

EDHEC RISK AND ASSET MANAGEMENT RESEARCH CENTRE

Edhec -1090 route des crêtes - 06560 Valbonne - Tel. +33 (0)4 92 96 89 50 - Fax. +33 (0)4 92 96 93 22

Email: research@edhec-risk.com – Web: www.edhec-risk.com

Hedge Fund Indices from an Academic Perspective: Reconciling Investability and Representativity

Lionel Martellini

Professor of Finance at the Edhec Graduate School of Business
Scientific Director of the Edhec Risk and Asset Management Research Centre

Mathieu Vaissié

Research Engineer with the Edhec Risk and Asset Management Research Centre

Felix Goltz

Research Engineer with the Edhec Risk and Asset Management Research Centre

November 2004



Abstract

Following a growing concern among investors about the quality of hedge fund index return data, and given the lack of capacity and transparency specific to that industry, this paper questions from an academic perspective whether it is feasible or not to design hedge fund benchmarks satisfying all defining properties for a good index. In particular, in an attempt to test whether achieving investability necessarily comes at the cost of representativity, as sometimes claimed by hedge fund index providers, we borrow from the asset pricing literature the concept of *factor replicating portfolios* and apply it to the benchmarking of hedge fund style returns. Our results suggest that it is actually possible to construct representative indices based on a limited number of funds, except perhaps in the case of Equity Market Neutral strategies, provided that i) these funds are suitably selected and ii) an optimal portfolio is designed with the objective of replicating the common trend in hedge fund returns for a given strategy. A wide range of tests are performed that show that high correlation of the factor replicating portfolios with the common factor of returns for each strategy is remarkably robustness with respect to modifying the number of funds in the replicating portfolio, choosing less correlated funds, changing the frequency of rebalancing, using a static selection of funds or imposing constraints on weights.

Edhec is one of the top five business schools in France owing to the high quality of its academic staff (100 permanent lecturers from France and abroad) and its privileged relationship with professionals that the school has been developing since its establishment in 1906. Edhec Business School has decided to draw on its extensive knowledge of the professional environment and has therefore concentrated its research on themes that satisfy the needs of professionals.

Edhec pursues an active research policy in the field of finance. Its “**Risk and Asset Management Research Centre**” carries out numerous research programs in the areas of asset allocation and risk management in both the traditional and alternative investment universes.

One of the by-products of the bull market of the 90s has been the consolidation of hedge funds as an important segment of financial markets. It was recently announced that the value of the hedge fund industry worldwide had passed the \$1 trillion mark for the first time, with approximately 7,000 hedge funds in the world, around 1,000 of which were launched in 2003.¹

There are however a number of obstacles to the industrialization of the alternative investment industry. Its adoption by institutional investors will only come about if a serious effort is made in terms of transparency and rationalization of the investment management process and, above all, performance evaluation. Due to the scarcity of information, the logic of representativeness through market capitalization is difficult to apply to the alternative universe. As a result, finding a benchmark that is representative of a particular management universe is not a trivial problem. The different indices available on the market are constructed from different data, according to diverse selection criteria and methods of construction, and they evolve at differing paces (see Amenc and Martellini (2003) and Vaissié (2004)). Because of this heterogeneity, investors cannot rely on competing hedge fund indices to obtain a “true and fair” view of hedge fund performance. Investors are therefore at a loss when selecting benchmarks.

One of the key reasons explaining such lack of homogeneity in hedge fund index return data is the fact that none of these existing indices is fully representative. In other words, this is a sample size problem: a number of funds that should be part of an index are not included in the index. Because of the lack of regulation on hedge fund performance disclosure, existing databases only cover a relatively small fraction of the hedge fund population. Probably only a little more than half of existing hedge funds choose to self-report their performance to one of the major hedge fund databases. As simple evidence of the fact that existing indices are not fully representative of the universe, it perhaps suffices to note that one of the most popular hedge fund indices, the EACM 100, does not account for more than a tiny percentage of all existing hedge funds (100 among more than 7,000 funds).

The concern over existing hedge fund indices not being representative of the universe has been intensified by the recent launch of several investable hedge fund indices (see table 1). Indices provided by S&P, HFR, CSFB/Tremont or MSCI are among the best-known examples. The principal objective of these investable “indices” is to allow a broad range of investors, access to alternative investment strategies at low cost. The objective of these “indices” therefore differs from that of maximizing the representativity dimension by covering the largest possible number of funds. Instead, the priority is to choose a limited number of funds that are open to new investors and that guarantee a minimum investment capacity. In other words, these indices are not intended to be used as a reference for the hedge fund market but to provide liquidity where it is crucially lacking. Obviously, the presence of these constraints suggests that investable hedge fund indices will be even less representative of the universe than non-investable indices.

It has often been argued (e.g., Kündi, Lodeiro, Meier and Ruckstuhl (2004)) that two distinct purposes of indices should be distinguished: an index can be used as a benchmark for investments in specific styles, instruments or locations; or it can be used as an investment vehicle. On the one hand, indices that act as benchmarks have to be unambiguous, verifiable, accountable and representative. On the other hand, an investable index should enjoy the same properties, and, in addition, be investable. It is important to note that these requirements should be achieved at the same time: if an investable index fails at enjoying the defining properties for an index (unambiguous, verifiable, accountable and representative), it should not be called an index but rather a fund of hedge funds.

¹ These numbers have been extracted from the 2004 *Alternative Fund Service Review Survey*, as reported in the weekly publication *International Fund Investment*, issue 116, May 17, 2004.

Table 1: Overview of Major Investable Hedge Fund Indices

Index Provider	Launch date	Strategy / Fund Weighting	Nbr of funds in the index	Rebalancing Frequency	Pricing Frequency
CSFB/Tremont	Aug-03	V.W. / V.W.	60	Semi Annually	Monthly
Dow Jones	Nov.-03	n.a. / E.W.	35	Quarterly *	Daily
FTSE	Apr.-04	I.W. / I.W.	40	Annual * *	Daily
HFRX	Mar-03	V.W / *	n.a.**	Quarterly	Daily
		Adj. Median Asset Weighted /			
MSCI	Jul.-03	E.W.	110	Quarterly	Daily
S&P	May-02	E.W. / E.W.	40	Annual * * *	Daily

* Fund weightings are optimized to maximize correlation with their group.

**Optimal number of funds for strategy replication is determined through Monte Carlo simulation

* Additions or deletions can occur without notice at the complete and absolute discretion of Dow Jones.

** Funds may be added/deleted more frequently in response to changing market conditions or fund specific events.

*** Annual at the strategy level and periodically at the fund level.

The success of investable hedge fund strategy indices, and their differentiated positioning with funds of hedge funds, will therefore strongly depend on the capacity of index providers to improve the investability of their indices without sacrificing the representativity dimension. This is not a trivial task because to be fully representative an index has to cover the whole universe or a whole strategy, including closed funds. Examining how modern portfolio theory and factor analysis techniques can be used to build investable, yet representative, hedge fund indices is the focus of this paper.

The issues regarding the benchmarking of hedge fund returns have recently received a significant amount of attention in the literature. Understanding the nature of risks associated with different hedge fund strategies are actually challenging because of the complex nature of the strategies and limited disclosure requirements faced by hedge funds. In particular, since hedge fund returns exhibit non-linear option-like exposures to traditional asset classes (Fung and Hsieh (1997, 2000)), standard asset pricing models offer limited help in evaluating the performance of hedge funds. The importance of taking into account such option-like features has been underlined by recent research. In particular, Fung and Hsieh (2001) and Mitchell and Pulvino (2001) stress the importance of taking into account option-like features while analyzing the performance of “trend-following” and “risk-arbitrage” strategies, respectively. More recently, Agarwal and Naik (2004) build on these insights and extend our understanding of hedge fund risks to a wide range of equity-oriented hedge fund strategies. They characterize the risk exposures of hedge funds using buy-and-hold and option-based strategies, and show that a large number of equity-oriented hedge fund strategies exhibit payoffs resembling a short position in a put option on the market index.

There are actually two possible ways to try and adapt standard asset pricing models to analyze returns on portfolios that exhibit a non-linear dependency to standard asset classes. The first approach consists in using a nonlinear APT model (see in particular Bansal and Viswanathan (1993) or Bansal, Hsieh and Viswanathan (1993)). The other method, which has been used by Glosten and Jagannathan (1994), as well as in the aforementioned papers on hedge fund benchmarking, is to include new regressors with non-linear exposure to standard asset classes, e.g., returns on option positions, to proxy for dynamic trading strategies in a linear regression. Apart from portfolios of options, there actually exists another set of natural candidates for portfolios exhibiting non-linear exposure to traditional asset classes. Such natural candidates are hedge fund indices, which are already very commonly used by funds of hedge fund managers to benchmark individual hedge fund returns (see for example Lhabitant (2001) for an example of style and VaR analysis based on hedge fund indices). In a related paper, Amenc and Martellini (2003) have suggested to construct “indices of indices” so as to design a benchmark with degrees of representativity and stability that are

significantly higher than those of the indices available on the market.² As a result of the aggregation process, the resulting portfolio of indices, however, was based on a large number of funds, including a number of closed funds, and therefore was not investable.

Our contribution to this literature is to investigate from an academic perspective whether designing sound (i.e., both representative and investable) hedge fund indices is a feasible task given the specific features of that industry, in particular the lack of capacity and transparency. To test whether or not investability should necessarily come at the cost of representativity, we use hereafter a well-known methodology in asset pricing literature based on the concept of *factor replicating portfolios* (see for example Huberman, Kandel and Stambaugh (1987)). More specifically, our results suggest that it is actually possible to construct representative indices based on a limited number of funds, provided that i) these funds are suitably selected and ii) an optimally designed portfolio is designed with the objective of replicating the common trend in hedge fund returns for a given strategy.

The rest of the paper is organized as follows. In section 1, we introduce the methodology used in the design of representative indices based on a limited number of funds. In section 2, we present the results of the base case experiment. In section 3, we conduct a number of robustness tests. In section 4, we present our conclusions and relegate to a dedicated appendix the description of the strategies covered in this paper.

1. Construction Methodology

In what follows, we use an approach similar to the one developed by Fung and Hsieh (1997) to form for each hedge fund strategy an investable portfolio replicating the performance of the common component in the funds' returns.³ This approach is based on the use of factor analysis techniques to extract the best possible one-dimensional summary of a set of competing individual funds following a given strategy, and design what replicating portfolios for the factor explaining the largest amount of information. The resulting indices can be thought of as the best possible one-dimensional summaries of information conveyed by a variety of hedge fund following a given style, in the sense of the largest fraction of the variance explained.

1.1. Conceptual Background

More specifically, our methodology is based on the concept of *factor replicating portfolios (FRP)*. In finance, the use of statistical techniques to model asset returns have been extensive, especially in the context of factor models. Going back to Feeney and Hester (1964) and Lessard (1973), or in more recent years, Black and Litterman (1992), or Chan, Karceski and Lakonishok (1998), principal component and factor analysis have been used to examine the existence of common movements in asset returns. They are seen as alternatives to fundamental approaches which relate the factors influencing financial asset returns to macroeconomic measures such as inflation, interest rate rates and market indices, or to company specifics such as size, book to market ratio or dividend yield.

A great deal of statistical factor-type of analysis has been performed for testing the Arbitrage Pricing Theory (Ross (1976)). In this context, historical returns are used to estimate orthogonal statistical factors and their relationship with the original variables. The construction of replicating portfolios for the statistical factors has been formalized by Huberman, Kandel and Stambaugh (1987).

The standard approach for constructing factor replicating portfolios uses the factor loadings in the stock selection process (Fama and French (1992)). The stocks are ranked according to their loadings on a particular factor, then a self-financed portfolio is set up with long positions on the stocks with the highest loadings on that factor, and short positions on the stocks with the smallest loadings. An alternative for constructing factor mimicking portfolios proposed by Fung and Hsieh (1997) in the context of hedge fund performance analysis considers, in the selection stage, only the assets (hedge funds) that are highly correlated to the principal component (implicit factor) for which

² This methodology has led to the design of EDHEC index of indices series (available at www.edhec-risk.com). Since their inception, an increasing number of institutional investors and funds of hedge fund managers have used these indices as reference benchmarks.

³ One difference with Fung and Hsieh (1997) is that they perform a factor analysis of a sample of hedge fund returns covering a mix of strategies, while we perform a separate factor analysis on each strategy sample.

the replica is constructed. Having selected the funds, their portfolio weights are optimized so as to deliver the maximal correlation of the replicating portfolio returns with the corresponding principal component (see also Alexander and Dimitriu (2004) for a use of the same methodology in a different context). This second method is more efficient in that the portfolio construction is based on optimization, as opposed to an arbitrary method such as equal weighting.

At the intuitive level, the aim of the methodology is to use a small sample of funds to design a replicating portfolio for the return on a given strategy. To this end, the natural selection criterion is the loading of individual funds on the *first principal component* (see below for a formal definition). The higher the loading of a fund on the first principal component, the higher will be its contribution to the common trend in hedge fund returns following a given strategy. Given that the first *eigenvector* corresponding to the first principal component is determined as to maximize the variance of the corresponding linear combination of fund returns, high factor loadings will be allocated to funds which have been highly correlated with their group over the calibration period. Such funds should be the most representative of their group. This allows us to achieve a satisfactory level of representativity, while ensuring for the constraint of investability based on a small number of funds.

1.2. Construction Methodology

While the hedge fund universe comprises a wide range of strategies, we have chosen in this paper to focus on the segments of the industry that are most important in terms of assets under management and in terms of accessibility by investors⁴. These strategies are: CTA Global Macro, Long/Short equity, Equity Market Neutral, Convertible Arbitrage, Event Driven (see the Appendix for a description of these strategies).

We now formalize these intuitions and provide more details on the procedure that we follow.

1.2.1. Factor Analysis of Hedge Fund Returns

Starting with a database of hedge fund returns, we are extracting the combination of individual funds that capture the largest possible fraction of the information contained in the data.⁵ Technically speaking, this amounts to using the first component of a Principal Component Analysis (PCA) of funds returns as a candidate for a pure style index. Note that the first component typically captures a large proportion of cross-sectional variations because various hedge funds following a given strategy tend to be at least somewhat positively correlated.

The PCA of a time-series involves studying the correlation matrix of successive shocks. Its purpose is to explain the behavior of observed variables using a smaller set of unobserved implied variables. From a mathematical standpoint, it involves transforming a set of K correlated variables into a set of orthogonal variables, or implicit factors, which reproduces the original information present in the correlation structure. Each implicit factor is defined as a linear combination of original variables. Define R as the following matrix:

$$R = (R_{tk})_{\substack{1 \leq t \leq T \\ 1 \leq k \leq n}}$$

R actually contains standardized, as opposed to raw, returns. It is better to conduct PCA on standardized returns (so that they all have mean zero and variance one) because this removes differences in variances caused by leverage differences (see Fung and Hsieh (1997)). For example, two funds employing the exact same trading strategy but different leverage will have different return variances.

We have n variables, i.e. returns for n different individual funds, and T observations of these variables, where T is the number of months in our case.

⁴ According to CSFB Tremont, these strategies made up 91 percent of assets under management in the hedge fund industry at the end of 2003. Likewise, in the CISDM database, funds in these strategies constitute 85 percent of total assets under management by single hedge funds.

⁵ In this paper, we use the CISDM database (see section 2.1).

$$R_{tk} = \sum_{i=1}^n \sqrt{\lambda_i} U_{ik} V_{ti} \quad (1)$$

where

$(U) = (U_{ik})_{1 \leq i, k \leq n}$ is the matrix of the n eigenvectors⁶ of $R'R$.

$(U^T) = (U_{ki})_{1 \leq k, i \leq n}$ is U transposed.

$(V) = (V_{ti})_{\substack{1 \leq t \leq T \\ 1 \leq i \leq n}}$ is the matrix of the n eigenvectors RR'

Note that these n eigenvectors are orthogonal. λ_i is the eigenvalue (ordered by degree of magnitude) corresponding to the eigenvector U_i . Denoting $S_{ik} = \sqrt{\lambda_i} U_{ik}$ the principal component sensitivity of the k^{th} variable to the i^{th} factor, and $V_{ti} = F_{ti}$, one can equivalently write equation (1)

$$R_{tk} = \sum_{i=1}^n S_{ik} F_{ti}$$

where the n factors F_{ti} are a set of orthogonal variables. One may use the method to describe each variable as a linear function of a reduced number of factors. To that end, one needs to select a number of factors I such that the first I factors capture a large fraction of asset return variance, while the remaining part can be regarded as statistical noise

$$R_{tk} = \sum_{i=1}^I \sqrt{\lambda_i} U_{ik} V_{ti} + \varepsilon_{tk} = \sum_{i=1}^I S_{ik} F_{ti} + \varepsilon_{tk} \quad (2)$$

where some structure is imposed by assuming that the residuals ε_{tk} are uncorrelated one to another. The percentage

of variance explained by the first I factors is given by $\frac{\sum_{i=1}^I \lambda_i}{\sum_{i=1}^N \lambda_i}$.

By taking $I=1$ in equation (2) this method can be used to generate "the best one-dimensional" summary of a set of individual funds.

1.2.2. Building Factor Replicating Portfolios

Following Fung and Hsieh (1997), we suggest using the following two-stage methodology for building factor-replicating portfolios (FRPs).

- Selection stage: For each strategy, we form a portfolio using 10 hedge funds in the corresponding category that are most correlated to the first principal component in the first 3-year calibration period (here, January 1998 to December 2000).
- Optimization stage: The portfolio weights are chosen so that the portfolio returns have maximal correlation with the corresponding principal component.

In the base case experiment, this two-stage procedure is repeated every year, and the performance of factor replicating portfolios is examined during the out-of-sample period ranging from January 2001 to December 2003.

⁶ An eigen vector is a vector that is scaled by a linear transformation, but not moved. The scaling factor is the eigen value.

2. Base Case Results

We first describe the results of a base case experiment and report the results of a series of robustness tests to the next section.

2.1. Factor Analysis

For each hedge fund strategy, we perform a principal component analysis (PCA) on all funds in the CISDM data base that have 3 years of returns history before the chosen starting date of January 2001. The CISDM Database provides information on 1,800 active hedge funds and 600 active CTAs and managed futures.⁸

The average correlation for all funds following a given strategy is given in table 2.

Table 2: Average correlation between all funds in the database –data from 01/1998 through 12/2003

Strategy	Average Correlation
Convertible	
Arbitrage	0.30
CTA Global	0.20
Event Driven	0.36
Equity Long / Short	0.28
Equity Market	
Neutral	0.11

As can be seen from table 2, funds following a given strategy tend to be positively correlated on average. The lowest average correlation is recorded in the Equity Market Neutral category, suggesting that these funds' performance is not significantly affected by common exposure to systematic factors.

For each hedge fund strategy, we now perform 3 successive PCAs, one on the calibration period ranging from January 1998 to December 2000, one on the calibration period ranging from January 1999 to December 2001, and one on the calibration period ranging from January 2000 to December 2002. The percentage of explanation of the cross-sectional return variance by the first factor for each PCA is given in table 3.

Table 3: Percentage of variance explained by 1st Principal Component

	01/98- 12/00	01/99- 12/00	01/00- 12/02	Average
Convertible Arbitrage	39%	31%	35%	35%
CTA Global	30%	32%	36%	33%
Event Driven	46%	32%	35%	38%
Equity Long / Short	40%	38%	35%	38%
Equity Market Neutral	23%	24%	24%	24%

This analysis confirms that Equity Market Neutral hedge funds follow purer alpha strategy compared to other types of hedge funds, so that a smallest percentage of their performance can be explained by a common factor.

⁸ For more information about the CISDM database, refer to the following website: <http://cisdm.som.umass.edu/resources/database.shtml>.

2.2. Building Factor Replicating Portfolios

The next step is to form factor replicating portfolios (FRPs) that represent portfolios of hedge funds that closely replicate each factor. This is done as follows.

2.2.1. Selection Stage

For each strategy, we form a portfolio using the 10 hedge funds in the corresponding category that are most correlated to the first principal component in the calibration period (e.g., January 1998 to December 2000). Note that we focus on a very small number of funds to emphasize that investability does not necessarily come at the cost on representativity if the investable index is carefully designed. Even better results would be obtained by including more funds in the portfolio (e.g., 15). The following table (table 4) compares the average, minimum and maximum correlation with the corresponding factor for selected funds with the population of funds in the database. In table 4, as well as in all other tables featuring in-sample results, the numbers presented are the arithmetic average of the 3 quantities of interest over each 3-year calibration period corresponding to each of the 3 successive PCAs.

Table 4: In sample correlations of funds with the 1st Principal Component: Averages over three sub-periods (01/1998-12/2000, 01/1999-12/2001, and 01/2000-12/2002)

	CTA Global		Conv. Arb.		Event Driven		Market Neutral		Long/Short	
	All funds	Selected funds	All funds	Selected funds	All funds	Selected funds	All funds	Selected funds	All funds	Selected funds
Minimum	-0.45	0.88	-0.18	0.73	-0.11	0.80	-0.48	0.67	-0.71	0.84
Average	0.46	0.90	0.54	0.78	0.56	0.85	0.30	0.78	0.48	0.87
Median	0.51	0.89	0.58	0.77	0.63	0.84	0.31	0.79	0.61	0.87
Maximum	0.92	0.92	0.85	0.85	0.91	0.91	0.88	0.88	0.92	0.92

This table illustrates the importance of the selection stage. By randomly selecting funds, one may end up forming portfolios that are very weakly correlated to the common factor in hedge fund returns. The average correlation of selected funds is always significantly higher than that of all funds in the database.

2.2.2. Optimization Stage

The portfolio weights are chosen so that the portfolio returns have maximal correlation with the corresponding principal component. Short sales constraints are imposed since it is not possible to sell short hedge funds. Every year, we repeat the analysis, i.e., we perform a PCA on the past 3 years of normalized returns, and we select the 10 funds with the highest correlation.

As a result of the optimization process, we have the weights on the selected funds in the factor replicating portfolio. Two issues of concern need to be addressed. First, as we repeat the optimization exercise every year, the composition changes and we have to terminate some funds, invest in new funds and more generally adjust the weight allocated to each fund. Secondly, portfolios that maximize correlation might make use of extreme weights in a few highly correlated funds, which may yield concerns over taking on unnecessary fund specific risk.

Panel A of table 5 reports different measures of turnover of funds in the optimal portfolios. The results show that on average, turnover from one year to another is low for all five strategies. The maximum turnover over the period studies is also reasonable for all strategies.

To address the concern over the possibility of highly concentrated portfolios, we calculate the Herfindahl Index for the FRPs. The Herfindahl Index (HI) is a measure of concentration. It is given by the sum of the squares of the weights of all funds included in the FRP. Formally, let w_j denote the portfolio weight of the j -th fund. We then have

$$HI = \sum_{j=1}^{10} w_j^2$$

The functional form of the Herfindahl Index penalizes large individual weights. For example, a portfolio where 2 funds have a 25% allocation each and the rest is equal weighted will have a lower HI ($HI=0.16$) than a portfolio where one fund has 40%, one has 10%, and the rest is equal weighted ($HI=0.20$). Generally speaking, the index takes on values between $1/n$ and 1. In our case, possible values lie between 0.1 and 1. High values indicate high concentration, with value of one indicating the extreme case of one single fund in the portfolio with the rest of the weights being set to zero.

Panel B of Table 5 reports the average value of the Herfindahl Index in the FRPs for each strategy. The reported values of around 0.2 are reasonably low for all strategies (see the above example of weights that yield an HI of this magnitude).

Table 5: Annual turnover by fund and average concentration coefficient—data from 01/ 2001 through 12/2003

	Convertible Arbitrage	CTA Global	Event Driven	Equity Long / Short	Equity Market Neutral
Panel A					
Minimum turnover	2%	0%	0%	0%	1%
Average turnover	11%	5%	8%	7%	7%
Maximum turnover	26%	21%	26%	23%	19%
Panel B					
Average concentration	20%	11%	23%	22%	20%

* Calculated using the formula of the Herfindahl Index.

In table 6, we report the correlation between the factor replicating portfolios and the corresponding first principal component (PC1). For comparison purposes, we also report the correlation of CSFB/Tremont investable indices, as well as the average correlation of non investable indices with respect to the first component.⁹

Table 6: Correlations between FRPs and PCI - for out-of-sample results, data from 01/ 2001 through 12/2003

	In- sample	Out-of- sample	CSFB/Tremont investable indices	Average of non investable indices
Convertible Arbitrage	0.98	0.95	0.87	0.89
CTA Global	0.99	0.96	0.93	0.98
Event Driven	0.99	0.96	0.87	0.95
Equity Long / Short	0.98	0.95	0.72	0.90
Equity Market Neutral	0.98	0.63	0.31	0.16

An important feature from table 6 is that not only in-sample correlations are very high, but also they remain high out-of-sample. For all strategies except equity market neutral, the out-of-sample correlation of the replicating portfolios with the corresponding first principal component is higher than 0.95, a spectacular result showing that a high level of

⁹ We use the CSFB Tremont because they are amongst investable indices the ones with the longest available history. We use in parallel all the indices entering into the composition of the Edhec Hedge Fund Indices to calculate the average of non investable hedge fund indices (refer to www.edhec-risk.com for a detailed list of index providers).

representativity can be achieved even on the basis of a very limited number of funds (10 in this experiment). Also, the correlation of the factor replicating portfolios with the first principal component is always significantly higher than that of the corresponding CSFB/Tremont indices, as well as that of the average non investable index. In the case of equity market neutral strategy, the significantly less attractive result obtained is undoubtedly due to the fact that systematic components are less present and a larger fraction of these funds' returns is explained by specific stock picking skills. It is important, however, to note that the *out-of-sample* correlation of the FRPs with the first principal component is twice as high than that of the corresponding CSFB/Tremont investable index, and almost 4 times higher than that of the average non investable index. This suggests that, while it is challenging to build an index representing the universe of market neutral manager, our technology allows for much better results than what is obtained otherwise.

Since by construction we choose funds that are highly correlated with the first principal component, it is not surprising that we find high correlation of the FRPs with PC1. The first principal component, however, is extracted from a database that has limited coverage of the entire hedge fund universe. In order to further assess the representativity of the FRPs, we look at their correlation with different proxies for the complete fund universe for each strategy. The proxies that are available to us are the Edhec indices and a proprietary database of hedge fund returns. The first proxy ("Edhec non investable indices") gives us an index by strategy that can be thought of as the best one dimensional summary of the information contained in the competing indices for one strategy (for more information, see www.edhec-risk.com or Amenc and Martellini (2003)). The second proxy ("representative portfolios") consists of the returns of equally weighted portfolios of funds in each strategy. These portfolios are composed from a database that spans several databases and includes data for additional funds that is not available in any commercial hedge fund database.¹¹ Therefore, these portfolios may be considered as highly representative of the fund universe for a given strategy.

Table 7 reports the correlation coefficients for both the first principal component extracted from the CISDM database and the FRPs with these two proxies.

Table 7: Correlations of PC1 and FRPs with EDHEC non-investable indices and representative portfolios - data from 01/ 2001 through 12/2003

	PC1		FRPs	
	EDHEC non-investable indices	Representative portfolios	EDHEC non-investable indices	Representative portfolios
Convertible Arbitrage	0.92	0.98	0.90	0.93
CTA Global	0.98	0.99	0.95	0.95
Event Driven	0.99	0.94	0.96	0.91
Equity Long / Short	0.97	0.98	0.95	0.94
Equity Market Neutral	0.32	0.58	0.36	0.25

The results confirm that the FRPs achieve high representativity for most strategies. The first principal component extracted from the CISDM database is extremely correlated with both the corresponding EDHEC non investable index and the corresponding representative portfolio. This is true for all strategies to the notable exception of Equity Market Neutral. Moreover, the correlations between the FRPs and these two proxies range between 0.90 and 0.96, which is only slightly lower than correlation of the FRPs with the first principle component itself (between 0.95 and 0.96). This is all the more remarkable that maximizing correlation with the Edhec non investable indices and the representative portfolios were not part of the objective in the optimization procedure.

Overall, this strongly suggests that representativity can be addressed with a very limited number of funds, provided that an adequate method is used in the design of the portfolios. Again, the results are less impressive in the case of Equity Market Neutral.

¹¹ More precisely, the representative portfolios use returns data and strategy classifications for 5,479 funds. This database merges several commercial hedge fund databases such as Hedge Fund Net, Altvest and CISDM. In addition funds that are not covered by any available database are included. We thank François-Serge Lhabitant for providing this data to us.

3. Robustness Analysis

To test for the robustness of the methodology, we now deviate from the base case experiment in a number of directions.

3.1. Number of Selected Funds

Depending on availability of funds as well as other considerations, it may happen that we will need to use a number of funds in each investable portfolio ranging anywhere from 5 to 15. To assess the robustness of the procedure with respect to the number of funds in the portfolio, we repeat the base case experiment but select 5 and 15 funds, respectively, as opposed to 10 funds.

The correlation coefficients with the first principal components for FRPs with a different number of funds are shown in table 8.

Table 8: Correlation coefficients between FRPs and PC1 – for out-of-sample results, data from 01/ 2001 through 12/2003

Number of funds in the FRP	5 funds		10 funds (base case)		15 funds	
	In-sample	Out-of-Sample	In-sample	Out-of-Sample	In-sample	Out-of-Sample
Convertible Arbitrage	0.93	0.93	0.98	0.95	0.99	0.96
CTA Global	0.98	0.93	0.99	0.96	0.99	0.97
Event Driven	0.96	0.95	0.99	0.96	0.99	0.96
Equity Long / Short	0.96	0.93	0.98	0.95	0.99	0.97
Equity Market Neutral	0.97	0.46	0.98	0.63	0.99	0.62

As can be seen from the table, both in sample and out-of-sample correlations with the first principal component are very robust with respect to the number of funds in the factor replicating portfolios. Even when only five funds are used instead of ten, correlations are very high. The only case where correlation drops in an important way is when using only five funds for the Equity Market Neutral FRP. Increasing the number of funds in the FRPs from ten to fifteen yields only marginal increases (in the order of 0.01) for the out-of-sample correlation coefficients.

3.2. Rank of Selected Funds in Terms of Correlation with Principal Component

It is of course unlikely that the funds exhibiting the highest correlation with the first principal component will be available and open. The question then arises to check how dependent the results are on the specific choice of the funds with the highest correlation.

To assess the robustness of the procedure with respect to the choice of funds in the portfolio, we repeat the base case analysis by selecting less correlated funds. Instead of choosing the 10 most correlated funds, we select the funds ranked 10 to 20, 20 to 30, 30 to 40 and 40 to 50.

The results are shown in table 9.

Table 9: Correlations between FRPs and PCI - for out-of-sample results, data from 01/ 2001 through 12/2003

	Convertible Arbitrage		CTA Global		Event Driven		Equity Long / Short		Equity Market Neutral	
	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample	In-sample	Out-of-sample
Top 10	0.98	0.95	0.99	0.96	0.99	0.96	0.98	0.95	0.98	0.63
10 to 20	0.96	0.90	0.98	0.98	0.97	0.87	0.98	0.95	0.96	0.40
20 to 30	0.92	0.82	0.98	0.97	0.97	0.92	0.98	0.86	0.78	0.54
30 to 40	0.88	0.87	0.98	0.96	0.96	0.87	0.98	0.96	0.23	0.24
40 to 50*	0.78	0.72	0.97	0.96	0.94	0.90	0.98	0.88	-0.11	0.16

* Bottom 10 for Convertible Arbitrage and Equity Market Neutral FRPs

As can be seen from the numbers in table 9, the analysis is extremely robust with respect to the choice of funds in terms of their correlation with the first principal component. In particular, even if we consider funds ranked 30 to 40 in terms of their in-sample correlation with the first component, the out-of-sample correlation of the investable portfolio with respect to the first component is still high, except for Equity Market Neutral.

3.3. Rebalancing of Funds in the Portfolio

In the base case experiment, the selection procedure was dynamically updated. In other words, the 10 funds exhibiting the highest degree of correlation were selected every year. In practice, it would be rather costly to terminate managers and invest in new ones each year. To test the robustness of the results with respect to how actively is the portfolio of funds managed, we now re-do the base case analysis, except that we now keep the same 10 funds for the whole experiment, while re-optimizing the quantities every year. The results can be found in table 10.

Table 10: Correlations between FRPs and PCI - for out-of-sample results, data from 01/ 2001 through 12/2003

	In-sample	Out-of-sample
Convertible Arbitrage	0.98	0.97
CTA Global	0.99	0.96
Event Driven	0.98	0.94
Equity Long / Short	0.98	0.94
Equity Market Neutral	0.98	0.62

As can be seen from table 10, and the comparison with table 6, the performance of the whole procedure appears to be only very marginally affected by the fact that we select only at the initial dates the best possible candidates. In light of this result, we suggest in practice to exclude a fund from the investable portfolio (and replace it with the new one) only in case of a major deterioration of this fund's correlation with the first principal component. This allows us to exclude funds that have enjoyed a too large level of style shifting.

One might also wonder whether increasing the frequency would have a significantly positive impact on the results. To this end, we now re-do the base case analysis, except that we select and optimize funds every 3 months. The results are featured in table 11.

Table 11. Correlations between FRPs and PCI- for out-of-sample results, data from 01/ 2001 through 12/2003

	Quarterly Optimization	
	In-sample	Out-of-sample
Convertible Arbitrage	0.98	0.95
CTA Global	0.99	0.97
Event Driven	0.99	0.96
Equity Long / Short	0.97	0.96
Equity Market Neutral	0.97	0.42

As can be seen from table 10, and the comparison with table 6, the performance of the whole procedure is only very marginally affected by the fact that we dynamically update the procedure on a quarterly, as opposed to a yearly, basis.

In practice, it would be reasonable to rebalance the portfolio not at fixed intervals of time, but only when weights have significantly shifted. This is actually consistent with insights obtained from optimal control techniques that have been used in the literature on dynamic portfolio optimization in the presence of transaction costs (see for example Leland (1999) or El Bied, Martellini and Priaulet (2002)).

In table 12 below, we report the results of an experiment where we only perform rebalancing if weights have significantly changed. For example if the average absolute change in weights is less than $x\%$ (we take $x=5\%$ and $x=10\%$ in our analysis), we do not rebalance.

Table 12. Correlations between FRPs and PCI - for out-of-sample results, data from 01/ 2001 through 12/2003

	Controlled Rebalancing ($x=5\%$)		Controlled Rebalancing ($x=10\%$)	
	In-sample	Out-of-sample	In-sample	Out-of-sample
Convertible Arbitrage	0.98	0.96	0.96	0.96
CTA Global	0.99	0.97	0.99	0.97
Event Driven	0.99	0.95	0.98	0.92
Equity Long / Short	0.98	0.95	0.98	0.94
Equity Market Neutral	0.98	0.64	0.98	0.64

Again, we conclude that there is a great deal of robustness with respect to how strictly the rebalancing procedure is enforced.

3.4. Constraints on Weights

Because of capacity constraints, and also because of the desire to limit the concentration of the portfolio, we test the impact of weight constraints, i.e., we impose maximum and minimum holdings. In table 13 below, we repeat the base case experiment, except that we now impose that no weights should be lower than 5% or higher than 20%.

Table 13: Correlations between FRPs and PC1 - for out-of-sample results, data from 01/ 2001 through 12/2003

	In-sample	Out-of-sample
Convertible Arbitrage	0.95	0.97
CTA Global	0.99	0.97
Event Driven	0.98	0.93
Equity Long / Short	0.98	0.96
Equity Market Neutral	0.98	0.65

Again, it appears that the procedure is remarkably robust with respect to weight constraints.

4. Conclusions

This paper considers the performance of a two-stage methodology based on factor analysis techniques to design factor replicating portfolios that may be regarded as investable and representative indices for various hedge fund strategies. Our results strongly suggest that a portfolio of well-chosen funds adequately captures the return characteristics of a large set of funds in the universe, except in the case of Equity Market Neutral strategies. A wide range of robustness tests show that high correlation of the factor replicating portfolios with the common factor of returns for each strategy is remarkably stable with respect to modifying the number of funds in the replicating portfolio, choosing less correlated funds, changing the frequency of rebalancing, using a static selection of funds and imposing constraints on weights.

Appendix: Definition of Hedge Fund Styles

Convertible Arbitrage	Investment in convertible bonds. The strategy generally consists of buying convertible bonds and selling short common stocks of the same company.
CTA Global	Although this is not commonly done, it seems appropriate to regroup Global Macro and CTA funds as one type of strategy. This is because both of these strategies involve trend-following strategies that use a wide set of assets such as currencies, interest rate products and commodities. Due to the great flexibility of these two strategies, it is commonly accepted that they are hard to distinguish, which has led us to the regrouping in a single category labelled “CTA Global”.
Event Driven	Investment strategy that exploits price movements related to the anticipation of events affecting the life of the company (merger, acquisition, bankruptcy, etc.).
Long Short Equity	Involves investing mainly in equities and derivative instruments. The manager systematically uses short selling, but takes care to maintain a permanent overall net position that is long with either a low or a high beta.
Equity Market Neutral	Exploits inefficiencies in the market through balanced buying of undervalued securities and selling of overvalued securities enabling either a beta or a dollar neutral approach to be obtained.

References

- Agarwal, V., and N. Naik, 2004, Risk and portfolio decisions involving hedge funds, *Review of Financial Studies*, 17 (1), 63-98.
- Alexander, C., and A. Dimitriu, 2004, Sources of out-performance in equity markets, *Journal of Portfolio Management*, 30 (4), 170-185.

- Amenc, N. and L. Martellini ,2003, The brave new world of hedge fund indices, *Working Paper, Edhec Risk & Asset Management Research Centre*.
- Bansal, R., and S. Viswanathan, 1993, No arbitrage and arbitrage pricing, *Journal of Finance*, 48 (4), 1231-1262.
- Bansal, R., D. Hsieh, and S. Viswanathan, 1993, A new approach to international arbitrage pricing, *Journal of Finance*, 48 (5), 1719-1747.
- Black, F. and R. Litterman ,1992, Global portfolio optimization, *Financial Analysts' Journal*, 48 (5), 28-43.
- Chan, L., J. Karceski and J. Lakonishok, 1998, The risk and return from factors, *Journal of Financial and Quantitative Analysis*, 33(2), 159-188.
- El Bied, S., L. Martellini and P. Priaulet, 2002, Competing investment strategies in the presence of market frictions, *USC working paper*.
- Fama, E. and K. French, 1992, The cross section of expected stock returns, *Journal of Finance*, 47 (2), 427-467.
- Feeney, G., and D. Hester, 1964, Stock market indices: a principal components analysis, *Working paper, Cowles Foundation*.
- Fung, W., and D. Hsieh, 1997, Empirical characteristics of dynamic trading strategies, *Review of Financial Studies*, 10 (2), 275-302.
- Fung, W., and D. A. Hsieh, 2000, Performance characteristics of hedge funds and CTA funds: natural versus spurious biases, *Journal of Financial and Quantitative Analysis*, 35 (3), 291-307.
- Fung, W., and D. Hsieh, 2001, The risk in hedge fund strategies: theory and evidence from trend followers, *Review of Financial Studies*, 14 (2), 313-341.
- Hubermann, G., S. Kandel and R. Stambaugh, 1987, Mimicking portfolios and exact arbitrage pricing, *Journal of Finance*, 42 (1), 1-9.
- Glosten, L.R., and R. Jagannathan, 1994, A contingent claim approach to performance evaluation, *Journal of Empirical Finance*, 1 (2), 133-160.
- Leland, H., 1999, Optimal asset rebalancing in the presence of transaction costs, *U.C. Berkeley Working Paper*.
- Lhabitant, F., 2001, Hedge fund investing: a quantitative look inside the black box, *Journal of Financial Transformation*, 1, 82-90.
- Lessard, D., 1973, International portfolio diversification, a multivariate analysis for a group of Latin countries, *Journal of Finance*, 28 (3), 619-633.
- Mitchell, M., and T. Pulvino, 2001, Characteristics of risk in risk arbitrage, *Journal of Finance*, 56 (6), 2135-2175.
- Ross, S., 1976, The Arbitrage Theory of Capital Asset Pricing", *Journal of Economic Theory*, 13 (3), 341-360.
- Ruckstuhl, A., P. Meier, S. Lodeiro and O. Kündig, 2004, Funds of hedge funds indices: properties, purpose and representativeness, *Working paper, ZHW*.
- Vaissié, M., 2004, Are all hedge fund indices created equal? Part 1 and Part 2, *Alternative Investments Quarterly*, 10 and 11.