Hedge fund performance, classification with machine learning, and managerial implications

BJM BRITISH JOURNAL of MANAGEMENT
BRITISH ACADEMY OF MANAGEMENT

British Journal of Management, Vol. 0, 1–25 (2025)
DOI: 10.1111/1467-8551.70011

Hedge Fund Performance, Classification with Machine Learning, and Managerial Implications

Emmanouil Platanakis,[1] Dimitrios Stafylas,[2] Charles Sutcliffe[3]
and Wenke Zhang[4]
[1]School of Management, University of Bath, Bath, BA2 7AY, UK, [2]School for Business and Society, University of York, York, YO10 5GD, UK, [3]The ICMA Centre, Henley Business School, University of Reading, Reading, RG6 6DL, UK, and [4]College of Business, Arts and Social Sciences, Brunel University of London, Middlesex, Uxbridge, UB8 3PH, UK
Corresponding author: email: c.m.s.sutcliffe@gmail.com

Prior academic research on hedge funds focuses predominantly on fund strategies in relation to market timing, stock picking and performance persistence, among others. However, the hedge fund industry lacks a universal classification scheme for strategies, leading to potentially biased fund classifications and inaccurate expectations of hedge fund performance. This paper uses machine learning techniques to address this issue. First, it examines whether the reported fund strategies are consistent with their performance. Second, it examines the potential impact of hedge fund classification on managerial decision-making. Our results suggest that for most reported strategies there is no alignment with fund performance. Classification matters in terms of abnormal returns and risk exposures, although the market factor remains consistently the most important exposure for most clusters and strategies. An important policy implication of our study is that the classification of hedge funds affects asset and portfolio allocation decisions, and the construction of the benchmarks against which performance is judged.

Introduction

During the last decade, hedge funds (HFs) have received significant attention from both academic researchers and practitioners. As of the second quarter of 2024, the total assets under management for the HF industry were almost USD$5.1 trillion (BarclayHedge, 2024). Each HF declares its investment strategy, which is both advertised to potential investors and used by databases when reporting HF performance.[1] Investors seek to achieve a diversified portfolio when making asset allocation decisions, which rely on the expected risk and return of possible investments and their correlations with each other. The formation of these expectations for HFs is heavily influenced by the reported past performance of the different HF strategies supplied by the available databases (e.g. Agarwal, Arisoy and Naik, 2017; Karehnke and Roon, 2022). Therefore, the classification of HFs into particular strategies by databases has an important influence on investment decisions.

While databases generally classify HFs according to the strategy declared by the HF itself, HFs sometimes diverge from, adjust or cease to follow their declared strategy. In consequence, such HFs are classified by the databases as an inappropriate strategy. Since performance differs as between strategies, this leads to investors forming inaccurate expectations when evaluating HFs, and this prevents them from forming portfolios that match their objectives. As there is no universal classification scheme for HF strategies, database vendors employ different classifications when forming HF

We are grateful to the editor, an associate editor and three anonymous reviewers of this journal, as well as conference participants at the 2023 Annual Meeting of the European Financial Management Association (EFMA) 2023 and the BAFA Annual Conference 2023 for their useful comments and suggestions that helped us improve the paper significantly.

[1]Due to the private partnership nature of HFs, information disclosure is not regulated by the Securities and Exchange Commission, and the inclusion of a HF in a database is a voluntary decision taken by its managers. This can lead to history bias, as only the more successful HF managers are motivated to report their performance. Even if a fund manager decides to report their performance, this could be limited to only one database. Hence, HF data are both biased, patchy and fragmented, which constitutes a major issue for researchers and investors.

indices, leading to differences between indices compiled by different databases that claim to measure the same strategy. This was documented by Amenc and Martellini (2003), who showed that indices for the same strategy have very low correlations. This inconsistency has an impact on investment decisions and performance against benchmarks.

Studies have analysed the investment attractiveness of HF indices (Brooks and Kat, 2002), the survivorship and selection biases of HF indices (Fung and Hsieh, 2002a) and the measurement and interpretation issues of HF indices (Brittain, 2001; Schneeweis, Kazemi and Martin, 2002; Stafylas, Anderson and Uddin, 2017). However, the use of inconsistent indices or benchmarks constitutes a challenge for finance managers and investors because they lead to ambiguous performance rankings (Dybvig and Ross, 1985; Dahlquist and Soderlind, 1999). Although the HF literature is vast, there is only limited research on fund classification and its implications for financial decision-making and asset and portfolio allocation. Our study addresses this lacuna.

The limited research related to the classification of HFs and its importance for investors' decisions serves as the main motivation of our paper. We shed light on the following questions: (i) Does HF classification vary according to specific performance features (risk and return characteristics)? (ii) If yes, does this have any economic significance and policy implications for finance managers and investors? To the best of our knowledge, we are the first to examine these classification issues when evaluating HF performance in terms of excess (abnormal) returns and exposures. Using data from *Morningstar* and exploiting machine learning (ML) techniques, we examine the classification of HFs into different strategies based on their risk and return characteristics.[2] We use ML techniques to classify HFs into ten clusters, depending on their specific features. Then we analyse the performance of these clusters in terms of their excess returns (*Jensen's* alpha) and systematic exposures using various asset pricing models. We then compare them to ten broad strategies based on the strategies reported by database vendors. We use several robustness checks and document the superiority of the ML approach in clustering over the traditional portfolio approach.

The main findings of our research include the following: (i) our results indicate that the classification of HFs by databases has only a modest relationship with risk and return; (ii) there are differences in excess returns and exposures between funds within the broad reported strategies, and clusters based on performance features; (iii) for both the broad reported strategies and our classification based on performance features, the market factor remains the most important risk exposure; (iv) funds that deviate from the others outperform their non-deviating peers; and finally (v) ML that considers higher moments as well such as skewness and kurtosis, produces clusters with lower correlations compared to the traditional portfolio approach.[3] We use several robustness checks that verify our findings.

We make several important contributions to the HF literature, with implications for financial decisions and policy-making. First, using ML techniques, we investigate the accuracy of the classification of HFs into broad strategies by the funds themselves and the data providers. Contrary to previous studies, where the grouping or strategy classification of HFs is mostly based on the views of database vendors or researchers, which may be subject to biases, we apply a statistically based clustering approach that uses each fund's return and risk characteristics to produce a more accurate classification. Second, we show that ML classification is superior to the traditional portfolio approach, producing classes that are more self-homogeneous and have lower correlations between them. Third, we compare the performance of the broad strategies used by database vendors and the previous literature with that of our clustered strategies. We do this by employing widely used empirical asset pricing models to analyse HF performance in terms of excess returns and systematic exposure. We find that the database strategy classifications are also heterogeneous in terms of factor sensitivities. Therefore, we find that database HF strategy classifications are a poor guide to HF performance. Our findings will help investors to classify HFs into strategies that are exposed to the same risk factors, and, therefore, we have a similar expected performance. This has implications for asset allocation by HF investors.

The rest of this article is organized as follows. The next section presents the related literature and the hypothesis development. The third section presents our data and methodology. The fourth section presents our results and discussion, and the fifth section concludes.

Literature review and hypotheses

Our work is centred around agency theory and information asymmetry. The former examines the impact of the conflict of interest between agents and principals, for instance, managers and shareholders (see Jensen and Smith, 1985) or bondholders and stockholders (Smith

[2]Morningstar is widely used in academic studies in empirical finance and HFs (see among others Prather and Middleton, 2006; Baibing *et al.*, 2017; Cui *et al.*, 2019). Morningstar contains both live and dead funds.

[3]Funds classified as 'remaining' belong to the dominant cluster, which represents the largest proportion of funds within each strategy. In contrast, funds assigned to any other cluster outside the dominant one for a given strategy are labelled as 'deviating'.

and Warner, 1979), among others. In our study, the relationship between investors and fund managers fits into this agency theory framework. Combined with the information asymmetry where fund managers have better and more timely information than investors (see Noe, 1988; Brennan and Hughes, 1991; Nachman and Noe, 1994), we shed light on fund managers' behaviour when they report funds' strategy and performance.

It is well known in the finance literature that inefficient benchmarks can result in misleading assessments (Dybvig and Ross, 1985; Dahlquist and Soderlind, 1999), due to the joint hypothesis testing problem (Li, Xu and Zhang, 2016). The classification of HFs and the selection of a relevant benchmark is a non-trivial process for investors. However, in investment practice and the academic literature, the strategy reported by HFs is taken for granted. For instance, when examining HF performance and systematic exposure, authors such as Getmansky (2004), Chen, Han and Pan (2021), Osinga, Schauten and Zwinkels (2021), Kuvandikov, Pendleton and Goergen (2022) and Karehnke and Roon (2022) use the classification scheme of the databases. Jawadi and Khanniche (2012), Meligkotsidou and Vrontos (2014) and Ferland and Lalancette (2021) use HF indices provided by databases. Other authors, such as Capocci and Hubner (2004), Joenvaara and Kosowski (2021) and Liang, Sun and Teo (2022) use more than one database and perform a mapping between the HF strategies provided by the databases.[4] Other authors, such as Agarwal and Naik (2004) and Kosowski, Naik and Teo (2007), map their data into broader classifications – directional, relative value, security selection and multi-process funds. Barès, Gibson and Gyger (2003) use a classification based on the asset class, investment process and geographical region provided by the fund manager.

Although important, previous studies do not question the validity of the HF strategy classifications employed by databases. This issue can have a significant impact on managerial decision-making with respect to the performance evaluation[5] of pension funds, endowment funds and other institutional investors. This is an important issue, as most institutional investors have policies related to the type and category of financial assets in which they are prepared to invest, particularly their riskiness. It is common knowledge that HFs have the flexibility to change their investment style without changing their declared strategy,[6] and a few strategies such as *Global Macro* and *Multi-strategy* are not well defined or easily replicated. Hence, HF indices, used as benchmarks for HF performance, might well be unsuitable for sound financial decision-making.

There are a few studies that examine strategy distinctiveness and fund performance using HF indices. For instance, Panopoulou and Voukelatos (2022) show that fund managers who deviate most from their peers have higher systematic and idiosyncratic risk without offering sufficiently higher returns. But Sun, Wang and Zheng (2012) found a positive relationship between strategy distinctiveness and subsequent performance. It is noticeable that both studies rely on prior literature classifications of HF (e.g. Joenvaara *et al.*, 2019; Brown and Goetzmann, 1997, 2003).[7]

Overall, existing HF classification practices are problematic for investors, database vendors and researchers. In their initial prospectus, HF managers may claim they follow a certain strategy, but later switch to another strategy when running their funds without publicizing this change in strategy. There is a strong need for a universally agreed-upon way of classifying HF into particular strategy groups. Mappings based on the information published by HF can be misleading and, if not performed with appropriate due diligence, produce misleading results. We address the issue of this potential subjectivity. Based on the foregoing discussion, our hypothesis related to the accuracy of strategy classification is as follows:

H 1. *Reported HF strategies are determined by specific features that describe the fund's characteristics.*

In the academic literature, there are many studies that deal with HFs' dynamic nature in terms of their exposure and returns (e.g. Giannikis and Vrontos, 2011; Chen, Han and Pan, 2021), changes in their asset and portfolio allocations (e.g. Patton and Ramadorai, 2013; Ferland and Lalancette, 2021) and significant exposure to specific factors (e.g. Meligkotsidou and Vrontos, 2014). Bali, Brown and Caglayan (2012) find that systematic risk is a significant factor in explaining the dispersion of cross-sectional HF returns. Similarly, strategies that attempt to be market neutral have exposure to market-wide risk factors (Duarte, Longstaff and Yu, 2007). Patton (2009) proposes five neutrality concepts for HFs (e.g. mean neutrality, variance neutrality, value-at-risk neutrality, tail neutrality and complete neutrality), and they document that even the so-called 'market neutral strategies' are not really neutral as

[4]Because the same strategies have different descriptions, for example relative value versus convertible arbitrage, event driven versus distressed securities and macro versus global macro; the mapping process involves grouping similar strategies into a single (broad) strategy.

[5]In terms of abnormal returns (Jensen's alpha) and risk exposure.

[6]There is an information asymmetry and agency theory conflict of interest between fund managers and investors.

[7]These classifications were HF indices of declared strategies.

approximately one-quarter of the funds have significant exposure to market risk.

Most of the above studies find common risk factors such as the market, commodities and credit are shared by many fund strategies. Factors related to the default spread and VIX (CBOE volatility index) are also economically important (Avramov, Barras and Kosowski, 2013); and other studies (Bali, Brown and Caglayan, 2011; 2014) find that macroeconomic risk factors, such as the default spread, term spread, short-term interest rates, equity market index and inflation rate are powerful determinants of HF returns. Other studies, such as Agarwal, Arisoy and Naik (2017) and Stafylas, Anderson and Uddin (2018), use macroeconomic variables and market uncertainty to explain HF returns over time. Investor sentiment or market psychology also has an important role in explaining HF returns (see Kellard *et al.*, 2017; Zheng, Osmer and Zhang, 2018; Osinga, Schauten and Zwinkels, 2021), as fund managers adjust the exposure of their portfolios to changes in market sentiment. Finally, another branch of the literature examines the timing ability of HFs (Chen and Liang, 2007; Cai and Liang, 2012; Cao *et al.*, 2013), showing that fund managers have timing skills.[8] Almost all previous studies examining different aspects of HF performance[9] take as given the reported classification of the funds. Hence, many conclusions and managerial decisions may be based on inconsistent and misleading classifications.

Despite the fact that there is an enormous literature which examines fund performance, fund characteristics, fund managers' skills, etc., there has been no examination of how HF strategies can be determined objectively from return data, rather than relying on statements by HF managers. There is also the problem that the data for such a statistical analysis of HFs comes from many different databases. We suggest a statistical approach which uses HF returns and their features, for example mean return, standard deviation, skewness and kurtosis. Practitioners can then use the resulting HF classification to develop and revise their portfolio allocation strategies and policies. Consequently, our second hypothesis concerning the impact of classification on managerial decision-making is proposed below:

H 2. *Hedge fund classification based on the first four moments of HF returns has the potential to improve investment decision-making.*

Data and methodology

We analyse the *Morningstar* database, which contains both live and dead funds and is one of the most widely used in the HF literature. The inclusion of dead funds addresses the problem of survivorship bias. As most HF databases came into existence in the early to mid-1990s, we consider net-of-fees monthly returns from January 1995 to August 2021. Similar to Ibbotson, Chen and Zhu (2011), Bali, Brown and Caglayan (2011), Stafylas, Anderson and Uddin (2018) and Chen, Han and Pan (2021), we exclude the first 12 monthly returns to minimize instant history bias. Other studies, such as Ackermann, McEnally and Ravenscraft (1999), exclude the first 24 or more returns; however, the exclusion of more returns can lead to truncated database bias.[10] The initial sample consists of 20 reported strategies of North American HFs, with no funds of funds. We aggregate the 20 strategies into ten broad strategies based on strategy descriptions from various sources; for example *Morningstar* and the classifications of other authors such as Baibing, Ji and Kai-Hong (2017) and Cui, Yao and Satchell (2019).[11] To prevent large funds dominating and thus biasing our results, we follow previous studies (Prather and Middleton, 2006; Bali, Brown and Caglayan, 2011; Cumming, Dai and Johan, 2015; Panopoulou and Voukelatos, 2022) and calculate total returns as the cross-sectional equally weighted mean return across funds for each strategy and broad HF category.

It is well known that fund managers smooth or massage their returns to mask risk (see Racicot, Théoret and Gregoriou, 2021; Racicot and Théoret, 2022, among others). When HF returns are based on non-market valuations of the underlying assets, there is an issue with illiquid markets and managed returns, leading to smoothed and biased returns. Return smoothing creates positive serial correlation in returns and reduces the variance and correlations with other assets. To address these issues, we use desmoothed returns[12]; Table 1 presents the correlation matrix (panel A) and basic statistics (panel B) for the ten broad HF strategies. The

[8]Other studies such as Bollen and Whaley (2009), Billio *et al.* (2012) and O'Doherty *et al.* (2015) consider methodological issues and structural breaks in HF returns via the use of advanced econometric methods. They show that funds' risk factors change over time, and that funds who can switch their exposure over time, outperform their peers.

[9]For example, cross-sectional variations in returns in relation to market-related risk factors (Fung and Hsieh, 1997, 2001, 2004; Agarwal, Daniel and Naik, 2003), macroeconomic variables (Avramov *et al.*, 2013; Bali *et al.*, 2014; Stafylas *et al.*, 2018), persistence (Baquero *et al.*, 2005; Stulz, 2007; Jagannathan *et al.* 2010) or as portfolio diversifiers (Denvir and Hutson, 2006; Eling, 2009; Platanakis *et al.*, 2019; Newton *et al.*, 2021).

[10]The initial sample was 2648 funds. After selecting US only funds with the base currency and removing the funds of funds, there were 1354 funds in the sample. After dealing with the instant history bias, we ended up with 1250 funds.

[11]In Online Appendix D, we provide a table with the groupings.

[12]For desmoothing the returns, we use the standard autoregressive (AR(1)) desmoothing process of Geltner (1993). In our robustness check section, we repeat our whole analysis considering the reported returns as well.

Table 1. Summary statistics

	Funds	Debt	Equity	Event driven	Multi-strategy	Systematic Futures	Volatility	Macro	Currency	Long Only	Others
Panel A: Correlation matrix											
Debt	240	1.000									
Equity	362	0.437	1.000								
		(6.063)									
Event driven	170	0.547	0.875	1.000							
		(8.163)	(22.586)								
Multi-strategy	123	0.468	0.904	0.898	1.000						
		(6.618)	(26.336)	(25.489)							
Systematic futures	152	−0.041	0.051	0.002	0.178	1.000					
		(−0.506)	(0.632)	(0.026)	(2.261)						
Volatility	37	0.217	0.409	0.334	0.386	0.099	1.000				
		(2.770)	(5.601)	(4.425)	(5.222)	(1.240)					
Macro	84	0.186	0.431	0.462	0.543	0.484	0.067	1.000			
		(2.359)	(5.973)	(6.511)	(8.072)	(6.911)	(0.841)				
Currency	13	−0.184	−0.049	−0.060	0.044	0.352	−0.220	0.277	1.000		
		(−2.333)	(−0.614)	(−0.749)	(0.554)	(4.700)	(−2.824)	(3.600)			
Long only	29	0.430	0.697	0.752	0.761	0.173	0.259	0.616	−0.054	1.000	
		(5.947)	(12.152)	(14.255)	(14.673)	(2.277)	(0.343)	(9.760)	(−0.678)		
Others	40	0.480	0.721	0.780	0.752	−0.024	0.215	0.407	−0.027	0.549	1.000
		(6.826)	(12.984)	(15.592)	(14.243)	(−0.295)	(2.755)	(5.560)	(−0.333)	(8.198)	
Panel B: Summary statistics											
Mean return	0.915	1.141	0.834	0.844	1.049	0.773	0.993	0.826	0.697	0.698	
Median	0.870	1.302	0.806	0.898	0.796	0.861	0.714	0.489	0.846	0.717	
St. Dev	3.021	3.165	2.644	2.001	3.861	2.501	2.634	3.354	2.518	2.315	
Skewness	3.903	−0.643	−1.452	−1.282	0.592	−0.675	0.210	1.246	−0.383	−0.993	
Kurtosis	75.587	5.915	11.681	12.291	4.210	5.172	4.335	6.989	5.229	13.789	
Jarque-Bera	71063.02	135.334	1117.091	1238.704	38.188	56.144	26.109	294.969	74.129	792.254	
Probability	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Sharpe ratio	0.243	0.303	0.247	0.332	0.225	0.270	0.308	0.192	0.205	0.285	
Tracking error	2.970	1.660	2.138	2.525	4.152	2.799	2.817	3.819	2.566	2.405	

Note: In this table, panel A, presents the correlation matrix of the ten broad hedge fund strategy desmoothed returns, which are the cross-sectional equal-weighted mean return across funds; *t*-Statistics are presented in parentheses. Panel B presents the monthly return statistics (%) of the mean, median, standard deviation (SD), skewness and kurtosis, along with their *Jarque-Bera* tests for normality for each broad strategy. The Sharpe ratio (which is the risk-free excess return to the portfolio standard deviation) and tracking error (with respect to the market index) are presented as well. For desmoothing the returns, the standard autoregressive (AR(1)) desmoothing process of Geltner (1993) is used.

highest (lowest) correlation is between the *Equity* and *Multi-Strategy* (*Volatility* and C*urrency*) broad strategies at 0.904 (−0.220). The *Currency* strategy has a low correlation with most HF broad strategies, as does the *Systematic Futures* strategy. Overall, most broad strategies have modest correlations that are statistically significant at the 5% level. With regard to the summary statistics, the *Equity* and *Systematic Futures* strategies provide the highest mean return (1.141 and 1.049, respectively) and among the highest standard deviations (3.165 and 3.861, respectively). The *Debt* strategy presents the highest skewness and kurtosis, along with one of the highest standard deviations. The *Systematic Futures* and *Macro* strategies have relatively low skewness and kurtosis. Using the *Jarque–Bera* test, we reject the normality of returns for the ten broad strategies at the 1% level.

We use three ML methods – support vector machines (SVM), random forests (RF) and *K*-means clustering. The first two are supervised learning, as they are trained on a back history of observations and the corresponding actual classifications. SVMs separate the observations based on their distance from a hyperplane, while RF uses an ensemble of decision trees to solve the classification problem. These two types of ML are suitable for our research as both work well in classifying observations, even when the number of observations is limited. The third ML method, *K*-means clustering, can be used as either supervised or unsupervised learning, making it suitable for situations where the back history lacks actual classifications, with classification based on a threshold distance.[13]

The following sections contain details of each method, followed by the empirical asset pricing models we use for HF performance evaluation.

Support vector machines

An SVM is a supervised ML model that computes either linear or non-linear boundaries between two classes. It finds the hyperplane that maximizes the distance from it to the nearest observation on each side (the margin). For multi-dimensional tasks that cannot be linearly separated, an SVM transforms the input data into a higher dimensional space by kernel functions that make the input data linearly separable.

In a classification setting, given a training set (x_k, y_k) $(k = 1,\ 2, \ldots, n)$ with a binary response $y_k \in \{-1,\ 1\}^n$, $w^\top x_k + b$ denotes the hyperplane that separates the sample data by maximizing the margin, w denotes a vector of coefficients of the input variables and b is the intercept. The distance (margin) of each point from the hyperplane is computed as

$$\frac{y_k\left(w^\top x_k + b\right)}{\| w\|_2}, \tag{1}$$

where $\| w\|_2$ is the ℓ_2 norm, that is $\| w\|_2 := \sqrt{w_1^2 + \cdots + w_n^2}$. The optimal classification model that maximizes the margin is obtained by solving the following quadratic optimization problem:

$$\min \| \mathbf{w} \|_2^2, \tag{2}$$

$$\text{s.t. } y_k\left(w^\top x_k + b\right) \geq 1\ \forall\ k = 1, \ldots,\ n, \tag{3}$$

$$w \in \mathbb{R}^p,\ b \in \mathbb{R}. \tag{4}$$

However, when the sample cannot be linearly separated, slack variables are introduced, leading to the following formulation (Cortes and Vapnik, 1995):

$$\min_{\xi, w, b} \| \boldsymbol{w} \|_2^2 + C \sum_{k=1}^{n} \xi_k^2, \tag{5}$$

$$\text{s.t. } y_k\left(w^\top x_k + b\right) \geq 1 - \xi_k,\ k = 1, \ldots, n, \tag{6}$$

$$\xi_k \geq 0,\ k = 1, \ldots,\ n, \tag{7}$$

where ξ_k is the slack (error) variable for observation k and C is a tuning weight that defines the trade-off between the minimization of the error and the maximization of the margin, with larger values of C representing a higher penalty for misclassification.

For more complex problems of multi-classification, the data can be mapped to a higher dimensional space through a mapping function $\Phi(x_k)$, which allows for linear classification in the new feature space. Based on the mapping function, the kernel function $K\ (x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ $(i, j = 1,\ 2, \ldots, n)$ is introduced for solving the quadratic programming problem, such that

$$\min \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i y_i \alpha_i \alpha_j K\left(X_i, X_j\right) - \sum_{j=1}^{n} \alpha_j, \tag{8}$$

$$\text{s.t. } \sum_{i=1}^{n} y_i\ \alpha_i = 0,\ \forall\ i = 1, \ldots, n, \tag{9}$$

$$0 \leq \alpha_i \leq C,\ \forall\ i = 1, \ldots, n, \tag{10}$$

Commonly used kernel functions include the polynomial, sigmoid and Gaussian kernels. In this approach, SVM identifies the hyperplane that separates every pair of classes, neglecting observations in the other classes.

[13]In Online Appendix E, we present details of the use of the SVM, along with a flowchart for tuning the parameters.

Random forests

RF is an ensemble method based on multiple decision trees. It uses bagging to generate many new training sets, which it uses to form different decision trees to separate the training set into classes. To classify new observations, RF selects the class indicated by the majority of the decision trees. By combining the predictions of multiple decision trees, RF has a better performance than a single classifier. RF also improves the performance of each decision tree by artificially restricting the set of features considered for each recursive split. The advantage of RF is its capability to capture complex data interactions with a relatively low bias if the tree grows sufficiently deep. RF is less prone to the overfitting problem and generally achieves a superior performance to decision trees.

Suppose there are N observations. The process of generating an RF starts by creating B bootstrap samples from the training data, where each sample consists of $n < N$ randomly chosen observations from the training set, with replacement. Then, random decision trees $T_b\,(b = 1,\ 2, \ldots,\ B)$ grow by randomly selecting m variables, picking the best variable among them and splitting the node into two sub-nodes. This process is repeated for the two sub-nodes until the minimum node size is reached. To predict with new data x, the regression function of the b_{th} RF tree is

$$\hat{f}^{B}_{\mathrm{rf}}\ (x) = \frac{1}{B}\sum_{b=1}^{B} T_b\,(x)\,. \tag{11}$$

In the regression, the RF model does not explicitly represent the error term and constant term in the mathematical equation. Instead, it implicitly incorporates them in the ensemble of decision trees $T_b(x)$ that constitute the model. The main objective of this model is to minimize the error between the predicted and actual values of the dependent variable by generating a diversified set of decision trees that can make precise predictions. The constant term of each decision tree $T_b(x)$ is the average of the dependent variable values of the training samples that fall within the leaf nodes. The error term for each decision tree $T_b(x)$ is the difference between the predicted and actual values of the dependent variable for each sample in the training set. The final error term for the RF regression model is the average of the error terms for all the decision trees in the forest.

For classification, the RF model identifies the best results by majority voting, which assigns a sample on the basis of the most frequent class assignment. The RF classification $\hat{C}^{B}_{\mathrm{rf}}(x)$ is formulated as

$$\hat{C}^{B}_{\mathrm{rf}}\ (x) = \ \text{majority vote}\ \left\{\hat{C}_b\,(x)\right\}_1^B. \tag{12}$$

K-*means clustering*

K-means clustering is one of the most commonly used unsupervised ML methods for partitioning a given data set into K groups, where K is pre-determined. Observations are classified by calculating their distance to the group centroids. The fundamental idea of K-means clustering is to minimize the within-cluster variation, which is defined as the sum of the squared Euclidean distances between each observation and its centroid. Formally, the distance function is as follows:

$$D\ (C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2, \tag{13}$$

where $x_i\ (i = 1,\ 2\ldots, n)$ is an observation belonging to a cluster $C_k\ (k = 1,\ 2\ldots, n)$ and μ_k is the mean value of observations assigned to the cluster C_k.The total within-cluster variance is the aggregation of the sum of squared distances in each cluster. It indicates the goodness of model performance, where a smaller value indicates a more accurate result. Formally, it is defined as follows:

$$T\ (C) = \sum_{k=1}^{n} D\,(C_k) = \sum_{k=1}^{n}\sum_{x_i \in C_k} (x_i - \mu_k)^2. \tag{14}$$

The first step of K-means clustering is to choose the number of clusters K. Then, the centroid of each cluster is randomly selected, and each observation is assigned to the closest cluster centroid. The centroid is updated by calculating the new mean of all the observations in the cluster iteratively to minimize the within-cluster variance. The iteration stops when the centroid and observations of the newly formed cluster stop changing.

Asset pricing models

We consider three empirical asset pricing models to obtain the alphas and factor betas for the broad strategies and clusters – (i) Carhart's four-factor momentum model (FF4) (Carhart, 1997), (ii) Fama and French's five-factor model (FF5) (Fama and French, 2015) and (iii) Fung and Hsieh's seven-factor model (FH7) (Fung and Hsieh, 2004).[14] More specifically, we apply the following models:

$$\begin{aligned}\mathrm{FF5}: R_{it} -\ R_{Ft} &= \alpha_i + b_i\,(R_{Mt} - R_{Ft}) + s_i SMB_t \\ &\quad + h_i HML_t + \delta_i MOM_t + e_{it},\end{aligned} \tag{15}$$

[14]The underlying factors of these models have been widely used not only in financial economics but in the HF literature (see, among others, Carhart, 1997; Fung and Hsieh, 2004; Capocci, 2009; Stafylas *et al.*, 2017).

Table 2. The clustering and proportions of 20 hedge fund strategies based on desmoothed data

C1	Proportion of funds	C2	Proportion of funds	C3	Proportion of funds	C4	Proportion of funds	C5	Proportion of funds	C6	Proportion of funds	C7	Proportion of funds	C8	Proportion of funds	C9	Proportion of funds	C10	Proportion of funds
GLM	5.23%	NON	32.49%	LSD	0.23%	DEA	26.51%	VOY	11.95%	CUY	14.87%	VOY	13.99%	SYF	34.09%	LOO	17.24%	LOE	5.45%
VOY	4.10%	COA	32.18%	NON		CUY	23.08%	DEA	6.04%	BME	13.33%	MEA	12.76%	EMN	33.40%	LSD	15.79%	BME	4.44%
DEA	4.03%	SLSE	27.08%	DEA		LOO	22.22%	COA	4.94%	EVD	9.79%	DIA	7.78%	DIS	30.12%	NON	15.74%	VOY	4.10%
LOD	3.50%	MEA	25.77%	LOD		GLM	21.29%	LSD	4.29%	LSD	9.73%	LOD	7.24%	GLM	27.76%	EVD	15.64%	SLSE	3.92%
DIS	3.42%	LSE	25.25%	BME		DIA	20.96%	LOO	3.07%	GLM	9.60%	EVD	6.86%	CUY	27.69%	BME	15.56%	COA	3.29%
MEA	3.06%	SYF	23.97%	LOE		SLSE	20.59%	LOD	2.57%	VOY	9.56%	LSD	6.74%	NON	26.40%	LOE	15.53%	EVD	2.02%
CUY	2.56%	EMN	23.58%	EMN		LSE	20.47%	BME	2.22%	LOO	9.20%	MUY	6.64%	LOD	25.70%	DIS	14.29%	LSE	1.77%
LSD	2.45%	MUY	23.38%	LSE		SYF	20.45%	MUY	1.55%	COA	9.14%	COA	6.22%	DIA	25.15%	LSE	13.83%	GLM	1.62%
MUY	2.14%	DIA	23.35%	SLSE		LOE	20.44%	CUY	1.54%	SLSE	8.95%	DIS	5.90%	SLSE	25.12%	MEA	13.52%	DIS	1.55%
EVD	2.02%	LOE	23.30%	EVD		MUY	20.21%	NON	1.52%	MUY	8.85%	EMN	5.28%	LSE	24.43%	MUY	13.27%	LSD	1.23%
SLSE	1.72%	LOD	23.13%	DIS		EVD	19.58%	EVD	1.31%	SYF	8.51%	DEA	5.03%	LOO	23.75%	DIA	13.17%	MUY	1.18%
LOO	1.53%	BME	22.22%	COA		DIS	19.25%	MEA	1.02%	DIS	7.76%	LOE	4.77%	LOE	23.71%	LOD	12.38%	NON	1.02%
LSE	1.47%	GLM	20.53%	DIA		EMN	18.87%	GLM	0.67%	LSE	7.65%	LSE	4.56%	EVD	23.51%	EMN	11.70%	DEA	0.67%
SYF	1.24%	EVD	19.27%	MEA		MEA	18.62%	DIS	0.62%	LOD	7.48%	BME	4.44%	DEA	22.82%	VOY	11.60%	SYF	0.57%
DIA	1.20%	LSD	19.08%	MUY		LSD	18.47%	EMN	0.57%	DIA	7.19%	NON	4.06%	MUY	22.79%	DEA	9.40%	EMN	0.57%
LOE	1.09%	LOO	18.77%	SYF		COA	18.46%	LSE	0.57%	DEA	7.05%	GLM	3.99%	LSD	22.07%	GLM	9.32%	LOO	0.38%
NON	1.02%	DEA	18.46%	VOY		BME	17.78%	SYF	0.43%	MEA	6.89%	LOO	3.83%	BME	20.00%	SLSE	8.82%	LOD	0.23%
COA	0.73%	CUY	17.95%	GLM		LOD	17.76%	SLSE	0.25%	EMN	5.66%	CUY	3.59%	MEA	18.37%	CUY	8.72%	DIA	
EMN	0.38%	DIS	17.08%	CUY		VOY	17.41%	LOE	0.14%	LOE	5.59%	SLSE	3.55%	COA	16.64%	SYF	8.61%	MEA	
BME		VOY	16.38%	LOO		NON	14.72%	DIA		NON	3.05%	SYF	2.14%	VOY	10.92%	COA	8.41%	CUY	

Note: This table shows the proportion of hedge funds in each new cluster using supervised learning. The odd columns show the reported classifications (symbols) of the hedge funds in each new cluster. The even columns show the proportions of hedge funds in each original strategy classified to each new cluster. DEA (Debt Arbitrage), LOD (Long Only Debt), LSD (Long/Short Debt), BME (Bear Market Equity), LOE (Long-Only Equity), EMN (Equity Market Neutral), LSE (Long/Short Equity), SLSE (Small Cap Long/Short Equity), EVD (Event Driven), DIS (Distressed Securities), COA (Convertible Arbitrage), DIA (Diversified Arbitrage), MEA (Merger Arbitrage), MUY (Multi-Strategy Funds), SYF (Systematic Futures), VOL (Volatility), GLM (lobal Macro), CUY (Currency), LOO (Long-Only Other), NON (Other (no names)).

Table 3. The number of individual funds in each cluster based on the desmoothed data

Morningstar strategy	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Convertible Arbitrage	3				3		20		11	1
Currency	6		1		4		1		1	
Debt Arbitrage	4		4		2		1	13	3	
Distressed Securities	5			7	17					
Diversified Arbitrage	3				6		1		3	
Equity Market Neutral	11			1	25		9		2	
Event Driven	15			1	29		9		8	5
Global Macro	22		1		36		5	1	2	14
Long-Only Debt	9		5		33		15	2	10	
Long-Only Equity	13				31		25		1	1
Long-Only Other	6		2	1	10		3		5	1
Long/Short Debt	21	2	9	2	32		25	6	31	9
Merger Arbitrage	8		2		2		10		1	
Multistrategy	23		8	1	52		26		9	3
Systematic Futures	35		1		102	2	8		2	2
Long/Short Equity	46		5		97		34		4	4
Small Cap Long/Short Equity	18				31		4		3	1
Volatility	2		1		4		21	1	2	6
Other (no names)	7				20		11		1	1
Bear Market Equity							2		1	
Sum	257	2	39	13	536	2	230	23	100	48

Note: This table shows the relationship between the 20 Morningstar strategies and the ten new clusters formed using unsupervised *K*-means clustering. The sums are the total number of funds in each cluster. For desmoothing the returns, the standard autoregressive (AR(1)) desmoothing process of Geltner (1993) is used.

$$\begin{aligned} \text{FF5}: R_{it} - R_{Ft} &= \alpha_i + b_i\,(R_{Mt} - R_{Ft}) + s_i SMB_t \\ &\quad + h_i HML_t + r_i RMW_t + c_i CMA_t \\ &\quad + e_{it}, \qquad (16) \end{aligned}$$

$$\begin{aligned} \text{FH7}: R_{it} - R_{Ft} &= \alpha_i + \eta_i PTFSBD_t + \kappa_i PTFSFX_t \\ &\quad + \rho_i PTFSCOM_t + b_i SP500_t \\ &\quad + s_i SIZESPR_t + \gamma_i \Delta BOND_t \\ &\quad + \theta_i \Delta CRSPR_t + e_{it}, \qquad (17) \end{aligned}$$

where R_{it} is the month t return on one of the portfolios from a classification of HFs; R_{Ft} is the risk-free rate (three-month Treasury bills); R_{Mt} is the return on a value-weighted market index; $(R_{Mt} - R_{Ft})$ is the market risk premium; and SMB_t (small minus big) and HML_t (high minus low) are the size and value factors, respectively. RMW_t (robust minus weak) is the profitability factor, CMA_t (conservative minus aggressive) is the investment factor and RMW_t is the difference between returns on diversified portfolios of stocks with robust and weak profitability. More details of the construction of these portfolios can be found in Fama and French (2015).

$PTFSBD_t$ is the return on a bond lookback straddle; $PTFSFX_t$ is the return on a currency lookback straddle; $PTFSCOM_t$ is the return on a commodity lookback straddle; $SP500_t$ is the return on the S&P500 index; $SIZESPR_t$ is the return difference between the Russell 2000 index and the S&P500 index; $\Delta BOND_t$ is the change in the yield of 10-year bonds; $\Delta CRSPR_t$ is the change in the difference between the *Baa* corporate bond yield and the 10-year treasury bond yield; and α_i is the alpha of the fund (selectivity skill) after controlling for the underlying risk factors. The coefficients to be estimated across these three models are α_i, b_i, s_i, h_i, δ_i, r_i, c_i, κ_i η_i, , ρ_i, γ_i and θ_i and e_{it} is the error term. We apply the three ML models and examine the impact of different classifications on HF abnormal returns (alpha) and factor exposures.

Results and discussion

First, we test our hypothesis (*H1*) and examine whether the reported HF strategies are consistent with HF performance. Second, we investigate the economic significance of our results for investors and test our second hypothesis (*H2*) regarding the potential impact of classification on managerial decision-making via its effect on abnormal returns and factor exposures.

Classification

The key features we use to describe the performance of HFs are the mean, variance, skewness and kurtosis of returns. To eliminate the influence of different magnitudes and units, we standardize these four variables so that they are suitable for comparative evaluation. We apply cross-validation (Bergmeir, Hyndman and Koo, 2018; Kaniel *et al.*, 2023; DeMiguel *et al.*, 2023) to

Table 4. Summary statistics of clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Panel A: Correlation matrix										
Cluster_1	1.000									
Cluster 2	−0.056	1.000								
	(−0.348)									
Cluster 3	0.654	0.066	1.000							
	(5.401)	(0.415)								
Cluster 4	0.072	−0.027	0.101	1.000						
	(0.452)	(−0.168)	(0.636)							
Cluster 5	0.922	0.017	0.696	0.032	1.000					
	(14.841)	(0.107)	(6.055)	(0.200)						
Cluster 6	0.068	0.034	0.069	0.039	0.099	1.000				
	(0.429)	(0.214)	(0.433)	(0.244)	(0.622)					
Cluster 7	0.840	0.010	0.918	0.063	0.889	0.101	1.000			
	(9.686)	(0.063)	(14.493)	(0.394)	(12.139)	(0.634)				
Cluster 8	0.564	0.094	0.941	0.054	0.629	0.083	0.857	1.000		
	(4.268)	(0.590)	(17.358)	(0.339)	(0.050)	(0.520)	(10.371)			
Cluster 9	0.752	0.084	0.957	0.113	0.809	0.107	0.967	0.912	1.000	
	(7.127)	(0.527)	(20.662)	(0.709)	(8.624)	(0.670)	(23.693)	(13.922)		
Cluster 10	0.215	−0.166	−0.430	−0.180	0.023	−0.078	−0.231	−0.429	−0.338	1.000
	(1.373)	(−1.053)	(−2.976)	(−1.143)	(0.146)	(−0.487)	(−1.483)	(−2.971)	(−2.242)	
Panel B: Summary statistics										
Mean return	1.166	1.471	0.801	1.176	0.965	0.932	0.790	0.309	0.602	1.697
Median	0.923	0.798	0.813	0.752	0.998	0.790	1.024	0.571	0.697	0.854
SD returns	2.431	5.099	3.312	4.968	2.402	6.679	2.275	4.659	2.955	4.178
Skewness	0.560	5.161	−3.707	3.757	−0.185	−4.495	−2.516	−8.737	−4.661	3.196
Kurtosis	4.224	31.288	35.458	29.799	3.233	57.245	18.593	106.384	42.828	19.513
Jarque–Bera	36.728	1586.838	14780.2	6261.90	2.553	40186.7	3579.43	119096.8	18543.9	3605.79
Probability	0.000	0.000	0.000	0.000	0.279	0.000	0.000	0.000	0.000	0.000
Sharpe Ratio	0.405	0.272	0.187	0.218	0.327	0.040	0.268	0.039	0.159	0.372
Tracking error	2.322	5.248	3.445	4.370	2.089	17.951	2.100	3.793	2.524	4.423
mCFSR	1.440	0.270	0.420	N/A	0.877	N/A	0.588	0.251	0.354	N/A

Note: In this table, panel A, presents the correlation matrix of the ten cluster desmoothed returns, which are the cross-sectional equal-weighted mean return across funds; *t*-Statistics are presented in parentheses. Panel B presents the monthly return statistics (%) of the mean, median, standard deviation (SD), skewness and kurtosis, along with their *Jarque–Bera* tests for normality for each broad strategy. The Sharpe ratio (which is the risk-free excess return to the portfolio standard deviation) and tracking error (with respect to the market index) are presented as well. The mCFSR *(modified Cornish Fisher Sharpe Ratio)* is presented for each cluster across the entire sample. It is computed as the excess portfolio return to a modified version of the value-at-risk at probability level α. N/A is due to mVaR < 0, and hence the calculation of the mCFSR is not possible. The clusters are formed using the *k-means* algorithm. For desmoothing the returns, the standard autoregressive (AR(1)) desmoothing process of Geltner (1993) is used.

decide the optimal hyperparameter of each ML model. To evaluate the classification accuracy, we utilize the accuracy ratio[15] that calculates the probability of correctly classified funds.

We investigate whether the 20 HF strategies employed by Morningstar[16] and using desmoothed returns are determined by fund performance. If so, this supports hypothesis *H1*, which has the implication that there should be few significant correlations between returns on the different HF strategies. However, Table 1 (panel A) shows that almost three-quarters of these correlations are significantly different from zero. We apply the three ML methods using four features (e.g. mean return, risk, skewness and kurtosis) as the input variables, and the 20 labels (strategies) as the corresponding output to assess whether HFs are classified in the same way as Morningstar. We use RF, which is efficient in dealing with redundant information, and the least affected by data quality when compared to the other two ML methods. Moreover, empirical studies show that the best results are obtained if 20–30% of the data are used for

[15]We do not use the receiver operator characteristic curve because it is only suitable for binary classification. In this multi-classification problem, we apply the precision-recall curve as a robustness check on classification accuracy (Davis and Goadrich, 2006; Tharwat, 2020).

[16]The broad strategies are those used in the academic literature based on vendors' strategy descriptions (e.g. Morningstar, Eurekahedge, etc.) and other authors' classifications (such as Baibing *et al.*, 2017; Cui *et al.*, 2019) using similar databases.

Table 5. Ten broad strategies – FF4

Dependent variable	Debt	Equity	Event driven	Multi-strategy	Systematic futures	Volatility	Macro	Currency	Long Only	Others
C	0.506**	0.451**	0.342**	0.471**	0.764**	0.494*	0.637**	0.625**	0.267*	0.290
	(2.593)	(5.284)	(2.698)	(4.530)	(3.605)	(2.488)	(4.462)	(3.709)	(2.116)	(1.370)
MKT_RF	0.284**	0.572**	0.366**	0.215**	0.057	0.190**	0.194**	−0.020	0.283**	0.267**
	(4.875)	(22.366)	(8.248)	(6.697)	(0.676)	(3.049)	(5.741)	(−0.408)	(7.147)	(3.810)
SMB	0.046	0.304**	0.201**	0.145**	−0.047	0.034	0.048	0.045	0.01	0.036
	(1.220)	(11.203)	(6.488)	(4.461)	(−0.666)	(0.369)	(1.085)	(0.699)	(0.234)	(0.808)
HML	0.033	0.033	0.052	0.052	0.08	−0.07	−0.073	0.092	0.038	−0.147
	(0.558)	(0.745)	(0.876)	(1.118)	(1.238)	(−0.952)	(−1.479)	(1.524)	(0.933)	(−1.130)
MOM	−0.028	−0.011	−0.053	−0.022	0.165**	0.013	0.041	0.064	0.043	−0.177**
	(−1.086)	(−0.659)	(−1.770)	(−1.142)	(3.501)	(0.372)	(1.242)	(1.445)	(1.270)	(−2.765)
R^2	0.202	0.876	0.566	0.361	0.039	0.106	0.128	0.017	0.234	0.503
F-statistic	19.899	558.62	102.655	44.551	3.214	5.966	11.53	1.329	24.11	38.684
Prob(*F*-stat)	0.000	0.000	0.000	0.000	0.013	0.000	0.000	0.259	0.000	0.000
Durbin–Watson	1.794	1.933	2.011	1.973	1.977	1.827	2.277	1.938	2.101	1.615
F-stat BG	0.561	2.133	0.772	0.671	1.271	0.609	2.152	1.952	2.403	1.573
Prob (*F*-stat)	0.873	0.015	0.679	0.779	0.235	0.833	0.014	0.028	0.006	0.106
Obs*R^2	6.954	24.927	9.493	8.284	15.332	7.672	25.130	22.961	27.809	18.657
Prob chi^2	0.861	0.015	0.660	0.763	0.224	0.810	0.014	0.028	0.006	0.097

Note: This table provides the results of the ten broad strategies in terms of alphas and exposures using the FF4 model employing desmoothed returns. Each broad strategy consists of a cross-sectional, equally weighted mean return across funds. The risk-free (RF) return is the one-month Treasury bill rate from the Fama and French online data library (Ibbotson Associates). SMB is small minus big, HML is high minus low book-to-market capitalization, and MOM is momentum. * and ** denote significance at the 5% and 1% levels, respectively. The *t*-statistics are in parentheses. The Newey–West (HAC) estimator is used to deal with any residual autocorrelation and heteroskedasticity.

testing, and the remaining 70–80% of the data for training (Gogas, Papadimitriou and Agrapetidou, 2018; Huang and Yen, 2019; Petropoulos *et al.*, 2020). So, in this study, 75% of the sample is used for training the model, with the remaining observations used for testing. After learning from the training, only 25% of the RF classifications agreed with those of Morningstar. For robustness, we employ the RF and SVM which show that the accuracy of the original classification is 0.25 and 0.19, respectively. This implies that the reported classification of three-quarters of the HFs differs from that based on HF performance. Consequently, hypothesis *H1* does not hold for the 20 HF strategies.

Following the low level of agreement between the two supervised ML methods and Morningstar, we applied unsupervised *K*-means clustering to classify the HFs and compared the results with the ten broad strategies, that is we set the number of clusters $K = 10$.[17] *K*-means clustering is efficient at non-linear classification and is suitable for large data sets compared to other clustering models, such as hierarchy clustering.

In Table 2, we present the proportion of HFs in each Morningstar strategy that is classified to each of the 10 *K*-means clusters using supervised learning. The *Long-Only Equity* (LOE) and *U.S. Small Cap Long/Short Equity* (SLSE) account for large proportions of *K*-means clusters 1, 2, 3, 4, 6, 8 and 9 and 1, 3, 4, 7, 8 and 9, respectively.

We also classify the individual funds directly, regardless of any prior knowledge of the 20 strategy classifications, that is unsupervised learning using *K*-means clustering. To avoid the influence of inconsistent time periods across individual funds, we calculate the features based on the whole period when the fund was active. We exclude HFs with missing information, which leaves 1250 HFs. Table 3 shows the number of individual funds in each of the new clusters, and the first column shows their original Morningstar label. To confirm our findings in Table 2, we use the results in Table 3 to calculate the adjusted rand index (ARI) (Hubert and Arabie, 1985) between the original Morningstar labels and the predicted labels of the individual HFs. The ARI measures the similarity between two clusterings, taking into account the differences in the number of clusters by adjusting for chance agreement (In Online Appendix G, we present the general form).

The value of ARI is 0.033, which is very low, suggesting that the original strategies are inadequate for HF classification based on the four characteristics of fund performance. This clustering result supports our earlier finding using RF and SVM, that hypothesis *H1* is not valid. Therefore, to the extent that HFs are following their declared strategy, their strategy is a poor guide to performance, which raises questions about the usefulness of the reported strategies.

[17] *K* was set to 10 to have the same number of broad strategies, for comparison reasons.

Table 6. Ten broad strategies – FF5

Dependent variable	Debt	Equity	Event Driven	Multi-strategy	Systematic Futures	Volatility	Macro	Currency	Long Only	Others
C	0.475*	0.474**	0.274*	0.426**	0.704**	0.470*	0.64**	0.596**	0.283*	0.322
	(2.210)	(5.335)	(1.988)	(3.951)	(3.333)	(2.533)	(4.485)	(3.675)	(1.983)	(1.371)
MKT_RF	0.298**	0.565**	0.397**	0.239**	0.065	0.167**	0.193**	−0.008	0.268**	0.311**
	(4.537)	(23.651)	(7.847)	(8.673)	(0.750)	(2.827)	(5.572)	(−0.149)	(5.757)	(4.355)
SMB	0.053	0.29**	0.218**	0.144**	0.026	0.069	0.045	0.05	0.038	0.047
	(1.045)	(9.273)	(6.345)	(4.174)	(0.265)	(0.743)	(0.937)	(0.637)	(0.698)	(0.641)
HML	0.038	0.064	0.054	0.023	−0.163	0.004	−0.128*	−0.031	0.023	−0.047
	(0.425)	(1.240)	(0.619)	(0.406)	(−1.855)	(0.047)	(−2.170)	(−0.435)	(0.361)	(−0.385)
RMW	0.025	−0.044	0.059	0.014	0.206	0.142	−0.004	0.03	0.061	−0.066
	(0.354)	(−1.068)	(1.109)	(0.346)	(1.875)	(1.384)	(−0.068)	(0.324)	(0.903)	(−0.809)
CMA	−0.006	−0.033	0.003	0.091	0.276	−0.353**	0.106	0.228	−0.062	−0.007
	(−0.061)	(−0.602)	(0.033)	(1.413)	(1.527)	(−2.640)	(1.242)	(1.828)	(−0.591)	(−0.091)
R^2	0.200	0.877	0.560	0.364	0.023	0.148	0.127	0.020	0.233	0.414
F-statistic	15.73	448.48	79.793	35.91	1.486	6.947	9.102	1.285	19.034	21.461
Prob(*F*-stat)	0.000	0.000	0.000	0.000	0.194	0.000	0.000	0.27	0.000	0.000
Durbin–Watson	1.783	1.920	1.991	1.982	2.038	1.868	2.296	1.981	2.077	1.384
F-stat BG	0.576	2.080	0.729	0.463	1.135	0.501	2.141	1.966	2.306	3.547
Prob (*F*-stat)	0.861	0.018	0.723	0.935	0.331	0.912	0.015	0.027	0.008	0.000
Obs*R^2	7.164	24.429	9.004	5.787	13.806	6.384	25.092	23.184	26.861	36.838
Prob chi^2	0.847	0.018	0.703	0.927	0.313	0.896	0.014	0.026	0.008	0.000

Note: This table provides the results of the ten broad strategies in terms of alphas and exposures using the FF5 model and employing desmoothed returns. Each broad strategy consists of a cross-sectional, equally weighted mean return across funds. The risk-free (RF) return is the one-month Treasury bill rate from the Fama and French online data library (Ibbotson Associates). SMB is small minus big, HML is high minus low book-to-market capitalization, RMW is profitability and CMA is investment. * and ** denote significance at the 5% and 1% levels, respectively. The *t*-statistics are in parentheses. The Newey–West (HAC) estimator is used to deal with any residual autocorrelation and heteroskedasticity.

The above results suggest that the reported strategies are inadequate for fund classification based on the four characteristics of fund performance. A question that naturally arises is why this difference exists. In the HF industry, there is no universal classification scheme, and no transparency, as fund managers are not obliged to disclose information publicly and may have a variety of reasons for deviating from their stated fund strategy. The reasons for this deviation, or style drift, seem to be related to risk and return (see Schwindler and Oehler, 2006; Koenig and Burghof, 2022), whether or not a fund is based in Delaware for US funds (Cumming, Dai and Johan, 2015) or the legal conditions in the country for non-US funds (Cumming, Dai and Johan, 2013). These deviations may be related to agency theory, which acknowledges the conflict of interest between agents and principals (see, for instance, Lambert, 2001). Fund managers may give priority to their own interests (e.g. bonuses at the end of the financial year) and present a better image to investors via return smoothing and misreporting returns (Agarwal, Daniel and Naik, 2011; Cassar and Gerakos, 2011; Bollen and Pool, 2009).

Table 4 presents the correlation matrix and the summary statistics for the ten clusters. The correlation coefficients are, on average, lower than for the broad strategies presented in Table 1 (0.28 versus 0.32). Using the *Jarque–Bera* test, we reject the normality of returns for the ten clusters at the 1% level. Clusters 2, 4 and 10 have the highest means (1.471, 1.176 and 1.697, respectively) and also among the highest Sharpe ratios. Cluster 8 has the lowest mean (0.309). It also has the highest kurtosis and the lowest skewness. This is due to the inclusion of many funds from the *Debt* strategy, which have similar risk characteristics.[18] Cluster 5 has one of the lowest skewness and kurtosis measures, which is because it consists mainly of funds following the *Equity* and *Systematic Futures* strategies. Clusters 1 and 7 have moderate skewness and kurtosis since they have larger proportions of funds with similar risk characteristics; however, cluster 7 has negative skewness because it contains more funds from the *Equity* strategy. A few clusters have extreme tail risk and low Sharpe ratios, while other clusters have a more balanced risk-return profile. Based on a modified Sharpe ratio, which takes into account higher moments (*modified Cornish Fisher Sharpe*

[18]Debt-focused hedge funds frequently employ a high level of leverage, which amplifies both returns and risk, increasing their volatility. Fixed-income securities are illiquid (Asness *et al.*, 2001) and liquidity shocks could lead to extreme price movements contributing to high skewness and kurtosis. In addition, they have high leverage, require more ‘intellectual capital’ and produce high skewness and kurtosis (Duarte *et al.*, 2007). Debt-focused hedge funds have a high sensitivity to interest rates and credit spreads (Fung and Hsieh, 2002b, 2003) and are especially exposed to tail risk. Malkiel and Saha (2005) document that fixed income funds have high kurtosis and skewness.

Table 7. Ten broad strategies – FH7

Dependent variable	Debt	Equity	Event Driven	Multi-strategy	Systematic Futures	Volatility	Macro	Currency	Long Only	Others
C	0.559**	0.51**	0.403**	0.540**	0.869**	0.543**	0.657**	0.667**	0.347**	0.437*
	(3.825)	(6.539)	(4.028)	(7.020)	(4.124)	(2.982)	(5.082)	(3.988)	(2.790)	(2.000)
PTFSBD	−0.01	−0.003	−0.008	−0.013	0.041*	−0.011	0.005	−0.018	0.002	0.006
	(−0.919)	(−0.602)	(−1.048)	(−1.591)	(2.541)	(−1.233)	(0.669)	(−1.568)	(0.248)	(0.991)
PTFSFX	−0.01	0.003	−0.004	0.005	0.048**	−0.018	0.018*	0.086**	0.007	0.002
	(−1.002)	(0.927)	(−0.829)	(1.419)	(4.052)	(−1.917)	(2.163)	(6.335)	(1.039)	(0.279)
PTFSCOM	−0.016	−0.005	−0.011	0.00	0.044*	0.005	0.036**	−0.003	0.015	−0.021
	(−1.712)	(−1.210)	(−1.460)	(−0.020)	(2.577)	(0.399)	(2.792)	(−0.197)	(1.637)	(−1.928)
SP500	0.225**	0.554**	0.324**	0.175**	0.091	0.124	0.213**	0.020	0.246**	0.249**
	(5.552)	(26.196)	(10.130)	(6.799)	(1.122)	(1.926)	(6.505)	(0.388)	(6.197)	(3.778)
SIZESPR	0.052	0.37**	0.201**	0.121**	0.007	0.046	0.08	0.06	0.021	0.023
	(1.872)	(12.832)	(6.464)	(4.054)	(0.112)	(0.556)	(1.576)	(0.899)	(0.564)	(0.506)
ΔBOND	−0.056*	0.006	−0.036*	−0.014	−0.056	−0.021	−0.03	−0.048	−0.044*	−0.005
	(−1.999)	(0.459)	(−2.134)	(−1.070)	(−1.569)	(−0.942)	(−1.692)	(−1.701)	(−2.526)	(−0.369)
ΔCRSPR	−0.106	−0.043**	−0.113**	−0.101**	−0.094*	−0.024	−0.052**	−0.045	−0.115**	−0.076*
	(−1.768)	(−3.398)	(−5.106)	(−3.337)	(−2.170)	(−1.010)	(−2.771)	(−1.529)	(−4.292)	(−2.120)
R^2	0.268	0.876	0.622	0.453	0.180	0.136	0.181	0.228	0.294	0.467
F-statistic	16.287	313.822	73.425	36.902	9.751	4.455	9.862	13.159	18.524	18.741
Prob(*F*-stat):	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Durbin–Watson	1.842	1.996	2.094	2.017	1.879	1.895	2.324	1.973	2.044	1.463
F–stat BG	0.488	0.949	1.306	1.583	1.315	0.878	1.694	1.941	1.895	2.608
Prob (*F*-stat)	0.921	0.029	0.214	0.096	0.209	0.570	0.067	0.029	0.034	0.004
Obs*R^2	6.133	23.15	15.882	19.052	15.99	11.042	20.312	23.05	25.55	29.212
Prob chi^2	0.909	0.027	0.197	0.087	0.192	0.525	0.061	0.027	0.032	0.004

Note: This table provides the results of the ten broad strategies in terms of alphas and exposures using the FH7 model and employing desmoothed returns. Each broad strategy consists of a cross-sectional, equally weighted mean return across funds. The risk-free (RF) return is the one-month Treasury bill rate from the Fama and French online data library (Ibbotson Associates). PTFSBD is the return on a bond lookback straddle, PTFSFX is the return on a currency lookback straddle, PTFSCOM is the return on a commodity lookback straddle, SP500 is the return on the S&P500, SIZESPR is the return difference of the Russell 2000 and the S&P500 index, ΔBond is the monthly change in the 10-year treasury security yield, and ΔCRSPR is the change in the difference between the BAA and 10-year treasury security yield. * and ** denote significance at the 5% and 1% levels, respectively. The *t*-statistics are in parentheses. The Newey–West (HAC) estimator is used to deal with any residual autocorrelation and heteroskedasticity.

Ratio or mCFSR), some clusters display a better ability to balance risk and return, particularly when dealing with extreme tail events. Cluster 1 has the highest mCFSR (1.440), and cluster 8 has the lowest (0.251).[19] Overall, the summary statistics highlight that the clusters are distinct groups of funds with different performance profiles. A few clusters, such as clusters 6 and 8, have extreme tail risk and low Sharpe ratios; while other clusters, such as 1 and 5, have a more balanced risk-return profile.

Performance

In order to examine our second hypothesis *H2* regarding managerial decision-making, we employ three of the most popular asset pricing models (FF4, FF5 and FH7) and compare the abnormal returns and systematic risk of the ten broad HF strategies, and those formed by *K*-means clustering using fund performance. With the market as the most important factor, our findings are consistent with previous studies regarding the exposures of trend-following funds (Fung and Hsieh, 2004) and exposure to various factors (Agarwal, Daniel and Naik, 2003; Meligkotsidou and Vrontos, 2014; Fama and French, 2015; Stafylas, Anderson and Uddin, 2018).

Table 5 presents the results using the FF4 model.[20] All but one (*Others*) broad strategy delivers statistically significant excess returns at the 5% level or higher. The highest excess return is for *Systematic Futures* at 0.764% (*t*-statistic = 3.605), and the lowest is for *Long Only* at 0.267% (*t*-statistic = 3.709). All but *Systematic Futures* and *Currency* have a statistically significant exposure to the market factor. Three of the strategies have a significant exposure to size (SMB), and two have a significant exposure to momentum (MOM).

When using the FF5 model, our results in Table 6 show that almost all the broad strategies have a signif-

[19]In section ‘Deviating versus Remaining Funds’, we discuss the mCFSR measure.

[20]In all models, the *Newey–West (HAC - heteroskedasticity and autocorrelation consistent)* estimator is used to deal with any residual autocorrelation and heteroskedasticity.

Table 8. K-*means clusters – FF4*

Dependent variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
C	0.693**	1.462	0.472*	0.853	0.417**	0.836*	0.318**	−0.047	0.251	1.515**
	(6.424)	(1.818)	(2.255)	(1.897)	(4.699)	(2.275)	(2.699)	(−0.115)	(1.193)	(5.754)
MKT_RF	0.329**	0.050	0.153*	0.247	0.399**	−0.087	0.342**	0.344	0.328**	−0.033
	(10.309)	(0.460)	(2.199)	(1.769)	(16.978)	(−0.806)	(7.699)	(1.550)	(3.249)	(−0.460)
SMB	0.142**	−0.229	0.196**	0.049	0.176**	−0.249	0.112**	0.025	0.061	0.321*
	(3.372)	(−0.626)	(3.664)	(0.388)	(5.946)	(−1.224)	(4.031)	(0.439)	(1.508)	(2.402)
HML	0.033	0.281	0.121	−0.014	0.000	−0.008	0.123	0.159	0.1	−0.202*
	(0.928)	(0.114)	(1.336)	(−0.120)	(0.001)	(−0.049)	(1.863)	(1.735)	(1.233)	(−2.120)
MOM	0.006	0.230	−0.034	0.045	0.044	0.061	−0.027	−0.021	−0.029	−0.132
	(0.231)	(1.032)	(−0.735)	(0.543)	(1.964)	(0.655)	(−1.286)	(−0.532)	(−0.853)	(−1.820)
R^2	0.455	0.020	0.119	0.047	0.659	0.025	0.574	0.136	0.309	0.090
F-statistic	65.787	0.215	10.608	2.322	152.323	2.023	105.892	10.04	29.152	6.721
Prob(*F*-stat)	0.000	0.929	0.000	0.058	0.000	0.091	0.000	0.000	0.000	0.000
Durbin–Watson	1.933	2.119	1.765	1.937	2.020	1.967	1.947	1.678	1.726	2.101
F-stat BG	1.971	0.171	1.627	1.258	0.458	0.525	1.273	2.027	1.137	3.741
Prob (*F*-stat)	0.027	0.999	0.083	0.247	0.938	0.898	0.234	0.023	0.331	0.000
Obs*R^2	23.174	3.189	19.373	15.251	5.698	6.519	15.353	23.659	13.818	40.772
Prob chi^2	0.026	0.994	0.079	0.228	0.931	0.888	0.223	0.023	0.313	0.000

Note: This table provides the results of the ten *K*-means algorithm clusters in terms of alphas and exposures using the FF4 model and employing desmoothed returns. Each cluster consists of a cross-sectional, equally weighted mean return across funds. The risk-free (RF) return is the one-month Treasury bill rate from the Fama and French online data library (Ibbotson Associates). SMB is small minus big, HML is high minus low book-to-market capitalization and MOM is momentum. * and ** denote significance at the 5% and 1% levels, respectively. The *t*-statistics are in parentheses. The Newey–West (HAC) estimator is used to deal with any residual autocorrelation and heteroskedasticity. The clusters are formed using the *K*-means algorithm.

icantly positive constant and that *Systematic Futures* (*Long Only*) continues to deliver the highest (lowest) excess returns to investors, equal to 0.704% (0.283%). The investment (CMA) factor is statistically significant once, and the small minus big (SMB) and/or high minus low (HML) factors are significant for four strategies. The market factor continues to be significant for eight strategies. As for the FF4 model, *Systematic Futures* and *Currency* have very low R^2 values and *Equity* has the highest R^2.

As Table 7 shows, there are similar results in terms of excess returns when using the FH7 model, with all ten constants being significantly positive. Returns on the S&P500 index (*SP500*) have a significant positive coefficient for seven strategies, and the credit spread factor ($\Delta CRSPR$) also has a significant negative coefficient for seven strategies. The regressions for three strategies (*Event Driven*, *Systematic Futures* and *Macro*) have five significant factors. *Equity* still has the highest R^2, while the low R^2 values for *Systematic Futures* and *Currency* have increased.

Table 8 has our results when applying *K*-means clustering based on HF performance measures for the FF4 model. All except clusters 4, 8 and 9 provide excess returns to investors, ranging from 0.318 (*t*-statistic = 2.699) to 3.082 (*t*-statistic = 2.207). Cluster 2 has a particularly low R^2 (0.020) with a *Prob F-stat* (0.215), which indicates that the FF4 model does not explain this cluster's returns. This might be due to the low number of funds in this cluster. This is also the case with other clusters, such as cluster 6, which means that the underlying funds are different in terms of their return and risk characteristics. Overall, the most common exposure is to the market factor.

Table 9 provides the results for the FF5 model in relation to the *K*-means clusters. Clusters 1, 4, 5, 7 and 10 provide statistically significant excess returns to investors. The highest is from cluster 10 at 1.497 (*t*-statistic = 6.037) and the lowest is from cluster 7 at 0.245 (*t*-statistic = 1.987). The most common exposure is the market factor, followed by the SMB. The significance of the other factors (e.g. HML, RMW and CMA) is less common.

Table 10 has the results of applying the FH7 model to the *K*-means clusters. All but clusters 2 and 8 provide statistically significant excess returns to investors. The highest is from cluster 10 at 1.466 (*t*-statistic = 5.603), and the lowest is from cluster 9 at 0.345 (*t*-statistic = 2.422). The particularly low R^2 for cluster 2 (0.065) and cluster 4 (0.064) shows that the underlying model cannot explain their returns. Cluster 1 has statistically significant positive coefficients for *PTFSFX* at 0.024 (*t*-statistic = 3.544), *PTFSCOM* at 0.020 (*t*-statistic = 2.138), *SP500* at 0.353 (*t*-statistic = 10.372) and *SIZESPR* at 0.177 (*t*-statistic = 4.232). ΔCRSPR has a significantly negative coefficient at −0.061 (*t*-statistic = −3.262). The bond-oriented risk factors ($\Delta BOND$ and $\Delta CRSPR$) both have statistically significant negative

Table 9. K-*means clusters – FF5*

Dependent variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
C	0.688**	1.730	0.390	0.987*	0.418**	0.704	0.245*	−0.124	0.142	1.497**
	(6.617)	(1.623)	(1.845)	(2.2052)	(4.967)	(1.825)	(1.987)	(−0.273)	(0.590)	(6.037)
MKT_RF	0.336**	0.028	0.185**	0.21	0.395**	−0.071	0.371**	0.363	0.37**	0.037
	(10.458)	(0.199)	(2.666)	(1.551)	(15.629)	(−0.612)	(8.122)	(1.603)	(3.307)	(0.688)
SMB	0.121**	−0.383	0.235**	−0.033	0.189**	−0.033	0.142**	0.127	0.129**	0.112
	(2.767)	(−0.731)	(3.440)	(−0.275)	(5.234)	(−0.173)	(4.305)	(1.616)	(2.815)	(1.042)
HML	0.001	0.357	0.096	0.013	−0.05	−0.089	0.092	0.168	0.057	−0.219*
	(0.021)	(0.165)	(0.873)	(0.100)	(−1.112)	(−0.536)	(0.990)	(1.304)	(0.464)	(−2.280)
RMW	−0.043	−0.398	0.117	−0.348*	0.033	0.545*	0.093*	0.217	0.179*	−0.381**
	(−0.959)	(−0.644)	(1.463)	(−2.031)	(0.756)	(2.342)	(2.063)	(1.792)	(2.117)	(−2.644)
CMA	0.118	−0.310	0.001	−0.111	0.053	−0.352	0.028	−0.237	−0.035	0.606*
	(1.681)	(−1.482)	(0.005)	(−0.337)	(0.835)	(−0.942)	(0.306)	(−1.601)	(−0.298)	(2.459)
R^2	0.463	0.039	0.122	0.058	0.654	0.060	0.578	0.153	0.324	0.166
F-statistic	54.058	0.297	8.712	2.332	118.655	4.022	85.936	9.152	24.869	10.715
Prob(*F*-stat)	0.000	0.912	0.000	0.044	0.000	0.001	0.000	0.000	0.000	0.000
Durbin–Watson	1.969	2.103	1.781	1.995	2.076	1.955	1.934	1.664	1.704	2.111
F-stat BG	1.689	0.247	1.363	1.152	0.393	0.844	1.059	2.093	1.062	3.182
Prob (*F*-stat)	0.068	0.993	0.183	0.321	0.966	0.605	0.395	0.018	0.393	0.000
Obs*R^2	20.131	4.624	16.446	14.133	4.922	10.385	12.922	24.447	13.004	35.583
Prob chi^2	0.065	0.969	0.172	0.293	0.961	0.582	0.375	0.018	0.369	0.000

Note: This table provides the results of the ten *K*-means algorithm clusters in terms of alphas and exposures using the FF5 model and employing desmoothed returns. Each cluster consists of a cross-sectional, equally weighted mean return across funds. The risk-free (RF) return is the one-month Treasury bill rate from the Fama and French online data library (Ibbotson Associates). SMB is small minus big, HML is high minus low book-to-market capitalization, RMW is profitability and CMA is investment. * and ** denote significance at the 5% and 1% levels, respectively. The *t*-statistics are in parentheses. The Newey–West (HAC) estimator is used to deal with any residual autocorrelation and heteroskedasticity. The clusters are formed using the *K*-means algorithm.

coefficients for clusters 3 and 6. The equity-oriented risk factors (*SP500* and *SIZESPR*) have statistically significant coefficients for clusters 1, 5 and 7.

Deviating versus remaining funds. We examine some further aspects, such as analysing whether funds that deviate, perform better or worse than funds that do not deviate. Separately for each HF strategy and across the entire sample, we classify the underlying funds as 'remaining' and 'deviating'. The funds defined as 'remaining' are the ones that belong to the dominant cluster, that is the cluster that accumulates the greatest proportion of funds in each fund strategy, while the rest of the funds, that is funds that belong to any other cluster except the dominant cluster for each strategy, are defined as 'deviating'. Table 11 (panel A) reports the proportions of the funds assigned to the dominant cluster per strategy. We observe that they vary from about 23% and 36% for *Long/Short Debt* and *Long-Only Other*, respectively, to about 67% for *Bear Market Equity* and *Systematic Futures*. On average, the mean proportion of funds assigned to the dominant cluster across all strategies is about 50%.

Continuing our analysis, beyond the strategy types of the US funds, we examine the effects of some other characteristics, such as the assets under management (AUM) and manager tenure. Considering individual funds, we find that (Table 11, panel B) in 12 out of 20 cases (strategies), although having a weak statistical significance, deviating funds are those with higher AUM. This might happen because of decreasing return opportunities for large funds that cannot scale up their trading without eroding the opportunity (Getmansky, 2004) or diseconomies of scale (Agarwal, Daniel and Naik, 2003, 2009). In terms of manager tenure, with strong statistical significance (Table 11, panel C), we find that in most cases (15 out of 20) remaining funds have longer manager tenure. Managers with longer tenure tend to have a lower exposure to market risk (see Chevalier and Ellison, 1999; Clare *et al.*, 2022) and might be less inclined to change the risk-return profile of the fund.

Due to the non-linear nature of HFs and the ability of our study to conduct classification based on their performance measured with higher statistical moments (mean, standard deviation, skewness and kurtosis), we use the modified Cornish Fisher Sharpe ratio ($mCFSR_{\alpha}$),[21] which employs higher moments to measure and compare performance between the remaining and deviating funds.

Table 12 reports the annualized mCFSRs for both the remaining and deviating funds for each strategy. We observe that the deviating funds outperform the remaining funds for 14 of the 19 HF strategies.[22] For exam-

[21] In Online Appendix F, we explain the computation of $mCFSR_{\alpha}$.

[22] We compare 19, instead of 20, HF strategies since $mVaR_{\alpha} < 0$ for the deviating funds of the *Long/Short Debt* strategy.

Table 10. K-*means clusters – FH7*

Dependent variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
C	0.729**	1.687	0.621**	0.989*	0.482**	0.861*	0.395**	0.148	0.345*	1.466**
	(7.142)	(1.621)	(4.724)	(2.296)	(5.772)	(2.451)	(4.580)	(0.517)	(2.422)	(5.603)
PTFSBD	0.007	−0.010	−0.029	0.016	0.007	0.038*	−0.011	0.02	0.003	0.033*
	(1.138)	(−0.818)	(−1.797)	(1.149)	(1.857)	(2.050)	(−1.609)	(0.850)	(0.204)	(2.134)
PTFSFX	0.024**	0.014	−0.004	−0.005	0.016**	0.034	−0.003	−0.031	−0.019	−0.011
	(3.544)	(0.387)	(−0.538)	(−0.264)	(4.120)	(1.333)	(−0.766)	(−1.620)	(−1.841)	(−0.595)
PTFSCOM	0.020*	−0.037	−0.011	−0.012	0.009	−0.029	−0.013	−0.021	−0.018	0.02
	(2.138)	(−0.909)	(−0.910)	(−0.495)	(1.428)	(−0.877)	(−1.902)	(−1.266)	(−1.908)	(0.867)
SP500	0.353**	−0.207	0.034	0.171	0.394**	−0.081	0.285**	0.156	0.225**	0.086
	(10.372)	(−1.295)	(0.651)	(1.591)	(14.779)	(−0.693)	(10.180)	(1.380)	(4.110)	(1.062)
SIZESPR	0.177**	0.001	0.099*	−0.006	0.237**	−0.141	0.112**	−0.014	0.043	0.28*
	(4.232)	(0.007)	(2.194)	(−0.049)	(7.007)	(−0.921)	(3.764)	(−0.262)	(0.965)	(2.291)
ΔBOND	−0.007	−0.065	−0.075*	0.022	−0.026	−0.302**	−0.021	−0.134	−0.076	0.031
	(−0.423)	(−1.068)	(−2.1052)*	(0.423)	(−1.836)	(−3.483)	(−1.515)	(−1.483)	(−1.849)	(0.609)
ΔCRSPR	−0.061**	−0.167	−0.227**	−0.082	−0.059**	−0.195*	−−0.106**	−0.357*	−0.192**	0.038
	(−3.262)	(−1.288)	(−−2.951)	(−1.479)	(−4.573)	(−−2.201)	(−4.393)	(−2.367)	(−3.033)	(0.812)
R^2	0.525	0.065	0.309	0.064	0.6800	0.119	0.661	0.373	0.497	0.070
F-statistic	49.308	0.337	19.888	1.808	94.864	5.981	86.882	21.379	36.385	2.884
Prob(*F*-stat)	0.000	0.931	0.000	0.088	0.000	0.000	0.000	0.000	0.000	0.006
Durbin–Watson	1.879	2.075	1.829	1.880	2.043	1.979	2.042	1.650	1.769	2.158
F-stat BG	2.063	0.205	1.485	1.304	0.201	0.681	0.987	2.017	0.733	4.801
Prob (*F*-stat)	0.019	0.997	0.129	0.220	0.998	0.769	0.462	0.024	0.718	0.000
Obs*R^2	24.389	4.217	17.938	16.008	2.554	8.492	12.149	23.820	9.186	50.703
Prob chi^2	0.018	0.979	0.118	0.191	0.998	0.746	0.434	0.022	0.687	0.000

Note: This table provides the results of the ten *K*-means algorithm cluster strategies in terms of alphas and exposures using the FH7 model and employing desmoothed returns. Each cluster consists of a cross-sectional, equally weighted mean return across funds. The risk-free (RF) return is the one-month Treasury bill rate from the Fama and French online data library (Ibbotson Associates). PTFSBD is the return of the bond lookback straddle, PTFSFX is the return of the currency lookback straddle, PTFSCOM is the return of the commodity lookback straddle, SP500 is the return of the S&P500, SIZESPR is the return difference of the Russel 2000 and the S&P500 index, ΔBond is the monthly change of the 10-year treasury security and ΔCRSPR is the change in difference of the BAA and 10-year treasury security. * and ** denote significance at the 5% and 1% levels, respectively. The *t*-statistics are in parentheses. The Newey–West (HAC) estimator is used to deal with any residual autocorrelation and heteroskedasticity. The clusters are formed using the *K*-means algorithm.

Table 11. Manager tenure, AUM and outperformance for deviating and remaining funds

Panel A		Panel B			Panel C		
		Remaining funds	Deviating funds		Remaining funds	Deviating funds	
Strategy		Manager tenure	Manager tenure	Outperformance	Average AUM	Average AUM	Outperformance
Debt Arbitrage	50.00%	11.8	10.7	R	106.30	235.26	D(*)
Long-Only Debt	44.59%	4.0	6.6	D(***)	135.26	1023.26	D
Long/Short Debt	23.36%	10.7	8.6	R(*)	29,950.70	2686.70	R(***)
Bear Market Equity	66.67%	8.8	11.4	D(&)	58.87	92.13	D(&)
Long-Only Equity	43.66%	12.2	7.6	R(***)	102.90	116.12	D
Equity Market Neutral	52.08%	10.5	8.2	R(*)	229.60	397.21	D
Long/Short Equity	50.53%	12.1	13.1	D	286.08	127.59	R(***)
Small Cap Long/Short Equity	54.39%	14.8	11.4	R(**)	110.85	52.47	R(**)
Event driven	43.28%	13.3	14.9	D	156.99	189.12	D
Distressed Securities	58.62%	15.0	11.4	R(***)	336.92	183.56	R
Convertible Arbitrage	52.63%	14.9	12.2	R(**)	205.63	302.57	D(&)
Diversified Arbitrage	46.15%	12.5	10.8	R(&)	2110.48	181.80	R(&)
Merger Arbitrage	43.48%	16.9	14.2	R	177.93	162.78	R
Multi-strategy	42.62%	10.2	9.8	R	1369.10	602.10	R
Systematic Futures	67.11%	13.8	12.6	R	11,140.34	415.04	R
Volatility	56.76%	4.3	6.8	D(*)	670.87	2846.33	D
Global Macro	44.44%	12.6	9.8	R	1238.00	29572.06	D
Currency	46.15%	17.9	13.1	R(&)	145.21	285.93	D(&)
Long-Only Other	35.71%	9.5	6.0	R(**)	701.83	8811.56	D
Other (no names)	50.00%	10.0	8.8	R(&)	146.30	196.96	D

Note: In this table, panel A reports the proportions of the funds assigned to the dominant cluster per hedge fund strategy. Panel B reports the mean manager tenure in years for the remaining and deviating funds in each hedge fund strategy. Panel C reports the average assets under management (AUM) for the remaining and deviating funds in each hedge fund strategy. The funds defined as 'remaining' are the ones that belong to the dominant cluster, that is the cluster that accumulates the greatest proportion of funds in each fund strategy, while the rest of the funds, that is funds that belong to any other cluster except the dominant cluster for each strategy, are defined as 'deviating'. D(*/**/***) and R(*/**/***) denote statistically significant outperformance for the deviating and remaining funds, respectively. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively, based on a two-sample t-statistic with unequal means and variances. Outperformance without statistical significance for the deviating and remaining funds is denoted as D and R, respectively. D(&) and R(&) denote outperformance in terms of the sample mean for the deviating and remaining funds, respectively, but no statistical tests were conducted due to the insufficient number of observations (<10) for the variable of interest.

Table 12. Outperformance (mCFSR) for deviating and remaining funds

Strategy	Remaining funds (R) mCFSR	Deviating funds (D) mCFSR	Outperformance
Debt Arbitrage	0.15	0.68	D(***)
Long-Only Debt	0.80	0.46	R(***)
Long/Short Debt	0.63	N/A	N/A
Bear Market Equity	−0.05	0.42	D(***)
Long-Only Equity	0.46	0.54	D(*)
Equity Market Neutral	3.36	1.61	R(***)
Long/Short Equity	0.74	0.93	D(*)
Small Cap Long/Short Equity	0.49	0.62	D
Event Driven	0.48	0.70	D(**)
Distressed Securities	0.23	0.94	D(***)
Convertible Arbitrage	0.24	0.21	R
Diversified Arbitrage	0.45	0.62	D(*)
Merger Arbitrage	0.71	1.78	D(***)
Multi-strategy	0.74	0.85	D
Systematic Futures	0.48	1.01	D(***)
Volatility	0.27	0.86	D(***)
Global Macro	0.57	1.86	D(***)
Currency	0.72	0.30	R(**)
Long-Only Other	0.31	0.90	D(***)
Other (no names)	0.99	0.59	R(*)

Note: This table reports the annualized modified Cornish Fisher Sharpe ratio (mCFSR) for the remaining and deviating funds in each hedge fund strategy. D(*/**/***) and R(*/**/***) denote statistically significant outperformance for the deviating and remaining funds, respectively. *, ** and *** indicate significance at the 10%, 5% and 1% level, respectively, based on 1000 bootstrapped values of the mean, standard deviation, skewness and kurtosis for computing 1000 values of mCFSR per hedge fund strategy and corresponding p-values. Outperformance without statistical significance for the deviating and remaining funds is denoted as D and R, respectively. N/A is due to mVaR < 0, and hence the calculation of the mCFSR is not possible.

ple, the annualized mCFSR for the deviating (remaining) funds is 0.68 (0.15), 0.42 (−0.05), 0.70 (0.48), 0.94 (0.23) and 1.78 (0.71), for *Debt Arbitrage, Bear Market Equity, Event Driven, Distressed Securities* and *Merger Arbitrage*, respectively. In order to examine whether the observed differences are statistically significant, we use bootstrapping, a statistical resampling technique, to estimate the sampling distribution of the statistic. Based on 1000 bootstrapped values of the mean, standard deviation, skewness and kurtosis, we computed 1000 values of mCFSRs per HF strategy and, consequently, the corresponding p-values. For 12 of these 14 strategies, the outperformance of the deviating funds, relative to the remaining funds, is statistically significant. We also consider structural breaks, specifically the two periods from 2000 to 2015, and the period from 2016 to the end of the sample. The results (see Tables H1 and H2 in Online Appendix H) confirm our findings. In both the first period and second periods, most deviating fund strategies outperformed the remaining strategies, that is 12 out of 20 for the first period, and 11 out of 18 (18 cases are examined due to $mVaR_{\alpha} < 0$) for the second period. There is outperformance of the deviating funds for both periods, but overall, for the second period, the mCFSRs are lower than for the first period (for both deviating/remaining funds), and this is probably due to the Dood–Frank Act. Our findings are in alignment with the previous literature in terms of performance (see Cumming, Dai and Johan, 2020). In Online Appendix H, we present the findings for the two periods under examination.

The outperformance of the deviating funds may be due to their innovation and activism, making them stand out from the herd by producing a stronger performance. For example, Sun, Wang and Zheng (2012) document that skilled HF managers tend to adopt unique investment strategies that result in superior performance, while DesJardine and Durand (2020) show that HF activism results in immediate increases in market value and profitability. Further, Brav *et al.* (2018) show that firms targeted by activists improve their innovation efficiency following HF intervention, while Xiao and Shen (2024) show that firms with a stronger commitment to exploration vis-à-vis exploitation in innovation search increase the likelihood of HF activism.

Additional analysis. We compare our approach with the traditional approach of sorting funds into portfolios using a 2×2×2×2 independent sort. Exploiting this portfolio benchmark,[23] we consider the in-sample vari-

[23]Sorting the data into bins is one of the most common statistical methodologies in portfolio analysis, especially when examining the cross-sectional relationship between two or more variables. It is a nonparametric technique, which means it does not make any assumptions about the nature of the cross-sectional relationships between the variables under examination. Thus, we use it as the benchmark compared with the ML approach.

Table 13. Correlation matrix using desmoothed returns

Panel A	HHHH	HHHL	HHLH	HHLL	HLHH	HLHL	HLLH	HLLL	LHHH	LHHL	LHLH	LHLL	LLHH	LLHL	LLLH	LLLL
HHHH	1															
HHHL	0.455	1														
	(8.205)															
HHLH	0.53	0.628	1													
	(10.042)	(12.964)														
HHLL	0.565	0.808	0.838	1												
	(11.005)	(22.046)	(24.646)													
HLHH	0.204	0.49	0.214	0.315	1											
	(3.347)	(9.02)	(3.525)	(5.338)												
HLHL	0.422	0.781	0.561	0.713	0.495	1										
	(7.477)	(20.094)	(10.895)	(16.32)	(9.147)											
HLLH	0.463	0.589	0.818	0.796	0.243	0.549	1									
	(8.397)	(11.711)	(22.841)	(21.11)	(4.031)	(10.556)										
HLLL	0.419	0.69	0.709	0.762	0.411	0.622	0.657	1								
	(7.41)	(15.325)	(16.15)	(18.897)	(7.252)	(12.769)	(13.984)									
LHHH	0.348	0.541	0.709	0.653	0.359	0.482	0.612	0.571	1							
	(5.968)	(10.332)	(16.128)	(13.833)	(6.181)	(8.829)	(12.441)	(11.173)								
LHHL	0.18	0.773	0.313	0.454	0.354	0.527	0.338	0.468	0.324	1						
	(2.943)	(19.59)	(5.298)	(8.187)	(6.071)	(9.948)	(5.772)	(8.497)	(5.510)							
LHLH	0.437	0.554	0.888	0.719	0.174	0.499	0.71	0.653	0.575	0.321	1					
	(7.803)	(10.684)	(31.02)	(16.606)	(2.841)	(9.253)	(16.177)	(13.842)	(11.281)	(5.453)						
LHLL	0.415	0.824	0.695	0.754	0.403	0.67	0.641	0.715	0.625	0.656	0.629	1				
	(7.328)	(23.332)	(15.533)	(18.424)	(7.065)	(14.500)	(13.405)	(16.414)	(12.87)	(13.959)	(13.011)					
LLHH	0.307	0.518	0.407	0.516	0.338	0.513	0.402	0.379	0.341	0.335	0.346	0.457	1			
	(5.188)	(9.738)	(7.167)	(9.684)	(5.765)	(9.609)	(7.058)	(6.586)	(5.822)	(5.718)	(5.914)	(8.248)				
LLHL	0.195	0.557	0.38	0.416	0.211	0.442	0.384	0.364	0.286	0.593	0.370	0.500	0.396	1		
	(3.187)	(10.764)	(6.592)	(7.343)	(3.475)	(7.926)	(6.678)	(6.285)	(4.800)	(11.82)	(6.407)	(9.276)	(6.925)			
LLLH	0.465	0.524	0.899	0.758	0.153	0.502	0.837	0.64	0.612	0.253	0.836	0.595	0.378	0.368	1	
	(8.439)	(9.886)	(32.926)	(18.695)	(2.488)	(9.314)	(24.579)	(13.387)	(12.444)	(4.197)	(24.507)	(11.899)	(6.557)	(6.367)		
LLLL	0.263	0.566	0.412	0.568	0.269	0.474	0.472	0.555	0.416	0.468	0.351	0.590	0.352	0.410	0.391	1
	(4.381)	(11.019)	(7.263)	(11.077)	(4.481)	(8.635)	(8.59)	(10.715)	(7.346)	(8.502)	(6.018)	(11.745)	(6.050)	(7.219)	(6.823)	

Note: This table presents the correlation matrix of the 16 fund groups of desmoothed returns; each group is the cross-sectional equal-weighted mean return across funds; t-Statistics are presented in parentheses, below the correlation coefficients. A traditional approach of sorting funds into portfolios using a 2×2×2×2 independent sort is followed; 16 different portfolios are formed considering the in-sample variables (e.g. mean return, standard deviation, skewness and kurtosis). Where H and L denote high and low portfolios, sorting accordingly.

Table 14. Common exposures – ML and 2×2×2×2

	ML approach	Common exposures	2×2×2×2 approach	Common exposures
FF4	Market	50%	Market	100%
	Size	50%	Size	56%
	HML	10%	MOM	31%
Average		**37%**		**62%**
FF5	Market	50%	Market	94%
	Size	50%	Size	75%
	HML	10%	HML	13%
	RMW	40%	RMW	19%
	CMA	10%	CMA	13%
Average		**32%**		**43%**
FF7	PTFSBD	10%	PTFSBD	13%
	PTFSFX	20%	PTFSFX	31%
	PTFSCOM	10%	PTFSCOM	38%
	SP500	40%	SP500	100%
	SIZESPR	50%	SIZESPR	56%
	ΔBOND	20%	ΔBOND	19%
	ΔCRSPR	70%	ΔCRSPR	75%
Average		**31%**		**47%**

Note: This table provides the common exposures (in terms of the total number of exposures) for the FF4, FF5 and FF7 employing the machine learning (*K*-means algorithm) and the traditional portfolio approach (2×2×2×2 independent sorting of funds into portfolios). Each group is the cross-sectional equal-weighted mean return across funds. HML is high minus low book-to-market capitalization; MOM is momentum, RMW is profitability and CMA is investment; PTFSBD is return of bond lookback straddle, PTFSFX is return of the currency lookback straddle, PTFSCOM is return of commodity lookback straddle, SP500 is return of the S&P500, SIZESPR is return difference of the Russel 2000 and the S&P500 index, ΔBond is monthly change of the 10-year treasury security and ΔCRSPR is change in difference of the BAA and 10-year treasury security.

Table 15. CAPM alpha and beta

	Alpha cluster 1 (low)	Alpha cluster 2 (high)
Alpha metric	0.21	0.55
	Beta cluster 1 (Low)	Beta cluster 2 (high)
Beta metric	0.15	0.81

Note: This table provides the high/low alphas and betas from applying the machine learning approach to these two metrics. Each cluster is the cross-sectional equal-weighted mean return across funds within the underlying cluster. The clusters are formed using the *K*-means algorithm. The estimation of the CAPM alphas and betas is based on a rolling 36-month window similar to other studies, such as Cui, Yao and Satchell (2019).

ables (e.g. mean return, standard deviation, skewness and kurtosis), and so we construct 16 different portfolios. Table 13 presents the correlation matrix, where all the coefficients are statistically significant. In addition, the average correlation among the portfolios is much higher (0.502) compared to our ML classification (0.265), which means that our approach is better in classifying funds based on their performance characteristics. In addition, compared to the 2×2×2×2 portfolios, the ML classification has relatively fewer common significant exposures for all the asset pricing models under consideration. Table 14 shows that for the ML approach (2×2×2×2 approach) and employing the FF4 model, the average common exposure count (in terms of the total number of exposures) is 37% (62%); for the FF5 model, this is 32% (43%); and for the FF7 model, it is 31% (47%). The previous results support the superiority of the ML approach to classification.

To see the extent to which the results are mechanical, we apply the ML algorithms to two metrics: the CAPM alpha and beta, using a rolling 36-month window regression similar to other studies (e.g. Cui, Yao and Satchell, 2019). Table 15 shows that when applying the ML approach, we obtain clusters with low alphas (0.21) or betas (0.15), and others with high alphas (0.55) and betas (0.81). In general, we observe that there are some clusters with a high CAPM alpha or beta, and others with low alphas or betas. While some separation is expected, the results show that the clustering does not merely sort on extreme values of alpha or betas individually but rather identifies some economically meaningful groupings.[24] The previous suggests that the ML approach captures deeper patterns in return characteristics, and the observed dispersion is not solely a mechanical artifact.

Overall, our main findings are robust to the different asset pricing models we use. However, we also use a battery of robustness checks. First, we consider the time variability of the coefficients and consider the whole period, excluding the two crisis periods (subprime and pandemic). Second, we consider structural breaks over the sample period. Third, we check the stability of our

[24]In general, hedge funds provide positive alphas to investors and have some market exposure (even those that are so called market neutral funds – e.g. see Duarte *et al.*, 2007, among others).

classification procedure. Fourth, we replicate the whole analysis using the raw returns of the funds. Segmenting the time periods and incorporating additional fund characteristics alters both the number and composition of the funds. As a result, the datasets are substantively different and should be treated as distinct. Our results verify our main findings concerning *H1* and *H2*. Online Appendices A to C present our detailed results.

Managerial implications. The classification of HFs based either on performance (return) features or on reported broad strategies has implications for the portfolio construction process when using HFs as portfolio diversifiers (see Platanakis, Sakkas and Sutcliffe, 2019; Newton *et al.*, 2021) and when dealing with different client profiles. We show that, especially for portfolio construction classification studies (e.g. Chen et al., 2021), classification matters when assessing the likely future performance of stand-alone HFs, and when HFs are used as portfolio diversifiers. Our analysis using RF and SVM has found that the classification of HFs using their past performance (mean, variance, skewness and kurtosis of returns) differs from their reported HF classifications, contradicting Hypothesis 1.

We also examine the potential impact of HF classification on managerial decisions by examining the abnormal returns and factor exposures of HFs classified by their performance, and by HF databases. Using the FF4 model, there are four clusters in Table 8 with the same three significant coefficients (constant, Market and SMB) as *Equity, Event Driven* and *Multi-strategy* in Table 5. This leaves seven strategies in Table 5, which do not match any group in Table 8 in terms of their significant variables. A portfolio manager looking to invest in an HF that is sensitive to a particular set of factors would probably make a different decision if they use past performance, rather than the reported strategy. For the FF5 model, only the *Equity, Event Driven and Multi-Strategy* in Table 6 have the same three significant coefficients (constant, Market and SMB) as clusters 1, 5 and 7 in Table 9. Again, this gives seven strategies in Table 6 that do not match any of those in Table 9. For the FH7 model, as presented in Table 7, *Equity* and *Multi-Strategy* have the same statistically significant coefficients (constant, market, SIZESPR and ΔCRSPR) as cluster 7 in Table 10. The *Others* strategy and cluster 9 have the same three significant coefficients (constant, market and ΔCRSPR). Six strategies in Table 7 do not match with any of those in Table 10.

Across Tables 5 –10, regardless of the asset pricing model, there is evidence that HF indices formed using reported strategies have different performance and factor sensitivities from those formed using ML using past performance. This does not support Hypothesis 2 that reported strategies are a good guide to performance.

Conclusions

We investigate whether the HF classifications used by databases produce strategy classes that are homogenous in terms of risk and return. We use three ML methods to test whether the reported HF classifications correspond to classifications based on HF performance. We find considerable differences between the rival classifications, with three-quarters of HFs assigned to a different strategy. This suggests that the database classifications are not very helpful for investors who are interested in risk and return when building their portfolios.

We also examined the economic significance of our finding of major differences between reported HF classifications and classifications based on performance. We compared the performance of ten HF classifications used by databases with that of our ten clusters formed using *K*-means clustering. We computed the abnormal returns and factor exposures of these two alternative classifications using various asset pricing models. There is evidence that HF indices formed using reported strategies have different market sensitivities from those formed using HF performance. We also find that the market factors SMB and ΔCRSPR remain the most important risk exposures for HFs. Moreover, we find that deviating funds outperform their non-deviating peers.

Future research could study the application of different statistical classification techniques or consider higher moments of HF returns. A limitation is that performance may not be the only criterion for defining a strategy, and the nature of operations (e.g. hedging or investing in ETFs) could be considered by future research. Future research could apply the Lasso (Tibshirani, 1996) and elastic net (Zou and Hastie 2005) techniques for variable selection and regression trees (Breiman *et al.*, 1984); and bagging (Breiman, 1996) to reduce the variance of the predictor when examining HF returns or double/debiased ML (Chernozhukov *et al.*, 2018).

Overall, the classification problem is a major issue not only in investment decisions but also in the broader management-related context; for instance, corporate finance (e.g. when comparing companies). A company's financial profile could be categorized with some objective criteria, such as the mean distance or a classification algorithm similar to our work that considers higher moments. Finally, the issue of different classifications of the same company has been successfully applied in other areas, such as business strategy classification (e.g. Hill, 1988; Kald et al., 2002; Landini, Arrighetti and Bartoloni, 2020).

Conflicts of interest

The authors declare no conflicts of interest.

References

Ackermann, C., R. McEnally and D. Ravenscraft (1999). 'The performance of hedge funds: risk, return and incentives', *Journal of Finance*, **54** pp. 833–874.

Agarwal, V., E. Arisoy and N. Naik (2017). 'Volatility of aggregate volatility and hedge fund returns', *Journal of Financial Economics*, **125**, pp. 491–510.

Agarwal, V., N. Daniel and N. Naik (2003). 'Flows, performance, and managerial incentives in hedge funds', Working Paper [Online], accessed 24 January 2022, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=424369 .

Agarwal V., N. D. Daniel, N. Y. Naik (2009). 'Role of management incentive and discretion in hedge fund performance', *Journal of Finance*, **64** pp. 2221–2256.

Agarwal, V., N. Daniel and N. Naik (2011). 'Do hedge funds manage their reported returns?', *Review of Financial Studies*, **24** pp. 3281–3320.

Amenc, N. and L. Martellini (2003). 'The Alpha and Omega of Hedge Fund Performance Measurement', [Online]Accessed 21 July 2022, https://risk.edhec.edu/sites/risk/files/the_alpha_and_omega.pdf .

Asness, C., R. Krail and J. Liew (2001). 'Do hedge funds hedge?', *Journal of Portfolio Management*, **28** pp. 6–19.

Avramov, D., L. Barras and R. Kosowski (2013). 'Hedge fund return predictability under the magnifying glass', *Journal of Financial and Quantitative Analysis*, **48**, pp. 1057–1083.

Baibing, L., L. Ji and T. Kai-Hong (2017). 'The market liquidity timing skills of debt-oriented hedge funds', *European Financial Management*, **23** pp. 32–54.

Bali, T., S. Brown and M. Caglayan (2011). 'Do Hedge Funds' Exposures to Risk Factors Predict Their Future Returns?', *Journal of Financial Economics*, **101**, pp. 36–68.

Bali, T., S. Brown and M. Caglayan (2012). 'Systematic risk and the cross section of hedge fund returns', *Journal of Financial Economics*, **106** pp. 114–131.

Bali, T., S. Brown and M. Caglayan (2014). 'Macroeconomic risk and hedge fund returns', *Journal of Financial Economics*, **114**, pp. 1–19.

Baquero, G., J. Ter Horst and M. Verbeek (2005). 'Survival, look-ahead bias, and persistence in hedge fund performance', *Journal of Financial and Quantitative Analysis*, **40** pp. 493–517.

BarclayHedge (2024). 'Assets Under Management Report', assessed 20 October 2024. https://www.barclayhedge.com/solutions/assets-under-management/.

Barès P.-A., R. Gibson and S. Gyger (2003). 'Performance in the hedge funds industry: an analysis of short-and long-term persistence', *The Journal of Alternative Investments*, **6**, pp. 25–41.

Bergmeir, C., R. J. Hyndman and B. Koo (2018). 'A note on the validity of cross-validation for evaluating autoregressive time series prediction', *Computational Statistics & Data Analysis*, **120**, pp. 70–83.

Billio, M., M. Getmansky and L. Pelizzon (2012). 'Dynamic risk exposures in hedge funds', *Computational Statistics and Data Analysis*, **56**, pp. 3517–3532.

Bollen, N. and V. Pool (2009). 'Do hedge fund managers misreport returns? Evidence from the pooled distribution', *Journal of Finance*, **64**, pp. 2257–2288.

Bollen, N. and R. Whaley (2009). 'Hedge fund risk dynamics: implications for performance appraisal', *Journal of Finance*, **64**, pp. 985–1035.

Brav, A., W. Jiang, S. Ma and X. Tian (2018). 'How does hedge fund activism reshape corporate innovation?', *Journal of Financial Economics*, **130** pp. 237–264.

Breiman, L. (1996). 'Bagging predictors.' *Machine Learning'*, **24**, pp. 123–140. https://doi.org/10.1007/BF00058655

Breiman, L., J. Friedman, R. Olshen and C. Stone (1984). 'Construction of trees from a learning sample', in *Classification and Regression Trees', pp.297-317*. New York: Routledge. https://doi.org/10.1201/9781315139470

Brennan, M. and P. Hughes (1991). 'Stock prices and the supply of information', *Journal of Finance*, **46** pp. 1665–1691.

Brittain, W. H. (2001). 'Hedge Funds and the Institutional Investor', *Journal of International Financial Management and Accounting*, **12**, pp. 225–234.

Brooks, C. and H. Kat. (2002). 'The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors', *The Journal of Alternatives Investments*, **5**, pp. 26–44.

Brown, S. and W. Goetzmann (1997). 'Mutual fund styles', *Journal of Financial Economics*, **43**, pp. 373–399.

Brown, S. and W. Goetzmann (2003). 'Hedge funds with style', *Journal of Portfolio Management*, **29** pp. 101–112.

Cai, L. and B. Liang (2012). 'On the dynamics of hedge fund strategies', *The Journal of Alternative Investments*, **14**, pp. 51–68.

Cao, C., Y. Chen, B. Liang and A. W. Lo (2013). 'Can hedge funds time market liquidity?', *Journal of Financial Economics*, **109**, pp. 493–516.

Capocci, D. (2009). 'The persistence in hedge fund performance: extended analysis', *International Journal of Finance and Economics*, **14** pp. 233–255.

Capocci D., G. Hübner (2004). 'Analysis of Hedge Fund Performance', *Journal of Empirical Finance*, **11**, pp. 55–89.

Carhart, M. (1997). 'On persistence in mutual fund performance', *Journal of Finance*, **52**, pp. 57–82.

Cassar, G. and J. Gerakos (2011). 'Hegde funds: pricing controls and the smoothing of self-reported returns', *Review of Financial Studies*, **24** pp. 1698–1734.

Chen Y., B. Han, J. Pan (2021). 'Sentiment trading and hedge fund returns', *Journal of Finance*, **76** pp. 2001–2033.

Chen, Y. and B. Ling (2007). 'Do market timing hedge funds time the market?', *Journal of Financial and Quantitative Analysis*, **42**, pp. 827–856

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, H. Hansen, W. Newey and J. Robins (2018). 'Double/debiased machine learning for treatment and structural parameters', *Econometrics Journal*, **21** pp. C1–C68, https://doi.org/10.1111/ectj.12097

Chevalier, J. and G. Ellison (1999). 'Are some mutual fund managers better than others? Cross-sectional patterns in behaviour and performance', *Journal of Finance*, **54** pp. 875–899.

Clare A., M. Sherman, N. O'Sullivan, J. Gao, S. Zhu (2022). 'Manager characteristics: predicting fund performance', *International Review of Financial Analysis*, **80**, art. 102049.

Cortes, C. and V. Vapnik (1995). 'Support-vector networks', *Machine learning*, **20** pp. 273–297.

Cui, W., J. Yao and S. Satchell (2019). 'Trapped in diversification – another look at the risk of fund of hedge funds', *European Journal of Finance*, **25** pp. 1055–1076.

Cumming, D., N.A. Dai and S Johan (2020). 'Dodd–Franking the hedge funds', *Journal of Banking and Finance*, **119**, 105216.

Cumming, D., D. N., and S. Johan (2013). '*Hedge Fund Structure Regulation and Performance around the World*', New York: Oxford University Press.

Cumming, D., N. Dai and S. Johan (2015). 'Are hedge funds registered in Delaware different?' *Journal of Corporate Finance*, **35**, pp. 232–246.

Dahlquist, M. and P. Soderlind (1999). 'Evaluating portfolio performance with stochastic discount factors', *Journal of Business*, **72** pp. 347–383.

Davis, J. and M. Goadrich (2006). 'The relationship between Precision-Recall and ROC curves'. In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 233–240. New York: Association for Computing Machinery.

DeMiguel, V., J. Gil-Bazo, F. J., Nogales and A. A. Santos (2023). 'Machine learning and fund characteristics help to select mutual funds with positive alpha', *Journal of Financial Economics*, **150** art. 103737.

Denvir, E. and E. Hutson (2006). 'The performance and diversification benefits of funds of hedge funds', *Journal of International Financial Markets Institutions and Money*, **16**, pp. 4–22.

DesJardine, M. R. and R. Durand (2020). 'Disentangling the effects of hedge fund activism on firm financial and social performance', *Strategic Management Journal*, **41** pp. 1054–1082.

Duarte, J., F. Longstaff and F. Yu (2007). 'Risk and return in fixed-income arbitrage: nickels in front of steamroller?' *Review of Financial Studies*, **20** pp. 769–811.

Dybvig, P. H. and S. A. Ross (1985). 'The analytics of performance measurement using a security market line', *Journal of Finance*, **40** pp. 401–416.

Eling, M. (2009). 'Does hedge fund performance persist? Overview and new empirical evidence', *European Financial Management*, **15**, pp. 362–401.

Fama, E. and K. French (2015). 'A five-factor asset pricing model', *Journal of Financial Economics*, **116** pp. 1–22.

Ferland, R. and S. Lalancette (2021). 'Portfolio choices and hedge funds: a disappointment aversion analysis', *European Journal of Finance*, **27** pp. 679–705.

Fung, W. and D. A. Hsieh (1997). 'Empirical Characteristics of Dynamic Trading Strategies: the Case of Hedge Funds', *Review of Financial Studies*, **10**, pp. 275–302.

Fung, W. and D. A. Hsieh (2001). 'The risk in hedge fund strategies: theory and evidence from trend followers', *Review of Financial Studies*, **14**, pp. 313–341.

Fung, W. and D. A. Hsieh (2002a). 'Hedge-fund benchmarks: information context and biases', *Financial Analysts Journal*, **58**, pp. 22–34.

Fung, W. and D. A. Hsieh (2002b). 'The risk in fixed-income hedge fund strategies.' *The Journal of Fixed Income*, **12** pp. 16–27.

Fung, W. and D. A. Hsieh, (2003). 'The risk in hedge fund strategies: alternative alphas and alternative betas'. In Lars Jaeger (ed.) *Managing the Risks of Alternative Investment Strategies*, pp. 72–87. London: Euromoney Institutional Investors PLC.

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

Geltner, D. M. (1993). 'Estimating market values from appraised values without assuming an efficient market', *Journal of Real Estate Research*, **8** pp. 325–345.

Getmansky, M. (2004). 'The life cycle of hedge funds: fund flows, size and performance', *Quarterly Journal of Finance*, **2**, pp. 1–53.

Giannikis, D. and I. Vrontos (2011). 'A Bayesian approach to detecting nonlinear risk exposures in hedge fund strategies', *Journal of Banking and Finance*, **35**, pp. 1399–1414.

Gogas, P., T. Papadimitriou and A. Agrapetidou (2018). 'Forecasting bank failures and stress testing: a machine learning approach'. *International Journal of Forecasting*, **34** pp. 440–455.

Hill, C. (1988). 'Corporate control type, strategy, size, and financial performance', *Journal of Management Studies*, **25** pp. 403–417.

Huang, Y. P. and M. F. Yen (2019). 'A new perspective of performance comparison among machine learning algorithms for financial distress prediction', *Applied Soft Computing*, **83**, art. 105663.

Hubert, L. and P. Arabie (1985). 'Comparing partitions', *Journal of classification*, **2**, pp. 193–218.

Ibbotson, R., P. Chen and K. Zhu (2011). 'The ABC of hedge funds: alphas, betas and costs', *Financial Analysts Journal*, **67** pp. 15–25.

Jagannathan, R., A. Malakhov and D. Novikov (2010). 'Do hot hands exist among hedge fund managers? An empirical evaluation', *Journal of Finance*, **65**, pp. 217–255.

Jawadi, F. and S. Khanniche (2012). 'Modeling hedge fund exposure to risk factors', *Economic Modelling*, **29** pp. 1003–1018.

Jensen, M. and C. Smith (1985). 'Stockholder, manager and creditor interests: applications of agency theory', *Recent Advances in Corporate Finance*, **1** pp. 93–131.

Joenvaara, J., M. Kaupilla, R. Kosowski and P. Tolonen (2019). 'Hedge fund performance: Are stylized facts sensitive to which database one uses?', *Critical Finance Review*, **9**, pp. 271–327.

Joenväärä J., R. Kosowski (2021). 'The effect of regulatory constraints on fund performance: new evidence from UCITS hedge funds', *Review of Finance*, **25**, pp. 189–233.

Kald M., F. Nilsson, B. Rapp (2000). 'On strategy and management control: the importance of classifying the strategy of the business', *British Journal of Management*, **11** pp. 197–212.

Kaniel, R., Z. Lin, M. Pelger and S. Van Nieuwerburgh (2023). 'Machine-learning the skill of mutual fund managers', *Journal of Financial Economics*, **150** pp. 94–138.

Karehnke P., F. De Roon (2022). 'Spanning tests for assets with option-like payoffs: the case of hedge funds', *Management Science*, **66** pp. 5969–5989.

Kellard, N., Y. Millo, J. Simon and O. Engel (2017). 'Close communications: hedge funds, brokers and the emergence of herding', *British Journal of Management*, **28**, pp. 84–101.

Koenig, L. and H. P. Burghof (2022). 'The investment style drift puzzle and risk-taking in venture capital', *Review of Corporate Finance*, **2**, pp. 527–585.

Kosowski, R., N. Naik and M. Teo (2007). 'Do hedge funds deliver alpha? A Bayesian and bootstrap analysis', *Journal of Financial Economics*, **84** pp. 229–264.

Kuvandikov, A., A. Pendleton and M. Goergen (2022). 'Activist hedge funds and takeovers: their effects on employment and performance', *British Journal of Management*, **33** pp. 346–368.

Lambert, R. A. (2001). 'Contracting theory and accounting', *Journal of Accounting and Economics*, **32**, pp. 3–87.

Landini, F., A. Arrighetti and E. Bartoloni (2020). 'The sources of heterogeneity in firm performance: lessons from Italy', *Cambridge Journal of Economics*, **44** pp. 527–558.

Li, H., Y. Xu and X. Zhang (2016). 'Hedge fund evaluation under the stochastic discount factor framework', *Journal of Financial and Quantitative Analysis*, **51** pp. 231–257.

Liang, H., L. Sun and M. Teo (2022). 'Responsible hedge funds', *Review of Finance*, **1** pp. 1–49.

Malkiel, B. and A. Saha (2005). 'Hedge funds: risk and return', *Financial Analysts Journal*, **61** pp. 80–88.

Meligkotsidou, L. and I. Vrontos, (2014). 'Detecting structural breaks in multivariate financial time series: evidence from hedge fund investment strategies', *Journal of Statistical Computation and Simulation*, **84**, pp. 1115–1135.

Nachman, D. and T. Noe (1994). 'Optimal design of securities under asymmetric information', *Review of Financial Studies*, **7** pp. 1–44.

Newton D., E. Platanakis, D. Stafylas, C. Sutcliffe, X. Ye (2021). 'Hedge fund strategies, performance and diversification: a portfolio theory and stochastic discount factor approach', *British Accounting Review*, **53**–, art. 101000.

Noe, T. (1988). 'Capital structure and signalling game equilibria', *Review of Financial Studies*, **1** pp. 331–355.

O' Doherty, M., N. E. Savin and A. Tiwari (2015). 'Evaluating hedge funds with pooled benchmarks', *Management Science*, **62**, pp. 69–89.

Osinga, A. J., M. Schauten and R. Zwinkels (2021). 'Timing is money', *Journal of Empirical Finance*, **62**, pp. 266–281.

Panopoulou, E. and N. Voukelatos (2022). 'Should hedge funds deviate from the benchmark?', *Financial Management*, **1**, pp. 1–29.

Patton, A. (2009). 'Are "market neutral" hedge funds really market neutral?' *Review of Financial Studies*, **22** pp. 2495–2530.

Patton, A. and T. Ramadorai (2013). 'On the high-frequency dynamics of hedge fund risk exposures', *The Journal of Finance*, **68**, pp. 597–635.

Petropoulos, A., V. Siakoulis, E. Stavroulakis and N. E. Vlachogiannakis (2020). 'Predicting bank insolvencies using machine learning techniques', *International Journal of Forecasting*, **36** pp. 1092–1113.

Platanakis, E., A. Sakkas and C. Sutcliffe (2019). 'Harmful diversification: evidence from alternative investments', *British Accounting Review*, **51** pp. 1–23.

Prather, L. and K. Middleton (2006). 'Timing and selectivity of mutual fund managers: an empirical test of the behavioural decision-making theory', *Journal of Empirical Finance*, **13** pp. 249–273.
Racicot, F. É. and R. Théoret (2022). 'Tracking market and non-traditional sources of risks by procyclical and countercyclical hedge fund strategies under extreme scenarios: a nonlinear VAR approach', *Financial Innovation*, **8**, pp. 1–56.
Racicot F.-É.r., R. Théoret, G. N. Gregoriou (2021). 'The response of hedge fund higher moments to macroeconomic and illiquidity shocks', *International Review of Economics and Finance*, **72**, pp. 289–318.
Schneeweis, T., H. Kazemi and G. Martin (2002). 'Understanding hedge fund performance: research issues revisited – Part I', *The Journal of Alternative Investments*, **5**, pp. 6–22.
Schwindler, O. and A. Oehler (2006). 'Style analysis of funds of hedge funds: measurement of asset allocation and style drift', *Quantitative Finance*, pp. 145–169.
Smith, C. and J. Warner (1979). 'On financial contracting: an analysis of bond covenants', *Journal of Financial Economics*, **7** pp. 117–161
Stafylas, D., K. Anderson and M. Uddin (2017). 'Hedge fund index-engineering methodologies: a comparison and demonstration', *Applied Economics*, **50**, pp. 596–612.
Stafylas, D., K. Anderson and M. Uddin (2018). 'Hedge fund performance attribution under various market conditions', *International Review of Financial Analysis*, **56**, pp. 221–237.
Stulz, R. M. (2007). 'Hedge funds: past, present, and future', *Journal of Economic Perspectives*, **21**, pp. 175–194.
Sun, Z., A. Wang and L. Zheng (2012). 'The road less travelled: strategy distinctiveness and hedge fund performance', *Review of Financial Studies*, **25** pp. 96–143.
Tharwat, A. (2020). 'Classification assessment methods', *Applied computing and informatics*, **17** pp. 168–192.
Tibshirani, R. (1996). 'Regression shrinkage and selection via the Lasso', *Journal of the Royal Statistical Society*, Series B (Methodological), **58** pp. 267–288. http://www.jstor.org/stable/2346178
Xiao, F. and Y. Shen (2024). 'Wolves at the door to the unknown: innovation search and hedge fund activism', *Research Policy*, **53** art. 104934.
Zheng, Y., E. Osmer and R. Zhang (2018). 'Sentiment hedging: how hedge funds adjust their exposure to market sentiment', *Journal of Banking and Finance*, **88**, pp. 147–160.
Zou, H. and T. Hastie (2005). 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **67** pp. 301–320.

Emmanouil Platanakis, PhD, is an associate professor of finance at the University of Bath, School of Management and holds a PhD in finance from Henley Business School, University of Reading. Dr. Platanakis' research focuses on portfolio theory, investments, fintech and business education. His research work has been published in a wide variety of leading academic journals such as the *European Journal of Operational Research, British Journal of Management, Tourism Management, Insurance: Mathematics and Economics, International Journal of Forecasting, European Financial Management and European Journal of Finance*, among others.

Dimitrios Stafylas, PhD. He is an assistant professor of finance at the University of York, with industry experience. He has a PhD in finance, an MBA and an MSc in net-centric systems. His research interests lie in empirical finance and corporate finance. He has published in journals such as *European Journal of Finance*, *International Review of Financial Analysis, British Accounting Review, Research in International Business and Finance* and *Applied Economics*. His work has been presented at international academic conferences.

Charles Sutcliffe, PhD. He is an emeritus professor of finance at the ICMA Centre, Henley Business School, University of Reading. He has published in a wide range of refereed journals and is also the author of 11 books. He has acted as a consultant to the Financial Services Authority, the Securities and Investments Board, H.M. Treasury, Cabinet Office, Corporation of London, United Nations, London Stock Exchange and London International Financial Futures and Options Exchange.

Wenke Zhang, PhD. She is a lecturer in finance, banking and accounting at Brunel Business School, Brunel University of London. Her research interests include banking, risk management and machine learning models. She holds a PhD from the University of Bath and a master's degree from the University of Bristol. Her work has been published in prestigious journals such as the *European Journal of Operational Research*. Wenke also serves as a referee for journals including the *International Review of Financial Analysis, European Journal of Finance* and *Review of Quantitative Finance and Accounting*.

Supporting Information

Additional supporting information can be found online in the Supporting Information section at the end of the article.