) for the July 2000 to June 2010 timeframe. Results in columns 2 to 7 are for equation 1 of the VECM (hedge fund index series as dependent variable): k = number of coefficients estimated including intercepts and trends (cointegrating equations and VAR), $\hat{R}^2 = R^2$ adjusted for k, F denotes the joint significance. Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBIC) are the univariate information criteria, MSBIC is the modified Schwarz criterion for the overall model. Other acronyms as follows: AR = absolute return, CA = convertible arbitrage, DS = dedicated short, ED = event driven, EH = equity hedge, EMN = equity market neutral, G = global, M = macro, MA = merger arbitrage, MD = market directional, RV = relative value. Additional acronyms for HFNI: CTA = CTA/managed futures, FIA = fixed income arbitrage, FS = finance sector, LO = long only/bias, LS = long/short equity.

Table 7.10: Aggregated results for Granger from VECMs for HFRX and HFNI

	D(BABDIDX)	D(GSCI)	D(MSCIW)	all
D(HFRX_AR)	0.813	9.041··	18.386···	44.543···
D(HFRX_CA)	24.072···	29.250···	30.026···	109.137···
D(HFRX_DS)	0.316	3.082	36.013···	57.435···
D(HFRX_ED)	2.816	5.434	160.430···	206.998···
D(HFRX_EH)	2.765	18.122···	147.532···	232.609···
D(HFRX_EMN)	1.084	3.778	8.978··	11.755
D(HFRX_G)	5.956	22.041···	105.329···	198.847···
D(HFRX_M)	2.425	22.378···	6.979·	35.037···
D(HFRX_MA)	3.533	3.785	26.698···	39.343···
D(HFRX_MD)	5.655	28.817···	148.929···	249.886···
D(HFRX_RV)	13.499···	35.764···	83.400···	173.325···
	D(BABDIDX)	D(GSCI)	D(MSCIW)	all
D(HFNI_CTA)	7.339·	22.560···	3.386	38.625···
D(HFNI_DS)	2.329	11.791···	178.256···	245.375···
D(HFNI_ED)	2.599	7.990··	172.395···	237.074···
D(HFNI_EH)	1.012	30.632···	269.882···	381.471···
D(HFNI_EM)	15.398···	47.124···	205.869···	356.462···
D(HFNI_EMN)	1.224	27.316···	21.318···	79.600···
D(HFNI_FIA)	14.989···	21.428···	32.946···	90.934···
D(HFNI_FS)	4.349	18.468···	120.193···	128.140···
D(HFNI_G)	10.849··	41.561···	121.251···	268.762···
D(HFNI_LO)	0.837	20.467···	368.284···	488.210···
D(HFNI_LS)	1.578	31.214···	182.168···	270.453···
D(HFNI_M)	5.062	28.363···	47.068···	118.476···
D(HFNI_RV)	7.975··	40.012···	131.861···	229.247···

Results are for HFRX (Hedge Fund Research indices) and HFNI (HedgeFund.Net indices) for the July 2000 to June 2010 timeframe. The aggregated overview is limited to the first equation of the VECM (hedge fund index series). The row headings give the dependent variable, the column headings are the predictor variables. The 'all' column gives the overall significance of the three series in predicting changes in the first differenced series (D(...)) of the hedge fund index. The test statistic is the chi-squared distributed Wald. Statistical significance is as follows: · denotes significance at 10% level, ·· denotes significance at 5% level, and ··· denotes significance at 1% level (significance of the temporal ordering of the movements in the series). All strategic acronyms as in Table 7.9.

Table 7.9 shows the regression statistics and Table 7.10 displays the results from Block significance tests for the individual hedge fund index equations of the VECMs. The results are for the first equation from the VEC system of equations (i.e. the function for the hedge fund index series). From Table 7.9, the regression statistics reveal a high explanatory power of the model for all index series with the exception of CTA/managed futures, macro, (equity) market neutral and merger arbitrage. The overall fit is significant for all series. There are differences between hedge fund index series from the two different providers. Some of this is attributable to index weighting and re-balancing. However, it is possible that some indices differ with respect to their index composition and attribution of hedge funds to a particular classification.

In this context, the differences between HFRX_M and HFNI_M (Macro index) are particularly pronounced. The goodness-of-fit is increased substantially for the HFNI_M series when CTAs are not classified together with other Macro funds (as is the case for HFRX). As with the pure style indices in section 7.4, there are large differences between the VECMs for different indices, both in terms of the predictive power of the single equation as well as overall model fit: macro funds/CTAs and market neutral strategies provide the best diversification benefits whereas directional strategies such as dedicated short, equity hedge and long/short equity are closely matched observing their own past observations as well as innovations to the three other index series.

The results are confirmed in Table 7.10. There is strong evidence for a lead-lag relationship between the first-differenced series of the MSCIW and changes in hedge fund index levels (CTAs/managed futures are the only exception). There is some evidence of exogeneity for equity market neutral funds. As before, there are differences between database providers, as well as between HFRX/HFNI and the pure style indices from this research. However, it is confirmed from impulse response tests and variance decomposition (with changing Cholesky ordering) that changes in hedge fund index levels follow changes to standard asset indices rather than vice versa. As before, this does not necessarily imply that changes are induced but merely expresses the temporal ordering of changes.

There are several possible explanations for the disparities of style and commercial indices, which are discussed in turn. Detailed inclusion criteria for HFRX and HFNI are unknown, although there is some information on reporting style and minimum AUM. It is likely that some of the minimum requirements are similar to those in section 5.3. However, minimum track records are much shorter (e.g. 24 months for HFRX), resulting in a smaller number of funds to be excluded from the sample. Conversely, HFRX excludes smaller funds from entering the sample (minimum AUM). Even though the index weights were chosen to best replicate a particular strategy, the re-balancing intervals are shorter for HFRX than for the pure style indices from Chapter 6 (quarterly vs. annual). This may lead to some trending behaviour in the pure style indices over shorter intervals.

The factor axis methodology introduced differentiates between the common and individual return component of single manager hedge funds. The upshot is an orthogonal factor model where extracted common vectors are orthogonal representations of distinct strategies. Cluster and correlation analysis are used in classifying hedge funds of the HFR database. Consequently, strategic clusters of hedge funds result in index series that are highly correlated with other strategic representations. On the one hand, it is unknown what degree of communality is required for hedge funds to belong to the same specific cluster. On the other hand, hedge funds classified differently may share some common return component that is disguised by their unique return components. Because factor extraction was repeated every month on the basis of the past 120 observations for all hedge funds in the sample, all information was considered from the moment it became available.

In Chapter 6, the ‘inception date’ varies for different strategies. As the sample size of single manager hedge funds fulfilling the minimum criteria increased, a larger number of extracted orthogonal factors was required to account for the co-variance between the sample funds. Thus, some strategies initially shared the same track record until factor extraction necessitated a split of the hedge funds within a grouping (recall that the first factor extraction revealed five strategic groupings whereas the final estimation period suggested ten). Defunct hedge funds exiting and new funds entering the sample caused index composition to change radically over time (only some funds remained as constituents for the same index throughout the entire observation period of July 2000 to June 2010). This problem is less pronounced for hedge fund index vendors that allow history-backfilling and bias towards survivors. In the case of HFRX, inclusion of funds with shorter track records unlikely biases against underperforming funds, whereas the minimum requirements with respect to assets-under-management does (compare sections 3.5.2 and 3.5.3, Amin & Kat, 2003; Gregoriou, 2006).[19] Index composition in Chapter 6 includes hedge funds from the graveyard database but favours long-surviving funds.

Construction of investable indices using the methodology introduced in this research was subject to several important caveats. Firstly, lockdown periods aggravated continuous re-balancing of index constituents and required longer looking-ahead intervals. Secondly, the analysis included hedge funds that are currently closed-for-investments. Lastly, even though the maximum number of constituents was limited, some indices consisted of a large number of hedge funds, some of which only represented a small proportion of the total investment in terms of constituent weight. Minimum investment amounts at the single manager hedge fund level pose problems for the construction of investable indices. Despite these limitations, it still might be possible to create investable indices that closely track the performance of the style indices introduced here.

This chapter has provided guidelines to estimate a correctly specified multivariate model, estimate the short-term and long-term relationship between variables of the model, and identify those hedge fund strategies that offer the best diversification

[19] It is noteworthy that both HFRX as well as HFNI include equity market neutral as a separate classification. It is unlikely that this research was biased against these funds on the basis of the sample selection criteria (minimum track record), since related research suggests that mean survival times for equity market neutral funds are longer than for other strategies (e.g. Gregoriou, 2005).

benefits in a hypothesised portfolio. All results are subject to the underlying assumptions (e.g. lag specification in the first-differenced terms, number of cointegrating equations and rank of the cointegrating relationship, Cholesky-ordering for error variance analysis etc.). In addition, a treatment of the original series to invoke stationarity (i.e. return series), HAC or asymptotic normality of the error-term (log transformation) was waived in favour of building a VEC model from the level index series. The upshot is retention of the maximum amount of information from the original variables. Treatment of the series prior to entering the multivariate models is likely to yield different results.

All results have been reproduced for the standalone HFR database and are included in the appendix in Table A.34 through Table A.43 as well as Figure A.1 and Figure A.2. The results by and large confirmed the findings for the joint HFR and HFN database and did not warrant a separate discussion.

CHAPTER 8: CONCLUDING REMARKS

Principal axis and factor rotation were used to extract common factors from hedge funds reporting to the HFR and HFN databases. The stepwise procedure yielded statistical clusters of hedge funds that load on one common factor only. Continuously re-balanced indices were created from the CFPs and labelled according to the predominant strategies of hedge funds within the CFPs. The estimated series of the indices were regressed against asset-based factors and factors representing simplistic trading strategies. Ten classifications were identified that subsume a significant proportion of the sample hedge funds. For the remaining hedge funds, no communalities with other hedge funds could be ascertained.

This has led to two conclusions: firstly, the majority of hedge funds follow a broad strategic theme that is common to all hedge funds within a classification. The long-term return of hedge funds is a function of the contribution from the common factor and the specific factor representative of the unique trading style of the manager. Secondly, the remaining hedge funds operate in niche markets / sectors or employ a specialised investment approach that is not easily replicated (even though the statistical factor model revealed one factor portfolio representing hedge funds with a finance-sector focus). The results were persistent throughout time and different macroeconomic cycles. Considering the relatively small sample size as a result of the minimum requirements for hedge funds to enter, it is unlikely that the ten classifications describe the entire spectrum of investment strategies. However, they are an indicator of the predominant investment themes over the past 20 years. The regression analysis has provided some initial indications as to how those investment themes might be replicated.

In contrast to other statistical factor models, the extracted factors from principal axis explained a significant proportion of the *co*-variance between all hedge funds across the two databases (PCA, for example, seeks to explain the variance for a singular hedge fund). The specific component not explained by the extracted factors was diversified away in portfolios. The upshot was that all factors contributing towards explaining the communalities between hedge funds were retained. The number of retained common factors was determined by running parallel analyses with 100 replications. While the results have not been discussed in detail, the degree of communality for the majority of single manager funds (i.e. the proportion of the returns explained by the common factor) is well in excess of 50 percent. Thus, a single unobserved factor, following rotation of the initial eigenvector estimates, explains a large proportion of the return variation for hedge funds within CFPs.

The specific statistical properties of hedge fund return series have been accounted for to the degree possible. Using the residuals from correctly specified univariate $ARMA$-models eliminated the problems of serially correlated returns. Additionally, factor axis methodology was less susceptible to non-normality of the return frequency distribution compared to other statistical factor models. For the weighted index series representing the balanced return of the 10 CFPs, the impact from serial correlation and non-normality was mitigated.

Using 120 estimation windows allowed for the creation of a continuously re-balanced index series to be used in further analysis. Index creation was unbiased since past performance was used to estimate the communalities between, and hence, attribution of single manager hedge funds to particular CFPs. The one-month-ahead performance was then used to determine the individual fund's performance contribution. The portfolio weights for every month were selected so as to maximise the correlation of the weighted series with the extracted factor. The index series was re-balanced in 12-monthly intervals; however, the series could be adjusted to reflect shorter re-balancing intervals since the composition of the CFPs and portfolio weights of the constituents were estimated for every month.

The resulting index series for the ten CFPs covered a 120 month timeframe. The period under observation allowed for estimation of the impact from the liquidity and credit crisis of 2007 and the following period of economic recovery. Throughout the observation timeframe, the CFPs were robust with respect to their persistence and composition (i.e. single manager funds within one CFP were likely to belong to the same classification in the following period, except for defunct funds or those exhibiting style drift). The methodology allowed for increased strategic diversity as an increasing sample of hedge funds results in additional classifications. The track record for the pure style indices was long enough to be used in multivariate analysis and to make statistically significant inferences.

In the VEC models, the index level series was used to differentiate between the long-term and short-term dependencies between the hedge fund index series and the equity, bond and commodity market proxy. It is perhaps surprising to find that not only is current performance of hedge fund indices related to past performance (as is confirmed in related research), but also to the lagged movements of the BABDIDX, MSCIW and, to a lesser degree, the GSCI series. Simultaneous coefficient estimation removed the bias from *a priori* determination of externality/endogeneity of the model variables. It also accounted for the possibility that the lead-lag relationship between hedge funds and other assets might be reversed at times: whilst there is no conclusive proof that such a causal relationship exists (e.g. CTAs inducing changes in commodity markets), it is reasonable to draw some inferences with respect to the temporal ordering of markets responding to externalities.

Illiquidity in hedge fund returns is evidenced by the significance of their exposure to their own past performance (i.e. past performance linking in with present and future performance). This type of trending behaviour is, however, not unique to the hedge fund index series and can be ascertained for the other three indices as well. In addition, the exposure to past performance varied greatly between different strategies. For example, whilst none of the lagged coefficients was significant for the M index series, trending behaviour was observed for EH. Lagged predictors from other asset classes suggest that publicly available information was not priced in a timely manner. This may be due to trading in illiquid assets. Here too the results vary greatly between different hedge fund classifications. For example, systemic shocks to the MSCIW series took particularly long to work their way through the error variance of the LS index series. However, the significance of that impact was difficult to gauge.

Looking at the residual diagnostics from the multivariate model allowed for identification of structural breaks resulting from cataclysmic events that severely impacted on the performance of all index series. Such an event could be identified for the months following the downfall of the Lehman Brothers bank and the subsequent credit crisis. Interestingly, all markets quickly reverted back to an equilibrium suggesting that the repercussions from such events are often short-term. Removing the two months following the Lehman crisis from analysis proved sufficient to invoke stability and variance consistency in the residual series (this is in contrast to allowing for a new persistence regime representing a novel equilibrium between assets).

To the degree possible, the model specifications were decided on using unbiased information criteria. Thus, no significant relationship between variables was ignored. However, some assumptions were made such as maximum lag length for autoregressive models and the ordering of variables in extended diagnostics. The assumptions were based on related research and are economically justifiable. Some lags were excluded that might have been jointly significant in explaining the variation in index levels across the models. However, there were practical reasons to limit the number of estimated coefficients to retain model degrees of freedom. In a similar vein, coefficient restrictions were imposed where such restrictions did not significantly impair the informative value of the cointegration equations.

All the results from factor modelling and vector autoregression were reproduced for the HFR database to account for survivorship bias in the HFN database. A separate discussion was largely omitted as the main findings have been confirmed. Extended results for both the combined HFR / HFN database as well as standalone HFR are available from the author on request. The empirical results and interpretations are only representative of the two databases included here. The methodologies and models introduced may be applied to other databases and are likely to yield somewhat different results. It is argued that HFR and HFN represent a large enough combined database to justify the general inferences in this research. It stands to reason that the inferences from analysing hedge fund style indices are not easily transferred to single manager hedge funds. It is often the specific return component, which gets diversified away in hedge fund portfolios or FoHFs that sets them apart from the performance of other assets. However, for investors seeking pure style representations to complement existing portfolios, the research here presented is a useful framework providing empirical results for the HFR and HFN database.

This study has refrained from making definitive recommendations to investors regarding portfolio composition and the selection of particular hedge fund styles based on the empirical results. One reason is that no differentiation was made between non-investable hedge funds and hedge funds open for investment. As a consequence, style indices may not be easily replicated. This also represents one of the possible future extensions of this research: the construction of investable indices. Alternatively, the index regression results from section 6.4 present a different approach. The return on pure style indices may be replicated using a combination of passive indices and simplistic trading strategies. Confirming these results with respect of out-of-sample fit is another avenue of future investigation.

The practitioner may draw some general inferences from the results of this research, Firstly, the application of advanced statistical factor models allows fund managers to identify communalities across hedge funds and to construct portfolios accordingly. This will be of particular interest if the objective is portfolio diversification. Secondly, factor models should be used as part of the due diligence in the investment selection process for FoHFs or pension funds: to identify differences between the stated investment objective and implemented trading strategy (style drift) and to identify those investments that mitigate persistent risks of the existing portfolio. Thirdly, the results from asset-based factor regression in Chapter 6 provide some hints to which hedge fund style is not easily replicated and may be indicative of managerial skill. The multivariate VECM and associated causality test in Chapter 7 confirm the initial interpretations from Chapter 6: CTAs and macro hedge funds offer superior diversification benefits in the context of a traditional asset portfolio, whilst outperforming their peers on an average monthly performance level. Owed to the detailed explanations of the methodology section, the results are replicable for any sub-sample.

The extension of the empirical results to other major hedge fund databases is of further interest, in particular TASS and CISDM. It is not unreasonable to assume that the results from this research are somewhat confirmed for other databases. Many hedge funds report to more than one database vendor and the number of double-reporters is likely to be high. However, an increased sample size improves on statistical significance and inclusion of other databases eliminates any associated selection bias. Since TASS includes defunct and derelict funds, analysis of that database is of particular interest.

Lastly, significant selection bias resulted from the sample selection criteria outlined in section 5.3. The minimum number of observations is likely to have precluded many hedge funds from consideration. Considering average survival and attrition rates, the majority of hedge funds were not entered into the samples (the samples may also be skewed in terms of strategic representativeness). However, the focus of this research has been with time series analysis and the inherent limitations of including series with short performance histories. Rather than assuming that the style indices are representative of the entire universe of hedge fund investments, they may be regarded as the central strategic themes prevalent in the hedge fund industry. The index series may be used to classify hedge funds with shorter performance histories by conducting regression or correlation analysis.

It is suggested that the methodologies and empirical results of this research are of benefit not only to quantitative researchers but to practitioners as well. It is likely that on-going research into hedge funds will bring with it a multitude of novel approaches to classify and assess hedge funds. This research contributes to the continuous efforts of developing quantitative models for hedge funds.

SOURCES

Ackermann, C., McEnally, R. & Ravenscraft, D., 1999. The performance of hedge funds: Risk, return, and incentives. *Journal of Finance*, **54**(3), June, 833–874. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=2176473&site=ehost-live. Accessed: 03 July 2013.

Agarwal, V., Daniel, N. D. & Naik, N. Y., 2009. Role of managerial incentives and discretion in hedge fund performance. *Journal of Finance*, **64**(5), 2221–2256. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=44343123&site=ehost-live. Accessed: 03 July 2013.

Agarwal, V. & Kale, J. R., 2007. On the relative performance of multi-strategy and funds of hedge funds. *Journal of Investment Management*, **5**(3), Fall, 41–63. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=31274977&site=ehost-live. Accessed: 03 July 2013.

Agarwal, V. & Naik, N. Y., 1999. On taking the 'Alternative' route: Risks, rewards, style and performance persistence of hedge funds. *SSRN eLibrary.* [Online] Available: http://ssrn.com/paper=150388. Accessed: 03 July 2013.

Agarwal, V. & Naik, N. Y., 2000a. Generalized style analysis of hedge funds. *Journal of Asset Management*, **1**(1), 93–109.

Agarwal, V. & Naik, N. Y., 2000b. Multi-Period performance persistence analysis of hedge funds. *Journal of Financial & Quantitative Analysis*, **35**(3), 327–342. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=3617811&site=ehost-live. Accessed: 03 July 2013.

Agarwal, V. & Naik, N. Y., 2000c. Performance evaluation of hedge funds with option-based and buy-and-hold strategies. *SSRN eLibrary.* [Online] Available: http://ssrn.com/paper=238708. Accessed: 03 July 2013.

Agarwal, V. & Naik, N. Y., 2004. Risks and portfolio decisions involving hedge funds. *Review of Financial Studies*, **17**(1), 63–98. [Online] Available: http://www.jstor.org/stable/1262669. Accessed: 03 July 2013.

Aggarwal, R. K. & Jorion, P., 2010. Hidden survivorship in hedge fund returns. *Financial Analysts Journal*, **66**(2), 1–6. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=48578551&site=ehost-live. Accessed: 03 July 2013.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716–723.

Amenc, N., El Bied, S. & Martellini, L., 2003. Predictability in hedge fund returns. *Financial Analysts Journal*, **59**(5), 32–46. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=11003324&site=ehost-live. Accessed: 03 July 2013.

Amenc, N. & Goltz, F., 2008. Revisiting the limits of hedge fund indices: A comparative approach. *Journal of Alternative Investments*, **10**(4), 50–63. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=31822336&site=ehost-live. Accessed: 03 July 2013.

Amenc, N., Malaise, P., Martellini, L. & Vaissié, M., 2004. *Fund of hedge fund reporting - A return-based approach to fund of hedge fund reporting.* Unpublished research report. Edhec Risk and Asset Management Research Centre, February. [Online] Available: http://www.edhec-risk.com/features/Fund%20of%20Hedge%20Fund%20Reporting%20Survey/attachments/FoHF%20Reporting%20Discussion%20Paper.pdf. Accessed: 03 July 2013.

Amenc, N. & Martellini, L., 2001. *The brave new world of hedge fund indexes.* Unpublished research report. Edhec Risk and Asset Management Research Centre. [Online] Available: http://www.edhec-risk.com/edhec_publications/RISKArticle1036431597535828l5/attachments/bnwhfi.pdf. Accessed: 03 July 2013.

Amenc, N. & Martellini, L., 2003. *Desperately seeking pure style indices.* Unpublished research report. Edhec Risk and Asset Management Research Centre. [Online] Available: http://www.edhec-risk.com/edhec_publications/copy_of_Desperately%20Seeking%20Pure%20Style%20Indices/attachments/desperately_seeking_pure_style_indexes_251103_new.pdf. Accessed: 03 July 2013.

Amenc, N., Martellini, L. & Vaissié, M., 2003a. Benefits and risks of alternative investment strategies. *Journal of Asset Management*, **4**(2), 96–118. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=10289319&site=ehost-live. Accessed: 03 July 2013.

Amenc, N., Martellini, L. & Vaissié, M., 2003b. *Indexing hedge fund indexes.* Unpublished research report. Edhec Risk and Asset Management Research Centre, December. [Online] Available: http://docs.edhec-risk.com/EAID-2007/doc/Indexing_Hedge_Fund_Indexes.pdf. Accessed: 04 July 2013.

Amin, G. S. & Kat, H. M., 2003. Welcome to the dark side: Hedge fund attrition and survivorship bias over the period 1994-2001. *Journal of Alternative Investments*, **6**(1), Summer, 57–73. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=10253546&site=ehost-live. Accessed: 03 July 2013.

Ammann, M. & Moerth, P., 2008a. Impact of fund size and fund flows on hedge fund performance. *Journal of Alternative Investments*, **11**(1), Summer, 78–96. [Online] Available:

http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=33119598&site=ehost-live. Accessed: 03 July 2013.

Ammann, M. & Moerth, P., 2008b. Performance of funds of hedge funds. *Journal of Wealth Management*, **11**(1), 46–63. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=32153282&site=ehost-live. Accessed: 03 July 2013.

Amo, A.-V., Harasty, H. & Hillion, P., 2007. Diversification benefits of funds of hedge funds: Identifying the optimal number of hedge funds. *Journal of Alternative Investments*, **10**(2), 10–22. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=27172720&site=ehost-live. Accessed: 03 July 2013.

Anand, G., Maier, T., Kutsarov, I. & Storr, M., 2011. Importance of tactical strategy allocation on fund-of-hedge-funds allocations. *Journal of Wealth Management*, **14**(2), Fall, 49–58. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=64406302&site=ehost-live. Accessed: 04 July 2013.

Anderson, T. W. Darling, D. A. Darling, 1952. Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics*, **23**(2), June, 193–212. [Online] Available: http://www.jstor.org/stable/2236446. Accessed: 13 November 2013.

Anderson, T. W. Darling, D. A. Darling, 1954. A Test of Goodness of Fit. *Journal of the American Statistical Association*, **49**(268), December, 765–769. [Online] Available: http://www.jstor.org/stable/2236446. Accessed: 13 November 2013.

Ang, A., Rhodes-Kropf, M. & Zhao, R., 2008. Do fund-of-funds deserve their fees-on-fees? *Journal of Investment Management*, **6**(4), Winter, 34–58. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=36151325&site=ehost-live. Accessed: 03 July 2013.

Aragon, G. O., 2007. Share restrictions and asset pricing: Evidence from the hedge fund industry. *Journal of Financial Economics*, **83**(1), January, 33–58. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=23353134&site=ehost-live. Accessed: 03 July 2013.

Asness, C. S., Krail, R. & Liew, J. M., 2001. Do hedge funds hedge? *Journal of Portfolio Management*, **28**(1), April, 6–19. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=6539543&site=ehost-live. Accessed: 03 July 2013.

Baghai-Wadji, R. & Klocker, S., 2007. *Performance and style shifts in the hedge fund industry.* Unpublished research report. Vienna University of Economics and Business Administration, January. [Online] Available: http://www.efmaefm.org/0EFMAMEETINGS/EFMA%20ANNUAL%20MEETINGS/2007-Vienna/Papers/0571.pdf. Accessed: 03 July 2013.

Baquero, G., Horst, J. R. ter & Verbeek, M., 2005. Survival, look-ahead bias, and persistence in hedge fund performance. *Journal of Financial & Quantitative Analysis*, **40**(3), 493–517. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=18228374&site=ehost-live. Accessed: 04 July 2013.

Barès, P. A., Gibson, R. & Gyger, S., 2003. Performance in the hedge funds industry: An analysis of short and long-term performance. *Journal of Alternative Investments*, **6**(3), 25–41. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=36151325&site=ehost-live. Accessed: 03 July 2013.

Barry, R., 2002. *Hedge funds: A walk through the graveyard.* Research paper no. 25. Macquarie University Applied Finance Centre, September. [Online] Available: http://ssrn.com/paper=333180. Accessed: 03 July 2013.

Beckers, S., Curds, R. & Weinberger, S., 2007. Funds of hedge funds take the wrong risks. *Journal of Portfolio Management*, **33**(3), 108–121. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=24949287&site=ehost-live. Accessed: 03 July 2003.

Beltratti, A. & Morana, C., 2008. Aggregate hedge funds' flows and returns. *Applied Financial Economics*, **18**(21), 1755–1764. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=36151325&site=ehost-live. Accessed: 03 July 2013.

Billio, M., Getmansky, M. & Pelizzon, L., 2007. *Phase-locking and switching volatility in hedge funds.* Working Paper Series No. 54/06. University Ca' Foscari of Venice, Deptartment of Economics Research, February. [Online] Available: http://ssrn.com/paper=948407. Accessed: 03 July 2013.

Blazsek, S. I. & Downarowicz, A., 2008. *Regime switching models of hedge fund returns.* Working Paper No 12/08. School of Economics and Business Administration, University of Navarra, November. [Online] Available: http://ideas.repec.org/p/una/unccee/wp1208.html. Accessed: 03 July 2013.

Bollen, N. P. B. & Pool, V. K., 2008. Conditional return smoothing in the hedge fund industry. *Journal of Financial & Quantitative Analysis*, **43**(2), 267–298. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=32502678&site=ehost-live. Accessed: 03 July 2013.

Bollen, N. P. B. & Pool, V. K., 2009. Do hedge fund managers misreport returns? Evidence from the pooled distribution. *Journal of Finance*, **64**(5), 2257–2288. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=44343122&site=ehost-live. Accessed: 03 July 2013.

Bollen, N. P. B. & Whaley, R., 2009. Hedge fund risk dynamics: Implications for performance appraisal. *Journal of Finance*, **64**(2), 985–1035. [Online] Available:

http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=36923037&site=ehost-live. Accessed: 03 July 2013.

Boyson, N. M., 2003. *Do hedge funds exhibit performance persistence? A new approach.* Working Paper. Northeastern University and Purdue University, October. [Online] Available: http://www.edhec-risk.com/research_news/choice/RISKReview1076512980486833155/attachments/Boyson%20persistence.pdf. Accessed: 04 July 2013.

Brandon, R. G. & Wang, S., 2013. Liquidity risk, return predictability, and hedge funds' performance: An empirical study. *Journal of Financial & Quantitative Analysis*, **48**(1), 219–244. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=87775043&site=ehost-live. Accessed: 04 July 2013.

Brooks, C., 2008. *Introductory econometrics for finance.* 2nd Edition. Cambridge: Cambridge University Press.

Brooks, C. & Kat, H. M., 2002. The statistical properties of hedge fund index returns and their implications for investors. *Journal of Alternative Investments*, **5**(2), 26–44. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=7510511&site=ehost-live. Accessed: 03 July 2013.

Brown, S. J. & Goetzmann, W. N., 2003. Hedge funds with style. *Journal of Portfolio Management*, **29**(2), Winter, 101–112. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=10017455&site=ehost-live. Accessed: 03 July 2013.

Brown, S. J., Goetzmann, W. N. & Ibbotson, R. G., 1999. Offshore hedge funds: survival and performance 1989 - 1995. *Journal of Business*, **72**(1), January, 91–117. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=1532669&site=ehost-live. Accessed: 03 July 2013.

Brown, S. J., Goetzmann, W. N. & Liang, B., 2004. Fees on fees in funds of funds. *Journal of Investment Management*, **2**(4), Winter, 39–56. [Online] Available: http://ssrn.com/abstract=335581. Accessed: 04 July 2013.

Brown, S. J., Goetzmann, W. N. & Park, J., 2001. Careers and Survival: Competition and Risk in the Hedge Fund and CTA Industry. *Journal of Finance*, **56**(5), October, pp. 1869–1886. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=5114967&site=ehost-live. Accessed: 14 November 2013.

Capocci, D., Corhay, A. & Hübner, G., 2005. Hedge fund performance and persistence in bull and bear markets. *European Journal of Finance*, **11**(5), 361–392. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=18685243&site=ehost-live. Accessed: 03 July 2013.

Capocci, D. & Hübner, G., 2004. Analysis of hedge fund performance. *Journal of Empirical Finance*, **11**(1), 55–89. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=22238873&site=ehost-live. Accessed: 03 July 2013.

Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance*, **52**(1), March, 57–82. [Online] Available: http://www.jstor.org/stable/2329556. Accessed: 04 July 2013.

Chan, N., Getmansky, M., Haas, S. M. & Lo, A. W., 2006. Do hedge funds increase systemic risk? *Economic Review*, **91**(4), 49–80. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=23788235&site=ehost-live. Accessed: 03 July 2013.

Chandler, B., 1994. *Managed futures - An investor's guide.* 1st Edition. Chichester: Wiley.

Chen, K. K. & Passow, A., 2003. *Quantitative selection of long-short hedge funds.* Research Paper No. 94. International Center for Financial Asset Management and Engineering, July. [Online] Available: http://ssrn.com/paper=470642. Accessed: 04 July 2013.

Christiansen, C. B., Madsen, P. B. & Christensen, M. 2004. A quantitative analysis of hedge fund style performance. In Schachter, B. (eds.). *Intelligent hedge fund investing.* London: Risk Books.

Chung, Y. P., Johnson, H. & Schill, M. J., 2006. Asset pricing when returns are nonnormal: Fama-French factors versus higher-order systematic comoments. *Journal of Business*, **79**(2), 923–940. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=19554769&site=ehost-live. Accessed: 03 July 2013.

Collins, D. P., 2003. Return of the long-term trend follower. *Futures: News, Analysis & Strategies for Futures, Options & Derivatives Traders*, **32**(4), 68. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9215175&site=ehost-live. Accessed: 03 July 2013.

Connor, G. & Korajczyk, R. A., 1986. Performance measurement with the arbitrage pricing theory. *Journal of Financial Economics*, **15**(3), 373–394. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=12243326&site=ehost-live. Accessed: 03 July 2013.

Connor, G. & Korajczyk, R. A., 1988. Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, **21**(2), 255–289. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=12249249&site=ehost-live. Accessed: 03 July 2013.

Cornish, E. A. & Fisher, R. A., 1938. Moments and cumulants in the specification of distributions. *Review of the International Statistical Institute*, **5**(4), January, 307–320. [Online] Available: http://www.jstor.org/stable/1400905. Accessed: 04 July 2013.

Cottier, P., 2000. *Hedge funds and managed futures - Performance, risks, strategies, and use in investment portfolios.* 3rd Edition. Bern: Haupt.

Csörgö, S. & Faraway J., 1996. The exact and asymptotic distributions of Cramer-von Mises statistics. *Journal of the Royal Statistical Society*, Series B, **58**(1), 221–234. [Online] Available: http://www.jstor.org/stable/2346175. Accessed: 12 November 2013.

Dickey, D. A. & Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**(366), 427–431. [Online] Available: http://www.jstor.org/stable/2286348. Accessed: 04 July 2013.

Ding, B. & Shawky, H. A., 2007. The performance of hedge fund strategies and the asymmetry of return distributions. *European Financial Management*, **13**(2), 309–331. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=24219625&site=ehost-live. Accessed: 03 July 2013.

Doornik, J. A. & Hansen, H., 2008. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics & Statistics*, **70**, 927–939. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=35393876&site=ehost-live. Accessed: 03 July 2013.

Durbin, J., 1970. *Distribution Theory for Tests Based on the Sample Distribution Function*. Philadelphia: Society for Industrial and Applied Mathematics.

Edwards, F. R. & Caglayan, M. O., 2001. Hedge fund performance and manager skill. *Journal of Futures Markets*, **21**(11), November, 1003–1028. [Online] Available: http://www3.interscience.wiley.com/cgi-bin/abstract/85510864/ABSTRACT?CRETRY=1&SRETRY=0. Accessed: 04 July 2013.

Edwards, F. R. & Liew, J., 1999. Hedge funds versus managed futures as asset classes. *Journal of Derivatives*, **6**(4), 45–64. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=1941741&site=ehost-live. Accessed: 03 July 2013.

Eling, M., 2009. Does hedge fund performance persist? Overview and new empirical evidence. *European Financial Management*, **15**(2), 362–401. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=36790892&site=ehost-live. Accessed: 03 July 2013.

Eling, M. & Faust, R., 2010. The performance of hedge funds and mutual funds in emerging markets. *INFOR*, **34**(8), 1993–2009. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=51293919&site=ehost-live. Accessed: 03 July 2013.

Engle, Robert F. & Granger, C. W. J., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, **55**(2), 251–276. [Online] Available: http://www.jstor.org/stable/1913236. Accessed: 11 November 2013.

Fama, E. F. & French, K. R., 1992. The cross-section of expected stock returns. *Journal of Finance*, **47**(2), June, 427–465. [Online] Available: http://www.jstor.org/stable/2329112. Accessed: 04 July 2013.

Fama, E. F. & French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33**(1), February, 3–56. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=12243094&site=ehost-live. Accessed: 04 July 2013.

Fama, E. F. & French, K. R., 1996. Multifactor explanations of asset pricing anomalies. *Journal of Finance*, **51**(1), March, 55–84. [Online] Available: http://www.jstor.org/stable/2329302. Accessed: 04 July 2013.

Favre, L. & Galeano, J.-A., 2002. Mean-modified value-at-risk optimization with hedge funds. *Journal of Alternative Investments*, **5**(2), 21. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=7510493&site=ehost-live. Accessed: 03 July 2013.

Francis, J. G. F., 1961. The QR transformation: A unitary analogue to the LR transformation - Part 1. *Computer Journal*, **4**(3), 265–271. [Online] Available: http://comjnl.oxfordjournals.org/content/4/3/265.abstract. Accessed: 04 July 2013.

Francis, J. G. F., 1962. The QR Transformation - Part 2. *Computer Journal*, **4**(4), 332–345. [Online] Available: http://comjnl.oxfordjournals.org/content/4/4/332.abstract. Accessed: 04 July 2013.

Frumkin, D. & Vandegrift, D., 2009. The effect of size, age, beta and disclosure requirements on hedge fund performance. *Journal of Derivatives & Hedge Funds*, **15**(3), 241–251. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=44630870&site=ehost-live. Accessed: 07 March 2013.

Fung, W. & Hsieh, D. A., 1997a. Empirical characteristics of dynamic trading strategies: the case of hedge funds. *Review of Financial Studies*, **10**(2), Summer, 275–302. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=4233734&site=ehost-live. Accessed: 04 July 2013.

Fung, W. & Hsieh, D. A., 1997b. Survivorship bias and investment style in the returns of CTAs. *Journal of Portfolio Management*, **24**(1), 30–41. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9711112836&site=ehost-live. Accessed: 03 July 2013.

Fung, W. & Hsieh, D. A., 1999. A primer on hedge funds. *Journal of Empirical Finance*, **6**(3), September, 309–331. [Online] Available: http://faculty.fuqua.duke.edu/~dah7/primer.pdf. Accessed: 17 March 2009.

Fung, W. & Hsieh, D. A., 2000. Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases. *Journal of Financial & Quantitative Analysis*, **35**(3), September, 291–307. [Online] Available: http://www.jstor.org/stable/2676205. Accessed: 04 July 2013.

Fung, W. & Hsieh, D. A., 2001. The risk in hedge fund strategies: theory and evidence from trend followers. *Review of Financial Studies*, **14**(2), Summer, 313–341. [Online] Available: http://www.jstor.org/stable/2696743. Accessed: 04 July 2013.

Fung, W. & Hsieh, D. A., 2002a. Asset-based style factors for hedge funds. *Financial Analysts Journal*, **58**(5), 16–27. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9406088&site=ehost-live. Accessed: 03 July 2013.

Fung, W. & Hsieh, D. A., 2002b. Benchmarks of hedge fund performance: Information content and measurement biases. *Financial Analysts Journal*, **58**, 22–34. [Online] Available: http://ssrn.com/abstract=278744. Accessed: 04 July 2013.

Fung, W. & Hsieh, D. A., 2002c. Risk in fixed-income hedge fund styles. *Journal of Fixed Income*, **12**(2), 6–27. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=7528752&site=ehost-live. Accessed: 03 July 2013.

Fung, W. & Hsieh, D. A. 2003. The risk in hedge fund strategies: Alternative alphas and alternative betas. In Jaeger, L. (eds.). *The new generation of risk management for hedge funds and private equity investments.* London: Euromoney Institutional Investor PLC.

Fung, W. & Hsieh, D. A., 2004a. Extracting portable alphas from equity long/short hedge funds. *Journal of Investment Management*, **2**(4), 57–75. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=16214738&site=ehost-live. Accessed: 03 July 2013.

Fung, W. & Hsieh, D. A., 2004b. Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, **60**(5), 65–80. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=14654795&site=ehost-live. Accessed: 03 July 2013.

Fung, W. & Hsieh, D. A., 2011. The risk in hedge fund strategies: Theory and evidence from long/short equity hedge funds. *Journal of Empirical Finance*, **18**(4), 547–569. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=65044368&site=ehost-live. Accessed: 04 July 2013.

Füss, R., Kaiser, D. G. & Adams, Z., 2007. Value at risk, GARCH modelling and the forecasting of hedge fund return volatility. *Journal of Derivatives & Hedge Funds*, **13**(1), 2–25. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=30015162&site=ehost-live. Accessed: 03 July 2013.

Geltner, D. M., 1991. Smoothing in appraisal-based returns. *Journal of Real Estate Finance and Economics*, **4**(3), September, 327–345.

Geltner, D. M., 1993. Estimating market values for appraised values without assuming an efficient market. *Journal of Real Estate Research*, **8**(3), Summer, 325–345. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=4474449&site=ehost-live. Accessed: 03 July 2013.

Getmansky, M., Lo, A. W. & Makarov, I., 2004. An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics*, **74**(3), 529–609. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=15413827&site=ehost-live. Accessed: 03 July 2013.

Getmansky, M., Lo, A. W. & Mei, S. X., 2004. Sifting through the wreckage: Lessons from recent hedge-fund liquidations. *Journal of Investment Management*, **2**(4), 6–38. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=16214736&site=ehost-live. Accessed: 03 July 2013.

Giamouridis, D. & Ntoula, I., 2009. A comparison of alternative approaches for determining the downside risk of hedge fund strategies. *Journal of Futures Markets*, **29**(3), 244–269. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=36055988&site=ehost-live. Accessed: 03 July 2013.

Glorfeld, L. W., 1995. An improvement on Horn's parallel analysis methodology for selecting the correct number of factor to retain. *Educational & Psychological Measurement*, **55**(3), 377. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9507250655&site=ehost-live. Accessed: 03 July 2013.

Goltz, F., Martellini, L. & Vaissié, M., 2007. Hedge fund indices: Reconciling investability and representativity. *European Financial Management*, **13**(2), 257–286. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=24219627&site=ehost-live. Accessed: 03 July 2013.

Goodworth, T. R. J. & Jones, C. M., 2007. Factor-based, non-parametric risk measurement framework for hedge funds and fund-of-funds. *European Journal of Finance*, **13**(7), 645–655. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=27192989&site=ehost-live. Accessed: 03 July 2013.

Gorsuch, R. L., 1974. *Factor analysis.* 1st Edition. Philadelphia: Saunders.

Granger, C. W. J., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**(3), 424–438. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=6859639&site=ehost-live. Accessed: 03 July 2013.

Grecu, A., Malkiel, B. G. & Saha, A., 2007. Why do hedge funds stop reporting performance? *Journal of Portfolio Management*, **34**(1), 119–126. [Online] Available:

http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=48072854&site=ehost-live. Accessed: 03 July 2013.

Gregoriou, G. N., 2002. Hedge fund survival lifetimes. *Journal of Asset Management*, **3**(3), September, 237–252. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=8651563&site=ehost-live. Accessed: 03 July 2013.

Gregoriou, G. N., 2005. The market neutral and event driven cemetery. *Journal of Wealth Management*, **7**(4), 67–77. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=15932123&site=ehost-live. Accessed: 03 July 2013.

Gregoriou, G. N., 2006. Survival of micro hedge funds. *Derivatives Use, Trading & Regulation*, **12**(3), 209–218. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=23697190&site=ehost-live. Accessed: 04 July 2013.

Gregoriou, G. N., Hübner, G., Papageorgiou, N. & Rouah, F. D., 2007. Funds of funds versus simple portfolios of hedge funds: A comparative study of persistence in performance. *Journal of Derivatives & Hedge Funds*, **13**(2), 88–106. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=30015167&site=ehost-live. Accessed: 03 July 2013.

Gregoriou, G. N., Kooli, M. & Rouah, F. D., 2008. Survival of strategic, market defensive, diversified and conservative fund of hedge funds: 1994 - 2005. *Journal of Derivatives & Hedge Funds*, **13**(4), February, 273–286. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=30015152&site=ehost-live. Accessed: 03 July 2013.

Gregoriou, G. N. & Rouah, F., 2001. Last year's winning hedge fund as this year's selection: A simple trading strategy. *Derivatives Use, Trading & Regulation*, **7**(3), 269–274. [Online] Available:

http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=6897660&site=ehost-live. Accessed: 03 July 2013.

Hannan, E. J. & Quinn, B. G., 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, **41**(2), 190–195. [Online] Available: http://www.jstor.org/stable/2985032. Accessed: 04 July 2013.

Harri, A. & Brorsen, B. W., 2004. Performance persistence and the source of returns for hedge funds. *Applied Financial Economics*, **14**(2), 131–141. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=11900864&site=ehost-live. Accessed: 03 July 2013.

Hasanhodzic, J. & Lo, A. W., 2007. Can hedge-fund returns be replicated? The linear case. *Journal of Investment Management*, **5**(2), 5–45. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=26863845&site=ehost-live. Accessed: 03 July 2013.

Hayes, B. T., 2006. Maximum drawdowns of hedge funds with serial correlation. *Journal of Alternative Investments*, **8**(4), 26–38. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=20567387&site=ehost-live. Accessed: 03 July 2013.

Hedge Fund Research, 2013. *HFR industry report - Synopsis Q1 2013*, October. [Online] Available: https://www.hedgefundresearch.com/index.php?fuse=products-irglo&1372937127. Accessed: 04 July 2013.

Heidorn, T., Kaiser, D. G. & Roder, C., 2009. The risk of funds of hedge funds: An empirical analysis of the maximum drawdown. *Journal of Wealth Management*, **12**(2), 89–100. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=43862989&site=ehost-live. Accessed: 03 July 2013.

Horn, J., 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30**, 179–185. [Online] Available: http://dx.doi.org/10.1007/BF02289447. Accessed: 04 July 2013.

Horst, J. R. ter, Nijman, T. E. & Roon, F. A. de, 2004. Evaluating style analysis. *Journal of Empirical Finance*, **11**(1), 29–53. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=22238872&site=ehost-live. Accessed: 03 July 2013.

Huber, C. & Kaiser, H., 2004. Hedge fund risk factors with option-like structures: Examples and explanations. *Journal of Wealth Management*, **7**(3), 49–60. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=15267245&site=ehost-live. Accessed: 03 July 2013.

Ibbotson, R. G. & Chen, P., 2006. The A,B,Cs of hedge funds: Alphas, betas, and costs. *SSRN eLibrary.* [Online] Available: http://ssrn.com/paper=733264. Accessed: 04 July 2013.

Jaeger, L. & Wagner, C., 2005. Factor modeling and benchmarking of hedge funds: Can passive investments in hedge fund strategies deliver? *Journal of Alternative Investments*, **8**(3), 9–36. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=19464147&site=ehost-live. Accessed: 03 July 2013.

Jarque, C. M. & Bera, A. K., 1987. A test for normality of observations and regression residuals. *International Statistical Review / Revue Internationale de Statistique*, **55**(2), 163–172. [Online] Available: http://www.jstor.org/stable/1403192. Accessed: 04 July 2013.

Jegadeesh, N. & Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, **48**(1), March, 65–91. [Online] Available: http://www.jstor.org/stable/2328882. Accessed: 04 July 2013.

Jensen, M. C., 1968. The performance of mutual funds in the period 1945-1964. *Journal of Finance*, **23**(2), May, 389–416. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=4661558&site=ehost-live. Accessed: 03 July 2013.

Johansen, S. 1991. Estimation and Hypothesis testing of cointegration vectors in Gaussean vector autoregressive models. *Econometrica*, **59**(6), November, 1551–1580. [Online] Available: http://www.jstor.org/stable/2938278. Accessed: 11 November 2013.

Johansen, S., 1995. Likelihood-based inference in cointegrated vector autoregressive models. Edition. Oxford: Oxford University Press.

Jolliffe, I. T., 2004. *Principal component analysis.* 2nd Edition. New York: Springer New York.

Kaiser, H., 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**(3), 187–200. [Online] Available: http://dx.doi.org/10.1007/BF02289233. Accessed: 04 July 2013.

Kat, H. M. & Lu, S., 2002. An excursion into the statistical properties of hedge fund returns. *SSRN eLibrary.* [Online] Available: http://ssrn.com/paper=310227. Accessed: 04 July 2013.

Kat, H. M. & Miffre, J., 2008. The Impact of non-normality risks and tactical trading on hedge fund alphas. *Journal of Alternative Investments*, **10**(4), 8–22. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=31822333&site=ehost-live. Accessed: 03 July 2013.

Kat, H. M. & Palaro, H. P., 2006. *Hedge fund indexation the FundCreator way: Efficient hedge fund indexation without hedge funds.* Working paper no. 8. Alternative Investment Management Association (AIMA), December. [Online] Available: http://ssrn.com/paper=949057. Accessed: 04 July 2013.

Karush, W. 1939. *Minima of functions of several variables with inequalities as side constraints*. Dissertation. University of Chicago – Department of Mathematics.

Kelejian, H. H., 1982. An extension of a standard test for heteroskedasticity to a systems framework. *Journal of Econometrics*, **20**(2), 325–333. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=4998777&site=ehost-live. Accessed: 03 July 2013.

Knuth, D. E., 2009. *Introduction to combinatorial algorithms and Boolean functions.* 4th Edition. Upper Saddle River, New Jersey: Addison-Wesley.

Kooli, M., Amvella, S. P. & Gueyié, J.-P., 2005. Hedge funds in a portfolio context: A mean-modified value at risk framework. *Derivatives Use, Trading & Regulation*, **10**(4), 373–383. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=16853145&site=ehost-live. Accessed: 03 July 2013.

Kosowski, R., Naik, N. Y. & Teo, M., 2005. Do hedge funds deliver alpha? A bayesian and bootstrap analysis. *SSRN eLibrary.* [Online] Available: http://ssrn.com/abstract=829025. Accessed: 04 July 2013.

Kouwenberg, R., 2003. Do hedge funds add value to a passive portfolio? Correcting for non-normal returns and disappearing funds. *Journal of Asset Management*, **3**(4), 361–382. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9458309&site=ehost-live. Accessed: 03 July 2013.

Kugler, P., Henn-Overbeck, J. & Zimmermann, H., 2010. Style consistency of hedge fund indexes across providers. *Applied Financial Economics*, **20**(5), 355–369. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=48068714&site=ehost-live. Accessed: 03 July 2013.

Kuhn, H. W. & Tucker, A.W. 1951. *Nonlinear programming*. Berkeley: University of California Press.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**(1/2/3), 159–178. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=6067972&site=ehost-live. Accessed: 03 July 2013.

Levchenkov, D., Coleman, T. F. & Li, Y., 2009. Estimating a hedge fund return model based on a small number of samples. *INFOR*, **47**(1), 43–58. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=46794729&site=ehost-live. Accessed: 03 July 2013.

Lhabitant, F.-S., 2004. *Hedge Funds - Quantitative Insights.* 1st Edition. Chichester: John Willey.

Liang, B., 1999. On the performance of hedge funds. *Financial Analysts Journal*, **55**(4), July/August, 72–85. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=2176439&site=ehost-live. Accessed: 03 July 2013.

Liang, B., 2000. Hedge Funds: The living and the dead. *Journal of Financial and Quantitative Analysis*, **35**(3), September, 309–326. [Online] Available: http://www.jstor.org/stable/2676206. Accessed: 04 July 2013.

Liang, B., 2001. Hedge fund performance: 1990-1999. *Financial Analysts Journal*, **57**(1), January/February, 11–18. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=4141289&site=ehost-live. Accessed: 03 July 2013.

Liang, B., 2003a. *On the performance of alternative investments: CTAs, hedge funds, and funds-of-funds.* Unpublished research report. University of Massachusetts at Amherst - Department of Finance & Operations Management; China Academy of Financial Research (CAFR), April. [Online] Available: http://ssrn.com/paper=378140. Accessed: 04 July 2013.

Liang, B., 2003b. The accuracy of hedge fund returns. *Journal of Portfolio Management*, **29**(3), 111–122. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9769488&site=ehost-live. Accessed: 03 July 2013.

Liang, B. & Park, H., 2007. Risk measures for hedge funds: a cross-sectional approach. *European Financial Management*, **13**(2), 333–370. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=24219624&site=ehost-live. Accessed: 03 July 2013.

Lilliefors, H. W., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, **62**(318), 399–402. [Online] Available: http://www.jstor.org/stable/2283970. Accessed: 04 July 2013.

Liow, K. & Yang, H., 2005. Long-term co-memories and short-run adjustment: Securitized real estate and stock markets. *Journal of Real Estate Finance & Economics*, **31**(3), 283–300. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=18188677&site=ehost-live. Accessed: 03 July 2013.

Ljung, G. M. & Box, G. E. P., 1978. On a measure of lack of fit in time series models. *Biometrika*, **65**(2), August, 297–303. [Online] Available: http://www.jstor.org/stable/2335207. Accessed: 04 July 2013.

Lo, A. W., 2001. *Risk management for hedge funds: Introduction and overview.* Unpublished research report. MIT Sloan School of Management; National Bureau of Economic Research (NBER), June. [Online] Available: http://ssrn.com/paper=283308. Accessed: 04 July 2013.

Lòpez de Prado, M. M. & Peijan, A., 2004. Measuring loss potential of hedge fund strategies. *Journal of Alternative Investments*, **7**(1), 7–31. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=13681942&site=ehost-live. Accessed: 03 July 2013.

Loudon, G., John, Okunev, J. & White, D., 2006. Hedge hund risk factors and the value at risk of fixed income trading strategies. *Journal of Fixed Income*, **16**(2), 46–61. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=22731984&site=ehost-live. Accessed: 03 July 2013.

Lütkepohl, H., 2013. Reducing confidence bands for simulated impulse responses. *Statistical Papers*, **54**(4), March, 1131–1145. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=90574710&site=ehost-live. Accessed: 13 November 2013.

Lysandrou, P., 2011. The primacy of hedge funds in the subprime crisis. *Journal of Post Keynesian Economics*, **34**(2), 225–254. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=72108614&site=ehost-live. Accessed: 04 July 2013.

Malkiel, B. G. & Saha, A., 2005. Hedge funds: Risk and return. *Financial Analysts Journal*, **61**(6), November/December, 80–88. [Online] Available: http://ssrn.com/paper=872868. Accessed: 04 July 2013.

Mitchell, M. & Pulvino, T., 2001. Characteristics of risk and return in risk arbitrage. *Journal of Finance*, **56**(6), December, 2135–2175. [Online] Available: http://www.jstor.org/stable/2697819. Accessed: 04 July 2013.

Miura, R., Aoki, Y. & Yokouchi, D., 2009. A note on statistical models for individual hedge fund returns. *Mathematical Methods of Operations Research*, **69**(3), 553–577. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=40306219&site=ehost-live. Accessed: 03 July 2013.

Neter, J., Kutner, M. H., Nachtsheim, C. J. & Wasserman, W., 1996. *Applied linear statistical models.* 4th Edition. Chicago: Irwin.

Newey, W. K. & West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**(3), 703–708. [Online] Available: http://www.jstor.org/stable/1913610. Accessed: 04 July 2013.

Okunev, J. & White, D., 2002. *Smooth returns and hedge fund risk factors.* Unpublished research report. University of Otago, August. [Online] Available: http://otago.ourarchive.ac.nz/bitstream/handle/10523/1526/SmoothReturnsHedgeFundRiskFactors2002.pdf?sequence=3. Accessed: 04 July 2013.

Phillips, P. C. B. & Perron, P., 1988. Testing for a unit root in time series regression. *Biometrika*, **75**(2), 335–346. [Online] Available: http://www.jstor.org/stable/2336182. Accessed: 04 July 2013.

Preacher, K. J., MacCullum, R. C., 2003. Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, **2**(1), 13–43. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9228740&site=ehost-live. Accessed: 13 November 2013.

Ranaldo, A. & Favre, L., 2005. Hedge fund performance and higher-moment market models. *Journal of Alternative Investments*, **8**(3), Winter, 37–51. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=19464148&site=ehost-live. Accessed: 03 July 2013.

Rencher, A. C., 2002. *Methods of multivariate analysis.* 2nd Edition. Hoboken, NJ: Wiley-Interscience.

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13**(3), December, 341–360.

Schneeweis, T., Karavas, V. N. & Georgiev, G., 2002. *Alternative investments in the institutional portfolio.* Unpublished research report. Alternative Investment Management Association (AIMA), March. [Online] Available: http://www.intercontilimited.com/mfutsarchive/alt_inv_for_inst_port_may2002.pdf. Accessed: 04 July 2013.

Schneeweis, T., Kazemi, H. B. & Szado, E., 2012. Hedge fund return-based style estimation: A review of comparison hedge fund indices. *Journal of Alternative Investments*, **15**(2), 24–53. [Online] Available: http://ssrn.com/abstract=2205885. Accessed: 04 July 2013.

Schneeweis, T. & Spurgin, R., 1998. Multifactor analysis of hedge fund, managed futures, and mutual fund retrun and risk characteristics. *Journal of Alternative Investments*, **1**(2), 1–24.

Schwartz, G., 1978. Estimating the dimension of a model. *Annals of Statistics*, **6**(2), 461–464. [Online] Available: http://www.jstor.org/stable/2958889. Accessed: 04 July 2013.

Seck, D., 1996. The substitutability of real estate assets. *Real Estate Economics*, **24**(1), 75–95. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9604072844&site=ehost-live. Accessed: 03 July 2013.

Sharpe, W. F., 1992. Asset Allocation: Management style and performance measurement. *Journal of Portfolio Management*, **18**(2), Winter, 7–19. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=15200332&site=ehost-live. Accessed: 04 July 2013.

Spurgin, R., Martin, G. & Schneeweis, T., 2001. A method of estimating changes in correlation between assets and its application to hedge fund investment. *Journal of Asset Management*, **1**(3), 217. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=7016423&site=ehost-live. Accessed: 03 July 2013.

Sun, Z., Ashley, W. & Zheng, L., 2012. The road less traveled: Strategy distinctiveness and hedge fund performance. *Journal of Financial Studies*, **12**(1), January, 96–143.

Teo, M., Koh, F. & Koh, W. T., 2003. *Asian hedge funds: Return persistence, style, and fund characteristics.* Unpublished research report. Singapore Management University, June. [Online] Available: http://ssrn.com/paper=416960. Accessed: 04 July 2013.

Tsay, R. S., 2005. *Analysis of financial time series.* 2nd Edition. Hoboken, New Jersey: Wiley-Interscience.

Urzua, C. M. 1997. Omnibus tests for multivariate normality based on a class of maximum entropy distributions. In Fomby, T. B.; Hill, R. C.; Jeliazkov, I. R.; Carlos, J. (eds.). *Advances in Econometrics.* Greenwich, Conneticut: JAI Press.

Viebig, J. & Poddig, T., 2010. Does a contagion effect exist between equity markets and hedge funds in periods of extreme stress in financial markets? *Journal of Alternative Investments*, **13**(2), 78–103. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=54374740&site=ehost-live. Accessed: 03 July 2013.

West, K. D., 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, **61**(209), 631–653. [Online] Available: http://search.ebscohost.com.ez.sun.ac.za/login.aspx?direct=true&db=buh&AN=9502074354&site=ehost-live. Accessed: 03 July 2013.

APPENDICES

List of tables in the appendix

List of figures in the appendix

Appendix 1: Tables

Table A.1: Chronological list of literature (as per *research focus*)

Author (Year)	Description	Database(s)
	Factor Modeling	
Sharpe (1992)	Asset-class-factor model and application to mutual funds.	Jaye C. Jarrett & Company (Mutual funds)
Fung and Hsieh (1997a)	Statistical factor models; PCA; factor replicating portfolios; style analysis; PTFS.	TASS
Schneeweis and Spurgin (1998)	Multifactor models: Regressors include commodity indices and intermonth volatility to account for non-linearities.	Zurich/LaPorte*
Ackermann, McEnally and Ravenscraft (1999)	Single-Factor models and systematic risk.	HFR and MAR*
Agarwal and Naik (1999)	Asset-class-factor models and stepwise regression.	Indices from HFR
Brown, Goetzmann and Ibbotson (1999)	CAPM and Jensen's Alpha.	U.S. Offshore Funds Directory and MAR
Liang (1999)	Stepwise regression models; Multi-factor model including equity, fixed income, commodity and cash proxy.	HFR

Agarwal and Naik (2000b)	Single-Factor models, appraisal ratio and performance persistence.	HFR
Edwards and Caglayan (2001)	6-factor model including Fama-French, Momentum and a yield-curve proxy.	MAR*
Gregoriou and Rouah (2001)	CAPM.	Zurich/Laporte*
Mitchell and Pulvino (2001)	Uncovered index put options to estimate the return in risk-arbitrage funds.	CRSP firms that were delisted during the period 1963 to 1998 because of a merger or acquisition
Fung and Hsieh (2002a)	Updated version of the Fung and Hsieh (1997a) article: PTFS as proxies for non-linear factor exposure.	HFR indices
Boyson (2003)	Asset-class-factor model including Fama-French and Momentum factors.	TASS
Chen and Passow (2003)	Single-Factor CAPM and multifactor models to estimate Jensen's Alpha.	Single Manager funds from FoHF
Fung and Hsieh (2003)	Asset-based style factors (ABS) and portable alpha.	TASS
Kouwenberg (2003)	Stock index options as estimators of non-linear risk exposure	Zurich/Laporte*
Teo, Koh and Koh (2003)	Local indices to be used in asset-class-factor models for Asian hedge funds.	AsiaHedge and EurekaHedge

Agarwal and Naik (2004)	Differentiation between buy-and-hold and option-based factors in the spirit of 'location' and 'style' choices described in Fung and Hsieh (2002a).	HFR and CSFB/Tremont indices
Capocci and Hübner (2004)	Single-Factor CAPM and multifactor models to estimate Jensen's Alpha.	HFR and MAR*
Harri and Brorsen (2004)	Six-factor asset-class-factor model.	Zurich/Laporte*
Capocci, Corhay and Hübner (2005)	Many factors including buy-and-hold and option-based factors described in Agarwal and Naik (2004).	MAR*
Jaeger and Wagner (2005)	Regression Beta disguised as Alpha; Under-specification of models; Covered-call-writing strategy as non-linear proxies.	TASS
Kosowski, Naik and Teo (2005)	Extension of the Fung and Hsieh (2002a) on asset-class-factor models and PTFS for many databases.	TASS, HFR, CISDM and MSCI
Ranaldo and Favre (2005)	Higher-moment models and four-moment CAPM.	HFR
Aragon (2007)	Lagged market model, broad market model, option model and Fama-French-momentum model applied to hedge fund return series.	HFN
Ding and Shawky (2007)	Higher-moment models incorporating skewness.	CISDM
Goodworth and Jones (2007)	Factor dimensionality reduction techniques, factor co-linearity and stepwise models.	Simulated data

Hasanhodzic and Lo (2007)	Six-factor asset-class-factor model.	TASS
Ammann and Moerth (2008a)	Asset-class-factor models for FoHFs; Fund size and asset flows as estimator of fund performance.	TASS
Kat and Miffre (2008)	Non-normality; Regression Beta disguised as Alpha; Under/Miss-specification of models.	TASS
Eling (2009)	Comparing several asset-class-factor models and their application to hedge funds including CAPM and Fama-French extension.	CISDM*
Levchenkov, Coleman and Li (2009)	Application of option-based market timing strategies in modeling hedge fund returns.	Simulated data
Eling and Faust (2010)	Asset-class factor models for emerging markets hedge funds using various equity and bond proxies.	CISDM*
Fung and Hsieh (2011)	Long-Short Equity: exposure to asset-based factors and managerial alpha.	TASS, HFR, CISDM
	Statistical Factor Modeling	
Fung and Hsieh (1997a, 2002a, 2004b)	PCA and factor replicating portfolios. Extensions of earlier research and out-of-sample validation.	TASS, HFR, HFR indices, Zurich Capital Markets
Barès, Gibson and Gyger (2003)	Extracting eight components using PCA.	Financial Risk Management

Amenc and Martellini (2001, 2003)	PCA and equally-weighted portfolios of hedge fund indices.	Hedge fund indices from HFN, HFR, TASS, Altvest and MAR*
Brown and Goetzmann (2003)	Generalized style classifications and statistical clustering of hedge fund return series versus self-classification.	TASS
Christiansen, Madsen and Christensen (2004)	Analogous to earlier research conducted by Fung and Hsieh (1997a).	CISDM/MAR*
Goltz, Martellini and Vaissié (2007)	Factor portfolios from individual hedge fund returns and construction of 'pure style' indices; Shortcomings of existing indices.	CISDM*
Kugler, Henn-Overbeck and Zimmermann (2010)	Extracting components from index returns to test for consistency of style classifications across providers.	Indices from HFR, Altvest, Barclay, CDFB/Tremont, CISDM, Greenwich, HFN, Hennessee
	Non-normality	
Agarwal and Naik (2000c)	On third and fourth moments of the distribution.	HFR
Huber and Kaiser (2004)	Non-normality and resulting non-linearities in regression models.	Standard & Poor's hedge fund indices
Ranaldo and Favre (2005)	Higher-moment CAPM (co-moments).	HFR
Chan, Getmansky, Haas and Lo (2006)	Higher moments and non-linear risk exposure; Regime-switches.	TASS

Chung, Johnson and Schill (2006)	Fama-French factors as proxies for higher systematic moments.	Portfolios formed from common stocks included in the Center for Research in Security Prices
Kat and Miffre (2008)	Overstatement of hedge fund alphas and the risks from non-normality.	TASS
Anand, Maier, Kutsarov and Storr (2011)	Skewness and return distribution in different regimes.	TASS, HFN
	Value-at-Risk	
Favre and Galeano (2002)	Non-normality and Value-at-Risk estimation.	No empirical results
Kooli, Amvella and Gueyié (2005)	Cornish-Fisher approximation to estimate Value-at-Risk	CISDM indices
Hayes (2006)	Maximum-drawdown-at-risk and illiquidity in hedge funds.	HFR indices
Füss, Kaiser and Adams (2007)	$GARCH$-type Value-at-Risk.	Standard & Poor's hedge fund indices
Goodworth and Jones (2007)	Factor decomposition and non-parametric Value-at-Risk.	Simulated data
Liang and Park (2007)	Cornish-Fisher approximation for Value-at-Risk and expected shortfall.	TASS
	Autocorrelation and autoregression	

Asness, Krail and Liew (2001)	On the exposure of hedge fund returns to lagged market betas; Stale and managed prices.	CSFB/Tremont indices
Lo (2001)	Autocorrelation coefficients as proxies for illiquidity in hedge funds.	Selected hedge funds from Altvest
Kat and Lu (2002)	Skewness, kurtosis and first-order autocorrelation; Unsmoothing algorithm.	TASS
Okunev and White (2002)	Correcting for first and second-order autocorrelation coefficients.	CSFB/Tremont, HFR, Zurich and MSCI
Amenc, Malaise, Martellini and Vaissié (2004)	Significance of autocorrelation coefficients (Herfindahl index and Ljung-Box).	No empirical results
Getmansky, Lo and Makarov (2004)	Application of $MA[q]$-models to hedge fund return series.	TASS
Lòpez de Prado and Peijan (2004)	$ARMA[p,q]$-models applied to hedge fund indices.	HFR indices
Loudon, John, Okunev and White (2006)	On first and second order autocorrelation coefficients and elimination thereof.	HFR indices
Bollen and Pool (2008)	Conditional return-smoothing.	CISDM*
Blazsek and Downarowicz (2008)	Non-linear autoregressive models and conditional variances of hedge fund return series; Markov-switching $ARMA-GARCH$-model.	HFR indices
Bollen and Whaley (2009)	Sources of autocorrelation and significance of $AR[1]$-models	CISDM*

Giamouridis and Ntoula (2009)	$ARMA - GARCH$-model applied to hedge fund indices.	HFR indices
Miura, Aoki and Yokouchi (2009)	Application of $AR[p]$-models to hedge fund return series.	TASS
Brandon and Wang (2013)	Systematic liquidity risk for different hedge fund strategies.	TASS
Vector autoregression		
Amenc, El Bied and Martellini (2003)	Moving averages and lagged parameters to estimate hedge fund performance.	CSFB Tremont indices
Viebig and Poddig (2010)	VAR-model for various hedge fund styles and equity proxies; Hedge fund contagion.	Unknown
Hedge fund indexing		
Amenc and Martellini (2001)	On the heterogeneity of hedge fund indices from different providers (track record, distinct style classifications, weighting scheme).	Hedge fund indices from HFN, HFR, TASS, Altvest and MAR*
Brooks and Kat (2002)	Investability versus representativeness in hedge fund indices; Survivorship bias in index returns.	Indices from HFR, Zurich Capital Markets*, CAFB/Tremont, Hennessee, Van, Altvest, HFN
Fung and Hsieh (2002b)	Measurement bias and errors in hedge fund indexing.	TASS and HFR.
Amenc, Martellini and Vaissié (2003b)	Style drift and other data bias in hedge fund indices.	Hedge fund indices from HFN, HFR, TASS, Altvest, MAR* and

		others
Kat and Palaro (2006)	On investability of hedge fund indices and double-fee structure.	HFR and CISDM*, Composite indices from Edhec
Goltz, Martellini and Vaissié (2007)	Investability versus representativeness in hedge fund indices; Communalities with FoHFs; PCA and creation of FRPs.	CISDM*
Amenc and Goltz (2008)	Data bias in hedge fund indices versus other asset indices; PCA and creation of FRPs.	Investable indices from HFR, MSCI and S&P
Kugler, Henn-Overbeck and Zimmermann (2010)	Application of factor analysis to extract representative style indices.	Indices from HFR, Altvest, Barclay, CDFB/Tremont, CISDM, Greenwich, HFN, Hennessee
Schneeweis, Kazemi and Szado (2012)	Database selection bias in and statistical properties of hedge fund indices.	CISDM*, HFR and CSFB
	Survivorship bias and Attrition	
Fung and Hsieh (1997b)	Hedge fund survivorship bias 1994 – 1998: 1.3% (FoHFs) and 3.0% (single manager); Managed futures survivorship bias 1989 – 1995: 3.5%.	TASS
Ackermann, McEnally and Ravenscraft (1999)	Survivorship bias 1988 - 1995: 0.2%.	HFR

{Brown 1999 #3 /textcit	Hedge fund survival time 1989 – 1995: 2.5 years half-life.	TASS
Fung and Hsieh (2000)	Managed futures survivorship bias 1991 – 1997: 3.6%.	TASS
Edwards and Caglayan (2001)	Survivorship bias 1990 – 1998: 1.9%.	MAR*
Liang (2001)	Survivorship bias 1990 – 1999: 2.4%; Annual attrition rate 1990 – 1999: 8.4%.	TASS
Barry (2002)	Survivorship bias 1994 – 2001: 3.8%.	TASS
Gregoriou (2002)	Hazard function applied to Market Neutral and Event Driven funds.	Zurich Capital Markets*
Amin and Kat (2003)	Survivorship bias 1994 – 2001: 0.6% (FoHFs) to 1.9% (single manager).	TASS
Getmansky, Lo and Makarov (2004)	Survivorship bias 1977 – 2001: 4.1% (as performance difference between live and defunct funds; cut-off for graveyard funds: 1994).	TASS
Getmansky, Lo and Mei (2004)	Attrition rates 1994 – 2004: 5.2% (Convertible Arbitrage) funds to 14.4% (managed futures).	TASS
Gregoriou (2005)	Hedge fund survival times 1990 – 2001: 6.08 years (Market Neutral) and 7.50 years (Event Driven).	Zurich Capital Markets*
Malkiel and Saha (2005)	Survivorship bias 1996 – 2003: 4.4%.	TASS

Gregoriou (2006)	Differences in attrition rates for funds of different size and strategic classification.	HFR
Ibbotson and Chen (2006)	Survivorship bias 1995 – 2006: 2.8% (after accounting for back-filling).	TASS
Grecu, Malkiel and Saha (2007)	Hazard function and survival times.	TASS
Gregoriou, Kooli and Rouah (2008)	Attrition rates 1994 – 2005 for different FoHF classifications.	HFR
	Other data bias	
Liang (2000)	Database selection bias; Communalities between TASS and HFR (starting dates and number of funds reporting to both vendors).	TASS and HFR
Edwards and Caglayan (2001)	Instant-history / backfilling bias 1990 – 1998: 1.2%.	MAR*
Fung and Hsieh (2001)	Instant-history / backfilling bias 1994 – 1998: 1.4%.	TASS
Barry (2002)	Instant-history / backfilling bias 1994 – 2001: 1.4%.	TASS
Fung and Hsieh (2004b)	Database selection bias; Communalities between TASS, HFR and Zurich Capital Markets.	TASS, HFR, Zurich Capital Markets*
Baquero, ter Horst and Verbeek (2005)	Look-ahead bias for 1-year horizon: 3.8%. Bias effects resulting from methodology (minimum return observations).	TASS

Capocci, Corhay and Hübner (2005)	Instant-history / backfilling bias 1994 – 2002: 1.32%.	MAR*
Grecu, Malkiel and Saha (2007)	Database selection bias and self-selection bias: Non-reporters; When do funds stop reporting - underperformance versus exceeding capital requirements.	TASS
Aggarwal and Jorion (2010)	Instant-history bias from the merger of Tass and Tremont: 1.5%	TASS/Tremont
Eling (2009)	Instant-history bias 1996 – 2005 (monthly returns) considering various cut-offs: 0.03% - 0.08% (FoHFs) and 0.18% - 0.40% (single manager)	MAR/CISDM
	Other research	
Edwards and Liew (1999)	Hedge funds / managed futures and portfolio diversification.	MAR*
Lo (2001)	On phase-locking behavior.	Selected hedge funds from Altvest
Spurgin, Martin and Schneeweis (2001)	Time-varying correlation.	EACM and MAR
Schneeweis and Spurgin (1998)	Differentiation of managed futures and hedge funds based on exposure to dynamic trading strategies.	MAR*, EACM, Barclay, TASS, HFR
Schneeweis, Karavas and Georgiev (2002)	On phase-locking behavior.	EACM

Liang (2003a)	Differences between hedge funds and managed futures; Liquidity.	Zurich Capital Markets*
Brown, Goetzmann and Liang (2004)	Double-fee structure in FoHFs.	TASS
Fung and Hsieh (2004b)	Style drift and time-varying behaviour in regression alphas and betas.	TASS and HFR
Baghai-Wadji and Klocker (2007)	Style drift performance impact on hedge funds.	CISDM*
Beckers, Curds and Weinberger (2007)	Portfolio selection and FoHFs; Managerial alpha.	HFR, HFN, TASS, CISDM, Eurekahedge, MSCI, BGI
Billio, Getmansky and Pelizzon (2007)	On phase-locking behavior and beta-switching regime models.	CSFB Tremont indices
Gregoriou, Hübner, Papageorgiou and Rouah (2007)	Single manager versus FoHFs; double-fee structure.	HFR
Agarwal and Kale (2007)	Multi-Strategy fund performance versus FoHFs.	TASS
Amo, Harasty and Hillion (2007)	FoHFs in a terminal wealth framework.	TASS
Ammann and Moerth (2008b)	Attrition rates for FoHFs; double-fee structure.	TASS
Ang, Rhodes-Kropf and Zhao (2008)	Investor’s utility function and value added from FoHFs.	TASS

Agarwal, Daniel and Naik (2009)	Hedging against incentive fees; option-like incentive fee contracts.	CISDM*. HFR, MSCI, TASS
Heidorn, Kaiser and Roder (2009)	Downside protection and absolute returns in FoHFs.	TASS
* The MAR database was sold in March 2001 to Zurich Capital Markets and gifted to the University of Massachusetts CISDM in August 2002. The last columns denote the name of the database as depicted in the respective article.		

The list above does not claim to be exhaustive. It contains previous research into the nature of hedge fund investment that is deemed relevant to the research here presented. Note that literature is sorted first according to year of publication and second according to alphabetical order of the authors. Acronyms for the databases are as in the list of acronyms and abbreviations. All figures included are per annum unless otherwise specified.

Table A.2: Common investment strategies in hedge funds

Classification	***Description***
Convertible Arbitrage	Convertible Arbitrage funds exploit pricing anomalies between convertible bonds and underlying stocks. Convertible Bond strategies benefit from differences in the implicit volatility of the convertible bond and the volatility of the underlying asset. Fund managers estimate sensitivities of convertible bonds and compare them to sensitivities of the traded stock. In the case of rising stock prices, losses from the convertible bond are overcompensated for by gains of the short-position in the underlying stock, and vice versa. Convertible Arbitrage funds use borrowed capital to leverage their performance. Despite their classification as market neutral, Convertible Arbitrage strategies are not free of systemic risks.
Event Driven	Event Driven hedge funds place bets on the outcome of events that have an impact on the performance of a single asset or a specific sector. Fund managers either respond successfully to events that induce changes in the price of an asset or actively influence the outcome of such an event. Event Driven funds invest in assets that are, to a large extent, influenced by an occurrence rather than influenced by the development of the overall market. Consequently, the performance of Event Driven strategies is independent of broad market movements. The success of the strategy depends on the manager's ability to efficiently exploit arbitrage opportunities and to correctly assess the probability of a favourable outcome of an event. Such triggering events can be hostile or friendly take-over attempts, acquisitions, spin-offs, mergers, insolvency of a company, and the restructuring of companies or part of a company in distress. Distressed Securities funds invest in companies with financial or operational difficulties. The fund manager expects that the restructuring of a part of a company or the company as a whole will be successful. In some instances, fund managers choose to engage as consultants in the company to increase the probability of a successful restructuring attempt. Distressed Securities funds either purchase (convertible) shares at a discount or take over defaulted loans from companies in distress. Additionally, hedge funds often act as creditors to companies in distress. Because of the increased shortfall and default risks inherent to

	Distressed Securities, hedge funds rarely borrow capital to leverage the performance of their strategy. Often, Distressed Securities funds buy orphan equities or junk bonds below their nominal value. In some instances, fund managers use the opportunity of companies in financial crisis to seize control of the board by purchasing the majority of outstanding shares. Merger Arbitrage funds buy or sell shares of companies that have become targets of take-overs, mergers, or leveraged buyouts. Fund managers determine the probability of the success of such take-over or merger attempts, and invest accordingly. In pair trades, the fund manager purchases shares of the take-over target and short-sells shares of the overtaking company. By using borrowed capital or stock borrowing, the fund can leverage its performance. Besides financial and operational factors, successful mergers or acquisitions often depend on external factors such as an approval given by antitrust authorities. Most Merger Arbitrage funds hedge their positions with derivatives to limit shortfall risks. As with distressed securities, managers require superior knowledge of the industry or sector to accurately assess the probability of a successful merger or take-over.
Fixed Income Arbitrage	Fixed Income Arbitrage funds place bets on pricing disparities between fixed income bonds and their derivatives. When bonds are mis-priced (e.g., the market price does not correspond with the rating of the issuer), fund managers speculate on a correction in price in the near future. To generate factor neutral portfolios, fund managers hedge their exposure to widening yield spreads, changes in the yield curve, and other market related risks with derivative instruments and short-selling. Alternatively, fund managers may purchase and short-sell corporate or government bonds with similar interest premiums on the risk-free rate (Credit Spread Arbitrage). As with Convertible Bond Arbitrage funds, disparities between market prices and expected prices should exceed the costs of hedging. Yield Curve Arbitrage strategies speculate on disparities of the yield curve. Fund managers either hedge their portfolio exposure to parallel shifts of the yield curve, or place bets on widening/narrowing spreads of bonds with longer and shorter maturity to generate additional profits. Hedge funds use the Treasury-

	Eurodollar (TED) spread to play the difference between inter-bank debentures and government bonds of similar convexity and maturity. Capital Structure Arbitrage strategies speculate on yield spreads between junior and senior bonds of the same issuer.
(Global) Macro	Macro funds are amongst the oldest and most popular alternative investment funds. There is little limitation as to the markets that Macro funds invest in. Macro funds make extensive use of leverage, deal in futures and options on currencies and commodities, and exploit arbitrage opportunities through spread trades. In addition, many funds invest in emerging markets, distressed securities, or private equity. One can differentiate between Macro funds employing different trading techniques: discretionary (fundamental evaluation of market data) versus systematic (mathematical, algorithmic and technical models) analysis, top-down (as in Global Macro funds) or bottom-up, and quantitative versus fundamental approaches. From macroeconomic data and long-term trends, hedge fund managers try to estimate the development of international markets and to identify inefficiencies between them. The investment process is predicated on movements in underlying economic variables and their impact on equity, fixed income, currency and commodity markets. Because of their critical mass, Macro funds can often exploit over/under evaluation of assets more efficiently than other market participants. Emerging Market funds predict the performance of emerging economies from macroeconomic indicators. They achieve exceptional performance through long-only investments into undervalued equities, bonds, or sovereign debt. Currency funds invest exclusively into currency options and futures to benefit from cyclical exchange rate fluctuations, short-sell overvalued currencies, or speculate on interest rate differentials between currency markets through carry trades. Because of their focus on exclusive, closed-off, or illiquid markets, Macro funds often yield higher average returns, but also exhibit higher return volatility.
Long/Short Equity	Long/Short Equity funds use pair trades to exploit market inefficiencies by short-selling overvalued stocks and simultaneously entering long positions in undervalued stocks. In a bottom-up approach, managers identify stocks that have the potential to outperform their benchmark index in the near future (stock picking),

	whilst betting against high-net-worth stocks. The success of the investment strategy depends on the manager's skill to identify and exploit inefficiencies more quickly than other market participants. When using options on stocks, Long/Short Equity managers create factor-neutral portfolios with little or no systemic risk. In practice, however, it is found that Net-Long and Long-Only (Equity-Non-Hedge) are most common. Long/Short Equity funds usually leverage their investments. Some funds hedge their exposure additionally with options or futures on the underlying assets. Dedicated Short Bias funds bet exclusively on downward markets and hence fund managers use short-selling or purchase derivatives on stocks they expect to decrease in value in the near future. Because most funds place stronger weighting on the long component of the portfolio, Long/Short Equity funds are often thought of to be strongly correlated with equity indices.
Managed futures	Managed futures developed separately from the hedge fund industry. It is believed that managed futures originated in the mid 1960s, when Dunn and Hargitt became the first CTAs for managed futures accounts. Managed futures were developed as an alternative to direct investments in forward markets to allow private and institutional investors to access exclusive or closed-off commodity markets. Professional commodity traders were able to collect and distil relevant information and monitor markets efficiently. Contrary to most hedge funds, managed futures invest only in listed options and futures on commodities or currencies. Managed futures are subject to a number of legal restrictions regarding leverage, margin-to-equity ratios, and confidentiality of information as well as fund performance. Contrary to hedge funds, managed futures are overseen by national regulatory commissions that control the fund's adherence to local legislation. Managed futures are traditionally weakly correlated with equities and treasury bonds and thus provide diversification benefits to traditional portfolios.
Multi-Strategy	Multi-Strategy funds combine several individual strategies in one portfolio. One may distinguish between single-manager Multi-Strategy funds and FoHFs creating a portfolio from individual hedge funds. The former implement dynamic strategy allocation as market conditions change, whereas the latter allocate

	capital to several managers within a strategy (style-specific) or managers across strategies (multi-strategy) to generate diversification benefits. Multi-Strategy funds employ strategies within a general theme (e.g. relative value) or use both directional as well as market neutral strategies.
Relative Value (Arbitrage)	Relative Value Arbitrage funds bet on pricing discrepancies between related instruments including equity, fixed income, derivatives and other security types. According to the EMH, pricing discrepancies are expected to dissipate over time. By simultaneously purchasing and selling related securities, managers can benefit from correctly predicting a convergence in price of the two securities. They may use mathematical, fundamental or technical analysis to identify pricing anomalies. Because of pair trading and spread trades, Relative Value Arbitrage funds have a low correlation with standard asset indices. Convertible Arbitrage, Fixed Income Arbitrage and Equity Market Neutral are often subsumed under this classification.

The table above includes some of the more common strategic themes in hedge funds. Detailed descriptions of various sub-strategies are available at https://www.hedgefundresearch.com/index.php?fuse=indices-str&1291127795#m:ms (HFR) or http://www.hedgefund.net/def.php3 (HFN).

Table A.3: Statistical properties of the frequency distribution of returns (HFR)

	m	$\bar{\mu} \times 10^{-2}$	$\bar{\sigma} \times 10^{-2}$	$\bar{m}_3$	$\bar{m}_4$	$\bar{\chi}_{JB}$	$\%\chi_{JB}$	$\bar{d}_n$	$\%d_n$
EH	104	0.877	3.174	-0.600	9.366	483.6	96.2	0.122	94.2
ED	433	1.053	5.425	0.243	7.608	399.9	92.4	0.096	82.7
M	207	1.013	4.984	0.644	6.486	335.4	77.8	0.089	72.5
RV	125	0.784	2.916	-1.512	17.864	4100.7	96.8	0.159	100.0
Total	869	0.984	4.689	-0.015	9.026	926.9	90.0	0.106	84.1

Results are for the HFR database. The timeframe under consideration is April 1990 to June 2010. Here n denotes the number of funds in the subsample as per main strategy, $\bar{\mu}$ and $\bar{\sigma}$ give the sample mean and standard deviation, $\bar{m}_3$ and $\bar{m}_4$ represent the third and fourth moment used in calculating the Jarque-Bera test statistic ($\bar{\chi}_{JB}$) and $\%\chi_{JB}$ is the proportion of sample hedge funds with significant p-values (95% confidence level) for the Jarque-Bera test, , $\bar{d}_n$ is the average Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test statistic for all funds within the sample, $\%d_n$ is the proportion of funds with significant p-values for the Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test (i.e. the proportion of funds for which the null assumption of a normal distribution is violated). Acronyms for strategies as reported in the databases as follows: EH = equity hedge, ED = event driven, m = Macro, RV = relative value.

Table A.4: Serial correlation in consecutive returns (HFR)

	$k =$											
ED	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{p}_k$	0.271	0.122	0.091	0.069	0.053	0.031	0.056	0.034	0.006	0.017	-0.002	0.001
$\hat{p}_k'$	0.271	0.040	0.038	0.016	0.017	-0.001	0.040	-0.006	-0.005	0.002	-0.018	-0.009
$\%\chi_{LB}$	82.7	78.8	70.2	71.2	74.0	72.1	71.2	72.1	69.2	68.3	67.3	66.3
EH												
$\hat{p}_k$	0.154	0.049	0.031	0.024	-0.008	0.021	0.043	0.034	0.010	-0.005	-0.001	-0.019
$\hat{p}_k'$	0.154	0.008	0.017	0.002	-0.018	0.014	0.040	0.005	0.001	-0.016	0.001	-0.030
$\%\chi_{LB}$	49.2	48.7	48.3	46.9	45.7	47.6	49.2	47.6	47.3	46.0	46.7	47.1
M												
$\hat{p}_k$	0.054	-0.017	0.006	-0.026	0.015	-0.015	-0.004	0.026	0.040	0.019	0.011	-0.014
$\hat{p}_k'$	0.054	-0.033	0.004	-0.043	0.005	-0.033	-0.007	0.006	0.032	0.006	0.012	-0.023
$\%\chi_{LB}$	18.8	22.7	21.7	24.2	25.6	26.1	25.6	25.1	27.1	27.1	28.0	27.5
RV												
$\hat{p}_k$	0.296	0.154	0.115	0.079	0.042	0.032	0.055	0.031	0.026	0.034	0.019	0.012
$\hat{p}_k'$	0.296	0.016	0.049	0.005	0.002	-0.009	0.029	-0.007	0.002	0.018	-0.011	-0.010
$\%\chi_{LB}$	83.2	81.6	79.2	78.4	79.2	79.2	80.0	79.2	79.2	78.4	76.8	76.8

Results are for the HFR database. The timeframe under consideration is April 1990 to June 2010. Here $\hat{p}_m$ denotes the average autocorrelation coefficient across the subsample at lag m, $\hat{p}_m'$ is the average partial autocorrelation coefficient at distinct lag m and $P_{\chi_{LB}}$ is the proportion of funds with significant p-values for the Ljung-Box statistic (95% confidence). Acronyms for strategies as reported in the databases as follows: EH = equity hedge, ED = event driven, m = Macro, RV = relative value.

Table A.5: Regression results for CFPs (HFR)

CFPs	m	R^2	$\widehat{R}^2$	F	k
CTA	37	44.1	40.1	$10.940^{\cdots}$	9
ED	25	82.1	81.0	$73.628^{\cdots}$	8
EHf	21	71.3	69.8	$46.895^{\cdots}$	7
EHg	44	78.6	77.1	$51.106^{\cdots}$	9
EHv	79	89.0	88.1	$99.074^{\cdots}$	10
EM	25	78.8	77.3	$51.693^{\cdots}$	9
L	32	88.9	87.9	$87.524^{\cdots}$	11
LS	49	85.3	83.9	$63.031^{\cdots}$	11
M	32	51.1	46.1	$10.249^{\cdots}$	12
RV	39	70.8	67.6	$21.657^{\cdots}$	13

Results are for the HFR database. The timeframe under consideration is July 2000 to June 2010. All error estimates are HAC at every step of the regression algorithm. Significance for overall fit as follows: $^{\cdot}$ denotes significance at 10% level, $^{\cdot\cdot}$ denotes significance at 5% level, and $^{\cdots}$ denotes significance at 1% level. Regressors without coefficient estimates either do not enter the model or are removed. The initial critical value of the F distribution for F-to-enter = 3.9 and F-to-remove = 3.8 (this corresponds to a significance level of 5% for 120 observations). The intercept is always entered. The timeframe under consideration is July 2000 to June 2010 (n = 120 observations). Here k = number of coefficients including the intercept and $\hat{R}^2 = R^2$ adjusted for k. F denotes the joint significance for the regressors entered. CTA = CTA/managed futures, ED = event driven/distressed securities, EH – Finance = equity hedge - FV finance sector, EH – Growth = equity hedge – fundamental growth, EH – Value = equity hedge – fundamental value, EM = emerging markets, LS = long/short Equity – quantitative directional, L = long bias, M = macro system/trend, RV = relative value.

Table A.6: Regression coefficients (HFR)

CTA_t	$= 0.001 + 0.079PTFSCOM_t + 0.284GOLDUSD_t - 0.258SMB_{t-1} + 0.175HML_{t-1} - 0.122GSCI_{t-1} + 0.311HML_t + 0.621BABDIDX_{t-2} + 0.036PTFSSTK_{t-1}$
ED_t	$= 0.003 + 0.142MSCIEM_t + 0.273MSCIUS_t + 0.038WML_{t-3} - 0.005PTFSIR_{t-2} + 0.149SMB_t - 0.264USDIDX_t - 0.010PTFSIR_t$
EHf_t	$= 0.005 + 0.090MSCIEM_t - 0.062CBOEVIX_t + 0.084MSCIEXUS_{t-1} + 0.030PTFSSTK_t - 0.017PTFSCOM_{t-2} - 0.035WML_{t-3}$
EHg_t	$= -0.001 + 0.123MSCIEM_t - 0.028CBOEVIX_t + 0.175MSCIEXUS_t + 0.318BABDIDX_{t-2} + 0.160SMB_t + 0.007PTFSIR_{t-3} + 0.065GOLDUSD_t - 0.016PTFSBD_{t-2}$
EHv_t	$= 0.005 + 0.122MSCIEM_t + 0.261MSCIUS_t + 0.171SMB_t - 0.278USDIDX_t - 0.035CBOEVIX_t - 0.008PTFSIR_{t-2} + 0.012PTFSFX_{t-3} + 0.034PTFSSTK_t - 0.009PTFSIR_t$
EM_t	$= -0.001 + 0.303MSCIEM_t + 0.047GSCI_t + 0.412BABDIDX_{t-2} - 0.340USDIDX_t - 0.134SMB_t - 0.016PTFSIR_t + 0.053GSCI_{t-2} - 0.084HML_{t-2}$
L_t	$= -0.016 + 0.158MSCIEXUS_t + 0.412MSCIUS_t + 0.145HML_t - 0.017PTFSCOM_{t-2} - 0.022PTFSCOM_t + 0.035WML_{t-3} - 0.012PTFSFX_{t-3} + 2.294USMO10Y_t + 0.044MSCIEM_{t-2} - 1.225USMO10Y_{t-3}$
LS_t	$= 0.001 + 0.134MSCIEM_t + 0.315MSCIUS_t + 0.109WML_t + 0.215SMB_t + 0.031WML_{t-2} + 0.348BABDIDX_{t-2} - 0.071GOLDUSD_{t-2} - 0.038HML_{t-2} + 0.019PTFSSTK_{t-2} + 0.043GOLDUSD_{t-3}$
M_t	$= -0.002 + 0.056PTFSFX_t - 0.142SMB_{t-1} + 0.193HML_t - 0.197MSCIEM_{t-3} + 0.184GOLDUSD_t + 0.415BABDIDX_{t-2} + 0.220MSCIEXUS_{t-3} - 0.032CBOEVIX_t + 0.137HML_{t-1} + 0.016PTFSIR_{t-1} - 0.030PTFSCOM_{t-1}$
RV_t	$= -0.004 + 0.052MSCIEM_t + 0.058MSCIEXUS_{t-1} - 0.020CBOEVIX_t + 0.199BABDIDX_{t-3} - 0.006PTFSIR_t + 0.011PTFSFX_{t-1} + 0.143USTB3M_t + 0.064SMB_t + 0.177USBAA10Y_{t-3} - 0.005PTFSIR_{t-3} + 0.153USDIDX_{t-2} - 0.004PTFSIR_{t-2}$

Results are for the HFR database. The timeframe under consideration is July 2000 to June 2010. Time indices reflect lagged coefficients (e.g. $t-3$ is the 3-month lagged exposure). All regressors are in order as they entered the equation. Other acronyms as in Table A.5.

Table A.7: Statistical properties of the frequency distribution of style index returns (HFR)

	n	$\mu \times 10^{-2}$	$\sigma \times 10^{-2}$	m_3	m_4	χ_{LB}	$p_{\chi_{JB}}$	d_n	p_{d_n}
CTA	37	0.676	3.430	-0.081	2.984	0.1	0.936	0.051	#N/A
ED	25	0.423	2.957	-1.287	6.629	99.8	0.000	0.092	0.014
EHf	21	0.437	2.204	-0.130	7.132	86.4	0.000	0.093	0.011
EHg	44	0.349	2.645	-0.506	3.951	9.7	0.008	0.061	#N/A
EHv	79	0.432	3.138	-1.030	5.596	55.4	0.000	0.077	0.078
EM	25	0.359	3.252	-1.450	10.312	311.9	0.000	0.150	0.000
L	32	0.462	2.888	-1.031	5.806	61.2	0.000	0.095	0.010
LS	49	0.324	2.614	-0.386	2.720	3.4	0.183	0.070	#N/A
M	32	0.382	2.551	0.244	3.581	2.9	0.235	0.062	#N/A
RV	39	0.507	1.170	-1.028	5.235	46.5	0.000	0.080	0.055

Results are for the HFR database. The timeframe under consideration is July 2000 to June 2010. Statistical significance is denoted by accents: $\dot{}$ denotes significance at 10% level, $\ddot{}$ denotes significance at 5% level, and $\dddot{}$ denotes significance at 1% level. The mean return is denoted by μ, standard deviation is σ, the third and fourth moment of the distribution are m_3 and m_4, χ_{JB} is the Jarque-Bera test statistic and $p_{\chi_{JB}}$ is the associated p-value, d_n is the Kolmogorov-Smirnov-Lilliefors Goodness-of-Fit test statistic and p_{d_n} is the associated p-value. Other acronyms as in Table A.5.

Table A.8: Serial correlation in style indices (HFR)

	$k =$											
CTA	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{p}_k$	−0.009	−0.202	−0.051	−0.015	−0.081	−0.206	0.017	0.119	0.122	−0.015	0.020	−0.029
$\hat{p}_k'$	−0.009	−0.203	−0.057	−0.060	−0.110	−0.244	−0.054	0.003	0.087	−0.012	0.038	−0.060
χ_{LB}	0.009	5.135	5.463	5.489	6.326	11.824	11.861	13.732	15.703	15.732	15.787	15.903
ED												
$\hat{p}_m$	0.280	−0.024	0.063	0.161	0.018	−0.144	−0.056	−0.022	0.033	0.004	−0.016	−0.081
$\hat{p}_m'$	0.280	−0.111	0.112	0.118	−0.062	−0.128	0.007	−0.051	0.078	0.014	−0.010	−0.101
χ_{LB}	9.757	9.828	10.335	13.636	13.678	16.350	16.766	16.829	16.976	16.978	17.015	17.900
EHf												
$\hat{p}_m$	0.102	−0.151	−0.018	0.171	−0.003	−0.094	0.045	0.046	−0.098	0.000	−0.001	−0.056
$\hat{p}_m'$	0.102	−0.163	0.017	0.152	−0.044	−0.044	0.063	−0.013	−0.090	0.057	−0.054	−0.063
χ_{LB}	1.283	4.143	4.183	7.921	7.922	9.069	9.332	9.616	10.905	10.905	10.905	11.330
EHg												
$\hat{p}_m$	0.234	−0.005	0.017	0.053	−0.040	−0.079	−0.046	−0.113	−0.118	−0.102	−0.055	−0.114
$\hat{p}_m'$	0.234	−0.063	0.036	0.042	−0.065	−0.054	−0.020	−0.111	−0.065	−0.066	−0.028	−0.105
χ_{LB}	6.786	6.789	6.827	7.179	7.380	8.194	8.476	10.168	12.018	13.411	13.824	15.606
EHv												
$\hat{p}_m$	0.237	−0.061	0.067	0.186	0.010	−0.177	−0.018	−0.015	−0.045	−0.045	0.002	−0.033
$\hat{p}_m'$	0.237	−0.124	0.121	0.141	−0.065	−0.151	0.045	−0.084	0.008	0.015	0.002	−0.057
χ_{LB}	6.961	7.424	7.993	12.397	12.410	16.452	16.493	16.521	16.787	17.061	17.062	17.212
EM												
$\hat{p}_m$	0.314	0.217	0.090	0.178	−0.063	−0.119	−0.040	−0.078	−0.150	−0.087	−0.144	−0.181
$\hat{p}_m'$	0.314	0.131	−0.012	0.145	−0.181	−0.121	0.071	−0.084	−0.085	0.042	−0.155	−0.112
χ_{LB}	12.230	18.093	19.107	23.152	23.666	25.500	25.707	26.511	29.515	30.535	33.326	37.788

L												
$\hat{p}_m$	0.191	−0.048	0.096	0.149	−0.003	−0.247	−0.062	0.055	−0.076	−0.089	0.022	0.079
$\hat{p}'_m$	0.191	−0.088	0.129	0.104	−0.041	−0.246	0.008	0.026	−0.048	0.006	0.034	0.009
χ_{LB}	$4.\ddot{5}36$	$4.\dot{8}22$	5.995	$8.\dot{8}37$	8.838	$16.\ddot{7}27$	$17.\ddot{2}34$	$17.\ddot{6}26$	$18.\ddot{3}93$	$19.\ddot{4}58$	$19.\dot{5}22$	$20.\dot{3}75$
LS												
$\hat{p}_m$	0.157	0.006	−0.011	0.032	−0.047	−0.080	0.073	0.015	−0.075	−0.047	−0.020	−0.077
$\hat{p}'_m$	0.157	−0.019	−0.010	0.037	−0.059	−0.065	0.100	−0.017	−0.077	−0.016	−0.025	−0.076
χ_{LB}	$3.\dot{0}56$	3.061	3.077	3.210	3.492	4.320	5.010	5.040	5.793	6.093	6.149	6.963
M												
$\hat{p}_m$	−0.030	−0.188	−0.036	−0.056	0.018	−0.189	−0.005	0.111	0.092	−0.011	0.016	0.054
$\hat{p}'_m$	−0.030	−0.189	−0.050	−0.099	−0.006	−0.232	−0.033	0.015	0.079	−0.017	0.064	0.039
χ_{LB}	0.108	4.548	4.709	5.105	5.145	9.789	9.793	11.410	12.532	12.547	12.581	12.978
RV												
$\hat{p}_m$	0.282	0.088	0.080	0.122	0.072	−0.086	−0.040	−0.045	0.069	0.008	−0.039	−0.144
$\hat{p}'_m$	0.282	0.009	0.057	0.092	0.011	−0.130	0.005	−0.045	0.104	−0.014	−0.027	−0.154
χ_{LB}	$9.\dddot{8}95$	$10.\dddot{8}71$	$11.\dddot{6}70$	$13.\dddot{5}76$	$14.\ddot{2}40$	$15.\ddot{1}96$	$15.\ddot{4}01$	$15.\ddot{6}65$	$16.\dot{2}89$	$16.\dot{2}97$	16.499	$19.\dot{3}33$

Results are for the HFR database. The timeframe under consideration is July 2000 to June 2010. Statistical significance is denoted by accents: $\dot{}$ denotes significance at 10% level, $\ddot{}$ denotes significance at 5% level, and $\dddot{}$ denotes significance at 1% level. Autocorrelation coefficients at lag m are $\hat{p}_k$ and partial autocorrelation coefficients $\hat{p}'_k$, χ_{LB} denotes the significance of cumulative lags according to Ljung-Box. Other acronyms as in Table A.5.

Table A.9: Results for unit root tests (HFR)

	Δy_t			$\Delta^2 y_t$		
	$H_0: y_t \sim I(1)$		$H_0: y_t \sim I(0)$	$H_0: y_t \sim I(2)$		$H_0: y_t \sim I(1)$
μ	ADF_τ	PP_τ	$KPSS_{LM}$	ADF_τ	PP_τ	$KPSS_{LM}$
BABDIDX	0.058	0.044	1.301	-9.912^{***}	-10.623^{***}	0.133
GSCI	−2.064	−2.060	0.530	-8.216^{***}	-8.315^{***}	0.099
MSCIW	−1.738	−1.807	0.327	-8.230^{***}	-8.359^{***}	0.115
CTA	−1.029	−0.839	1.272	-9.441^{***}	-12.116^{***}	0.068
ED	−1.001	−0.942	1.049	-7.970^{***}	-8.024^{***}	0.076
EHf	0.200	0.190	1.230	-8.573^{***}	-8.272^{***}	0.140
EHg	−0.750	−0.688	1.177	-8.292^{***}	-8.266^{***}	0.080
EHv	−0.957	−0.955	1.058	-8.156^{***}	-8.189^{***}	0.072
EM	−1.517	−1.587	0.884	-7.841^{***}	-8.064^{***}	0.047
L	−1.075	−1.062	1.103	-8.560^{***}	-8.639^{***}	0.057
LS	−0.471	−0.647	1.135	-9.156^{***}	-9.167^{***}	0.091
M	−2.188	−2.301	1.269	-9.549^{***}	-12.653^{***}	0.281
RV	−0.482	−0.578	1.288	-7.818^{***}	-7.880^{***}	0.059
	Δy_t			$\Delta^2 y_t$		
	$H_0: y_t \sim I(1)$		$H_0: y_t \sim I(0)$	$H_0: y_t \sim I(2)$		$H_0: y_t \sim I(1)$
$\mu + \lambda t$	ADF_τ	PP_τ	$KPSS_{LM}$	ADF_τ	PP_τ	$KPSS_{LM}$
BABDIDX	−1.885	−1.779	0.167	-9.886^{***}	-10.591^{***}	0.120
GSCI	−2.011	−1.965	0.191	-8.216^{***}	-8.314^{***}	0.046
MSCIW	−1.940	−1.984	0.145	-8.208^{***}	-8.333^{***}	0.111
CTA	-3.600^{**}	-3.726^{**}	0.185	-9.405^{***}	-12.094^{***}	0.069
ED	−2.093	−2.058	0.142	-7.939^{***}	-7.993^{***}	0.077
EHf	−2.246	−2.446	0.074	-8.697^{***}	-8.412^{***}	0.059
EHg	−2.915	−2.811	0.127	-8.285^{***}	-8.255^{***}	0.047
EHv	−2.193	−2.141	0.140	-8.122^{***}	-8.153^{***}	0.072
EM	−2.317	−2.386	0.105	-7.805^{***}	-8.031^{***}	0.048
L	−2.211	−2.203	0.137	-8.521^{***}	-8.601^{***}	0.058
LS	−2.534	−2.525	0.110	-9.145^{***}	-9.149^{***}	0.061
M	-4.208^{***}	-3.789^{**}	0.120	-9.648^{***}	-13.584^{***}	0.091
RV	−2.110	−2.112	0.128	-7.779^{***}	-7.842^{***}	0.057

Results are for the HFR database and for the July 2000 to June 2010 timeframe. The table includes three test statistics from confirmatory analysis: Augmented Dickey-Fuller (ADF), Phillips-Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin ($KPSS$). The intercept is denoted by μ, the time trend is λt, Δ^k is the k-differenced term of y_t, H_o the null hypothesis for the test statistic and $I(k)$ an integrated process of order k. Other acronyms and designation of statistical significance according to Table A.5 and Table 6.5.

Table A.10: Results for the Johansen cointegration test (HFR)

CTA								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	36.962	1	0	2	yes	yes	yes
	SBIC*	37.166	1	0	0	yes	yes	yes
Δy_{t-1}	AIC	36.959	0	0	1	yes	yes	yes
	SBIC	37.464	0	0	0	yes	yes	yes
Δy_{t-2}	AIC	36.889	0	0	1	yes	yes	yes
	SBIC	37.828	0	0	0	yes	yes	yes
Δy_{t-3}	AIC*	36.848	1	1	1	yes	yes	yes
	SBIC	38.236	1	1	0	yes	yes	yes
ED								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.969	1	1	2	yes	yes	yes
	SBIC*	35.192	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	34.994	0	0	1	no	yes	yes
	SBIC	35.552	0	0	0	yes	yes	yes
Δy_{t-2}	AIC*	34.882	1	0	2	no	yes	yes
	SBIC	35.844	0	0	0	yes	yes	yes
Δy_{t-3}	AIC	34.907	1	1	2	yes	yes	yes
	SBIC	36.271	1	1	1	no	no	yes
EHf								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.877	1	1	2	yes	yes	yes
	SBIC*	35.156	1	1	0	yes	yes	yes
Δy_{t-1}	AIC	34.880	1	1	2	yes	yes	yes
	SBIC	35.509	1	1	0	yes	yes	yes
Δy_{t-2}	AIC*	34.826	1	1	2	yes	yes	yes
	SBIC	35.864	1	1	0	yes	yes	yes
Δy_{t-3}	AIC	34.857	1	1	2	yes	yes	yes
	SBIC	36.299	1	1	1	yes	yes	yes
EHg								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	AIC	34.736	1	0	2	yes	yes	yes
	SBIC*	34.951	1	0	0	yes	yes	yes
Δy_{t-1}	AIC	34.831	0	0	1	yes	yes	yes
	SBIC	35.311	0	0	0	yes	yes	yes
Δy_{t-2}	AIC*	34.730	1	1	1	no	yes	yes
	SBIC	35.703	0	1	0	yes	yes	yes
Δy_{t-3}	AIC	34.731	1	1	2	yes	yes	yes
	SBIC	36.131	1	1	0	yes	yes	yes

EHv								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	aic*	34.758	1	1	2	yes	yes	yes
	sbic*	35.024	1	1	0	yes	yes	yes
Δy_{t-1}	aic	34.803	0	1	1	no	yes	yes
	sbic	35.374	0	0	0	yes	yes	yes
Δy_{t-2}	aic	34.783	1	1	2	no	yes	yes
	sbic	35.753	0	0	0	yes	yes	yes
Δy_{t-3}	aic	34.782	1	1	2	yes	yes	yes
	sbic	36.158	1	1	1	no	no	yes
EM								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	aic	35.365	1	0	2	yes	yes	yes
	sbic*	35.601	1	0	0	yes	yes	yes
Δy_{t-1}	aic	35.390	0	0	2	yes	yes	yes
	sbic	35.950	1	1	0	no	yes	yes
Δy_{t-2}	aic*	35.274	0	0	1	yes	yes	yes
	sbic	36.219	0	0	0	yes	yes	yes
Δy_{t-3}	aic	35.344	0	1	1	yes	yes	yes
	sbic	36.730	0	1	0	yes	yes	yes
L								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	aic	34.768	1	1	1	yes	yes	yes
	sbic*	35.068	1	1	1	yes	yes	yes
Δy_{t-1}	aic	34.818	0	1	1	yes	yes	yes
	sbic	35.422	0	1	0	yes	yes	yes
Δy_{t-2}	aic	34.835	1	1	2	yes	yes	yes
	sbic	35.855	1	1	0	yes	yes	yes
Δy_{t-3}	aic*	34.747	1	1	2	yes	yes	yes
	sbic	36.159	1	1	1	no	yes	yes
LS								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	aic*	34.455	1	1	2	yes	yes	yes
	sbic*	34.762	1	1	0	yes	yes	yes
Δy_{t-1}	aic	34.602	0	0	2	no	yes	yes
	sbic	35.108	0	0	0	yes	yes	yes
Δy_{t-2}	aic	34.530	1	1	2	no	yes	yes
	sbic	35.540	1	1	0	yes	yes	yes
Δy_{t-3}	aic	34.564	1	1	2	yes	yes	yes
	sbic	35.989	1	1	0	yes	yes	yes

M								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	aic	36.091	1	1	1	yes	yes	yes
	sbic*	36.308	1	1	0	yes	yes	yes
Δy_{t-1}	aic	36.067	1	1	2	yes	yes	yes
	sbic	36.662	1	1	0	yes	yes	yes
Δy_{t-2}	aic*	36.006	1	2	2	yes	yes	yes
	sbic	37.050	1	2	0	yes	yes	yes
Δy_{t-3}	aic	36.044	1	2	2	yes	yes	yes
	sbic	37.491	1	2	0	yes	yes	yes
RV								
max Δy_{t-k}		Stats	λ_{trace}	λ_{max}	Rank CE	Trend	Intercept	Linear
none	aic	34.026	1	0	2	yes	yes	yes
	sbic*	34.229	1	0	0	yes	yes	yes
Δy_{t-1}	aic	34.022	0	0	2	yes	yes	yes
	sbic	34.577	0	0	0	yes	yes	yes
Δy_{t-2}	aic	33.930	1	1	2	yes	yes	yes
	sbic	34.968	1	1	0	yes	yes	yes
Δy_{t-3}	aic*	33.878	1	2	2	yes	yes	yes
	sbic	35.341	2	1	1	no	no	yes

Results are for the HFR database and for the July 2000 to June 2010 timeframe. All cointegration tests for groups consisting of hedge fund index and $BABDIDX$, $GSCI$ and $MSCIW$. The table shows the Akaike Information Criterion (AIC) and Schwarz Bayesian Information Criterion (SBIC) for several test specifications and different number of lags considered for the differenced series. Here $*$ denotes the selected model according to AIC and SBIC, λ_{trace} and λ_{max} denote the number of cointegrating relations according to trace and maximum eigenvalue respectively, Δy_{t-k} is the k-differenced series for all variables. Trend and Intercept are for CE. Other acronyms according to Table A.5.

Table A.11: Testing restrictions for the adjustment coefficients of the cointegrating relationship (HFR)

χ^2	CTA	ED	EHf	EHg	EHv	EM	L	LS	M	RV
(4)	$4.\ddot{9}57$	$6.\ddot{2}26$	$9.\dddot{2}37$	$7.\dddot{7}88$	$\mathbf{4.482}$	$3.\dot{0}02$	$\mathbf{3.292}$	$19.\dddot{6}96$	$\mathbf{2.371}$	$6.\ddot{9}22$
(3)	$\mathbf{0.486}$	$\mathbf{2.786}$	$6.\ddot{5}31$	$\mathbf{1.173}$	$\mathbf{3.471}$	$\mathbf{0.120}$	$5.\dot{5}50$	$\mathbf{0.607}$	$\mathbf{0.257}$	$\mathbf{2.371}$
(3)(4)	$4.\dot{9}84$	$9.\dot{0}24$	$16.\dddot{9}57$	$7.\ddot{8}28$	$\mathbf{7.535}$	$\mathbf{3.011}$	$8.\dot{1}19$	$26.\dddot{1}90$	$\mathbf{2.619}$	$8.\dot{7}96$
(2)	$23.\dddot{0}59$	$24.\dddot{2}74$	$23.\dddot{1}50$	$8.\dddot{7}30$	$24.\dddot{0}20$	$15.\dddot{6}92$	$15.\dddot{5}63$	$18.\dddot{3}66$	$21.\dddot{1}82$	$24.\dddot{5}90$
(2)(4)	$23.\dddot{4}63$	$27.\dddot{9}18$	$35.\dddot{8}26$	$25.\dddot{6}99$	$26.\dddot{7}49$	$19.\dddot{8}71$	$18.\dddot{6}79$	$45.\dddot{1}53$	$23.\dddot{5}09$	$29.\dddot{9}15$
(2)(3)	$24.\dddot{7}80$	$27.\dddot{2}57$	$37.\dddot{3}66$	$10.\dddot{5}95$	$28.\dddot{0}48$	$15.\dddot{7}32$	$25.\dddot{0}66$	$24.\dddot{3}14$	$27.\dddot{5}32$	$30.\dddot{9}69$
(2)(3)(4)	$25.\dddot{6}84$	#N/A	#N/A	$32.\dddot{0}34$	#N/A	$19.\dddot{8}82$	#N/A	#N/A	#N/A	#N/A
(1)	$\mathbf{0.001}$	$\mathbf{3.837}$	$15.\dddot{0}99$	$8.\dddot{4}99$	$\mathbf{3.343}$	$14.\dddot{3}67$	$10.\dddot{8}06$	$14.\dddot{1}90$	$18.\dddot{9}41$	$\mathbf{3.709}$
(1)(4)	$5.\dot{0}53$	$13.\ddot{2}69$	$15.\dddot{4}79$	$8.\ddot{5}74$	$12.\ddot{4}42$	$17.\dddot{4}45$	$19.\dddot{3}44$	$22.\dddot{0}55$	$20.\dddot{5}37$	$17.\dddot{2}52$
(1)(3)	$\mathbf{0.589}$	$\mathbf{6.154}$	$28.\dddot{4}43$	$8.\ddot{9}11$	$\mathbf{6.452}$	$15.\dddot{4}13$	$16.\dddot{4}57$	$18.\dddot{7}31$	$24.\dddot{8}59$	$\mathbf{5.877}$
(1)(3)(4)	$\mathbf{5.169}$	#N/A	#N/A	$8.\ddot{9}44$	#N/A	$20.\dddot{9}34$	#N/A	#N/A	#N/A	#N/A
(1)(2)	$24.\dddot{4}18$	$26.\dddot{3}27$	$48.\dddot{5}77$	$25.\dddot{3}06$	$26.\dddot{7}29$	$25.\dddot{1}44$	$27.\dddot{8}74$	$39.\dddot{2}04$	$46.\dddot{0}37$	$26.\dddot{2}92$
(1)(2)(4)	$28.\dddot{5}28$	#N/A	#N/A	$26.\dddot{5}50$	#N/A	$26.\dddot{0}87$	#N/A	#N/A	#N/A	#N/A
(1)(2)(3)	$24.\dddot{8}03$	#N/A	#N/A	$29.\dddot{0}08$	#N/A	$28.\dddot{4}46$	#N/A	#N/A	#N/A	#N/A
	1 CE	2 CE	2 CE	1 CE	2 CE	1 CE	2 CE	2 CE	2 CE	2 CE

	For 1 CE	For 2 CE
(1)	$\alpha_{11} = 0$	$\alpha_{11} = 0, \alpha_{12} = 0$
(2)	$\alpha_{21} = 0$	$\alpha_{21} = 0, \alpha_{22} = 0$
(3)	$\alpha_{31} = 0$	$\alpha_{31} = 0, \alpha_{32} = 0$
(4)	$\alpha_{41} = 0$	$\alpha_{41} = 0, \alpha_{42} = 0$

Results are for the HFR database and for the July 2000 to June 2010 timeframe. All exclusion tests for groups consisting of hedge fund index and$BABDIDX$, $GSCI$ and $MSCIW$. CE is the number of cointegrating equations selected according to Table 7.3, χ^2 is the relevant chi-squared distributed test statistic and α_{ij} is the coefficient on the cointegrating vector for the ith variable in the jth cointegrating relation. Other acronyms and designation of statistical significance according to Table A.5 and Table 6.5.

Table A.12: Testing restrictions for the parameters in the cointegrating vector (HFR)

χ^2	CTA	ED	EHf	EHg	EHv	EM	L	LS	M	RV
(8)	**2.047**	9.656	10.218	**2.284**	9.867	**0.202**	21.205	9.040	6.004	14.749
(7)	16.124	17.128	16.145	14.504	16.987	13.855	10.624	16.034	11.878	20.469
(7)(8)	24.224	32.722	47.480	15.406	34.424	14.682	41.206	33.350	33.600	46.811
(6)	7.730	14.986	16.886	**1.222**	13.423	5.656	14.714	6.440	13.545	14.888
(6)(8)	7.731	19.741	24.718	15.533	21.382	20.357	33.686	24.044	17.639	29.841
(6)(7)	21.994	30.737	48.200	14.714	30.869	14.457	29.624	30.931	27.613	38.320
(6)(7)(8)	26.017	42.044	50.131	17.487	43.008	21.787	49.875	40.517	34.693	53.305
(5)	**2.679**	6.863	7.450	4.647	8.124	**0.256**	18.197	15.235	19.240	12.914
(5)(8)	**3.769**	16.458	19.761	5.347	17.416	**0.256**	28.988	24.283	26.760	23.828
(5)(7)	16.856	31.955	36.573	17.599	34.259	19.336	42.382	37.496	38.129	43.943
(5)(7)(8)	33.551	48.584	60.332	30.125	50.869	33.168	57.161	54.831	54.445	59.218
(5)(6)	13.321	18.285	25.408	16.690	20.205	14.034	30.683	28.982	31.244	27.463
(5)(6)(8)	14.833	28.995	33.782	19.167	30.489	20.627	41.878	38.284	38.367	40.828
(5)(6)(7)	26.701	43.859	51.309	28.352	45.271	29.142	53.696	49.132	48.810	55.169
	1 CE	2 CE	2 CE	1 CE	2 CE	1 CE	2 CE	2 CE	2 CE	2 CE

	For 1 CE	For 2 CE
(5)	$\beta_{11} = 0$	$\beta_{11} = 0, \beta_{21} = 0$
(6)	$\beta_{12} = 0$	$\beta_{12} = 0, \beta_{22} = 0$
(7)	$\beta_{13} = 0$	$\beta_{13} = 0, \beta_{23} = 0$
(8)	$\beta_{14} = 0$	$\beta_{14} = 0, \beta_{24} = 0$

Results are for the HFR database and for the July 2000 to June 2010 timeframe. All exclusion tests for groups consisting of hedge fund index and $BABDIDX$, $GSCI$ and $MSCIW$. CE is the number of cointegrating equations selected according toTable 7.3, χ^2 is the relevant chi-squared distributed test statistic and β_{ij} represents the jth coefficient in the ith cointegrating relationship. Other acronyms and designation of statistical significance according to Table A.5 and Table 6.5.

Table A.13: Testing for combined restrictions on adjustment coefficient and vector parameters (HFR)

p_{χ^2} rank	CTA	ED	EHF	EHG	EHV	EM	L	LS	M	RV
1	(1)	(3)	(3)	(3)(8)	(1)	**(3)(5)(8)**	**(4)**	**(3)**	(3)	(3)
2	(1)(3)	**(1)(3)**	(5)	(3)	(3)	(5)(8)	(3)(4)	(6)	(3)(4)	**(1)(3)**
3	(3)	(1)	(4)	(6)	**(1)(3)**	(3)(8)	(3)	(3)(6)	(4)	(1)
4	(1)(8)	(3)(4)	(3)(8)	(3)(5)(8)	(3)(4)	(3)(5)	(4)(7)	(3)(8)	(3)(8)	(3)(4)
5	**(1)(3)(8)**	(4)	(8)	(8)	(4)	(3)	(7)	(8)	**(3)(4)(8)**	(4)
6	(1)(3)(5)	(1)(4)(5)	(3)(5)	**(3)(6)**	(1)(4)(5)	(8)	(1)	(1)(8)	(4)(8)	(1)(4)(5)
7	(3)(5)	(5)	(1)(4)	(1)(3)(4)(5)(8)	(1)(4)(8)	(5)	(2)(7)	(1)(3)(8)	(8)	(1)(4)(8)
p_{χ^2} rank	CTA	ED	EHF	EHG	EHV	EM	L	LS	M	RV
1	**0.001**	**2.786**	6.531	**2.394**	**3.343**	**0.420**	**3.292**	**0.607**	**0.257**	**2.371**
2	**0.589**	**6.154**	7.450	**1.173**	**3.471**	**0.256**	8.119	6.440	**2.619**	**5.877**
3	**0.486**	**3.837**	9.237	**1.222**	**6.452**	**0.288**	5.550	11.332	**2.371**	**3.709**
4	**2.049**	9.024	14.041	**5.558**	**7.535**	**0.402**	14.017	12.688	**6.719**	8.796
5	**3.757**	6.226	10.218	**2.284**	**4.482**	**0.120**	10.624	9.040	**9.588**	6.922
6	**3.950**	13.516	14.817	**4.268**	13.192	**0.202**	10.806	14.259	8.569	18.189
7	**2.732**	6.863	15.479	**9.177**	14.829	**0.256**	16.345	19.562	6.004	19.301
	1 CE	2 CE	2 CE	1 CE	2 CE	1 CE	2 CE	2 CE	2 CE	2 CE

	for 1 CE	for 2 CE
(1)	$\alpha_{11} = 0$	$\alpha_{11} = 0, \alpha_{12} = 0$
(2)	$\alpha_{21} = 0$	$\alpha_{21} = 0, \alpha_{22} = 0$
(3)	$\alpha_{31} = 0$	$\alpha_{31} = 0, \alpha_{32} = 0$
(4)	$\alpha_{41} = 0$	$\alpha_{41} = 0, \alpha_{42} = 0$
(5)	$\beta_{11} = 0$	$\beta_{11} = 0, \beta_{21} = 0$
(6)	$\beta_{12} = 0$	$\beta_{12} = 0, \beta_{22} = 0$
(7)	$\beta_{13} = 0$	$\beta_{13} = 0, \beta_{23} = 0$
(8)	$\beta_{14} = 0$	$\beta_{14} = 0, \beta_{24} = 0$

Results are for the HFR database and for the July 2000 to June 2010 timeframe. All exclusion tests for groups consisting of hedge fund index and $BABDIDX$, $GSCI$ and $MSCIW$. CE is the number of cointegrating equations and the results are ranked by the associated p-value of χ^2. Other acronyms and designation of statistical significance according to Table A.5 and Table 6.5.

Table A.14: Model statistics and diagnostics – CTA

Model statistics CTA					
	eq_1	eq_2	eq_3	eq_4	system
F	1.326	4.653	2.670	2.752	
R^2	0.142	0.368	0.250	0.256	
$\bar{R}^2$	0.035	0.289	0.157	0.163	
AIC	6.727	5.092	14.932	10.730	37.303
SBIC	7.056	5.420	15.261	11.059	38.735
k	14	14	14	14	56
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.612	13.050	6.641	4.981	26.283
χ_{JB} (corr)	0.896	16.248	4.540	9.025	30.708
χ_{JB} (cov)	1.322	13.695	10.230	6.200	171.180
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	67.338				
resid_2	28.960	10.137			
resid_3	69.917	38.117	64.357		
resid_4	40.958	43.847	57.506	33.135	
Joint					353.930
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	2.521	lag 5	43.105	lag 9	114.176
lag 2*	5.487	lag 6	63.878	lag 10	130.485
lag 3*	10.090	lag 7	74.534	lag 11	154.343
lag 4	22.429	lag 8	92.791	lag 12	166.751
	LM		LM		LM
lag 1	17.607	lag 5	27.200	lag 9	24.811
lag 2	14.961	lag 6	23.233	lag 10	19.464
lag 3	21.061	lag 7	13.647	lag 11	29.261
lag 4	17.387	lag 8	20.064	lag 12	15.089

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.15: Model statistics and diagnostics – ED

Model statistics ED					
	eq_1	eq_2	eq_3	eq_4	system
F	2.621	5.916	1.759	2.472	
R^2	0.263	0.446	0.193	0.252	
$\bar{R}^2$	0.162	0.370	0.083	0.150	
AIC	5.530	4.977	15.023	10.753	34.888
SBIC	5.883	5.329	15.375	11.105	36.532
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	19.612	6.333	19.483	9.200	54.627
χ_{JB} (corr)	9.377	5.789	8.127	5.415	28.707
χ_{JB} (cov)	8.203	6.376	14.955	6.708	216.110
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	62.916				
resid_2	32.276	13.298			
resid_3	61.122	26.835	78.675		
resid_4	59.292	32.279	60.205	50.861	
joint					341.871
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.732	lag 5	43.461	lag 9	113.210
lag 2*	3.762	lag 6	62.071	lag 10	130.624
lag 3*	8.322	lag 7	71.630	lag 11	148.028
lag 4	25.733	lag 8	85.576	lag 12	161.839
	LM		LM		LM
lag 1	20.146	lag 5	25.617	lag 9	32.988
lag 2	16.566	lag 6	24.723	lag 10	23.940
lag 3	26.369	lag 7	16.981	lag 11	25.441
lag 4	23.805	lag 8	19.179	lag 12	19.572

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.16: Model statistics and diagnostics – EHf

Model statistics EHf					
	eq_1	eq_2	eq_3	eq_4	system
F	4.038	3.938	3.095	2.914	
R^2	0.354	0.349	0.296	0.284	
$\bar{R}^2$	0.267	0.260	0.200	0.186	
AIC	4.787	5.138	14.886	10.709	34.676
SBIC	5.139	5.491	15.238	11.061	36.320
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	6.011	17.890	6.046	5.289	35.235
χ_{JB} (corr)	42.418	8.984	2.789	1.927	56.119
χ_{JB} (cov)	170.934	14.135	2.374	2.962	367.697
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	52.656				
resid_2	25.156	19.070			
resid_3	44.240	39.619	70.082		
resid_4	35.893	35.097	40.413	35.195	
joint					399.839
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.046	lag 5	44.093	lag 9	111.921
lag 2*	2.987	lag 6	62.126	lag 10	126.491
lag 3*	9.129	lag 7	84.454	lag 11	140.474
lag 4	27.410	lag 8	91.309	lag 12	156.996
	LM		LM		LM
lag 1	17.272	lag 5	22.243	lag 9	25.571
lag 2	12.296	lag 6	19.983	lag 10	15.263
lag 3	26.051	lag 7	29.511	lag 11	14.371
lag 4	30.718	lag 8	7.451	lag 12	19.241

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.17: Model statistics and diagnostics – EHg

Model statistics EHg					
	eq_1	eq_2	eq_3	eq_4	system
F	2.392	4.384	2.116	3.649	
R^2	0.245	0.373	0.223	0.332	
$\bar{R}^2$	0.143	0.288	0.118	0.241	
AIC	4.992	5.100	14.985	10.640	34.674
SBIC	5.345	5.452	15.337	10.992	36.318
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.755	4.573	3.759	12.522	22.609
χ_{JB} (corr)	3.446	4.426	6.333	7.712	21.917
χ_{JB} (cov)	3.188	2.951	14.624	3.923	147.383
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	41.735				
resid_2	40.395	19.399			
resid_3	55.688	53.510	69.604		
resid_4	47.310	50.627	61.349	38.991	
joint					350.939
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.075	lag 5	36.917	lag 9	116.997
lag 2*	2.257	lag 6	52.656	lag 10	131.762
lag 3*	6.646	lag 7	62.041	lag 11	154.321
lag 4	19.029	lag 8	90.597	lag 12	168.056
	LM		LM		LM
lag 1	10.403	lag 5	20.180	lag 9	27.186
lag 2	7.635	lag 6	16.417	lag 10	15.083
lag 3	13.920	lag 7	9.449	lag 11	23.684
lag 4	13.160	lag 8	28.063	lag 12	14.222

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.18: Model statistics and diagnostics – EHv

Model statistics EHv					
	eq_1	eq_2	eq_3	eq_4	system
F	2.862	5.255	1.877	2.327	
R^2	0.280	0.417	0.203	0.240	
$\bar{R}^2$	0.182	0.337	0.095	0.137	
AIC	5.658	5.028	15.010	10.768	34.838
SBIC	6.010	5.380	15.362	11.120	36.482
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	9.429	7.097	23.488	16.548	56.562
χ_{JB} (corr)	5.658	5.820	9.587	4.427	25.492
χ_{JB} (cov)	22.071	7.010	17.979	5.284	247.916
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	49.374				
resid_2	30.905	17.629			
resid_3	59.817	30.299	77.548		
resid_4	47.495	34.356	58.751	44.923	
joint					334.784
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.918	lag 5	44.504	lag 9	116.617
lag 2*	4.514	lag 6	66.166	lag 10	133.995
lag 3*	10.363	lag 7	76.493	lag 11	150.748
lag 4	27.482	lag 8	87.982	lag 12	161.513
	LM		LM		LM
lag 1	23.311	lag 5	26.410	lag 9	35.426
lag 2	19.569	lag 6	30.007	lag 10	23.800
lag 3	26.461	lag 7	17.548	lag 11	26.050
lag 4	25.172	lag 8	18.294	lag 12	18.306

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.19: Model statistics and diagnostics – EM

Model statistics EM					
	eq_1	eq_2	eq_3	eq_4	system
F	4.104	4.011	1.682	4.084	
R^2	0.358	0.353	0.186	0.357	
$\bar{R}^2$	0.271	0.265	0.075	0.270	
AIC	5.397	5.132	15.031	10.601	35.076
SBIC	5.750	5.484	15.384	10.953	36.720
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	39.608	3.989	38.566	0.066	82.230
χ_{JB} (corr)	23.496	4.300	11.753	1.261	40.810
χ_{JB} (cov)	21.278	4.224	16.808	3.681	201.539
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	48.049				
resid_2	54.800	19.141			
resid_3	54.203	54.595	77.179		
resid_4	40.785	41.569	52.963	31.438	
joint					350.309
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.147	lag 5	32.866	lag 9	91.001
lag 2*	3.060	lag 6	48.280	lag 10	104.873
lag 3*	6.788	lag 7	58.662	lag 11	126.685
lag 4	16.426	lag 8	73.051	lag 12	149.718
	LM		LM		LM
lag 1	13.366	lag 5	21.162	lag 9	18.137
lag 2	14.841	lag 6	17.786	lag 10	15.374
lag 3	15.029	lag 7	12.634	lag 11	25.798
lag 4	12.864	lag 8	16.178	lag 12	25.595

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.20: Model statistics and diagnostics – L

Model statistics L					
	eq_1	eq_2	eq_3	eq_4	system
F	3.123	3.773	3.286	2.577	
R^2	0.298	0.339	0.309	0.259	
$\bar{R}^2$	0.203	0.249	0.215	0.159	
AIC	5.563	5.153	14.868	10.742	34.671
SBIC	5.915	5.505	15.220	11.094	36.315
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	3.180	5.060	14.105	36.273	58.617
χ_{JB} (corr)	16.789	5.085	4.416	4.805	31.095
χ_{JB} (cov)	101.593	4.267	7.481	4.645	574.455
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	29.909				
resid_2	30.835	18.102			
resid_3	26.584	22.944	54.228		
resid_4	22.135	34.341	25.059	24.979	
joint					328.765
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	3.496	lag 5	51.005	lag 9	125.323
lag 2*	7.974	lag 6	68.321	lag 10	146.674
lag 3*	16.255	lag 7	81.701	lag 11	157.344
lag 4	37.721	lag 8	98.217	lag 12	169.229
	LM		LM		LM
lag 1	26.104	lag 5	20.472	lag 9	30.142
lag 2	21.230	lag 6	24.342	lag 10	24.756
lag 3	30.684	lag 7	17.840	lag 11	15.524
lag 4	25.954	lag 8	18.585	lag 12	16.742

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.21: Model statistics and diagnostics – LS

Model statistics LS					
	eq_1	eq_2	eq_3	eq_4	system
F	4.050	4.153	1.961	4.855	
R^2	0.355	0.361	0.210	0.398	
$\bar{R}^2$	0.267	0.274	0.103	0.316	
AIC	4.927	5.119	15.001	10.536	34.574
SBIC	5.279	5.472	15.353	10.888	36.218
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	0.152	4.989	9.239	15.633	30.013
χ_{JB} (corr)	1.857	4.143	7.596	11.547	25.143
χ_{JB} (cov)	0.440	3.304	19.074	4.721	201.728
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	35.162				
resid_2	32.951	20.784			
resid_3	38.722	45.729	70.148		
resid_4	33.617	46.846	47.746	24.477	
joint					310.239
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	0.989	lag 5	42.499	lag 9	110.504
lag 2*	3.156	lag 6	58.422	lag 10	123.149
lag 3*	5.766	lag 7	73.080	lag 11	145.004
lag 4	18.239	lag 8	85.860	lag 12	162.404
	LM		LM		LM
lag 1	11.672	lag 5	29.241	lag 9	25.802
lag 2	18.422	lag 6	17.331	lag 10	12.603
lag 3	9.546	lag 7	16.787	lag 11	22.427
lag 4	14.314	lag 8	13.920	lag 12	18.446

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.22: Model statistics and diagnostics – M

Model statistics M					
	eq_1	eq_2	eq_3	eq_4	system
F	3.103	3.866	2.086	2.628	
R^2	0.297	0.344	0.221	0.263	
$\bar{R}^2$	0.201	0.255	0.115	0.163	
AIC	5.205	5.145	14.988	10.737	36.134
SBIC	5.557	5.497	15.340	11.089	37.778
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.672	8.458	6.723	5.931	22.784
χ_{JB} (corr)	1.907	8.992	6.155	7.560	24.615
χ_{JB} (cov)	1.489	7.411	11.646	6.267	122.357
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	23.716				
resid_2	43.283	21.399			
resid_3	34.834	42.232	64.544		
resid_4	44.065	49.592	46.358	34.660	
joint					327.466
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	2.144	lag 5	42.445	lag 9	102.333
lag 2*	5.049	lag 6	58.813	lag 10	120.929
lag 3*	8.869	lag 7	68.986	lag 11	143.026
lag 4	22.654	lag 8	85.390	lag 12	154.477
	LM		LM		LM
lag 1	14.473	lag 5	26.801	lag 9	20.621
lag 2	22.185	lag 6	21.297	lag 10	21.747
lag 3	19.656	lag 7	13.320	lag 11	26.870
lag 4	18.462	lag 8	21.360	lag 12	17.167

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.23: Model statistics and diagnostics – RV

Model statistics RV					
	eq_1	eq_2	eq_3	eq_4	system
F	2.861	5.420	2.127	2.260	
R^2	0.280	0.424	0.224	0.235	
$\bar{R}^2$	0.182	0.346	0.119	0.131	
AIC	3.922	5.015	14.983	10.775	33.859
SBIC	4.274	5.367	15.336	11.127	35.503
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	23.013	4.739	27.778	3.900	59.429
χ_{JB} (corr)	6.258	6.153	8.471	1.430	22.313
χ_{JB} (cov)	5.192	4.719	14.488	7.731	182.676
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	39.476				
resid_2	42.189	10.572			
resid_3	39.725	28.704	67.874		
resid_4	32.607	31.917	40.330	31.235	
joint					324.000
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	3.186	lag 5	47.698	lag 9	115.065
lag 2*	6.134	lag 6	57.353	lag 10	131.630
lag 3*	10.262	lag 7	72.522	lag 11	150.098
lag 4	32.827	lag 8	89.022	lag 12	166.858
	LM		LM		LM
lag 1	30.981	lag 5	22.067	lag 9	33.267
lag 2	19.891	lag 6	16.271	lag 10	25.783
lag 3	23.610	lag 7	22.500	lag 11	27.237
lag 4	29.113	lag 8	21.322	lag 12	22.608

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.24: Model (incl. dummy) statistics and diagnostics – CTA

Model statistics CTA					
	eq_1	eq_2	eq_3	eq_4	system
F	1.228	5.671	4.408	4.270	
R^2	0.143	0.435	0.375	0.367	
$\bar{R}^2$	0.027	0.359	0.290	0.281	
AIC	6.743	4.996	14.768	10.585	36.813
SBIC	7.095	5.348	15.120	10.937	38.339
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.440	14.251	0.634	5.936	22.261
χ_{JB} (corr)	0.429	13.938	1.882	6.641	22.889
χ_{JB} (cov)	0.469	14.525	4.645	8.471	84.672
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	62.631				
resid_2	24.064	12.459			
resid_3	55.385	21.560	47.827		
resid_4	32.639	23.151	33.164	26.499	
Joint					295.153
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	4.014	lag 5	44.925	lag 9	122.865
lag 2*	12.170	lag 6	68.946	lag 10	132.458
lag 3*	16.538	lag 7	82.996	lag 11	149.605
lag 4	26.784	lag 8	96.968	lag 12	162.301
	LM		LM		LM
lag 1	21.332	lag 5	19.239	lag 9	26.862
lag 2	20.109	lag 6	26.396	lag 10	10.507
lag 3	13.901	lag 7	14.965	lag 11	17.757
lag 4	10.736	lag 8	15.859	lag 12	13.220

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.25: Model (incl. dummy) statistics and diagnostics – ED

Model statistics ED					
	eq_1	eq_2	eq_3	eq_4	system
F	6.978	5.919	3.386	4.903	
R^2	0.506	0.465	0.332	0.419	
$\bar{R}^2$	0.434	0.387	0.234	0.334	
AIC	5.146	4.958	14.850	10.517	34.342
SBIC	5.522	5.333	15.226	10.892	36.080
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	2.406	9.930	6.655	7.280	26.271
χ_{JB} (corr)	6.833	6.618	6.893	8.744	29.087
χ_{JB} (cov)	9.592	8.887	6.146	11.347	178.290
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	33.927				
resid_2	20.751	15.293			
resid_3	36.200	16.324	59.974		
resid_4	27.375	26.596	20.818	29.190	
joint					288.589
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	5.041	lag 5	49.216	lag 9	114.385
lag 2*	9.529	lag 6	64.895	lag 10	126.928
lag 3*	13.554	lag 7	75.072	lag 11	140.771
lag 4	33.632	lag 8	82.984	lag 12	152.690
	LM		LM		LM
lag 1	19.325	lag 5	17.771	lag 9	33.194
lag 2	17.548	lag 6	18.912	lag 10	13.636
lag 3	20.869	lag 7	11.982	lag 11	15.563
lag 4	23.971	lag 8	9.377	lag 12	13.236

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.26: Model (incl. dummy) statistics and diagnostics – EHf

Model statistics EHf					
	eq_1	eq_2	eq_3	eq_4	system
F	5.282	4.912	4.216	4.157	
R^2	0.437	0.419	0.383	0.379	
$\bar{R}^2$	0.354	0.334	0.292	0.288	
AIC	4.667	5.040	14.772	10.582	34.326
SBIC	5.042	5.416	15.148	10.958	36.064
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	19.327	7.843	11.115	3.824	42.108
χ_{JB} (corr)	62.764	4.908	7.079	2.755	77.506
χ_{JB} (cov)	164.444	6.233	8.242	4.999	305.957
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	61.199				
resid_2	24.248	21.951			
resid_3	39.781	24.734	52.897		
resid_4	43.021	29.236	24.656	37.566	
joint					390.491
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	2.580	lag 5	54.739	lag 9	126.023
lag 2*	12.213	lag 6	74.117	lag 10	139.318
lag 3*	18.028	lag 7	94.489	lag 11	151.494
lag 4	35.449	lag 8	100.619	lag 12	165.164
	LM		LM		LM
lag 1	16.567	lag 5	25.515	lag 9	27.740
lag 2	29.158	lag 6	21.840	lag 10	13.868
lag 3	24.837	lag 7	25.600	lag 11	12.487
lag 4	29.511	lag 8	6.800	lag 12	16.393

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.27: Model (incl. dummy) statistics and diagnostics – EHg

Model statistics EHg					
	eq_1	eq_2	eq_3	eq_4	system
F	3.956	5.468	4.697	5.059	
R^2	0.368	0.446	0.409	0.427	
$\bar{R}^2$	0.275	0.364	0.322	0.342	
AIC	4.832	4.994	14.729	10.503	34.280
SBIC	5.208	5.370	15.105	10.879	36.017
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	0.449	$4.\dot{7}05$	$6.\ddot{1}34$	$8.\ddot{7}83$	$20.\ddot{0}72$
χ_{JB} (corr)	2.260	4.085	4.317	$7.\ddot{4}55$	$18.\ddot{1}16$
χ_{JB} (cov)	3.203	4.546	$6.\ddot{5}31$	$4.\dot{9}61$	$138.\dddot{3}90$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	33.920				
resid_2	21.758	19.040			
resid_3	36.037	30.885	$40.\dot{3}34$		
resid_4	$43.\ddot{0}18$	28.029	31.590	$42.\ddot{0}52$	
joint					303.225
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	3.770	lag 5	$47.\dot{9}68$	lag 9	$125.\ddot{0}24$
lag 2*	7.131	lag 6	$66.\dot{3}91$	lag 10	134.311
lag 3*	11.703	lag 7	77.381	lag 11	150.584
lag 4	$30.\dot{1}49$	lag 8	95.765	lag 12	166.652
	LM		LM		LM
lag 1	$24.\dot{3}39$	lag 5	21.036	lag 9	$33.\dddot{5}16$
lag 2	19.926	lag 6	22.597	lag 10	11.785
lag 3	14.927	lag 7	13.541	lag 11	19.874
lag 4	21.445	lag 8	22.865	lag 12	17.409

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.28: Model (incl. dummy) statistics and diagnostics – EHv

Model statistics EHv					
	eq_1	eq_2	eq_3	eq_4	system
F	5.821	5.339	3.368	4.054	
R^2	0.461	0.440	0.331	0.373	
$\bar{R}^2$	0.382	0.357	0.233	0.281	
AIC	5.385	5.004	14.852	10.592	34.400
SBIC	5.760	5.380	15.228	10.968	36.137
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	4.778	9.731	8.503	14.525	37.537
χ_{JB} (corr)	9.721	6.543	7.849	7.289	31.401
χ_{JB} (cov)	23.481	9.243	6.264	9.981	196.778
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	23.544				
resid_2	18.449	19.691			
resid_3	33.996	17.734	59.880		
resid_4	23.147	27.318	21.880	28.966	
joint					296.219
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	5.468	lag 5	53.553	lag 9	123.411
lag 2*	11.418	lag 6	74.276	lag 10	136.338
lag 3*	15.233	lag 7	85.400	lag 11	149.783
lag 4	35.870	lag 8	91.870	lag 12	159.011
	LM		LM		LM
lag 1	24.693	lag 5	20.720	lag 9	34.251
lag 2	21.198	lag 6	26.247	lag 10	14.922
lag 3	17.591	lag 7	14.275	lag 11	16.848
lag 4	25.459	lag 8	9.181	lag 12	11.912

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.29: Model (incl. dummy) statistics and diagnostics – EM

Model statistics EM					
	eq_1	eq_2	eq_3	eq_4	system
F	7.024	5.512	3.590	5.188	
R^2	0.508	0.448	0.346	0.433	
$\bar{R}^2$	0.436	0.366	0.249	0.349	
AIC	5.148	4.990	14.830	10.493	34.625
SBIC	5.524	5.366	15.206	10.868	36.363
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	$56.\dddot{2}37$	$5.\dot{1}29$	$13.\dddot{2}52$	0.083	$74.\dddot{7}02$
χ_{JB} (corr)	$30.\dddot{9}51$	4.588	$10.\dddot{0}55$	1.937	$47.\dddot{5}32$
χ_{JB} (cov)	$31.\dddot{2}65$	$4.\dot{7}39$	$5.\dot{3}27$	$5.\dot{3}57$	$163.\dddot{0}56$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	33.956				
resid_2	$47.\dddot{4}98$	14.183			
resid_3	18.816	21.958	$51.\dddot{0}75$		
resid_4	27.429	32.700	23.492	31.653	
joint					$334.\ddot{1}69$
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	2.740	lag 5	38.742	lag 9	95.053
lag 2*	6.119	lag 6	51.749	lag 10	107.721
lag 3*	8.774	lag 7	62.603	lag 11	125.537
lag 4	24.657	lag 8	70.734	lag 12	144.117
	LM		LM		LM
lag 1	19.537	lag 5	21.149	lag 9	$29.\ddot{0}63$
lag 2	$23.\dot{9}38$	lag 6	16.891	lag 10	15.941
lag 3	14.364	lag 7	14.501	lag 11	$24.\dot{2}77$
lag 4	21.305	lag 8	12.565	lag 12	22.463

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.30: Model (incl. dummy) statistics and diagnostics – L

Model statistics L					
	eq_1	eq_2	eq_3	eq_4	system
F	4.550	4.549	4.066	3.527	
R^2	0.401	0.401	0.374	0.342	
$\bar{R}^2$	0.313	0.313	0.282	0.245	
AIC	5.421	5.072	14.786	10.642	34.345
SBIC	5.797	5.447	15.161	11.017	36.083
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	6.029	5.997	15.690	25.816	53.533
χ_{JB} (corr)	12.323	4.170	7.324	13.645	37.461
χ_{JB} (cov)	42.024	4.638	11.941	14.128	392.349
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	27.800				
resid_2	15.055	22.750			
resid_3	31.712	19.930	50.723		
resid_4	23.562	19.765	17.779	26.863	
joint					316.777
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	9.496	lag 5	59.314	lag 9	132.852
lag 2*	16.046	lag 6	77.003	lag 10	147.668
lag 3*	22.37	lag 7	90.207	lag 11	156.569
lag 4	39.745	lag 8	100.610	lag 12	169.920
	LM		LM		LM
lag 1	31.204	lag 5	25.490	lag 9	36.832
lag 2	26.337	lag 6	23.596	lag 10	19.035
lag 3	26.012	lag 7	17.914	lag 11	13.748
lag 4	22.716	lag 8	14.136	lag 12	19.490

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.31: Model (incl. dummy) statistics and diagnostics – LS

Model statistics LS					
	eq_1	eq_2	eq_3	eq_4	system
F	4.476	5.622	4.104	5.891	
R^2	0.397	0.453	0.376	0.464	
$\bar{R}^2$	0.308	0.372	0.285	0.385	
AIC	4.877	4.981	14.782	10.436	34.190
SBIC	5.252	5.357	15.158	10.811	35.928
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	0.085	4.498	2.187	8.643	15.414
χ_{JB} (corr)	1.650	3.570	2.083	10.874	18.177
χ_{JB} (cov)	0.447	4.067	1.346	4.658	113.518
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	30.633				
resid_2	23.914	18.135			
resid_3	28.341	27.927	47.826		
resid_4	35.776	31.697	26.779	31.108	
joint					306.551
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	3.619	lag 5	51.344	lag 9	124.757
lag 2*	6.345	lag 6	64.987	lag 10	135.630
lag 3*	8.913	lag 7	76.119	lag 11	149.494
lag 4	25.597	lag 8	90.293	lag 12	163.065
	LM		LM		LM
lag 1	22.466	lag 5	31.594	lag 9	40.862
lag 2	26.605	lag 6	19.153	lag 10	13.774
lag 3	11.737	lag 7	14.797	lag 11	19.530
lag 4	20.937	lag 8	19.390	lag 12	17.183

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.32: Model (incl. dummy) statistics and diagnostics – M

Model statistics M					
	eq_1	eq_2	eq_3	eq_4	system
F	2.874	5.318	3.815	4.389	
R^2	0.297	0.439	0.359	0.392	
$\bar{R}^2$	0.194	0.356	0.265	0.303	
AIC	5.221	5.006	14.809	10.561	35.672
SBIC	5.596	5.382	15.185	10.937	37.410
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.654	5.878	2.141	2.647	12.320
χ_{JB} (corr)	1.569	5.424	4.531	2.968	14.491
χ_{JB} (cov)	1.133	5.461	3.967	3.910	47.629
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	23.850				
resid_2	37.409	19.007			
resid_3	35.584	13.978	43.476		
resid_4	37.703	30.224	22.838	26.598	
joint					279.318
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	4.466	lag 5	47.193	lag 9	112.341
lag 2*	11.696	lag 6	68.259	lag 10	128.005
lag 3*	16.979	lag 7	80.608	lag 11	145.289
lag 4	27.949	lag 8	93.801	lag 12	157.144
	LM		LM		LM
lag 1	21.061	lag 5	24.175	lag 9	22.506
lag 2	26.190	lag 6	25.187	lag 10	17.920
lag 3	18.864	lag 7	15.814	lag 11	20.620
lag 4	14.383	lag 8	19.111	lag 12	16.258

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.33: Model (incl. dummy) statistics and diagnostics – RV

Model statistics RV					
	eq_1	eq_2	eq_3	eq_4	system
F	6.238	5.696	3.561	3.810	
R^2	0.478	0.456	0.344	0.359	
$\bar{R}^2$	0.402	0.376	0.247	0.265	
AIC	3.616	4.975	14.833	10.615	33.435
SBIC	3.992	5.351	15.209	10.990	35.172
k	16	16	16	16	64
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.851	$6.\ddot{8}18$	$7.\ddot{2}27$	3.889	$19.\ddot{7}85$
χ_{JB} (corr)	0.215	$5.\dot{9}72$	$5.\dot{8}92$	4.097	$16.\ddot{1}76$
χ_{JB} (cov)	0.027	$5.\dot{5}11$	$5.\dot{1}43$	$9.\dddot{4}61$	$96.\dddot{9}32$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	19.491				
resid_2	29.458	14.251			
resid_3	32.682	20.678	$53.\dddot{7}09$		
resid_4	19.882	26.903	20.457	31.761	
joint					286.464
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	3.377	lag 5	42.245	lag 9	108.186
lag 2*	9.252	lag 6	54.398	lag 10	119.617
lag 3*	12.802	lag 7	69.429	lag 11	132.069
lag 4	27.648	lag 8	81.811	lag 12	147.995
	LM		LM		LM
lag 1	23.229	lag 5	19.200	lag 9	$30.\dot{6}47$
lag 2	23.034	lag 6	16.072	lag 10	16.624
lag 3	22.088	lag 7	19.506	lag 11	16.338
lag 4	19.874	lag 8	15.554	lag 12	19.184

Results are for HFR and HFN for the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.34: Model statistics and diagnostics – CTA (HFR)

Model statistics CTA					
	eq_1	eq_2	eq_3	eq_4	system
F	1.462	4.551	2.400	2.586	
R^2	0.155	0.363	0.231	0.244	
$\bar{R}^2$	0.049	0.283	0.135	0.150	
AIC	6.277	5.100	14.958	10.746	36.880
SBIC	6.605	5.428	15.287	11.074	38.312
k	14	14	14	14	56
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	0.053	12.976	6.605	5.178	24.813
χ_{JB} (corr)	0.335	15.530	4.502	9.020	29.387
χ_{JB} (cov)	1.076	13.216	10.872	6.181	152.169
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	59.799				
resid_2	26.094	10.262			
resid_3	63.345	42.982	66.553		
resid_4	43.210	44.918	56.438	32.129	
joint					342.527
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	2.027	lag 5	40.007	lag 9	105.416
lag 2*	5.161	lag 6	62.472	lag 10	121.314
lag 3*	8.619	lag 7	72.234	lag 11	145.351
lag 4	20.045	lag 8	87.470	lag 12	156.722
	LM		LM		LM
lag 1	15.243	lag 5	24.482	lag 9	20.156
lag 2	15.370	lag 6	23.860	lag 10	18.185
lag 3	18.557	lag 7	11.352	lag 11	27.666
lag 4	15.271	lag 8	16.473	lag 12	12.934

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.35: Model statistics and diagnostics – ED (HFR)

Model statistics ED					
	eq_1	eq_2	eq_3	eq_4	system
F	2.489	6.061	1.766	2.487	
R^2	0.253	0.452	0.194	0.253	
$\bar{R}^2$	0.151	0.377	0.084	0.151	
AIC	5.500	4.966	15.022	10.751	34.959
SBIC	5.852	5.318	15.374	11.104	36.602
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	$22.\dddot{6}20$	$6.\ddot{3}58$	$17.\dddot{2}58$	$5.\dot{8}94$	$52.\dddot{1}30$
χ_{JB} (corr)	$8.\ddot{2}03$	$5.\dot{8}50$	$7.\ddot{8}04$	4.180	$26.\ddot{0}37$
χ_{JB} (cov)	4.087	$6.\ddot{4}77$	$16.\dddot{0}79$	$6.\ddot{8}32$	$185.\dddot{9}35$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	$61.\dddot{7}76$				
resid_2	34.906	13.145			
resid_3	$58.\dddot{6}70$	28.128	$79.\dddot{0}43$		
resid_4	$59.\dddot{3}00$	34.969	$58.\dddot{2}38$	$50.\dddot{0}27$	
joint					$343.\dddot{1}23$
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.636	lag 5	43.107	lag 9	113.622
lag 2*	3.499	lag 6	61.175	lag 10	131.540
lag 3*	7.772	lag 7	70.772	lag 11	148.369
lag 4	25.252	lag 8	85.329	lag 12	163.952
	LM		LM		LM
lag 1	20.271	lag 5	$25.\dot{9}45$	lag 9	$33.\dddot{4}84$
lag 2	16.270	lag 6	$23.\dot{6}06$	lag 10	$24.\dot{5}47$
lag 3	$27.\ddot{9}78$	lag 7	17.106	lag 11	$25.\dot{2}24$
lag 4	$24.\dot{4}74$	lag 8	20.047	lag 12	20.820

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.36: Model statistics and diagnostics – EHf (HFR)

Model statistics EHf					
	eq_1	eq_2	eq_3	eq_4	system
F	4.527	3.934	3.165	2.950	
R^2	0.381	0.348	0.301	0.286	
$\bar{R}^2$	0.297	0.260	0.206	0.189	
AIC	4.804	5.139	14.880	10.705	34.857
SBIC	5.157	5.491	15.232	11.058	36.501
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	14.459	16.763	2.847	2.503	36.572
χ_{JB} (corr)	65.396	8.825	1.804	1.785	77.809
χ_{JB} (cov)	202.646	12.396	1.218	2.208	395.063
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	58.758				
resid_2	28.340	17.399			
resid_3	48.785	40.320	70.318		
resid_4	35.235	34.480	36.999	32.889	
joint					411.566
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.128	lag 5	43.298	lag 9	109.305
lag 2*	3.154	lag 6	62.082	lag 10	123.123
lag 3*	8.985	lag 7	82.104	lag 11	137.563
lag 4	26.064	lag 8	89.144	lag 12	155.640
	LM		LM		LM
lag 1	17.198	lag 5	23.585	lag 9	26.131
lag 2	11.718	lag 6	20.641	lag 10	13.817
lag 3	23.246	lag 7	25.988	lag 11	14.497
lag 4	29.160	lag 8	7.789	lag 12	20.690

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.37: Model statistics and diagnostics – EHg (HFR)

Model statistics EHg					
	eq_1	eq_2	eq_3	eq_4	system
F	3.630	3.023	2.439	4.169	
R^2	0.312	0.274	0.234	0.343	
$\bar{R}^2$	0.226	0.184	0.138	0.260	
AIC	4.974	5.230	14.954	10.606	34.788
SBIC	5.303	5.558	15.283	10.935	36.220
k	14	14	14	14	56
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.840	4.751	6.064	13.903	26.558
χ_{JB} (corr)	4.365	3.863	6.244	8.035	22.508
χ_{JB} (cov)	3.031	2.931	13.892	4.400	121.730
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	40.015				
resid_2	41.163	21.485			
resid_3	48.967	56.120	74.535		
resid_4	33.412	43.455	53.714	26.544	
joint					347.156
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	0.633	lag 5	37.426	lag 9	112.514
lag 2*	3.357	lag 6	50.649	lag 10	128.761
lag 3*	8.446	lag 7	60.991	lag 11	155.727
lag 4	17.120	lag 8	87.978	lag 12	170.540
	LM		LM		LM
lag 1	9.786	lag 5	25.745	lag 9	28.159
lag 2	13.535	lag 6	17.383	lag 10	20.875
lag 3	17.757	lag 7	14.155	lag 11	32.057
lag 4	13.916	lag 8	28.376	lag 12	19.438

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.38: Model statistics and diagnostics – EHv (HFR)

Model statistics EHv					
	eq_1	eq_2	eq_3	eq_4	system
F	2.703	5.531	1.843	2.304	
R^2	0.269	0.429	0.200	0.238	
$\bar{R}^2$	0.169	0.352	0.092	0.135	
AIC	5.637	5.006	15.014	10.770	34.836
SBIC	5.989	5.359	15.366	11.122	36.480
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	$10.\dddot{5}17$	$6.\ddot{9}63$	$20.\dddot{5}85$	$11.\dddot{9}10$	$49.\dddot{9}76$
χ_{JB} (corr)	$5.\dot{9}53$	$5.\dot{7}43$	$8.\ddot{9}71$	4.162	$24.\dddot{8}28$
χ_{JB} (cov)	$15.\dddot{6}69$	$6.\ddot{9}66$	$18.\dddot{7}00$	$5.\dot{7}89$	$214.\dddot{1}43$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	$48.\dddot{6}09$				
resid_2	30.729	18.308			
resid_3	$58.\dddot{8}21$	31.505	$77.\dddot{8}23$		
resid_4	$47.\dddot{1}47$	35.436	$58.\dddot{2}99$	$44.\ddot{2}01$	
joint					$334.\ddot{5}78$
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.972	lag 5	42.481	lag 9	113.030
lag 2*	4.121	lag 6	63.381	lag 10	130.764
lag 3*	10.019	lag 7	74.149	lag 11	147.847
lag 4	26.346	lag 8	85.006	lag 12	158.878
	LM		LM		LM
lag 1	22.158	lag 5	$25.\dot{1}04$	lag 9	$34.\dddot{4}41$
lag 2	18.334	lag 6	$28.\ddot{0}79$	lag 10	$23.\dot{9}03$
lag 3	$27.\ddot{0}41$	lag 7	17.438	lag 11	$26.\dot{0}90$
lag 4	$24.\dot{4}75$	lag 8	17.344	lag 12	17.867

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.39: Model statistics and diagnostics – EM (HFR)

Model statistics EM					
	eq_1	eq_2	eq_3	eq_4	system
F	3.252	4.180	1.943	2.290	
R^2	0.289	0.343	0.195	0.223	
$\bar{R}^2$	0.200	0.261	0.095	0.125	
AIC	5.709	5.130	15.003	10.774	35.347
SBIC	6.038	5.458	15.332	11.103	36.780
k	14	14	14	14	56
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	57.792	4.718	28.485	0.085	91.079
χ_{JB} (corr)	16.073	5.109	9.920	1.713	32.815
χ_{JB} (cov)	16.388	4.799	8.643	5.060	212.291
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	83.326				
resid_2	68.562	12.115			
resid_3	80.536	55.088	73.726		
resid_4	72.291	53.460	71.319	46.895	
joint					387.922
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	0.916	lag 5	37.089	lag 9	99.026
lag 2*	2.617	lag 6	54.370	lag 10	116.034
lag 3*	6.473	lag 7	65.559	lag 11	138.555
lag 4	19.145	lag 8	77.451	lag 12	156.412
	LM		LM		LM
lag 1	11.208	lag 5	19.778	lag 9	20.812
lag 2	12.794	lag 6	18.424	lag 10	16.975
lag 3	18.273	lag 7	11.583	lag 11	23.980
lag 4	14.383	lag 8	12.269	lag 12	19.253

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.40: Model statistics and diagnostics – L (HFR)

Model statistics L					
	eq_1	eq_2	eq_3	eq_4	system
F	3.569	3.882	3.129	2.440	
R^2	0.327	0.345	0.298	0.249	
$\bar{R}^2$	0.235	0.256	0.203	0.147	
AIC	5.501	5.143	14.883	10.756	34.775
SBIC	5.853	5.496	15.235	11.108	36.419
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	3.182	4.699	7.519	45.170	60.571
χ_{JB} (corr)	28.286	5.078	2.242	5.701	41.307
χ_{JB} (cov)	214.747	3.988	4.526	4.415	649.527
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	31.923				
resid_2	37.659	21.756			
resid_3	29.755	31.880	57.857		
resid_4	25.928	39.681	29.421	29.333	
joint					342.813
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	3.162	lag 5	48.857	lag 9	122.039
lag 2*	6.987	lag 6	66.666	lag 10	149.797
lag 3*	14.801	lag 7	80.040	lag 11	161.230
lag 4	35.079	lag 8	96.043	lag 12	172.966
	LM		LM		LM
lag 1	27.265	lag 5	20.238	lag 9	27.338
lag 2	25.279	lag 6	24.015	lag 10	29.234
lag 3	28.279	lag 7	16.910	lag 11	16.036
lag 4	24.699	lag 8	17.592	lag 12	16.482

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.41: Model statistics and diagnostics – LS (HFR)

Model statistics LS					
	eq_1	eq_2	eq_3	eq_4	system
F	3.861	4.173	1.879	4.965	
R^2	0.344	0.362	0.203	0.403	
$\bar{R}^2$	0.255	0.275	0.095	0.322	
AIC	4.913	5.118	15.010	10.527	34.569
SBIC	5.265	5.470	15.362	10.879	36.213
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	0.380	4.408	12.243	11.707	28.738
χ_{JB} (corr)	1.766	3.362	8.371	10.930	24.430
χ_{JB} (cov)	0.694	2.847	20.934	4.119	177.321
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	36.133				
resid_2	33.977	22.857			
resid_3	39.372	43.668	71.992		
resid_4	33.751	46.295	48.466	23.918	
joint					316.647
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	1.028	lag 5	43.456	lag 9	111.921
lag 2*	3.268	lag 6	58.669	lag 10	123.399
lag 3*	6.001	lag 7	73.534	lag 11	145.052
lag 4	18.581	lag 8	86.134	lag 12	161.951
	LM		LM		LM
lag 1	12.110	lag 5	29.697	lag 9	26.970
lag 2	18.025	lag 6	16.556	lag 10	11.421
lag 3	9.567	lag 7	17.051	lag 11	22.341
lag 4	14.772	lag 8	13.744	lag 12	17.999

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.42: Model statistics and diagnostics – M (HFR)

Model statistics M					
	eq_1	eq_2	eq_3	eq_4	system
F	3.063	3.836	2.082	2.662	
R^2	0.294	0.343	0.221	0.266	
$\bar{R}^2$	0.198	0.253	0.115	0.166	
AIC	5.195	5.147	14.988	10.734	36.126
SBIC	5.547	5.500	15.340	11.086	37.769
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	1.382	$8.\ddot{3}00$	$6.\ddot{6}78$	$5.\dot{9}10$	$22.\dddot{2}70$
χ_{JB} (corr)	1.330	$8.\ddot{8}12$	$6.\ddot{1}48$	$7.\ddot{3}87$	$23.\dddot{6}77$
χ_{JB} (cov)	1.083	$7.\ddot{2}17$	$11.\dddot{7}52$	$6.\ddot{2}37$	$121.\dddot{6}05$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	22.649				
resid_2	$43.\ddot{4}24$	21.793			
resid_3	31.718	$41.\ddot{8}36$	$64.\dddot{2}74$		
resid_4	$44.\ddot{2}84$	$50.\dddot{1}18$	$46.\dddot{7}91$	33.993	
joint					$326.\ddot{0}46$
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	2.349	lag 5	42.412	lag 9	102.238
lag 2*	5.285	lag 6	58.949	lag 10	120.400
lag 3*	9.254	lag 7	68.852	lag 11	142.595
lag 4	22.760	lag 8	84.882	lag 12	154.561
	LM		LM		LM
lag 1	14.654	lag 5	$26.\ddot{5}86$	lag 9	21.329
lag 2	22.247	lag 6	21.813	lag 10	21.698
lag 3	19.800	lag 7	13.303	lag 11	$27.\ddot{0}62$
lag 4	18.430	lag 8	21.386	lag 12	18.073

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.43: Model statistics and diagnostics – RV (HFR)

Model statistics RV					
	eq_1	eq_2	eq_3	eq_4	system
F	2.519	5.432	2.046	2.593	
R^2	0.255	0.425	0.218	0.261	
$\bar{R}^2$	0.154	0.347	0.111	0.160	
AIC	3.925	5.014	14.992	10.741	33.928
SBIC	4.277	5.366	15.344	11.093	35.571
k	15	15	15	15	60
Diagnostics					
Normality	comp_1	comp_2	comp_3	comp_4	joint
χ_{JB}(chol)	$33\dddot{.}239$	3.955	$26\dddot{.}586$	0.530	$64\dddot{.}310$
χ_{JB} (corr)	$5\dot{.}516$	$5\dot{.}251$	$8\ddot{.}790$	0.024	$19\ddot{.}581$
χ_{JB} (cov)	0.129	4.368	$18\dddot{.}119$	3.013	$150\dddot{.}039$
Heterosked.	resid_1	resid_2	resid_3	resid_4	joint
resid_1	$38\dot{.}094$				
resid_2	$46\dddot{.}462$	11.692			
resid_3	$37\dot{.}110$	33.163	$70\dddot{.}963$		
resid_4	33.915	$37\dot{.}577$	$43\ddot{.}175$	35.795	
joint					$331\ddot{.}297$
Additional multivariate diagnostics					
Autocorrel.	χ_{LB}		χ_{LB}		χ_{LB}
lag 1*	4.066	lag 5	46.168	lag 9	112.135
lag 2*	7.268	lag 6	54.932	lag 10	128.422
lag 3*	12.131	lag 7	71.836	lag 11	147.363
lag 4	$31\dot{.}135$	lag 8	86.577	lag 12	165.418
	LM		LM		LM
lag 1	$32\dddot{.}456$	lag 5	$23\dot{.}684$	lag 9	$32\dddot{.}134$
lag 2	19.997	lag 6	15.174	lag 10	$24\dot{.}456$
lag 3	$23\dot{.}783$	lag 7	23.177	lag 11	$27\ddot{.}911$
lag 4	$28\ddot{.}401$	lag 8	22.153	lag 12	23.423

Results are for the HFR database and the July 2000 to June 2010 timeframe. The orthogonalization method used in the extensions of the Jarque-Bera statistics are the Cholesky of covariance (χ_{JB}(chol)), the square root of correlation (χ_{JB} (corr)) and the square root of covariance (χ_{JB} (cov)). The heteroskedasticity statistic is the White statistic without cross-terms. Serial correlation in the residual series is tested for up to the twelfth lag using multivariate extensions of Ljung-Box portmanteau test (χ_{LB}) and Lagrange multiplier (LM). Designation of statistical significance according to Table A.5.

Table A.44: Granger causality and block significance

Results are for HFR and HFN for the July 2000 to June 2010 timeframe.

	D(CTA)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(EM)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(CTA)		9.807	6.125	1.728	D(EM)		5.303	2.337	0.930
D(BABDIDX)	4.550		2.087	6.547	D(BABDIDX)	10.014		1.567	6.688
D(GSCI)	4.643	16.318		2.708	D(GSCI)	9.343	13.054		3.309
D(MSCIW)	0.754	7.887	5.352		D(MSCIW)	9.737	4.827	3.741	
all	10.869	27.018	13.952	11.985	all	19.713	22.051	11.529	10.388
	D(ED)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(L)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(ED)		26.714	1.055	1.623	D(L)		5.512	11.730	3.946
D(BABDIDX)	7.434		1.030	9.547	D(BABDIDX)	6.433		2.236	4.876
D(GSCI)	3.471	18.212		4.519	D(GSCI)	6.814	11.367		3.979
D(MSCIW)	3.740	12.766	1.688		D(MSCIW)	9.095	1.780	12.141	
all	13.062	45.836	8.503	13.997	all	22.489	21.982	20.248	11.410
	D(EHF)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(LS)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(EHF)		8.147	7.083	5.966	D(LS)		8.201	3.073	4.191
D(BABDIDX)	5.080		1.929	5.128	D(BABDIDX)	1.648		3.205	3.220
D(GSCI)	10.831	12.723		8.357	D(GSCI)	4.089	12.826		10.000
D(MSCIW)	12.415	2.283	5.138		D(MSCIW)	10.528	6.478	6.363	
all	26.660	22.962	13.627	16.107	all	17.534	23.723	9.848	19.632
	D(EHG)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(M)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(EHG)		8.656	6.768	3.905	D(M)		5.168	5.295	6.567
D(BABDIDX)	1.623		1.564	4.962	D(BABDIDX)	2.438		2.890	8.297
D(GSCI)	6.904	9.783		10.039	D(GSCI)	6.988	10.297		2.180
D(MSCIW)	5.741	11.098	4.610		D(MSCIW)	1.985	5.823	6.938	
all	14.748	21.925	13.073	21.258	all	12.468	19.837	14.734	16.462
	D(EHV)	D(BABDIDX)	D(GSCI)	D(MSCIW)		D(RV)	D(BABDIDX)	D(GSCI)	D(MSCIW)
D(EHV)		21.281	0.636	2.086	D(RV)		21.103	4.599	0.313
D(BABDIDX)	8.432		1.173	7.836	D(BABDIDX)	7.019		1.676	8.374
D(GSCI)	3.428	18.006		4.982	D(GSCI)	2.701	15.933		3.807
D(MSCIW)	5.730	12.080	2.866		D(MSCIW)	9.132	5.994	6.327	
all	15.704	38.668	9.064	12.198	all	15.799	39.781	12.604	13.116

Results are for HFR and HFN for the July 2000 to June 2010 timeframe.

Appendix 2: Figures

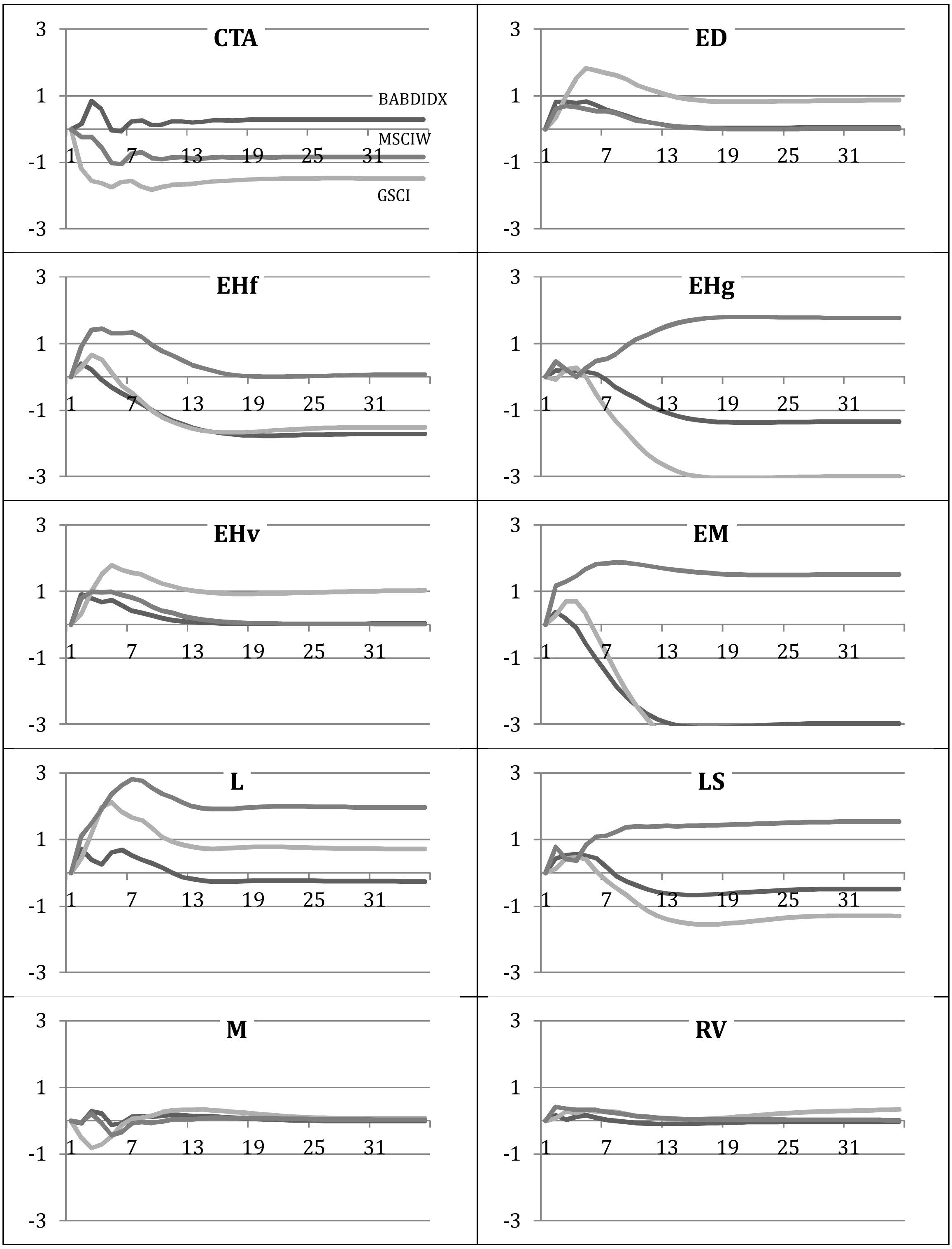

Figure A.1: Impulse response test (HFR)

Cholesky ordering: Hedge fund index, BABDIDX (dark grey), GSCI (light grey) and MSCIW (grey).

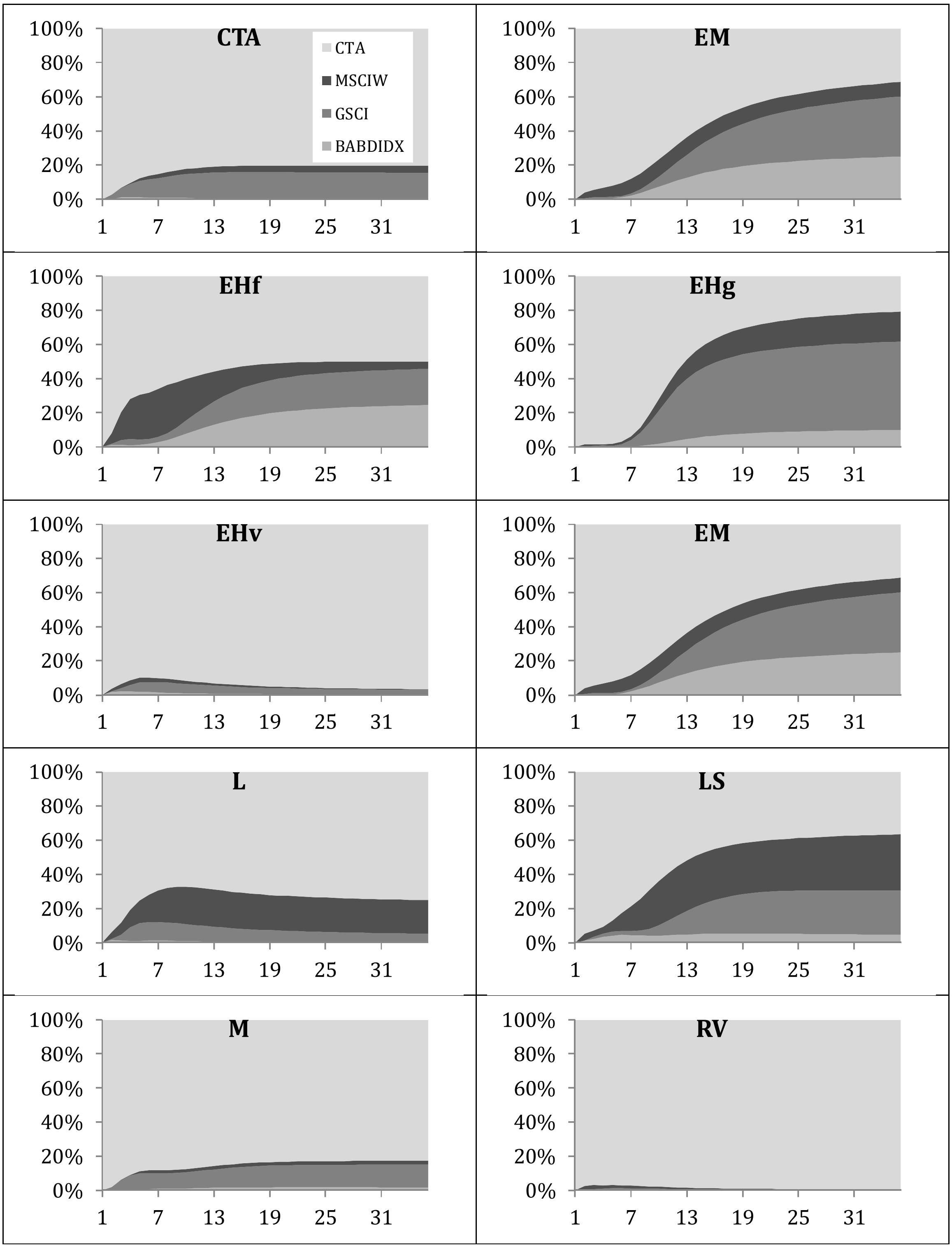

Figure A.2: Variance decomposition (HFR)

Cholesky ordering: Hedge fund index, BABDIDX, GSCI and MSCIW.

Appendix 3: QR method and estimation of eigenvalues

The basic idea behind the QR algorithm is that any non-singular matrix $\mathbf{T}$ can be expressed through $\mathbf{T} = \mathbf{QR}$ where $\mathbf{Q}$ is orthonormal and $\mathbf{R}$ is upper triangular.[20] This factorization is always possible. The method involves the following steps:

1. Factorize the given matrix $\mathbf{T} = \mathbf{QR}$
2. Multiply the two factors $\mathbf{Q}$ and $\mathbf{R}$ to obtain a new matrix $\mathbf{T_1} = \mathbf{RQ}$
3. Factorize the new matrix $\mathbf{T_1} = \mathbf{QR}$ and continue with step 1

The iterative process

$$\mathbf{T} = \mathbf{QR} \rightarrow \mathbf{T_1} = \mathbf{RQ} \qquad \text{...(A.1)}$$

$$\mathbf{T_1} = \mathbf{Q_1R_1} \rightarrow \mathbf{T_2} = \mathbf{R_1Q_1}$$

$$\vdots$$

$$\mathbf{T_k} = \mathbf{Q_kR_k} \rightarrow \mathbf{T_{k+1}} = \mathbf{R_kQ_k}$$

If the eigenvalues λ have distinct absolute values:

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| \qquad \text{...(A.2)}$$

[20] Note that $\mathbf{R}$ is not to be confused with the return matrix as defined in equation 4.14.. The notation was adopted from the name for the QR method.

and $\mathbf{T}$ is symmetric, then the matrix $\mathbf{T_p}$ converges to the diagonal form, where the elements are the eigenvalues of $\mathbf{T}$. For a detailed description of the methodology refer to Francis (1961, 1962).

Once the Eigenvalues are found, it is possible to determine the Eigenvectors by solving the following homogeneous system:

$$0 = (\mathbf{T} - s_i\mathbf{I})\mathbf{c_i} \qquad \text{...(A.3)}$$

where $\mathbf{T}$ is the original matrix, λ_i the i^{th} eigenvalue, $\mathbf{I}$ an identity matrix and $\mathbf{c_i}$ are the associated eigenvectors. For each eigenvalues with multiplicity one, there is only one eigenvector. Eigenvectors can be found by accumulating the transformations in the QR algorithm.