Evaluating Hedge Funds with Machine Learning-Based Benchmarks*

Tengjia Shu† Ashish Tiwari‡

July 23, 2023

ABSTRACT

The explanatory power of multi-factor models typically used to evaluate hedge fund performance is effectively zero for a sizable number of funds (so-called zero-R^2 funds). In this study, we explore alternative approaches based on several machine learning techniques to benchmark and evaluate individual hedge funds. In general, machine learning algorithms significantly improve the ability to track fund performance, especially for zero-R^2 funds. We find that a Bayesian ensemble-of-trees approach is particularly valuable in this context. The improvement in tracking performance enables more precise estimates of fund alphas, resulting in more accurate identification of superior funds. As a result, the proposed methodologies outperform the traditional multi-factor model in several contexts including real-time fund selection, and fund failure prediction. Our results offer compelling evidence that machine learning-based benchmark models can effectively capture the nonlinearities and interaction effects among risk factors, which are crucial for accurately characterizing the risks associated with hedge fund strategies. To further highlight the benefit of such methods we re-examine the well-established positive relationship between strategy distinctiveness and hedge fund performance. Our findings suggest that the observed relationship stems from benchmark model errors that contaminate inference based on the conventional factor models.

*We thank Nicolas Bollen, Cristian Tiu, Haibei Zhao and seminar participants at Concordia University, Lehigh University and the University of Iowa for their valuable feedback.

†Department of Finance, Tippie College of Business, University of Iowa, Iowa City, IA 52242; *Email: tengjia-shu@uiowa.edu; Ph: (319)335-0973*

‡Department of Finance, Tippie College of Business, University of Iowa, Iowa City, IA 52242; *Email: ashish-tiwari@uiowa.edu; Ph: (319)353-2185*

I. Introduction

Evaluating hedge fund performance is challenging in view of the dynamic nature of their strategies, operational flexibility in terms of asset class exposures, and the limited disclosure requirements. One of the primary challenges lies in determining the appropriate benchmark to be used in this context. Previous studies have made significant progress in addressing this issue. Nevertheless, despite these methodological advancements, the commonly used multi-factor benchmark models employed in hedge fund performance evaluation have a limited ability to accurately capture the risk exposures of a significant portion of these funds.

As Bollen (2013) documents, customized benchmark models based on the 7 factors used in the well-known Fung and Hsieh (2004) model have an R^2 insignificantly different from zero for more than 36% of a large sample of hedge funds. This result actually understates the severity of the problem as the model R^2 is effectively zero for over 50% of the youngest funds with fewer than 36 monthly return observations. Such zero-R^2 funds also have a much larger chance of failing, which is potentially even more worrisome. Motivated by these findings, in this study we explore several alternative approaches that rely on machine learning techniques to develop performance benchmarks for individual hedge funds. Specifically, we consider benchmarks based on Elastic Net, Random Forest, and a Bayesian ensemble-of-trees approach, namely, the Bayesian Additive Regression Trees or BART (Chipman, George, and McCulloch, 2010). As we subsequently describe in more detail, we find that machine learning methods are particularly well-suited to the challenge of benchmarking hedge fund performance.

An appropriate hedge fund benchmark should successfully track the performance of the hedge funds being evaluated. This naturally requires that the benchmark should capture (a) the inherently nonlinear and dynamic nature of hedge fund strategies, and (b) the potential interactions among the various risk factors that influence the strategy outcomes. It turns out that machine learning-based benchmarks perform very well on these dimensions. In addition to the ability to capture non-linear effects and interactions among the risk factors in a non-parametric fashion, the methodologies perform well even for funds with short histories. This is of course particularly relevant for hedge funds given the relatively high attrition rates in the industry.

We use the machine learning methods to model hedge fund returns and identify the benchmarks that optimally capture the risk-return profile of individual hedge funds. The benchmarks are based on the 7 factors in the Fung and Hsieh (2004) model that is widely used in this context. In addition to the Fung and Hsieh (2004) model we consider benchmarks based on three machine learning methods that provide increasing levels of flexibility in modeling hedge fund returns: Elastic Net, Random Forest, and Bayesian Additive Regression Trees (BART). [1] Among the various machine learning methods we consider, we find that the BART model excels in terms of capturing a substantial portion of the return variation in individual hedge fund monthly returns. Specifically, when incorporating the same set of 7 factors as the Fung and Hsieh (2004) model, the average fund R^2 achieved by the BART model reaches 52 percent, while the corresponding average fund R^2 based on the Fung and Hsieh (2004) model is notably lower at 34 percent. We next demonstrate the effectiveness of BART model-implied benchmarks in capturing the risk-return tradeoff for funds that exhibit practically zero R^2 values when evaluated against the Fung and Hsieh (2004) model. Specifically, we show that for such zero-R^2 funds the average R^2 based on the BART model is as high as 35 percent. This finding provides compelling evidence for the improved effectiveness of machine learning methods like BART in capturing the risks associated with funds whose strategies and return profiles are difficult to track using traditional linear factor models.

The increased tracking performance of the machine learning algorithms naturally results in more precise estimates of fund alphas, which should, in principle, improve one's ability to identify funds with superior performance. Accordingly, we next demonstrate the economic value of the machine learning methods in benchmarking and evaluating hedge fund performance. Specifically, we examine the performance of real-time strategies designed to invest in the top decile portfolio of funds based on the different evaluation methods/models considered.[2] We compare the performance of strategies based on the machine learning methods to a fund selection strategy that relies on the 7-factor Fung-Hsieh model. In general, we show that machine learning strategies outperform based on a variety of performance measures including average returns, Sharpe ratios, Fung-Hsieh model alphas, and the manipulation proof performance measure of Goetzmann, Ingersoll, Spiegel, and Welch (2007). Importantly, the superior performance of such strategies is also evident in the form of lower downside risk reflected in statistics like the Sortino ratio, maximum drawdown, and maximum monthly loss.

Next, we show that the proposed machine learning framework is valuable in the context of predicting fund failure. Specifically, consistent with prior literature we adopt the Cox proportional hazards model to predict fund failure. We show that indicators based on the each of the machine learning model-based alphas dominate competing measures, including those based on the widely used (Fung and Hsieh, 2004) model.

Finally, to further illustrate the advantage offered by the machine learning-based

[1]In addition to the elastic net model with the 7 factors, we also consider an extended version of the model that includes two-way interactions among the 7 factors.

[2]We also examine the performance of long-short strategies that invest in the top decile of funds while taking a (hypothetical) short position in the bottom decile of funds.

framework in a performance evaluation context, we re-examine the evidence regarding the well-documented positive relationship between hedge fund strategy distinctiveness and fund performance (see, e.g., Sun, Wang, and Zheng (2012)). Beyond its academic relevance, this result has obvious practical implications for investors' capital allocation decisions. Hence, it naturally serves as an important application for an improved performance evaluation methodology. Focusing on the Fung-Hsieh alphas of funds sorted by a measure of strategy distinctiveness confirms the documented positive relation. However, we show that the Fung and Hsieh (2004) model does a relatively poor job of characterizing the risks inherent in fund strategies whose returns appear to be distinctive or unique relative to their peers. Assessing fund alphas using the various machine learning model benchmarks we find no evidence of a relation between fund strategy distinctiveness and future performance. We conclude that the link between strategy distinctiveness and fund performance is largely an artifact of benchmark model error.

Collectively, our results serve to confirm the value of machine learning-based benchmarks in multiple contexts. In particular, they offer powerful evidence of the ability of machine learning methods to capture the non-linearities and interaction effects among risk factors that are important in fully characterizing the risks inherent in hedge fund strategies.

The rest of the paper proceeds as follows. Section II reviews the related literature while Section III describes the general methodological framework for hedge fund performance evaluation, including the use of machine learning methods in this context. Section IV presents the details of the machine learning algorithms we consider. Section V describes the data and Section VI presents the empirical results. Section VII re-examines the evidence on the relation between strategy distinctiveness and hedge fund performance, while Section VIII concludes.

II. Related Literature and Contribution

Our paper is related to the literature on hedge fund performance evaluation. Fung and Hsieh (1997) document the low correlation between the hedge fund returns and the traditional asset classes and propose three additional "style" factors to extend the Sharpe (1992) model for evaluating hedge fund performance. Motivated by this important early work subsequent studies attempt to improve the benchmark performance evaluation model by incorporating additional factors in the model. For example, Fung and Hsieh (2001) create "style" factors designed to mimic the characteristics of trend-following strategies in a number of asset classes. Similarly, building on the theoretical framework of Glosten and Jagannathan (1994), Agarwal and Naik (2004) propose a multi-factor model that incorporates option-based risk factors. The identification of the relevant risk factors is typically based on statistical techniques like factor analysis (e.g., Fung and Hsieh (1997)) or stepwise regression (e.g., Liang (1999), Fung and Hsieh (2000), Agarwal and Naik (2004), Bollen and Whaley (2009)).

A number of studies allow for variation in factor exposures over time as a way to improve model performance. For example, Bollen and Whaley (2009) employ an optimal changepoint regression framework which yields substantial improvements in performance over a model with constant parameters. Patton and Ramadorai (2013) model the dynamics of hedge funds' factor exposures using high frequency conditioning variables. They show that using daily conditioning information to model hedge fund risk exposures results in significantly improved model performance in terms of adjusted R^2 relative to a model with monthly conditioning information.

As shown by Bollen (2013), the traditional multi-factor modeling framework fails to account for the risk exposures of a substantial proportion of hedge funds. As a result, the seemingly high 'alpha' estimates for such funds may in fact represent compensation for exposure to "hidden" risk factors, rather than managerial skill. Motivated by this finding, and in the spirit of Box (1980), our paper recognizes that the earlier modeling innovations in the literature are useful but imperfect, in the sense that they do not fully capture the relevant risk exposures of hedge fund strategies. Accordingly, our paper fits in a growing literature related to benchmark misspecification in fund performance evaluation.

A number of recent papers explicitly recognize that all hedge fund models are misspecified and adopt different approaches to address this issue. For example, O'Doherty, Savin, and Tiwari (2016) propose a model pooling framework in which the benchmark model represents a weighted combination of the predictive return densities implied by several linear factor models. They show that the resulting model pooling-based benchmark model yields significant improvements in terms of fund selection and fund failure prediction. Ardia, Barras, Gagliardini, and Scaillet (2022) propose a framework that evaluates candidate models by their ability to capture the returns to hedge fund strategies. Specifically, the approach involves formally testing whether the models deliver similar alphas as the CAPM, a model which is of course not equipped to capture returns to alternative hedge fund strategies. Giglio, Liao, and Xiu (2021) develop a multiple testing procedure for false discovery rate (FDR) control that is robust to omitted factors in the benchmark model and missing returns data. They show that hedge fund returns display substantial correlations even after controlling for exposure to the standard Fung-Hsieh factors. By comparison to other FDR control methodologies, their methodology allows for the identification of a larger number of "skilled" funds (subject to a 5% false discovery rate) with superior out-of-sample performance.

Our paper contributes to this literature by adopting a Bayesian machine learning framework for benchmarking hedge fund performance. To our knowledge, this is the first study to implement a machine learning approach to develop customized benchmarks for hedge fund performance evaluation.[3] As we discuss in more detail in the next section, the methodology offers a number of advantages over conventional methods. The nonparametric and flexible nature of the methodology allows us to efficiently capture nonlinear effects and potentially high-order interactions among the relevant factors. Equally important, the resulting performance benchmarks are successful in capturing the performance of hedge funds with short return histories. This feature of the methodology is of course particularly important in the context of hedge funds.

III. Methodology

The objective of this paper is to develop a framework for performance evaluation of individual hedge funds that addresses the shortcomings of the existing approaches based on the use of linear factor models as benchmarks. Consistent with the typical performance attribution framework our main goal is to decompose the fund return into two parts - a portion that derives from exposure to a set of passive factors, and another that represents the benefits of active management. The latter, component, termed "alpha", is conventionally viewed as representing managerial skill.

Identifying a benchmark to properly capture the investing style of a hedge fund is challenging. First, hedge fund investment strategies are typically proprietary and not subject to any public disclosure requirements. Second, the prior literature has identified a common set of factors based on a general understanding of the nature of hedge fund strategies and the associated payoffs. Using a benchmark factor model with a common set of factors for all hedge funds can be problematic, as investing styles may differ considerably across funds even within a particular category, and the style followed by a fund may change over time. The challenge is to specify the proper set of factors for constructing the benchmark that closely mimics the systematic risk the fund is exposed to. Third, the factors identified in the prior literature may be correlated with each other, and fund returns may not always be linearly related to those established factors. A linear benchmark model only captures the fund's exposure to a linear combination of the systematic risk factors. The potentially missing non-linear features are likely to be important in characterizing a hedge fund's investing style.

[3] Wu, Chen, Yang, and Tindall (2021) apply machine learning methods to predict hedge fund returns. However, the focus of their study differs from ours as they are primarily interested in forecasting hedge fund returns.

To address the aforementioned concerns, we use several alternative machine learning algorithms to build flexible performance benchmarks for each hedge fund. This approach offers several advantages over the traditional linear model-based benchmark based on a common set of systematic risk factors. First, the machine learning-based benchmark models offer additional flexibility in terms of modeling hedge fund returns that typically feature nonlinear payoffs. Second, the tree-based machine learning methods that we consider can readily capture the non-linearities, and potentially high-order interaction effects among factors. This aspect of the methodology is particularly important in the context of factors with overlapping information that are hard to disentangle using a linear model.[4] Moreover, the machine learning framework is well-suited to identifying a sparse set of factors, out of the full set, which can help summarize a fund's exposure to systematic risks in a more efficient manner.

A. Fund Performance Evaluation: Conventional Approach

We start by reviewing the conventional approach to assessing fund performance by estimating its net alpha. Following Sharpe (1992), the conventional approach aims to decompose the fund's performance into a portion that is attributable to the fund's exposure to systematic risk factors, and a portion that is unrelated to sources of systematic risk. The latter is termed the 'alpha' and is indeed the key selling point of 'hedge' funds, that in principle, promise to deliver a return that is uncorrelated with traditional sources of risk. Consider a standard candidate factor model that characterizes the fund's exposure to the investing style factors. In the typical linear benchmark model framework, the conditional expectation of fund excess returns is assumed to be linear in the style (risk) factors. The net alpha $\alpha^*_{i,t}$ of fund i at time t is:

$$\alpha^*_{i,t} = E[r_{i,t}] - \beta^*_{i,t} E[f_{i,t}] \tag{1}$$

where $r_{i,t}$ is the excess return of the fund i, $f_{i,t}$ is a vector of benchmark factor realizations assigned to fund i, and $\beta^*_{i,t}$ is a vector of risk exposures for fund i to the factor vector $f_{i,t}$.

If we know the true model describing the return-generating process, then we can estimate the net alpha from the linear regression:

$$r_{i,t} = \alpha^*_{i,t} + \beta^*_{i,t} f_{i,t} + \epsilon^*_{i,t}, \quad \epsilon^*_{i,t} \overset{\text{iid}}{\sim} N(0, \sigma^2). \tag{2}$$

[4]In a recent study Shu and Tiwari (2022) demonstrate that these features are particularly valuable in an asset pricing context. In particular, their results highlight the advantage offered by the BART framework relative to traditional machine learning methods in addressing the challenge posed by the 'zoo' of factors that appear to be related to the cross-sectional variation in expected returns. For a more extensive recent review of the BART methodology, please refer to Hill, Linero, and Murray (2020)

A key part of the performance evaluation challenge is to identify the relevant factors, and estimate the factor exposures, $\beta^*_{i,t}$. The standard procedure, if we allow for the exposure $\beta^*_{i,t}$ to update over time, is to estimate $\hat{\beta}_{i,t}$ from a regression of the fund excess return on the benchmark model factors over the period from $t-s$ to $t-1$:

$$r_{i,\tau} = \alpha_{i,\tau} + \beta_{i,t} f_{i,\tau} + u_{i,\tau}, \quad \tau = t-s, ..., t-1. \tag{3}$$

The vector of estimated risk exposures, $\hat{\beta}_{i,t}$, can then be used to estimate the fund's net alpha:

$$\hat{\alpha}_{i,t} = r_{i,t} - \hat{\beta}_{i,t} f_{i,t}. \tag{4}$$

How close $\alpha_{i,t}$ is to $\alpha^*_{i,t}$ depends on the accuracy of the specified benchmark factor model set, $f_{i,t}$, in terms of tracking the fund's investing style.

As previously mentioned, a linear benchmark model with a fixed set of benchmark factors, $f_{i,t}$, common to all funds, may not be able to accurately capture the investing style of a specific fund. Furthermore, a linear factor model may be inappropriate given the potential nonlinearities and interaction effects among the factors, leading to significant bias in estimates of the factor exposures, $\hat{\beta}_{i,t}$, and thereby contaminating inference about the fund's alpha. In response to these issues, we consider several machine learning-based methods in order to better track the investing style for individual funds.

B. Fund Performance Evaluation: A Machine Learning-Based Approach

We now describe an alternative framework for fund performance evaluation based on competing machine learning models that rely on the identical set of factors used in the linear benchmark model.

To motivate a machine learning-based assessment of fund performance, consider a more general model to characterize hedge fund excess returns:

$$r_{i,t} = h^*(\alpha^*_{i,t}; \beta^*_{i,t}, f_{i,t}) + \epsilon_{i,t} \tag{5}$$

where $h^*(\alpha^*_{i,t}; \beta^*_{i,t}, f_{i,t})$ is a function consisting of two components. One component is the net alpha $\alpha^*_{i,t}$ that measures the value added by the hedge fund management team relative to a benchmark model. The other component is related to the factor attributes that reflect the funds risk exposures $\beta^*_{i,t}$, to the systematic risk factors, $f_{i,t}$. Notice that Equation (5) links the factor exposures, $\beta^*_{i,t}$, and the vector of factors, $f_{i,t}$ with a general functional form $h^*(\cdot)$ that allows for nonlinear terms of the systematic risk factors, which is a more flexible specification than Equation (1). It is also worth noting that the vector of benchmark factors $f_{i,t}$ is fund-specific, which contains up to k factors that mirror fund i's investment style.

The practical issues when applying the general model in Equation (5) are two-fold. First, without prior knowledge we cannot specify the exact functional form involving the factor terms ($f_{i,t}, f_{i,t}^2, f_{i,k_1,t} \times f_{i,k_2,t}$, etc.). Second, $\alpha_{i,t}^*$, and $\beta_{i,t}^*$ cannot be directly observed or extracted from the function $h^*(\alpha_{i,t}^*; \beta_{i,t}^*, f_{i,t})$. Multiple machine learning models such as tree-based methods, neural networks, etc., can capture the nonlinear effects and higher-order interactions among the risk factors $f_{i,t}$. However, the set of factors, $f_{i,t}$, that is specific to fund i is hard to identify based on the aforementioned machine learning methods. As for $\alpha_{i,t}^*$ and $\beta_{i,t}^*$, many machine learning models are not easily interpretable which makes it difficult to estimate $\alpha_{i,t}^*$ and $\beta_{i,t}^*$. Below we describe our proposed approach to addressing these issues in the context of several popular machine learning algorithms.

B.1 Penalized Linear Regression Framework

First, consider modeling hedge fund returns via a penalized linear regression model, for example, Elastic Net. We expand the linear benchmark model to allow for certain nonlinear terms involving two way interactions among the factors (e.g., $f_{i,k_1,t} \times f_{i,k_2,t}$) and higher order terms (e.g., $f_{i,t}^2$). Then, Equation (5) becomes a more general linear benchmark model as follows:

$$r_{i,t} = \alpha_{i,t}^* + \beta_{i,t}^* f_{i,t} + \lambda_{i,t}^* g_{i,t} + \Omega_{i,t} + \epsilon_{i,t}^*, \quad \epsilon_{i,t}^* \overset{\text{iid}}{\sim} N(0, \sigma^2), \tag{6}$$

where $\beta_{i,t}^* f_{i,t}$ represents the linear combination of the product of risk exposures $\beta_{i,t}^*$ and the original factors $f_{i,t}$, and $\lambda_{i,t}^* g_{i,t}$ represents the linear combination of the product of the risk exposures ($\gamma_{i,t}^*$) and the nonlinear extensions of the risk factors $g_{i,t}$. To guard against the problem of model over-fitting and the curse of dimensionality, we incorporate the penalty terms, $\Omega_{i,t}$, in the above model. Another advantage of introducing $\Omega_{i,t}$ is that it helps identify the terms in $f_{i,t}$ and $g_{i,t}$ that are genuinely related to the returns of fund i.

B.2 Non-parametric Regression Framework

Next, consider a more general, non-parametric machine learning model such as a decision-tree based model. In this case the evaluation of hedge fund performance proceeds in three steps. In the initial step, we approximate $h^*(\alpha_{i,t}^*; \beta_{i,t}^*, f_{i,t})$ with a machine learning method h:

$$r_{i,t} = h(f_{i,t}) + \epsilon_{i,t}, \ \epsilon_{i,t} \sim N(0, \sigma^2), \tag{7}$$

from which we can obtain the estimated fund return $\hat{r}_{i,t}$ as

$$\hat{r}_{i,t} = h(f_{i,t}). \tag{8}$$

In the second step, we approximate the original machine learning model with a more interpretable proxy linear model that can be used for performance attribution analysis. The main aim of this step is to obtain an estimate of a fund's factor risk exposures, which sets the stage for computing the fund's alpha. Specifically, we consider a linear projection of a fund's model-implied estimated return, $\hat{r}_{i,t}$, on the factor payoff space, and estimate the in-sample coefficient estimates $\hat{\beta}_{i,t}$ based on the following objective function:

$$\hat{\beta}_{i,t} = \arg\min_{\beta_{i,t}} \frac{1}{T} \sum_{\tau=t-s}^{t-1} \left[\hat{r}_{i,\tau} - \beta_{i,\tau} g(f_{i,\tau})\right]^2 + \Omega_{i,\tau}. \tag{9}$$

In the above equation, $\frac{1}{T}\sum_{t=1}^{T}\left[\hat{r}_{i,t} - \beta_{i,t} g(f_{i,t})\right]^2$ represents the mean squared error or deviation between estimates of the fund return as per the original machine learning model and the estimates based on the proxy model. The function $g(f_{i,t})$ incorporates the original seven factors, $f_{i,t}$, and nonlinear terms involving the factors.[5]. Similar to (6), we also incorporate the penalty terms $\Omega_{i,t}$.

In the third, and final step, we can estimate the net alpha at time t using data from $t-s$ to $t-1$ as[6]:

$$\hat{\alpha}_{i,t} = r_{i,t} - \hat{\beta}_{i,t} g(f_{i,t}) \tag{10}$$

where $f_{i,t}$ contains the risk factor related terms with nonzero fund exposures, $\hat{\beta}_{i,t}$.

In the next section, we describe in more detail the specific modelling approaches that we adopt for evaluating hedge fund performance.

IV. Models

A. Conventional Approach – Fung and Hsieh (2004) 7-Factor Model

We consider the widely used Fung and Hsieh (2004) 7-factor model as the primary linear benchmark model for performance attribution analysis (Equation (1)-(4)). The seven factors used in the Fung and Hsieh (2004) model include i) the market excess return (MKT), ii) the return on the equity size factor (SMB), iii) the Barclays Capital 7–10 year Treasury Index return in excess of the risk-free rate ($D10YR$), iv) the return spread of the Barclays Corporate

[5]For example, interactions among the factors ($f_{i,k_1,t} \times f_{i,k_2,t}$, etc.), or higher powers of the factors, e.g., $f_{i,t}^2$

[6]In the empirical analysis, we set $s = 24$.

Bond Baa Index over the 7–10 year Treasury Index ($DSPRD$), and the Fung and Hsieh (2001) factors representing Primitive Trend Following Strategies for v) bonds ($PTFSBD$), vi) foreign exchange rates ($PTFSFX$), and vii) commodities ($PTFSCOM$).

B. Machine Learning Approach I – Elastic Net Regression Model

As specific examples of a machine learning-based benchmark model in Equation (6), we consider two versions of the Elastic Net regression models. In one version, we only retain the 7 factors in the Fung and Hsieh (2004) model (and do not include any nonlinear terms of $f_{i,t}$), which we denote as *ENet*. To identify the truly relevant factor terms in $g_{i,t}$ for fund i, as well as to guard against the overfitting problem, we introduce L^1 and L^2 penalty terms – $\gamma_1|\beta_i|$ and $\gamma_2\beta^{'}_{i,t}\beta_{i,t}$, respectively. The resulting coefficient estimates $\hat{\beta}_{i,t}$ are optimized to approximate as closely as possible the machine learning model implied fund return estimations conditional on the benchmark factors $g(f_{i,t})$. Given the L^2-norm based penalty term $\gamma_2\beta^{'}_i\beta_{i,t}$, factors with greater contribution to characterizing the investing style of the fund are assigned larger values. The L^1-norm based penalty term $\gamma_1|\beta_{i,t}|$ limits the factor vector $f_{i,t}$ by zeroing out the coefficients for redundant factors.

In the other version, we add all 28 two-way interactions among the 7 factors,[7] which we denote as *ENet**.[8] These additional terms are designed to capture potential nonlinear payoffs such as market timing strategies. In both versions of the Elastic Net model, $\Omega_{i,t}$ contains L^1 and L^2 penalty terms for $\beta_{i,t}$ (and $\gamma_{i,t}$).

Therefore, the generalized equation in (6) can be re-expressed as follows:

$$r_{i,t} = \alpha^*_{i,t} + \beta^*_{i,t} f_{i,t} + \lambda^*_{i,t} g_{i,t} + \gamma_1|\beta_{i,t}| + \gamma_2\beta^{'}_{i,t}\beta_{i,t} + \epsilon^*_{i,t}, \quad \epsilon^*_{i,t} \overset{\text{iid}}{\sim} N(0,\sigma^2), \tag{11}$$

where $\lambda^*_{i,t} g_{i,t}$ is a nonzero term for *ENet** and a zero term for *ENet*.

The period t estimate of fund alpha based on *ENet* and *ENet** can be derived as

$$\hat{\alpha}_{i,t} = r_{i,t} - \hat{\beta}_{i,t} f_{i,t}, \tag{12}$$

where $\hat{\beta}_{i,t}$ is estimated from (11).

[7]The interaction terms include i) 21 interaction terms between any two of the 7 factors, and ii) the power terms of the 7 factors themselves.

[8]Technically, the model can contain many more nonlinear forms of the factors, but to keep the model interpretable and compact, we only consider the interactions as the representative nonlinear structure in this paper.

C. Machine Learning Approach II – Bayesian Additive Regression Trees and Random Forest

As specific examples of the more general machine learning-based benchmark model described earlier (Equation (7)-(10)), we consider two ensemble-of-trees models, namely, Bayesian Additive Regression Trees ($BART$) and Random Forest (RF).

C.1 Bayesian Additive Regression Trees (BART)

In the BART specification, we consider approximating $h^*(\alpha^*_{i,t}; \beta^*_{i,t}, f_{i,t})$ with a sum of m regression trees $h^*_i(\cdot) \approx h_i(\cdot) \equiv \sum_{j=1}^{m} g_{i_j}(\cdot)$, where each g_{i_j} denotes a regression tree:

$$r_{i,t} = \sum_{j=1}^{m} h_{i_j}(f_{i,t}) + \epsilon_{i,t},\ \epsilon_{i,t} \sim N(0, \sigma^2). \tag{13}$$

The BART model consists of two parts: a sum-of-trees model and a regularization prior on the parameters of that model. Let T denote the tree structure consisting of a set of interior nodes representing binary split decision rules of the form $\{f_{i,t} \leq c\}$ vs. $\{f_{i,t} > c\}$ for continuous $f_{i,t}$, and a set of terminal nodes. Further, let $M = \{\mu_1, \mu_2, ..., \mu_b\}$ denote a set of parameter values associated with each of the b terminal nodes of T. Each value of $f_{i,t}$ is associated with a single terminal node of T by the sequence of decision rules from top to bottom, and is then assigned the μ_i value associated with this terminal node. Thus,

$$r_{i,t} = h_i(f_{i,t}; T, M) + \epsilon_{i,t},\ \epsilon_{i,t} \sim N(0, \sigma_i^2) \tag{14}$$

is a single-tree model of the form considered by Chipman, George, and McCulloch (1998). Under Equation (14), the conditional mean of $r_{i,t}$ given $f_{i,t}$, $E(r_{i,t}|f_{i,t})$, equals the terminal node parameter μ_i assigned by the regression tree function, $h_i(f_{i,t}; T, M)$. With this notation, the sum-of-trees model in Equation (13) can be more explicitly expressed as

$$r_{i,t} = \sum_{j=1}^{m} h_i(f_{i,t}; T_j, M_j) + \epsilon_{i,t},\ \epsilon_{i,t} \sim N(0, \sigma_i^2). \tag{15}$$

Under Equation (15), $E(r_{i,t}|f_{i,t})$ equals the sum of all the terminal node values ($\mu_{i,j}$'s) assigned by the regression functions, $h_i(f_{i,t}; T_j, M_j)$'s. Furthermore, each $\mu_{i,j}$ represents a main effect when $h_i(f_{i,t}; T_j, M_j)$ depends on only one component of $f_{i,t}$ (i.e., a single variable), and represents an interaction effect when $h_i(f_{i,t}; T_j, M_j)$ depends on more than one component of $f_{i,t}$ (i.e., more than one variable). Thus, the sum-of-trees model can incorporate both nonlinear main effects and interaction effects. And, because Equation (15) may be based on trees of varying sizes, the interaction effects may be of varying orders. Importantly, unlike many other flexible models, BART does not require the researcher to explicitly specify the main effect and the potentially large set of interaction effects.

The BART model specification is completed by imposing a prior over all the parameters of the sum-of-trees model, namely, $(T_1, M_1), ..., (T_m, M_m)$ and σ. Specifically, the following serves as the regularization prior as it ensures that the contribution of each component of the model to the overall fit is small:

$$\begin{aligned} p((T_1, M_1), \cdots (T_m, M_m), \sigma) &= \Big[\prod_j p(T_j, M_j)\Big] p(\sigma) \\ &= \Big[\prod_j p(M_j|T_j) p(T_j)\Big] p(\sigma), \end{aligned} \tag{16}$$

and

$$p(M_j|T_j) = \prod_i p(\mu_{ij}|T_j), \tag{17}$$

where $\mu_{ij} \in M_j$. Under such priors, the tree components (T_j, M_j) are independent of each other and of σ, and the terminal node parameters of every tree are independent. The independence restrictions simplify the prior choice problem to the specification of prior forms for just the three key components, $p(T_j)$, $p(\mu_{ij}|T_j)$ and $p(\sigma)$. Following Chipman, George, and McCulloch (2010) we consider identical forms for all $p(T_j)$ and for all $p(\mu_{ij}|T_j)$. We use the default priors suggested by them as described below.[9]

The first element in the prior specification, $p(T_j)$, controls the depth of the nodes within the trees. The prior probability of a node of depth d being non-terminal is $\alpha(1 + d)^{-\beta}$ where $\alpha \in (0, 1)$, and $\beta \in (0, \infty)$. Default values of these hyperparameters recommended by Chipman, George, and McCulloch (2010) are: $\alpha = 0.95$, and $\beta = 2$. These values help regularize the model fit by ensuring that the influence of individual tree on the overall fit is relatively small. Another aspect of $p(T_j)$ concerns the distribution on the assignment of variables to be used for binary split decision rules ($\{f_{i,t} \leq c\}$ vs. $\{f_{i,t} > c\}$) at the interior nodes. We employ the default uniform prior on the candidate variables for this purpose. Next, consider the prior for the terminal node parameters, $\mu_{i,j}$'s. Note that under the sum-of-trees model structure, $E(r_{i,t}|f_{i,t})$ equals the sum of m $\mu_{i,j}$ values. So, the prior specification for $\mu_{i,j}$'s is chosen to ensure that a substantial prior probability is assigned to $E(r_{i,t}|f_{i,t})$ being between the minimum and maximum observed values based on the training dataset. Accordingly, the prior $p(\mu_{ij}|T_j)$ is specified as the conjugate Normal distribution $N(\mu_\mu, \sigma^2_\mu)$. The hyperparameter μ_μ is chosen

[9] In our empirical applications we employ the default priors while implementing BART via the R package *bartMachine* (Kapelner and Bleich, 2016).

as the midpoint of the range of observed values of the target (hedge fund return) variable in the training dataset. And, the hyperparameter σ_μ^2 is empirically chosen to ensure that $\mu_\mu \pm l\sqrt{m}\sigma_\mu$ will cover 95% of the observed values in the training dataset, for $l = 2$. The third component of the prior specification is the prior for the error variance, σ^2, which is specified as an inverse gamma distribution: $\sigma^2 \sim \text{InvGamma}(\nu/2, \nu\lambda/2)$. The hyperparameters ν and λ are chosen via a data-based approach. Specifically, the degrees of freedom parameter ν is set equal to 3, and the parameter λ is chosen such that 90% of the prior probability mass of σ lies below the residual standard deviation from a linear regression model for $E(r_{i,t}|f_{i,t})$. Chipman, George, and McCulloch (2010) confirm that the performance of BART is very robust with respect to hyperparameter choices. Therefore, hyperparameter tuning is not necessary in a particular application, which is useful in a real-world context in which prior beliefs may vary widely across investors. Of course, in addition to a data-driven prior approach, the BART framework allows for user-defined subjective priors, or theory-driven prior specifications.

C.2 Random Forest Regression Model

The Random Forest algorithm (Breiman (2001)) involves building an ensemble (*forest*) of decision trees to predict the outcome of interest. Each tree is trained on a random subset of the data, and features or predictor variables selected via bootstrapping, and the process is repeated many times resulting in a *forest* of trees. The repeated random sampling from the data and the variables helps ensure that the correlation among the trees is minimized. Given a set of B such decision trees, the Random Forests prediction is given by the average of the predictions provided by each of the trees.

C.3 Proxy Linear Model

In order to obtain the hedge fund alphas, we estimate a proxy linear model (described in Equation (9)) that approximates the expected excess hedge fund returns implied by the BART or Random Forest model. For the proxy model in Equation (9), $g_(f_{i,t})$ contains the original 7 factors in Fung and Hsieh (2004), and the 28 two-way interactions among the 7 factors. We set $\Omega_{i,t}$ to be a combination of L^1 and L^2 penalty terms. And $\hat{r}_{i,t}$ is the estimated return from either Random Forest or BART. Therefore, Equation (9) can be rewritten as

$$\hat{\beta}_{i,t} = \arg\min_{\beta_{i,t}} \frac{1}{T} \sum_{\tau=t-s}^{t-1} \left[\hat{r}_{i,\tau} - \beta_{i,\tau} g(f_{i,\tau})\right]^2 + \gamma_1 |\beta_{i,\tau}| + \gamma_2 \beta'_{i,\tau} \beta_{i,\tau}. \tag{18}$$

The estimated $\hat{\alpha_{i,t}}$ based on *BART* or *Random Forest* can then be derived based on Equation (10).

V. Data Description

A. Hedge Fund Data

We obtain data on individual hedge funds from the Lipper TASS database for the period from January 1994 to December 2021. For each sample fund, we collect the monthly returns, the reported trading strategy (*PrimaryCategory*), the currency of reported returns (*CurrencyCode*), information about the fund's use of a high-water mark provision (*HighWaterMark*), personal investment by managers (*PersonalCapital*), leverage (*Leveraged*), and lockup provision (*LockUpPeriod*). We convert all returns to excess returns by subtracting the monthly risk-free rate. We account for illiquidity-related smoothness in hedge fund returns by adjusting the reported returns following the procedure described by Getmansky, Lo, and Makarov (2004) to obtain the "un-smoothed" returns. In constructing the sample, we require fund returns to be denominated in U.S. dollars (i.e., *CurrencyCode* of "USD") and exclude funds that only report returns on a quarterly basis, or report only gross-of-fee returns. To mitigate the impact of backfill bias, we discard the first 24 months of returns for every fund from the time the fund first starts reporting to the database.

We group the sample funds into five broad hedge fund categories: directional funds (i.e., dedicated short bias, emerging markets, global macro, and managed futures), nondirectional funds (convertible arbitrage, equity market neutral, and fixed income arbitrage), semidirectional funds (event driven, long/short equity hedge, and multistrategy), fund of funds and other funds (option strategy, other and undefined). We remove funds that are not in any of the above 14 trading strategies. When conducting the fund-level performance analysis, we apply two selection rules. First, we require the minimum number of monthly return observations to be 61. Second, for most of our empirical analyses we also require a fund to have at least 60 months of consecutive returns. This results in a final sample with 4,202 funds, of which 791 are live and 3,411 are defunct.

Table I reports the summary statistics for the sample funds. The average fund has a mean (median) excess return of 0.42% (0.49%) per month with a standard deviation of 3.84%. The sample funds have negatively skewed returns with thick tails. We also summarize the funds by category. The major categories in our sample are *Semidirectional* (1,736 funds) and *Fund of Funds* (1,239 funds) which together account for over 70% of the sample. In terms of the primary categories, the Long/Short equity hedge fund category is the largest with 1,194 funds. There is considerable variation in the mean returns across different categories. The *Semidirectional* category has the highest mean excess return of 0.53%, while the corresponding figure for the *Fund of Funds* category is 0.21%. Consistent with the prior literature, the returns of funds in the *Nondirectional* and *Other* categories are characterized by relatively high kurtosis, suggesting they may exhibit higher than average losses and gains due to higher levels of gross leverage.

B. Factors

In constructing the various machine learning benchmarks we employ the seven factors from the Fung and Hsieh (2004) model. These include two equity factors downloaded from Kenneth French's website[10], namely, the the excess market return (MKT) and the return on the size factor (SMB). The set of factors also includes two bond market factors, namely, the Barclays Capital 7–10 year Treasury Index return in excess of the risk-free rate, and the return spread of the Barclays Corporate Bond Baa Index over the 7–10 year Treasury Index ($DSPRD$) available at the St. Louis Federal Reserve Bank's website,[11]. The factor set also includes three trend following factors for bonds ($PTFSBD$), foreign exchange rates ($PTFSFX$), and commodities ($PTFSCOM$), obtained from David Hsieh's website.[12] We also collect the risk-free rate from Kenneth French's website to calculate the excess returns of the funds. In the simulation exercise (described below), we also introduce a nonlinear term involving the traded liquidity factor (LIQ) (Pástor and Stambaugh, 2003), available through WRDS.[13]

Table II presents summary statistics for the seven Fung and Hsieh (2004) model factors. There is substantial variation in mean excess returns and standard deviations across the factors. Consistent with prior literature, the trend following factors have the highest standard deviations ranging from 14.00% to 19.11%, followed by the two fixed-income related factors. The mean and median return of the three trend following factors are low and negative, while the returns of the other factors are close to zero.

VI. Empirical Results

A. Simulation Exercise

Before conducting empirical analysis with real data, we first compare the performance of machine learning benchmark models with the linear benchmark Fung and Hsieh (2004) model based on a simulation analysis.

[10]`https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`
[11]`http://research.stlouisfed.org/fred2/`
[12]`https://people.duke.edu/~dah7/HFRFData.htm`
[13]https://wrds-www.wharton.upenn.edu/

To examine each model's performance in terms of the ability to incorporate and track nonlinear effects, we generate fund returns in accordance with a pre-specified model with nonlinear terms. Specifically, we simulate fund returns in accordance with a model that includes two additional terms in addition to the seven factors in the Fung and Hsieh (2004) model. One of these terms is the square of the market excess return (MKT), which captures a fund's potential market timing skill. The other additional term is in the form of the interaction between the market excess return and the liquidity factor from Pástor and Stambaugh (2003). So, for each fund i, we first estimate the following nine-factor ($\tilde{f}_{i,t}$) model

$$r_{i,t} = \alpha_i + \beta_i \tilde{f}_{i,t} + \epsilon_{i,t}, t = 1, ..., T, \tag{19}$$

and obtain the risk exposure estimates $\hat{\beta}_i$, and the residuals $\epsilon_{i,t}$.

We then generate simulated fund returns using a bootstrap exercise. In each iteration, we randomly sample K time points[14] with replacement from the time series of residuals, $\epsilon_{i,t}, t = 1, ..., T_i$, along with the corresponding contemporaneous factors $f_{i,t}$. The simulated fund returns can be expressed as:

$$r^*_{i,t} = \alpha^*_i + \beta_i \tilde{f}_{i,t} + \tilde{\epsilon}_{i,t}, t \in \tau_1, ..., \tau_K, \tag{20}$$

where α^*_i is a pre-specified value. We consider three values for α^*_i: i) zero ($\alpha^* = 0$, ii) the median value of the distribution of the estimated Fung-Hsieh alphas across all funds in the sample ($\alpha^* \approx 0.22\%$), and iii) the 95th percentile of the Fung-Hsieh estimated alphas across all funds ($\alpha^* \approx 0.99\%$).

With the simulated fund returns, we examine the performance of the benchmark models by fitting the models with fund returns along with Fung and Hsieh (2004) seven factors.[15] We document the R^2, estimated $\hat{\alpha}$ and its t-statistic for analysis. For each fund, we repeat the process for M times[16], and compute the mean of R^2 (average R^2), the median value of $\hat{\alpha}$ (median $\hat{\alpha}$), and count the proportion of iterations where $\hat{\alpha}$ is statistically significant based on the significance level of 0.05. For $\alpha^* = 0$, an alpha is considered significant if the p-value of the estimated alpha is lower than the significance level. For $\alpha^* \approx 0.22$ and $\alpha^* \approx 0.99$, an alpha is considered significant if the p-value of the estimated alpha is lower than the cutoff threshold and the estimated alpha is positive.

[14] We set K equal to T_i, the number of observations for the fund in the sample.

[15] Although the simulated returns are generated with liquidity factor, we do not include the liquidity factor when estimating the models. This design allows us to simulate a realistic real-world scenario in which the researcher uses an incomplete model with known parameters to model fund returns that are influenced by potentially unknown risk factors.

[16] In our paper, we set the number of iterations in the bootstrapping process $M = 500$

To highlight potential differences in the ability of different models to track the performance of funds with stronger or weaker linear risk structures, we conduct analysis based on zero-R^2 and nonzero-R^2 fund groups separately. We classify a fund as either a zero-R^2 or nonzero-R^2 fund by comparing its actual Fung and Hsieh (2004) model R^2 with the 95th percentile of the simulated R^2 distribution following Bollen (2013).[17] Within each fund group, we randomly draw 300 funds without replacement and conduct the simulation exercise. We then tabulate and summarize the three performance measures in Table A1.

Panel A of Table A1 reports the median value of the average R^2 across all funds. Consistent with the findings in Bollen (2013), funds classified as zero R^2 have effectively low R^2 on average. More pronounced is the dominance of the machine learning methods that could appropriately capture and incorporate nonlinear effects. In both zero R^2 and nonzero R^2 groups, *ENet**, *RF* and *BART* generate higher average R^2 values, as compared to the Fung and Hsieh (2004) model. It is also worth noting that the machine learning-based models also dominate in terms of the model R^2 values across all three cases with different pre-specified α^* values.

Next, we analyze the quality of the fund alpha estimates, and the Type I and Type II errors for each of the models considered. For each simulation with a pre-specified value of α^*, we first calculate the median estimated alpha for each fund. We then report the cross-sectional median alpha estimated across the funds in Panel B. In a similar vein, Panel C reports proportion of iterations with statistically significant estimated alpha values at the 5 percent level. As is clear from Panel C, when the true value of $\alpha^* = 0.22\%$, the machine learning models tend to have lower Type II error, whereas the Fung and Hsieh (2004) model tends to under- reject the null. The lower power of the Fung and Hsieh (2004) model against this particular alternative is not surprising given the model's inferior tracking performance. The estimated median $\hat{\alpha}$ values reported in Panel B of the table are also effectively close to the true alpha value of 0.22%. When the true value of $\alpha^* = 0$, the machine learning models tends to over-reject the null hypothesis with slightly higher type I error. These results suggest that the machine learning models are more sensitive and effective in detecting the funds with moderate levels of skill. For the case when funds display considerable skill and the true value of $\alpha^* = 0.99\%$, all models tend to have low Type II error rates. This suggests that all models are capable of detecting the funds with substantial skill, i.e., in cases where the departure(s) from the null are large economic terms. However, it is of interest that machine learning models outperform in terms of the

[17] Specifically, for each fund i we collect the realized R_i^2 value based on the Fung and Hsieh (2004) 7-factor benchmark model. We then obtain the 95th percentile critical value of R_i^2 by randomly simulating fund returns from a standard normal distribution under the null hypothesis consistent with the Fung and Hsieh (2004) seven-factor model, following Bollen (2013). We categorize a fund as a zero-R^2 fund if its realized R_i^2 is lower than the critical value, and nonzero-R^2 fund otherwise.

1st quartile of the distribution of the proportion of statistically significant alphas. This result further underscores that machine learning models do a better job at detecting better performing funds.

In summary, the simulation results confirm that flexible machine learning models that are capable of capturing nonlinear effects exhibit superior performance in terms of tracking hedge fund returns. Furthermore, the benefits offered by machine learning models are particularly pronounced when funds have only moderate levels of skill, i.e., in settings characterized by low signal-to-noise ratios.

B. Sample Evidence on Model Tracking Performance and Fund Alphas

In this sub-section we prsent evidence on the performance of the various benchmark models in terms of their ability to track the performance of hedge funds, as judged by the respective model R^2 values. Bollen (2013) documents that 36% of the hedge funds in his sample exhibit effectively zero R^2 based on customized linear factor models based on the Fung and Hsieh (2004) factors. Such zero-R^2 funds have higher alphas compared to non-zero R^2 funds, but they are also more prone to failure. These findings underscore the limitations of traditional benchmarks utilized for assessing hedge fund performance. Additional results documented by Bollen (2013) further highlight the importance of improving existing benchmark models in order to adequately characterize the systematic risks to which hedge funds are exposed.

We first demonstrate the significantly enhanced ability of the machine-learning-implied benchmark models in capturing the systematic risks of hedge funds, based on an analysis similar to Bollen (2013). Our primary findings present the fund R^2 values derived from the Fung and Hsieh (2004) benchmark model and the different machine learning-based benchmark models. Additionally, in the appendix, we present model R^2 values based on fund-specific benchmark models that incorporate a maximum of three out of the seven Fung and Hsieh (2004) model factors. These factors are identified using a stepwise linear regression procedure, as outlined by Bollen (2013).

In order to obtain the model R^2 values we estimate the machine learning models, i.e., *ENet*, *RF*, and *BART*, based on the realized fund returns in conjunction with the seven Fung and Hsieh (2004) factors. In addition, we estimate the *ENet** model, which incorporates the augmented set of factors (the seven risk factors plus the additional 28 two-way interaction terms). The *RF* and *BART* model-implied R_i^2 values are based on a proxy linear model calculated as:

$$R_i^2 = corr(r_{i,t} - \hat{r^*}_{i,t})^2, \tag{21}$$

where $\hat{r^*}_{i,t} = \hat{\beta}_i f_{i,t} + \bar{\alpha}_i = \hat{\beta}_i f_{i,t} + \frac{1}{T}\sum_{t=1}^{T} \hat{\alpha}_{i,t}$, and $\hat{\alpha}_{i,t} = r_{i,t} - \hat{\beta}_i f_{i,t}$ is estimated over the whole sample period from $t = 1$ to $t = T$ based on the fund return $r_{i,t}$ and factors $f_{i,t}$.

B.1 Model R^2

Table III presents the tracking performance of the benchmark models, in terms of the average fund R^2 across all hedge funds in the sample, as well as for funds within a category. The results are presented separately for the zero-R^2 and the nonzero-R^2 funds. Although not shown in the table, 593 out of 4,202 funds are identified as zero-R^2 funds. In the case of the zero-R^2 funds, it it is worth noting that the average R^2 values for the *ENet* and *RF* models are quite low at 7.2 percent and 4.5 percent, respectively, and in fact, slightly lower than the corresponding average Fung-Hsieh model R^2 of 9.3 percent. There are two primary limitations associated with the *ENet* model. One, it is unable to accurately capture nonlinear effects, which hampers its ability to reflect the complex relationships between fund returns and the relevant risk factors. And two, the penalty constraints imposed on (the coefficients of) the seven risk factors in the *ENet* model can potentially lead to model under-fitting when considering the in-sample fit over the entire sample period, compared to the unconstrained Fung and Hsieh (2004) model. By contrast, the more flexible *ENet** model performs better with an overall average R^2 at 17.8 percent, which is nearly twice that of the corresponding figure for the Fung and Hsieh (2004) model.

Although the tree-based *RF* model is in principle capable of capturing nonlinear effects, the model performs poorly in the case of the zero-R^2 funds with average fund R^2 value of only 4.5 percent. On the other hand the theoretical model structure of the *BART* model is more flexible that of *RF*. The BART model clearly dominates the other machine learning models and the Fung and Hsieh (2004) model with an overall average R^2 of 35.0 percent in the case of zero-R^2 funds. Note that this is nearly twice as high as the next best performing model, namely, *ENet**.

In the case of the nonzero-R^2 funds, as expected each of the benchmark models does a better job of tracking the hedge fund returns. The average fund R^2 value for the Fung and Hsieh (2004) model is 37.8 percent which is close to the corresponding figures for *Enet* (36.5 percent) and *RF* (30.2 percent) models. Once again the BART model dominates with an average fund R^2 of 54.4 percent, which significantly outpaces the corresponding figure of 43.5 percent for the second best performing model (*ENet**). It is also worth highlighting that these results remain qualitatively unchanged when we examine th erelative performance of the models across the various fund categories.

Furthermore, we perform the Diebold-Mariano test to assess the model predictions, and the results are summarized in Table A2. The Diebold-Mariano test compares the performance of the machine learning models to that of the Fung and Hsieh (2004) seven-factor model in terms of the models' abilities to fit the realized fund returns. For each fund, we calculate the residuals as the discrepancy between the realized fund returns and the estimated returns obtained from the benchmark models. We then compute the Diebold-Mariano test statistic to determine whether the machine learning models exhibit significantly lower residuals compared to the Fung and Hsieh (2004) 7-factor model. Table A2 reports the proportion of funds with a significant Diebold-Mariano statistic. Panel A of the table presents results for zero-R^2 funds while Panel B presents results for funds designated as nonzero-R^2 funds. Results in both Panel A and Panel B highlight the superiority of the *ENet** and *BART* models over the Fung-Hsieh 7-factor model, as evidenced by lower estimation residuals. In many fund categories, more than 90% of the funds exhibit a significant DM statistic, indicating the superior predictive performance of *ENet** and *BART*. Conversely, there are relatively few funds that demonstrate significantly lower residuals based on the *ENet* and *RF* models. These findings align with the results presented in Table III, reinforcing the consistency of our observations.

As a robustness check, following Bollen (2013), Table A3 presents results based on the optimal fund-specific model that yields the highest realized R^2 values using up to three (out of seven) factors. Following the procedure employed by Bollen (2013), we employ a step-wise regression approach for each fund to select the model with no more than three factors that yields the highest R^2. Next, we compute the 95th percentile of the simulated R^2 distribution under the null hypothesis as the critical value, which is then compared with the realized model R^2. Funds with model R^2 values below the 95th percentile critical value are identified as zero-R^2 funds.

In Table A3, we provide a summary of the average R^2 from both the optimal model and four machine learning models. Notably, based on the optimal fund-specific three-factor model, 690 out of 4,202 funds (16 percent) are classified as zero-R^2 funds. Furthermore, consistent with the results in Table III, the *ENet** and *BART* R^2 outperforms the Fung-Hsieh R^2 for both zero-R^2 and non-zero-R^2 fund groups. The outperformance applies to all fund categories. However, it should be noted that the fit of the *ENet** and *BART* models using up to three factors is less accurate compared to using all seven factors. This suggests that imposing a strict limitation on the number of factors could potentially compromise the overall performance of the models.

In summary, the above findings strongly indicate that machine learning models, which have the ability to capture factor interactions and nonlinear effects, outperform the linear Fung-Hsieh model in effectively capturing the systematic risks inherent in hedge fund strategies.

B.2 Estimates of Fund Alphas

Table IV presents estimates of hedge fund alphas based on the various benchmark models. First, consider the estimates for the zero-R^2 funds presented in Panel A of the table. The monthly alpha estimates range from 0.41 percent based on the Fung and Hsieh (2004) model to 0.51 percent for the BART model. The estimates based on the other 3 machine learning algorithms/models, particularly in the case of *Enet*, are quite close to that for the Fung and Hsieh (2004) model. In general, the results for the various fund categories are qualitatively similar.

Panel B of Table IV presents estimates estimates for the nonzero-R^2 funds. Overall, the monthly alpha estimates range from 0.17 percent based on the BART model to 0.30 percent for *ENet** and *RF* models. Focusing on the category-specific results in Panel B, we note that the Fung and Hsieh (2004) model-based alphas generally exceed estimates based on the BART model, with the sole exception of the *Fund of Funds* category. For example, the Fung-Hsieh average alpha for the *Directional Funds* is 0.30 percent, nearly 4 times the corresponding BART estimate of 0.08 percent. In general, estimates based on other machine learning methods also exceed the BART model-based estimates.

In interpreting the results in Table IV it is useful to recall that estimates of fund alphas are a measure of the average (excess) hedge fund returns that are unrelated to systematic risk factors. Hence, given the superior tracking performance of the BART model documented earlier, it is reasonable to conclude that the BART estimates are a more accurate reflection of fund alphas, i.e., the average excess hedge fund returns that are "skill-related."

C. Explanatory Power of Benchmark Models By Length of Performance Record

Hedge fund databases contain several funds with short-lived performance records due to the industry's notoriously high attrition rate. Additionally, successful hedge firms may voluntarily stop disclosing their results to data vendors. As a result, it is important to evaluate how sensitive a benchmark model's performance is to the duration of a fund's track record. Accordingly, in this subsection, we investigate how the explanatory power of the Fung and Hsieh (2004) model varies in relation to the length of funds' historical record, and compare these findings with the results obtained from the machine learning models. The analysis consistently reveals the superiority of the machine learning framework over the competing linear factor model framework. Notably, the advantage of the machine learning models is even more pronounced for funds with shorter historical records.

Table V presents the average R^2 and $\hat{\alpha}$ values derived from both the Fung and Hsieh (2004) model and the machine learning models. The table also provides the total number of funds categorized by their age groups. Consistent with previous research by Bollen and Whaley (2009), a significant portion of the funds in our dataset exhibit relatively short historical records. Notably, more than three-fourths of the funds have a history shorter than 156 months, while the entire sample period in our study spans 312 months. The median duration of performance track record for the funds in our sample is 76 months. It is evident from the table that that younger funds are more likely to be identified as zero-R^2 funds in comparison to their older counterparts suggesting that the Fung and Hsieh (2004) benchmark model performs poorly in tracking the performance of younger funds. The average R^2 value based on the Fung and Hsieh (2004) model is higher in the younger fund groups and decreases with fund age, for both zero- and nonzero-R^2 funds. The observed trend can be explained by the fact that as the return histories of funds become longer, the critical value of R^2 above (below) which a fund is categorized as a nonzero-R^2 (zero-R^2) fund decreases. Consequently, funds with longer return histories tend to have smaller average R^2 values across both groups.

For zero-R^2 funds, the average Fung-Hsieh R^2 remains consistently close to zero across all age groups. In contrast, nonzero-R^2 funds demonstrate higher average Fung-Hsieh R^2 values, ranging from 0.356 to 0.420. In terms of machine learning models, the *BART* model dominates the other models, and it significantly outperforms the Fung-Hsieh model displaying substantially higher R^2 values across all age groups. This divergence is particularly noticeable for younger funds. For instance, zero-R^2 funds with less than 60 months of return history exhibit an average BART R^2 value of 0.468, a significant improvement compared to the Fung-Hsieh R^2 value of 0.128. Even in the case of zero-R^2 funds with over 180 months of data, the average BART R^2 of 0.149 is still more than three times larger than the average Fung-Hsieh R^2 for the same age group. This pattern persists for nonzero-R^2 funds, with the average BART model R^2 consistently surpassing the average Fung-Hsieh R^2 across all age groups.

To summarize, the results presented in this sub-section confirm that younger funds are more likely to be classified as zero-R^2 funds following the procedure suggested by Bollen (2013). This can be attributed to the limitations of the linear factor model in adequately capturing the risk sources associated with funds with short histories. On the other hand, flexible machine learning models such as *BART* and *ENet** effectively address this issue by incorporating nonlinearities and interaction effects among the risk factors, even when considering a limited set of risk factor candidates. Consequently, these models offer a more accurate representation of the risk exposures of younger funds in particular, resulting in significantly higher R^2 values across funds of all age groups.

D. Trading Strategies

The findings from the previous section demonstrate the superiority of machine learning models in capturing nonlinear risk effects when characterizing hedge fund performance. A more relevant and practical concern for investors is whether these improved models can help them identify skilled fund managers in real time. Accordingly, in this section we assess the value of the machine learning model framework in identifying superior funds through a real-time exercise.

To this end, we investigate the performance of strategies that select funds into portfolios at the beginning of each year, based on their performance evaluated using one of the five benchmark models. The models used for this purpose are as follows: i) the Fung and Hsieh (2004) seven-factor model (*FH*), ii) the Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), iii) Elastic Net with an extended set of factors (the original 7 factors plus all two-way interaction terms involving the factors) (*Enet**), iv) Random Forest (*RF*), and v) Bayesian Additive Regression Trees (*BART*).[18]

D.1 Portfolios Based on Alpha t-statistics

We adopt a sorting method following Fama and French (2010) and O'Doherty, Savin, and Tiwari (2016) to create portfolios of hedge funds at the start of each year. Specifically, in January of each year, we rank the hedge funds into decile portfolios based on the t-statistics associated with their alphas estimated using one of the benchmark models. The funds are retained in the respective decile portfolios during the subsequent 12 months. We then evaluate the performance of the top decile portfolio using various performance measures. Although we also report the performance of the hypothetical long-short hedge portfolio, our analysis focuses primarily on the top decile portfolio since short selling hedge funds is impractical in real-world scenarios.

Turning next to the details of the fund selection process, consider that the t-statistic of a fund's estimated alpha at time t based on the prior d periods (from $\tau = t - d + 1$ to $\tau = t$) is

$$t_{\hat{\alpha},i,t} = \frac{\bar{\alpha}_{i,t}}{\sigma_\alpha/\sqrt{d}}, \tag{22}$$

where $\bar{\alpha}_{i,t}$ is the average fund return in excess of the return attributed to the benchmark factor model:

$$\bar{\alpha}_{i,t} = \frac{1}{d}\sum_{\tau=t-d+1}^{t} \hat{\alpha}_{i,\tau} = \frac{1}{d}\sum_{\tau=t-d+1}^{t} \left[r_{i,\tau} - \hat{\beta}'_{i,\tau} g(f_{i,\tau})\right]. \tag{23}$$

[18] As previously described, in order to estimate fund alphas using the estimates from tree-based algorithms (i.e., *RF* and *BART*) we rely on their corresponding interpretable proxy linear models.

In our analysis, we fix $d = 24$, which means that we compute the t-statistic based on the estimated alpha using data from the prior 24 months. For the Fung and Hsieh (2004) model, $\hat{\beta}_{i,\tau}$ is estimated based on the 7-factor model over a 36-month rolling window. For the machine learning models, we estimate each $\hat{\beta}_{i,\tau}$ over a period of 36 months consisting of a 24-month training period for model fitting, and a 12-month validation period for hyperparameter (i.e., γ_1 and γ_2 in Equations (13) and (18)) tuning. We then compute the estimate, $\hat{\alpha}_i$, and calculate its t-statistic following Equation (22) over the subsequent 24 months. We sort the funds based on the alpha t-statistics and form equal-weighted decile portfolios. We then track the performance of the decile portfolios over the subsequent year.[19]

Table VI presents the performance of the top decile portfolios (*P10*), and the hedge portfolios (*P1-P10*) based on the various benchmark models considered by us. We consider the portfolio performance in terms of the mean return in percent per year, standard deviation, annualized Sharpe ratio, information ratio, maximum portfolio drawdown, and maximum 1-month loss. We also report the annualized 7-factor alpha and its t-statistic, which we compute via the time-series regressions of the portfolio's realized excess returns on the Fung and Hsieh (2004) model factors. The last row in the table reports the manipulation-proof performance measure (MPPM) of Goetzmann, Ingersoll, Spiegel, and Welch (2007) based on the risk aversion level of 3.

To begin, consider the performance of the top decile portfolios labeled *P10.* It is evident that the machine learning models outperform the Fung and Hsieh (2004) model in terms of decile portfolio performance across multiple measures. The top decile portfolio based on the Fung and Hsieh (2004) model earns an average excess return of 4.494% per year. In contrast, the top decile portfolios based on the *ENet*, *ENet**, *RF* and *BART* models achieve average annual excess returns of 5.013%, 5.570%, 5.045% and 5.136%, respectively. The corresponding Sharpe ratios of the top decile portfolios based on the four machine learning models are 1.003, 1.160, 0.971 and 0.972, respectively. All of these Sharpe ratios comfortably exceed the corresponding value of 0.829 for the top decile portfolio based on the Fung and Hsieh (2004) model. Interestingly, the superior performance of the machine learning model-based portfolios is not accompanied by increased downside risk. In fact, the maximum drawdown and maximum 1-month loss figures for these machine learning model-based portfolios are lower compared to the equivalent figures for the top portfolio based on the Fung and Hsieh (2004) model. Additionally, the annualized

[19] For instance, we train the machine learning models with data spanning from January 1997 to December 1998, then validate these models using data from January 1999 to December 1999. At the start of January 2002, we compute the t-statistic of the alpha estimated from January 2000 to December 2001, sort funds into equal-weighted decile portfolios, and track their performance in the subsequent 12 months until December 2002.

Sortino ratio[20] for the top decile portfolio based on the Fung and Hsieh (2004) model is 1.318, whereas the top portfolios based on the four machine learning models deliver Sortino ratios of 1.584,1.901, 1.512 and 1.543, respectively. The reported seven-factor alpha estimates further confirm the efficacy of the machine learning model framework in evaluating fund performance. The top decile portfolios relying on *ENet*, *ENet**, *RF* and *BART* exhibit annualized alphas of 3.420%, 3.932%, 3.349% and3.395% (with corresponding t-statistic of 4.430, 4.430, 4.243 and 5.532, respectively). In economic terms, these alphas comfortably exceed the Fung-Hsieh model top decile portfolio alpha of 2.809%. The MPPM values for the top decile portfolios based on the four machine learning models are 2.921, 3.507, 2.922 and 3.000, respectively. These values all exceed the corresponding MPPM value of 2.334 for the top decile portfolio selected based on the Fung-Hsieh model alphas.

Next, consider the performance of the hypothetical hedge portfolio based on the long-short strategy. The hedge portfolio based on the Fung-Hsieh model exhibits a mean excess return of 0.205%. On the other hand, the long-short hedge portfolio based on machine learning models all have higher mean excess returns of around 1%. The outperformance of the hedge portfolios based on machine learning models is also quite pronounced when considering other performance measures like the Sharpe ratio, and the Sortino ratio. The four machine learning model-based hedge portfolios exhibit superior Sortino ratios of 0.303, 0.680, 0.424 and 0.248, surpassing the Sortino ratio of 0.083 for the Fung-Hsieh model-based hedge portfolio. Additionally, the alphas of the machine learning-based hedge portfolios are significantly higher compared to the Fung-Hsieh model-based hedge portfolio. The downside risk measures, such as the maximum drawdown and maximum 1-month loss, are consistently lower for the machine learning model-based hedge portfolios.

Figure 1 illustrates the cumulative growth of a $1 investment in the top decile portfolios during the period from January 1999 to December 2021. The top decile portfolios are determined based on either the Fung-Hsieh model-based alpha t-statistics, or the four machine learning-based alpha t-statistics. At the end of December 2021, an investor following the top decile strategy based on the *ENet** model would have witnessed their initial investment of $1 grow to $3.31. The cumulative wealth increase based on the investment strategies guided by *ENet*, *RF* and *BART* models is slightly lower, at $2.92, $2.99, and $2.94, respectively. In comparison, an investment in the Fung-Hsieh model-based top decile portfolios would have concluded the sample period with only $2.60. Notably, the growth rate of the Fung-Hsieh model top decile portfolio aligns with the other portfolios until the 2008 financial crisis. It closely follows the machine

[20]The Sortino ratio is calculated as the average excess portfolio return divided by the semi-variance.

learning-based top decile portfolios and even outperforms the four machine learning portfolios before 2002. However, its performance suffers during the post-crisis period and remains sluggish thereafter.

Additionally, we conduct a thorough evaluation of portfolio performance within each hedge fund category, including directional, nondirectional, semidirectional, fund of funds, and the 'other' category. This analysis ensures that the observed results are not driven by funds following a specific trading style. Table A5 presents the results of this analysis.

For directional funds (Panel A in Table A5), the strategy based on the *ENet* model demonstrates superior performance compared to the Fung-Hsieh model across several performance measures. Both the top decile portfolio and the hedge portfolio of *ENet* outperform the corresponding Fung-Hsieh portfolio in terms of mean excess return, Sharpe ratio, Sortino ratio, information ratio, and seven-factor alpha. Conversely, the performance of other machine learning methods falls short in comparison to *ENet* or the Fung and Hsieh (2004) model. These results indicate that the linear machine learning-based benchmark models perform better in characterizing the risk factors associated with directional funds and identifying funds with superior performance. T

When considering semidirectional funds (Panel B in Table A5), all four machine learning-based top decile portfolios outperform the Fung-Hsieh model-based top portfolio. The superiority of machine learning models becomes even more evident within nondirectional funds (Panel C in Table A5) and fund of funds (Panel D in Table A5), as all four machine learning-based top decile and hedge portfolios outperform their corresponding Fung-Hsieh model-based counterpart portfolios. The machine learning-based portfolios consistently outperform the Fung-Hsieh portfolios across various performance metrics. This suggests that the ability of machine learning methods to identify and capture nonlinear effects is of great importance for accurately modeling fund performance.

An exception to the generally superior performance of fund selection strategies based on machine learning models can be observed for funds categorized as "Others". The results indicate a preference for using the Fung and Hsieh (2004) model in capturing the risk exposures of these funds. In other words, a linear factor model appears to be sufficient for risk characterization and performance evaluation of such funds.

D.2 Subperiod Analysis

Bollen, Joenväärä, and Kauppila (2021) document a significant decline in hedge fund performance during the post-financial crisis period (2008–2016) compared to the period from

1997 to 2007. Motivated by their findings, this section explores the performance of benchmark models over different sample periods. Specifically, following Bollen, Joenväärä, and Kauppila (2021), the sample is split into two subperiods: the pre-financial crisis period (1998-2007) and the post-financial crisis period (2008-2021). The performance of strategies that select funds into portfolios based on the fund alphas estimated using the various benchmark models is examined in both sub-periods.

The performance of the selection strategies relying on the Fung-Hsieh model experiences a dramatic decline after 2008. In Panel A of Table VII, the top decile portfolio based on the Fung-Hsieh model demonstrates an excess return of 6.305%, an annualized Sortino ratio of 2.632, an information ratio of 2.260, and an annualized 7-factor alpha of 5.039% prior to 2008. However, during the post-2008 period, these metrics decline to 3.240% for excess return, 0.818 for Sortino ratio, 0.392 for information ratio, and 1.086% for the 7-factor alpha. In contrast, the decline in performance is relatively mild for the machine learning-based top portfolios. Furthermore, all four machine learning model-based top decile portfolios outperform the corresponding Fung-Hsieh model-based portfolio by a large margin after 2008.

Although the hedge portfolio is non-tradable, the performance improvement of the *ENet**-based hedge portfolios after 2008 is worth noting. For instance, the Sortino ratio of the *ENet** hedge portfolio increases from 0.465 pre-2008 to 0.843 post-2008, and the Sharpe ratio doubles after 2008. The annualized 7-factor alpha also increases from 1.592% pre-2008 to 4.474% post-2008. Additionally, it is worth highlighting that the MPPM of the four machine learning-based hedge portfolios all reverse in sign from being negative pre-2008, to turning positive post-2008. These improvements are primarily driven by the sharp decline in the performance of the bottom decile portfolios, indicating that the machine learning-based strategies are quite effective at identifying poorly performing funds. It is widely recognized that several hedge funds experienced failure during the financial crisis and the subsequent period. The machine learning model framework appears to be particularly valuable in screening out underperforming funds during the post-financial crisis period.

In summary, our results complement the findings in Bollen, Joenväärä, and Kauppila (2021). The performance of hedge fund portfolios selected based on the Fung-Hsieh 7-factor benchmark model deteriorates during the period following the financial crisis. On the other hand, machine learning-based models suffer much less from the performance decline. Despite the overall decline in hedge fund performance since 2008, machine learning-based models prove to be more efficient in identifying better-performing funds compared to the Fung-Hsieh model.

E. Failure Prediction

This subsection examines the role of the machine learning-implied benchmark models in predicting hedge fund failures. Fund failure is considered a more straightforward indicator of poor performance compared to returns-based analyses, as funds with persistently poor performance tend to cease reporting (Bollen, 2013). Accordingly, following Liang and Park (2010) and Bollen (2013), we examine the ability of the various benchmark models to predict fund failure using the Cox (1972) proportional hazards model. The analysis controls for other known determinants of fund failure, along with yearly fixed effects to capture time-series variation in failure rates caused by market conditions. Since some funds in the TASS database stop reporting their returns several months prior to failure, the hazard analysis is conducted on a fund-year basis, following O'Doherty, Savin, and Tiwari (2016).

It is assumed that the failure rate of a given fund depends on fund age t and a vector of covariates $z(t)$:

$$h(t, z) = h_0(t) \exp(\beta' z(t)), \tag{24}$$

where $h_0(t)$ is the baseline hazard rate that depends only on age t, and $z(t)$ is the set of time-varying covariates. A fund is identified as failing if it stops reporting its performance to the TASS database. Following Liang and Park (2010), we employ an additional performance filter requiring the fund's average excess return over the prior 12 months to be below the median value across all funds during the same period. Funds are considered "live" if they either remain in the database or were dropped from the database but had an average excess return over the prior 12 months above the median hedge fund return during the same period.

The primary question of interest is the informativeness of the performance relative to each benchmark model in predicting fund failures. Accordingly, we create failure indicators based on each benchmark model. For each of the models, (*FH*, *ENet*, *ENet**, *RF*, and *BART*), we rank sample funds at the beginning of each year based on the respective model-implied alpha t-statistics computed using the prior two years of fund returns. Assuming that poorly performing funds are more likely to fail, we create a model-specific failure indicator for each fund. The failure indicator equals one for funds ranked in the lowest quintile based on the fund alpha t-statistics derived from a particular model. Thus, based on each model, we obtain a failure indicator for each fund at the start of each year. Additionally, a ZeroR^2 flag indicator inspired by Bollen (2013) is created. A fund is flagged with a value of 1 if its realized R^2 from the 7-factor Fung and Hsieh (2004) model over the prior 60 months falls below the 95th percentile of the simulated R^2 distribution under the null hypothesis.

The remaining covariates in $z(t)$ are constructed for each fund-year following Liang and Park (2010), andO'Doherty, Savin, and Tiwari (2016). Dummy variables are included to identify funds with a high-water mark provision, personal investment by fund managers, leverage, and lockup provisions. To control for recent past performance, we include a fund's percentile rank based on the prior 12-month average excess returns relative to all funds (*return rank*). We also control for a fund's downside risk by computing the *expected shortfall*, which involves forming the monthly excess return distribution using all prior observations and averaging the returns that are below than the 5th percentile cutoff. We also include as controls dummy variables for the five broad investment styles (i.e., directional, nondirectional, semidirectional, fund of funds, and others). The log of fund age in months is also added as an explanatory variable, along with year-fixed effects. All variables are constructed and updated at the beginning of each year and are used to predict fund failure during the year. Thus, the model characterizes fund failures from January 1999 to December 2021.

The results of the fund failure analysis are presented in Columns (1) - (4) of Table VIII. The models in this case include one of the machine learning model-based faiure indicators, a failure indicator based on the Fung-Hsieh model, and the ZeroR^2 indicator, along with the aforementioned control variables and fixed effects. The table reports the parameter estimates and hazard ratios (in brackets) for each covariate. All four machine learning-based failure indicators are statistically significant predictors of hedge fund failure. The coefficient estimate is 0.563 for *ENet** with a corresponding hazard ratio of 1.757, suggesting that funds classified in the bottom quintile according to the *ENet** model have a 75.7% higher probability of failure compared to other funds, after controlling for the fund's prior returns and downside risk. In contrast, the indicators based on the Fung and Hsieh (2004) model and the ZeroR^2 classification (Bollen (2013)) show no predictive ability regarding fund failure, at the margin. Regarding the other covariates, the results align with prior studies. As expected, a higher prior return rank is significantly negatively related to the probability of failure, indicating that better-performing funds are more likely to survive. The estimated coefficient of *Shortfall* is significantly positive, suggesting that funds with higher downside risk are more prone to failure.

Next, we estimate the individual models that include only one failure indicator at a time, along with the control variables. Columns (5) - (10) of Table VIII report the coefficient estimates and hazard ratios of the respective failure indicators. The coefficient estimates for the machine learning-based indicators in columns (5) through (8) of the table remain significantly positive. The estimates range from 0.366 for the *Enet** model to 0.444 for the *RF* model. In economic terms, based on the hazard ratio, funds classified in the bottom quintile according to the *RF* model have a 55.9% higher chance of failure relative to other funds, after controlling for prior performance and downside risk. On the other hand, the ZeroR^2 indicator, when included by itself in the model with other control variables, remains statistically insignificant.[21] Finally, the Fung-Hsieh indicator has a coefficient of 0.099 which is statistically significant at the 1% level. Interestingly, the coefficient and the associated hazard ratio of 1.105 imply a notably weaker predictive value for the Fung-Hsieh failure indicator compared to the failure indicators based on the four machine learning-based models, in economic terms .

To summarize, the machine learning model-based approach yields better performance in terms of fund failure prediction compared to failure indicators based on the conventional linear factor models (e.g., the Fung and Hsieh (2004) model). Thanks to the ability to capture nonlinear and high-order interaction effects among risk factors, the machine learning framework can more accurately characterize the investment style and capture the risks to which funds are exposed. Consequently, the resulting alpha measure is an efficient means to identify the worst-performing funds that have an elevated probability of failing subsequently.

VII. Re-examining the Value of Hedge Fund Strategy Distinctiveness

The previous empirical findings demonstrate the effectiveness of machine learning-based benchmarks in capturing the investment styles of individual hedge funds, even for funds that the traditional Fung and Hsieh (2004) model fails to explain. These results also highlight the superior ability of the machine learning framework to identify skilled funds compared to the Fung and Hsieh (2004) benchmark model. Building upon these findings, we revisit a well-known result first documented by Sun, Wang, and Zheng (2012) to further illustrate the value of the machine learning framework.

In their study, Sun, Wang, and Zheng (2012) explore the relationship between a fund's strategy distinctiveness and its subsequent performance, motivated by the idea that skilled fund managers often employ innovative and unique trading strategies. They find striking results, showing that funds ranked in the top quintile of their Strategy Distinctiveness Index (SDI) outperform funds in the bottom quintile by 3.5% the following year, based on the 7-factor Fung and Hsieh (2001) alphas. Given the limitations of the traditional benchmark model, as discussed in this paper, it is important to reexamine this issue.

[21] We note that our results in this regard contrast with those of Bollen (2013), which we attribute to the difference in our respective samples and time periods. Furthermore, to align with the design of the machine learning model, we compute the ZeroR^2 flag indicator based on its R^2 and simulated critical value calculated from a regression of fund returns on the factors over the prior 60 months. In an unreported test where we follow Bollen (2013) and use the prior 24 months to construct the ZeroR^2 indicator, the associated p-value for its coefficient marginally clears the 5% significance level.

One key concern arising from our previous findings is the explanatory power of a linear factor model, especially when it comes to benchmarking the performance of funds that follow unique strategies defying the usual style classifications. The more unique a strategy appears to be, the more likely it is that the conventional benchmark model will fail to capture the relevant risks associated with the strategy. It becomes difficult to distinguish the effect of strategy distinctiveness on performance if the Fung and Hsieh (2004) model poorly captures the risks inherent in the fund's strategy. Funds following more distinctive or unique strategies relative to their peers are also exposed to potentially unique risks that may be overlooked by the benchmark model, which is otherwise well-suited for benchmarking typical or "average" funds.

To address this issue, we replicate the main findings of Sun, Wang, and Zheng (2012) and go further by assessing fund performance using both the machine learning model-based alphas and the Fung and Hsieh (2004) alphas. We also control for the effects of potentially omitted risks by separately analyzing zero-R^2 funds. For each fund, we construct the Strategy Distinctiveness Index (SDI) as described by Sun, Wang, and Zheng (2012). The SDI is estimated as (1 - *corr*), where *corr* represents the correlation between a fund's return and the average return of funds within the same investment style over the prior 24 months. Thus, the SDI reflects the lack of correlation between a fund's return and its peers within the same category. We create fund quintile portfolios based on the end-of-period SDI values and calculate the portfolio alphas relative to both the Fung and Hsieh (2004) model and the machine learning-based benchmark models. We update the portfolios every 6 months, and report the results over short- and long-term holding periods of 6 months and 24 months, respectively.[22]

We begin by examining the results for the full sample, as presented in Table IX. The table reports the R^2 and $\hat{\alpha}$ values of the quintile portfolios based on Fung and Hsieh (2004) and machine learning models.[23] The average fund's Fung-Hsieh model R^2 declines from 0.692 to 0.370 as we move from the lowest to the highest SDI portfolio. This result aligns with our previous findings, indicating that the Fung and Hsieh (2004) model fails to effectively capture the risks associated with high SDI funds. In contrast, the machine learning-based models are less vulnerable to this limitation. The average R^2 values of all five SDI portfolios based on the *ENet** and *BART* models consistently exceed the corresponding values for the Fung-Hsieh model. Remarkably, the average R^2 values for high SDI funds based on the *ENet** and the *BART* model are as high as 0.440 and 0.443, respectively. The other two machine learning models exhibit qualitatively similar or slightly higher R^2 relative to the Fung-Hsieh model. In other

[22] We also experimented with an alternative design by updating the portfolios every 3 months, while varying the holding period from 3 to 36 months. The results are qualitatively similar.

[23] The R^2 results of the portfolio based on the holding period of 24 months are qualitatively the same.

words, the machine learning models more accurately track the risks associated with strategies, including those that appear to be unique or distinctive. This is an important consideration as we next evaluate the alphas based on the machine learning model framework.

As observed, the monthly Fung-Hsieh alpha of the High-Low portfolio with a 6-month buy-and-hold period is 0.192% with a t-statistic of 1.687. The results are qualitatively similar and more pronounced at the longer holding period of 24 months. The Fung-Hsieh alpha is 0.186% with a t-statistic of 1.749. Overall, these results are qualitatively consistent with the findings of Sun, Wang, and Zheng (2012) and support the positive relationship between strategy distinctiveness and fund performance.

To delve deeper into this issue, we further investigate whether "strategy distinctiveness" serves as a proxy for omitted risks in the standard benchmark Fung and Hsieh (2004) model. To accomplish this, we control for the effect of omitted risks by conducting the aforementioned analysis separately for zero- and nonzero-R^2 funds identified using the method proposed by Bollen (2013). We expect that the positive relationship between SDI and performance will not persist in the nonzero-R^2 fund group, for which the benchmark Fung and Hsieh (2004) model adequately captures the risks and does not suffer from the "omitted risks" issue.

The results presented in Table X confirm our expectations. Panel A and B report the average R^2 and $\hat{\alpha}$ values for two groups of funds classified as zero- or nonzero-R^2 funds, respectively. In the zero-R^2 fund group, the average Fung-Hsieh R^2 values within each SDI portfolio group are notably lower compared to the R^2 values of the SDI portfolio based on nonzero-R^2 funds. Specifically, the Fung-Hsieh average R^2 value is 0.071 for the highest SDI portfolio in the zero-R^2 fund group. The corresponding average R^2 value for the highest SDI group of funds among the nonzero-R^2 funds is 0.496. As expected, the Fung and Hsieh (2004) model performs much better in explaining the performance of the nonzero R^2 funds. The machine learning models, on the other hand, generally improve the tracking (R^2) performance, but do not alter the pattern of higher (lower) average R^2 for nonzero (zero) R^2 funds.

The results of estimated alpha in Table X further support our hypothesis that the SDI measure serves as a proxy for the omitted risk. Notably, the significance of estimated alphas is only observed within the zero R^2 fund groups. In Panel B, the alphas are nearly indistinguishable across the three lowest SDI quintile portfolios, along with the High-Low portfolio. Moreover, it is worth noting that the average machine learning model R^2 does not decrease monotonically with the SDI level. Its value does not suffer from dramatic decrease from low to high SDI either. Additionally, the machine learning models consistently generate lower estimates of $\hat{\alpha}$ compared to Fung-Hsieh alpha. Therefore, for nonzero R^2 funds, when the strategy risks are adequately captured by the benchmark model, the apparent positive link between strategy distinctiveness and fund performance vanishes. This finding remains consistent across both the 6-month and 24-month holding periods.

In summary, our results establish that the the function of SDI is only effective when the benchmark model is unable to appropriately capture risk exposures of the funds. Machine learning-based benchmark models outperform in capturing the risks inherent in hedge fund strategies that appear to be distinctive or unique. Results based on the machine learning model alphas suggest that the documented positive relationship between strategy distinctiveness and fund performance largely stems from model errors that contaminate inference based on the conventional Fung and Hsieh (2004) model.

VIII. Concluding Remarks

The assessment of hedge fund performance is challenging given the flexibility of hedge fund strategies in terms of asset class exposures and leverage, along with the relative absence of disclosure requirements. This paper proposes the use of performance benchmarks based on machine learning techniques to address the challenge of hedge fund performance evaluation. In general, machine learning models offer the advantage of flexibility in in terms of estimating a benchmark that can successfully track hedge fund performance. The superior tracking performance of the machine learning models results in more precise estimates of fund alphas compared to traditional linear factor models. The precision of estimated fund alphas, in turn, leads to an improved ability to ex ante identify better performing funds, as well as funds likely to fail. We demonstrate that these features are valuable in the context of tracking the performance of individual hedge funds. Importantly, we show that the machine learning-based benchmark models can successfully characterize the risks of funds that have near-zero R^2 values with respect to traditional performance attribution models (e.g., Fung and Hsieh (2004) model), and funds with short return histories. In particular, a Bayesian ensemble-of-trees framework (BART) is particularly valuable in this context. A key reason for the success of the methodology is its ability to account for the nonlinearities and high-order interaction effects among risk factors that are important in determining hedge fund strategy payoffs. We further show that machine learning-based models offer significant advantage over the traditional approach in selecting superior performing funds in real-time. The machine learning methods we consider also dominate in terms of the ability to predict fund failure.

As a further illustration of the advantage offered by machine learning methods in a performance evaluation context, we re-examine the evidence regarding the well-documented positive relation between hedge fund strategy distinctiveness and fund performance. Our results suggest that the documented link between strategy distinctiveness and fund performance is an artifact of benchmark model error.

References

Agarwal, Vikas, and Narayan Y Naik, 2004, Risks and portfolio decisions involving hedge funds, *The Review of Financial Studies* 17, 63–98.

Ardia, David, Laurent Barras, Patrick Gagliardini, and Olivier Scaillet, 2022, Is it Alpha or Beta? A Formal Evaluation of Hedge Fund Models, *Swiss Finance Institute Research Paper*.

Bollen, Nicolas P.B., 2013, Zero-R^2 Hedge Funds and Market Neutrality, *The Journal of Financial and Quantitative Analysis* 48, 519–547.

Bollen, Nicolas PB, Juha Joenväärä, and Mikko Kauppila, 2021, Hedge fund performance: End of an era?, *Financial Analysts Journal* 77, 109–132.

Bollen, Nicolas PB, and Robert E Whaley, 2009, Hedge fund risk dynamics: Implications for performance appraisal, *The Journal of Finance* 64, 985–1035.

Box, George EP, 1980, Sampling and Bayes' inference in scientific modelling and robustness, *Journal of the Royal Statistical Society: Series A (General)* 143, 383–404.

Breiman, L., 2001, Random Forests, *Machine Learning* 45, 5–32.

Chipman, Hugh A, Edward I George, and Robert E McCulloch, 1998, Bayesian CART model search, *Journal of the American Statistical Association* 93, 935–948.

Chipman, Hugh A, Edward I George, and Robert E. McCulloch, 2010, BART: Bayesian additive regression trees, *The Annals of Applied Statistics* 4, 266–298.

Cox, David R, 1972, Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)* 34, 187–202.

Fama, Eugene F, and Kenneth R French, 2010, Luck versus skill in the cross-section of mutual fund returns, *The journal of finance* 65, 1915–1947.

Fung, William, and David A Hsieh, 1997, Empirical characteristics of dynamic trading strategies: The case of hedge funds, *The review of financial studies* 10, 275–302.

Fung, William, and David A Hsieh, 2000, Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases, *Journal of Financial and Quantitative analysis* 35, 291–307.

Fung, William, and David A Hsieh, 2001, The risk in hedge fund strategies: Theory and evidence from trend followers, *The review of financial studies* 14, 313–341.

Fung, William, and David A Hsieh, 2004, Hedge fund benchmarks: A risk-based approach, *Financial Analysts Journal* 60, 65–80.

Getmansky, Mila, Andrew W Lo, and Igor Makarov, 2004, An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529–609.

Giglio, Stefano, Yuan Liao, and Dacheng Xiu, 2021, Thousands of alpha tests, *The Review of Financial Studies* 34, 3456–3496.

Glosten, L.R., and R. Jagannathan, 1994, A contingent claims approach to performance evaluation, *Journal of Empirical Finance* 1, 133–160.

Goetzmann, William, Jonathan Ingersoll, Matthew Spiegel, and Ivo Welch, 2007, Portfolio performance manipulation and manipulation-proof performance measures, *The Review of Financial Studies* 20, 1503–1546.

Hill, Jennifer, Antonio Linero, and Jared Murray, 2020, Bayesian additive regression trees: A review and look forward, *Annual Review of Statistics and Its Application* 7, 251–278.

Kapelner, Adam, and Justin Bleich, 2016, bartMachine: Machine learning with Bayesian additive regression trees, *Journal of Statistical Software*.

Liang, Bing, 1999, On the performance of hedge funds, *Financial Analysts Journal* 55, 72–85.

Liang, Bing, and Hyuna Park, 2010, Predicting hedge fund failure: A comparison of risk measures, *Journal of Financial and Quantitative analysis* 45, 199–222.

O'Doherty, Michael S, Nathan Eugene Savin, and Ashish Tiwari, 2016, Evaluating hedge funds with pooled benchmarks, *Management Science* 62, 69–89.

Pástor, L'uboš, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political economy* 111, 642–685.

Patton, Andrew J, and Tarun Ramadorai, 2013, On the high-frequency dynamics of hedge fund risk exposures, *The Journal of Finance* 68, 597–635.

Sharpe, William F., 1992, Asset Allocation: Management Style and Performance, *Journal of Portfolio Management* 18, 7–19.

Shu, Tengjia, and Ashish Tiwari, 2022, Identifying Signals of the Cross Section of Stock Returns, *Available at SSRN:* *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3898282*.

Sun, Zheng, Ashley Wang, and Lu Zheng, 2012, The road less traveled: Strategy distinctiveness and hedge fund performance, *The Review of Financial Studies* 25, 96–143.

Wu, Wenbo, Jiaqi Chen, Zhibin Yang, and Michael L Tindall, 2021, A cross-sectional machine learning approach for hedge fund return prediction and selection, *Management Science* 67, 4577–4601.

Table I
Summary Statistics: Monthly Excess Returns of Hedge Funds

	Live	Defunct	Mean	Median	Std	Kurtosis	Skewness
All	791	3,411	0.42	0.49	3.84	-0.34	7.32
Dedicated Short Bias	0	23	-0.18	-0.40	6.61	0.18	6.14
Emerging Markets	61	267	0.54	0.65	6.28	-0.23	7.03
Global Macro	32	137	0.47	0.35	4.25	0.21	6.34
Managed Futures	2	2	0.17	0.13	3.25	-0.08	3.03
All Directional	95	429	0.48	0.50	5.62	-0.07	6.74
Event Driven	52	271	0.43	0.53	3.31	-0.49	8.24
Long/Short Equity Hedge	244	950	0.59	0.61	4.99	-0.01	5.68
Multi-Strategy	97	231	0.40	0.46	3.17	-0.57	8.32
All Semidirectional	393	1,452	0.53	0.57	4.37	-0.19	6.60
Convertible Arbitrage	7	90	0.42	0.46	3.70	-0.76	11.36
Equity Market Neutral	17	123	0.37	0.38	2.96	-0.32	8.40
Fixed Income Arbitrage	8	106	0.39	0.40	2.53	-0.75	12.64
All Nondirectional	32	319	0.39	0.41	3.02	-0.58	10.60
Fund of Funds	178	1,061	0.21	0.38	2.73	-0.60	7.09
Options Strategy	4	20	0.29	0.35	3.59	-0.27	14.68
Other	80	122	0.51	0.54	2.62	-0.45	10.65
Undefined	9	8	0.41	0.51	3.67	-0.19	3.46
All Others	93	150	0.48	0.52	2.79	-0.41	10.55

Note: This table reports the summary statistics for all sample hedge funds, and for funds grouped by category. The summary statistics include the mean and median monthly excess return (expressed as a percent) along with the standard deviation, skewness, and excess kurtosis. We also count the number of live and defunct funds in each category. Within a given category, the figures represent the equal weighted averages of the statistics across sample funds. The sample includes funds in the Lipper TASS database with at least 60 months of consecutive monthly net-of-fee returns, and a currency code of "USD" (U.S. dollar). The sample period extends from January 1994 to December 2021.

Table II
Summary Statistics of Factors

	Mean	Median	Std	Skewness	Kurtosis
MKT	0.68	1.31	4.29	-0.75	1.25
SMB	0.11	0.05	3.07	0.45	4.65
D10YR	-0.06	0.00	7.74	0.19	2.31
DSPRD	0.24	0.00	7.01	1.03	3.41
PTFSBD	-1.39	-4.49	15.71	1.35	2.29
PTFSFX	-1.06	-5.30	19.11	1.38	2.63
PTFSCOM	-0.50	-3.05	14.02	1.10	1.69

Note: This table reports the summary statistics for the (Fung and Hsieh, 2004) model risk factors. The summary statistics include the mean and median monthly returns expressed in percent, standard deviation, skewness, and excess kurtosis. The sample period extends from January 1994 to December 2021. For additional details regarding the factors please refer to the text.

Table III

Modelling Hedge Fund Returns: Fung and Hsieh (2004) Model vs. Machine Learning Approaches

	Panel A – Zero R^2 Funds						Panel B – Nonzero R^2 Funds					
Group	Obs.	FH	Enet	Enet*	RF	BART	Obs.	FH	Enet	Enet*	RF	BART
All	593	0.093	0.072	0.178	0.045	0.350	3609	0.378	0.365	0.435	0.302	0.544
Dedicated Short Bias	2	0.102	0.079	0.132	0.022	0.405	21	0.621	0.614	0.662	0.504	0.729
Emerging Markets	39	0.096	0.081	0.203	0.053	0.368	289	0.343	0.332	0.395	0.258	0.500
Global Macro	40	0.096	0.069	0.168	0.039	0.339	129	0.308	0.290	0.375	0.227	0.496
Managed Futures							4	0.268	0.262	0.389	0.240	0.407
Directional	81	0.096	0.075	0.184	0.045	0.355	443	0.345	0.332	0.401	0.260	0.509
Event Driven	41	0.101	0.081	0.177	0.053	0.355	282	0.357	0.344	0.420	0.283	0.531
Long/Short Equity Hedge	153	0.089	0.069	0.175	0.044	0.331	1041	0.413	0.400	0.460	0.335	0.559
Multi-Strategy	60	0.098	0.076	0.183	0.048	0.366	268	0.346	0.334	0.405	0.284	0.510
Semidirectional	254	0.093	0.072	0.177	0.047	0.343	1591	0.392	0.379	0.444	0.317	0.546
Convertible Arbitrage	15	0.087	0.072	0.171	0.041	0.322	82	0.322	0.309	0.397	0.242	0.535
Equity Market Neutral	61	0.087	0.066	0.165	0.047	0.337	79	0.269	0.247	0.336	0.192	0.461
Fixed Income Arbitrage	33	0.089	0.067	0.212	0.035	0.385	81	0.257	0.234	0.355	0.169	0.491
Nondirectional	109	0.088	0.067	0.180	0.042	0.349	242	0.283	0.263	0.363	0.201	0.496
Fund of Funds	66	0.108	0.083	0.192	0.045	0.381	1173	0.397	0.383	0.452	0.322	0.564
Options Strategy	8	0.090	0.069	0.117	0.038	0.349	16	0.382	0.363	0.470	0.285	0.656
Other	69	0.079	0.062	0.154	0.036	0.340	133	0.340	0.328	0.415	0.263	0.535
Undefined	6	0.176	0.151	0.298	0.094	0.422	11	0.388	0.379	0.432	0.340	0.521
Others	83	0.087	0.069	0.161	0.040	0.346	160	0.348	0.335	0.421	0.271	0.546

Note: This table reports the average R^2 values for individual hedge fund by category. The R^2's are derived from the Fung and Hsieh (2004) 7-factor benchmark model, and machine learning models, including Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions among the factors) (*Enet**), Random Forest model projected on a proxy linear model (*RF*), and Bayesian Additive Regression Trees model with a proxy model (*BART*). For detailed information about these models, please refer to the model description section in the text. We report the values separately for funds classified as zero R^2 and nonzero R^2 funds. To classify each fund, we compare its Fung-Hsieh model R^2 to the 95th percentile of the null distribution of R^2 values using simulations. Funds with Fung-Hsieh R^2 values below the 95th percentile of the null distribution are categorized as zero-R^2 funds, following the methodology outlined by Bollen (2013). The remaining funds are classified as nonzero-R^2 funds.

Table IV
Average Monthly Alphas of Zero- vs. Nonzero-R^2 Funds

	Panel A – Zero R^2 Funds						Panel B – Nonzero R^2 Funds					
Group	Obs.	FH	Enet	Enet*	RF	BART	Obs.	FH	Enet	Enet*	RF	BART
All	593	0.411	0.427	0.447	0.439	0.512	3609	0.199	0.231	0.305	0.304	0.169
Dedicated Short Bias	2	-0.081	-0.118	-0.019	-0.246	-0.286	21	0.359	0.291	0.246	0.018	0.193
Emerging Markets	39	0.244	0.318	0.442	0.159	0.700	289	0.233	0.277	0.434	0.216	-0.005
Global Macro	40	0.492	0.488	0.450	0.351	0.380	129	0.314	0.347	0.318	0.313	0.224
Managed Futures							4	0.048	0.100	0.369	0.246	0.682
Directional	81	0.359	0.391	0.435	0.243	0.518	443	0.260	0.296	0.391	0.236	0.077
Event Driven	41	0.351	0.387	0.429	0.464	0.588	282	0.253	0.276	0.384	0.359	0.235
Long/Short Equity Hedge	153	0.534	0.549	0.581	0.542	0.660	1041	0.239	0.294	0.348	0.371	0.186
Multi-Strategy	60	0.350	0.375	0.418	0.547	0.538	268	0.262	0.275	0.340	0.318	0.231
Semidirectional	254	0.461	0.481	0.518	0.531	0.620	1591	0.245	0.288	0.353	0.360	0.203
Convertible Arbitrage	15	0.470	0.466	0.463	0.484	0.340	82	0.299	0.315	0.290	0.159	0.036
Equity Market Neutral	61	0.345	0.351	0.319	0.393	0.416	79	0.229	0.267	0.335	0.276	0.217
Fixed Income Arbitrage	33	0.363	0.376	0.353	0.331	0.254	81	0.346	0.353	0.362	0.318	0.241
Nondirectional	109	0.368	0.375	0.349	0.387	0.356	242	0.292	0.312	0.329	0.250	0.164
Fund of Funds	66	0.204	0.194	0.230	0.265	0.277	1173	0.087	0.104	0.197	0.270	0.168
Options Strategy	8	0.333	0.318	0.306	0.312	0.381	16	0.051	0.122	0.302	0.263	0.029
Other	69	0.606	0.623	0.629	0.617	0.635	133	0.281	0.322	0.352	0.265	0.085
Undefined	6	-0.107	-0.022	-0.133	0.225	0.005	11	0.204	0.271	0.330	0.457	0.446
Others	83	0.528	0.547	0.543	0.560	0.565	160	0.253	0.298	0.346	0.278	0.104

Note: This table presents the average monthly alpha values (expressed in percent) for hedge funds by category. The table presents alphas derived from the Fung and Hsieh (2004) 7-factor benchmark model, and machine learning models, including Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). The *ENet* α and *ENet** α are estimated as the intercepts from the model. The *RF* α and *BART* α are estimated as the time-series average difference between the realized fund returns and the expected returns based on the respective proxy models, as explained in the text. For each model, we also report the values separately for funds within the zero R^2 and nonzero R^2 groups. To classify each fund, we compare its Fung-Hsieh R^2 to the 95th percentile of the null distribution of simulated R^2. Funds with Fung-Hsieh R^2 values below the 95th percentile of the null distribution are categorized as zero-R^2 funds, following Bollen (2013). The remaining funds are classified as nonzero-R^2 funds.

Table V
Summary Statistics of Zero- vs. Nonzero-R^2 Funds by Length of Performance Record

Panel A – Average Fund R^2

Group	Number		FH		Enet		Enet*		RF		BART	
	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero
$t \leq 60$	175	498	0.128	0.420	0.099	0.399	0.231	0.485	0.044	0.255	0.468	0.657
$60 < t \leq 72$	92	436	0.111	0.409	0.086	0.391	0.209	0.473	0.050	0.279	0.412	0.621
$72 < t \leq 84$	76	344	0.093	0.390	0.073	0.374	0.182	0.450	0.043	0.294	0.331	0.581
$84 < t \leq 108$	105	631	0.077	0.361	0.061	0.347	0.156	0.419	0.043	0.292	0.303	0.532
$108 < t \leq 132$	78	551	0.057	0.366	0.044	0.355	0.117	0.421	0.042	0.313	0.235	0.512
$132 < t \leq 168$	46	565	0.052	0.366	0.039	0.357	0.103	0.415	0.048	0.329	0.208	0.495
$t > 168$	21	584	0.036	0.356	0.027	0.349	0.072	0.404	0.043	0.339	0.145	0.458

Panel B – Average Monthly Alpha

Group	Number		FH		Enet		Enet*		RF		BART	
	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero
$t \leq 60$	175	498	0.367	0.132	0.372	0.168	0.363	0.268	0.441	0.316	0.478	0.141
$60 < t \leq 72$	92	436	0.359	0.157	0.380	0.179	0.408	0.247	0.400	0.328	0.428	0.126
$72 < t \leq 84$	76	344	0.469	0.203	0.480	0.232	0.489	0.292	0.395	0.313	0.674	0.187
$84 < t \leq 108$	105	631	0.426	0.206	0.441	0.239	0.475	0.311	0.446	0.309	0.450	0.206
$108 < t \leq 132$	78	551	0.452	0.170	0.487	0.212	0.516	0.307	0.467	0.286	0.503	0.125
$132 < t \leq 168$	46	565	0.384	0.233	0.407	0.259	0.485	0.317	0.471	0.279	0.638	0.170
$t > 168$	21	584	0.616	0.274	0.639	0.306	0.680	0.367	0.556	0.309	0.633	0.215

Note: This table presents information on funds grouped by the number of available observations over the sample period from 1994 to 2019. Panel A reports the number of observations and the average R^2 values for funds in each group, while Panel B reports the number of observations and the average monthly alphas (expressed in percent). The R^2 values and alphas are derived from the Fung and Hsieh (2004) model and machine learning models, including Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). More detailed descriptions of these models can be found in the model description section. Within each model, we also report the values separately for funds within zero R^2 and nonzero R^2 groups. To classify each fund, we compare its Fung-Hsieh R^2 to the 95th percentile of the null distribution of simulated R^2. Funds with Fung-Hsieh R^2 values below the 95th percentile of the null distribution are categorized as zero-R^2 funds, following the methodology outlined by Bollen (2013). The remaining funds are classified as nonzero-R^2 funds.

Table VI
Portfolios Formed on Past Performance Measures

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	4.494	0.205	5.013	1.141	5.570	2.146	5.045	1.398	5.136	0.865
Std. Dev	5.422	4.443	4.997	6.018	4.803	5.770	5.198	5.569	5.287	5.965
SharpeRatio	0.829	0.046	1.003	0.190	1.160	0.372	0.971	0.251	0.972	0.145
SortinoRatio	1.318	0.083	1.584	0.303	1.901	0.680	1.512	0.424	1.543	0.248
MaxDrawdown	16.611	17.111	14.945	23.248	15.047	21.257	18.328	18.782	13.843	22.852
Max1MLoss	-9.970	-5.159	-7.019	-6.370	-6.694	-5.407	-7.554	-5.820	-8.072	-6.751
Alpha	2.809	1.331	3.420	1.881	3.932	3.413	3.349	1.609	3.395	1.666
t value	(4.392)	(1.534)	(4.430)	(1.567)	(4.430)	(1.567)	(4.243)	(1.407)	(5.532)	(3.416)
IR	0.995	0.348	1.003	0.355	1.253	0.774	0.966	0.301	0.961	0.319
MPPM	2.334	-1.785	2.921	-1.104	3.507	-0.050	2.922	-0.766	3.000	-1.364

Note: This table presents the performance of hedge fund decile portfolios over the sample period from 1999 to 2021, based on the estimated alphas. In each January (month t), funds are sorted based on the alpha t-statistics computed using the estimated alpha over the previous 24 months. The alpha estimates are derived from the Fung and Hsieh (2004) model or machine learning models, specifically Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). Further details about the models can be found in the model description section. The equal-weighted decile portfolios are updated annually, and their performance is evaluated over the subsequent year (from month t to month $t + 11$). For each model, we report the performance of top decile portfolio (*P10*) and the high-minus-low hedge portfolio (*P10-P1*). The performance measures include the annualized mean excess returns and standard deviations (expressed in percent). Also reported are the annualized Sharpe ratios, Sortino ratios, maximum drawdown (in percent), and maximum 1-month loss (in percent), the annualized information ratios, the annualized Fung and Hsieh (2004) 7-factor alphas (expressed in percent, with t-statistics in parentheses), and the manipulation-proof performance measures (MPPM) of the portfolios, following the methodology outlined by Goetzmann et al. (2007), with a risk aversion level of 3.

Table VII

Portfolios Formed on Past Performance Measures — Pre- vs. Post-2008

Panel A – Pre-2008

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	6.305	-1.092	6.987	1.426	7.195	1.551	7.109	1.389	6.537	0.233
Std. Dev	4.738	3.462	5.403	6.827	4.195	5.654	4.920	6.753	5.263	6.696
SharpeRatio	1.331	-0.315	1.293	0.209	1.715	0.274	1.445	0.206	1.242	0.035
SortinoRatio	2.632	-0.525	2.251	0.340	3.492	0.465	2.829	0.360	2.222	0.061
MaxDrawdown	6.874	14.924	12.425	17.613	8.117	14.286	10.146	16.233	12.273	18.510
Max1MLoss	-3.245	-2.592	-7.019	-6.370	-5.064	-4.990	-5.762	-5.820	-6.646	-5.165
Alpha	5.039	-0.339	4.719	0.114	5.409	1.592	4.916	0.575	4.109	-0.553
t value	(6.307)	(-0.302)	(3.842)	(0.053)	(5.578)	(0.987)	(4.449)	(0.260)	(3.368)	(-0.262)
IR	2.260	-0.108	1.377	0.019	1.999	0.354	1.594	0.093	1.207	-0.094
MPPM	2.608	-4.614	3.182	-2.615	3.569	-2.270	3.387	-2.629	2.764	-3.767

Panel B – Post-2008

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	3.240	1.104	3.645	0.943	4.445	2.558	3.615	1.404	4.166	1.303
Std. Dev	5.836	5.006	4.673	5.410	5.172	5.865	5.358	4.599	5.302	5.420
SharpeRatio	0.555	0.220	0.780	0.174	0.859	0.436	0.675	0.305	0.786	0.240
SortinoRatio	0.818	0.404	1.135	0.274	1.314	0.843	0.953	0.497	1.167	0.400
MaxDrawdown	15.829	17.111	14.945	23.248	15.047	21.257	18.328	16.739	13.843	22.852
Max1MLoss	-9.970	-5.159	-6.698	-6.000	-6.694	-5.407	-7.554	-5.014	-8.072	-6.751
Alpha	1.086	3.270	1.962	2.895	2.293	4.474	1.516	1.973	2.197	2.829
t value	(1.278)	(2.896)	(2.706)	(2.207)	(2.749)	(3.461)	(1.795)	(1.528)	(2.624)	(2.100)
IR	0.392	0.888	0.829	0.677	0.843	1.061	0.550	0.468	0.804	0.644
MPPM	2.143	0.181	2.741	-0.056	3.464	1.493	2.600	0.528	3.162	0.304

Note: This table presents the performance of hedge fund decile portfolios before and after 2008, based on the estimated alphas. In each January (month t), funds are sorted based on the alpha t-statistics computed using the estimated alpha over the previous 24 months. The alpha estimates are derived from the Fung and Hsieh (2004) model or machine learning models, specifically Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). The equal-weighted decile portfolios are updated annually, and their performance is evaluated over the subsequent year (from month t to month $t+11$). Panels A and B display the performance for the periods 1999-2007 and 2008-2021, respectively. For each model, we report the performance of the top decile (*P10*) portfolio and the high-minus-low hedge portfolio (*P10-P1*). The performance measures include the annualized mean excess returns and standard deviations (expressed in percent). Also reported are the annualized Sharpe ratios, Sortino ratios, maximum drawdown (in percent), and maximum 1-month loss (in percent), the annualized information ratios, the annualized Fung and Hsieh (2004) 7-factor alphas (expressed in percent, with t-statistics in parentheses), and the manipulation-proof performance measures (MPPM) with a risk aversion level of 3, following Goetzmann et al. (2007).

Table VIII
Hedge Fund Failure Prediction

	Enet	Enet*	RF	BART	Enet	Enet*	RF	BART	FH	Low R^2
ML	0.584***	0.563***	0.490***	0.576***	0.389***	0.366***	0.444***	0.382***		
	[1.793]	[1.757]	[1.633]	[1.779]	[1.475]	[1.442]	[1.559]	[1.465]		
FH	-0.105	-0.070	-0.063	-0.077					0.099**	
	[0.901]	[0.932]	[0.939]	[0.926]					[1.105]	
Low R^2	0.192*	0.180*	0.147	0.177						0.141
	[1.212]	[1.197]	[1.158]	[1.193]						[1.151]
Age	0.001**	0.001*	0.001*	0.001*	0.000*	0.000*	0.001*	0.001*	0.001**	0.001*
	[1.001]	[1.001]	[1.001]	[1.001]	[1.000]	[1.000]	[1.001]	[1.001]	[1.001]	[1.001]
Shortfall	0.501	0.507**	0.523**	0.521**	0.356***	0.362***	0.354***	0.372***	0.416***	0.582**
	[1.651]	[1.661]	[1.687]	[1.684]	[1.428]	[1.437]	[1.425]	[1.451]	[1.516]	[1.790]
Returnrank	0.346	0.380***	0.364**	0.360**	0.430***	0.462***	0.392***	0.445***	0.561***	0.551***
	[1.413]	[1.462]	[1.439]	[1.434]	[1.537]	[1.587]	[1.481]	[1.560]	[1.752]	[1.734]
HighWaterMark	-0.173	-0.183*	-0.179*	-0.179*	-0.133***	-0.140***	-0.136***	-0.137***	-0.144***	-0.190***
	[0.841]	[0.833]	[0.836]	[0.836]	[0.875]	[0.869]	[0.873]	[0.872]	[0.866]	[0.827]
PersonalCapital	0.282	0.273**	0.273**	0.269**	0.059	0.054	0.055	0.052	0.055	0.273**
	[1.326]	[1.313]	[1.314]	[1.308]	[1.061]	[1.055]	[1.057]	[1.053]	[1.056]	[1.314]
Leveraged	-0.113	-0.118	-0.121	-0.123	-0.126**	-0.128**	-0.132**	-0.133***	-0.134***	-0.124
	[0.894]	[0.889]	[0.886]	[0.884]	[0.882]	[0.880]	[0.877]	[0.875]	[0.875]	[0.883]
LockUpPeriod	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]

Note: This table reports estimation results for the Cox (1972) proportional hazards models based on the Fung-Hsieh model-based and machine learning model-based failure indicators, as well as the zero R^2 indicator (Bollen (2013)). We estimate the models with failure indicators based on the Fung-Hsieh model and machine learning models based on the 7 Fung-Hsieh factors. The machine learning models include Elastic Net with the same set of 7 factors (*Enet*), Elastic Net with extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with proxy model (*RF*), and Bayesian Additive Regression Tree with proxy model (*BART*). For more details of the models, please refer to the model description section. Other variables in the hazard models include *return rank* that controls for recent past performance, *expected shortfall* that captures downside risks, fund age in months (*Age*), dummy variables for fund with high-water mark provision (*HighWaterMark*), personal investment by fund managers (*Personalinvestment*), leverage (*leverage*), and lockup provisions (*Lockupprovision*). The models also include style and year fixed effects (unreported). For each failure indicator the table reports parameter estimates and hazard ratios (in square brackets). Failing funds are those that cease reporting performance to the TASS database with the fund average excess return over the prior 12 months falling below the median average of all funds over the same period. The machine learning model failure indicators (*ML*) are based on the funds' (low) alpha t-statistics. The indicator equals 1 if the fund ranks in the lowest quintile based on the t-statistic of machine learning model-implied alpha over the prior 24 months. The *Fung-Hsieh* indicator (*FH*) is based on the low alpha t-statistic inferred from the Fung and Hsieh (2004) model. *Zero*R^2 indicator (*Low* R^2) is a dummy variable that equals 1 if the fund's Fung-Hsieh R^2 over the prior 60 months falls below the 95th percentile of the simulated R^2 distribution under the null.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table IX
Performance of Portfolios Sorted by Strategy Distinctiveness Index (SDI)

	R^2					6 months					24 months				
	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART
Low	0.692	0.691	0.732	0.668	0.728	0.109	0.109	0.191	0.184	0.118	0.106	0.108	0.109	0.084	0.114
						(1.112)	(1.163)	(2.155)	(1.995)	(1.257)	(1.148)	(1.217)	(1.258)	(0.965)	(1.280)
P2	0.734	0.733	0.767	0.710	0.764	0.161	0.161	0.136	0.082	0.171	0.145	0.145	0.099	0.051	0.152
						(1.806)	(1.889)	(1.681)	(1.000)	(1.999)	(1.769)	(1.839)	(1.289)	(0.676)	(1.924)
P3	0.756	0.755	0.789	0.721	0.784	0.189	0.189	0.198	0.113	0.197	0.146	0.146	0.165	0.067	0.153
						(2.396)	(2.506)	(2.714)	(1.601)	(2.614)	(2.000)	(2.078)	(2.411)	(1.022)	(2.169)
P4	0.737	0.737	0.777	0.711	0.764	0.257	0.257	0.242	0.241	0.263	0.191	0.191	0.145	0.132	0.195
						(3.962)	(4.144)	(4.213)	(4.173)	(4.225)	(3.153)	(3.277)	(2.654)	(2.406)	(3.332)
High	0.370	0.369	0.440	0.375	0.433	0.301	0.301	0.405	0.385	0.296	0.292	0.292	0.288	0.311	0.290
						(5.314)	(5.558)	(8.281)	(7.904)	(5.456)	(5.542)	(5.759)	(6.171)	(6.575)	(5.704)
High_Low						0.192	0.192	0.214	0.201	0.177	0.186	0.184	0.180	0.227	0.176
						(1.687)	(1.764)	(2.109)	(1.933)	(1.633)	(1.749)	(1.797)	(1.831)	(2.288)	(1.717)

Note: This table reports the Fung-Hsieh seven-factor model alphas and machine learning model-based alphas for the quintile portfolios sorted on the Strategy Distinctiveness Index (SDI). The alphas are expressed as percent per month and the associated t-statistics are reported in parentheses. Each fund's SDI is estimated following Sun, Wang, and Zheng (2012) as $(1 - corr)$, where *corr* is the correlation between a fund's return and the average return for funds belonging to the same style measured over the prior 24 months. The quintile portfolios are equal-weighted and updated every three months. The table reports the Fung-Hsieh model-implied alpha and machine learning model-implied alpha of the portfolios that are held for buy-and-hold time periods over the short term (6 months) and the long term (24 months). Also reported are the average R^2 for funds in each SDI quintile portfolio. The machine learning models include Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with proxy model (*RF*), and Bayesian Additive Regression Tree with proxy model (*BART*).

Table X

Performance of Portfolios Sorted by Strategy Distinctiveness Index (SDI): Zero-R^2 and Nonzero-R^2 Funds

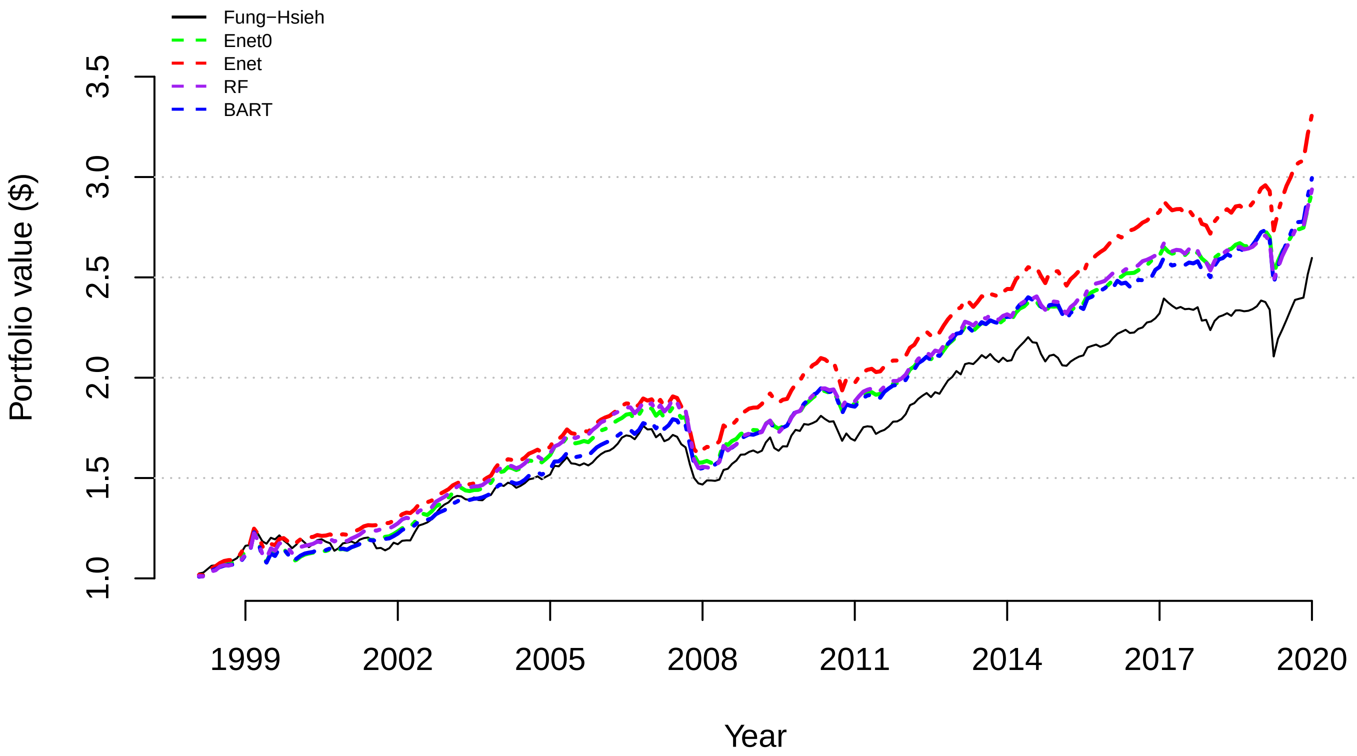
Panel A – Zero R^2 Funds

	R^2					6 months					24 months				
	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART
Low	0.589	0.589	0.621	0.571	0.621	0.173	0.173	0.285	0.262	0.129	0.132	0.134	0.186	0.159	0.111
						(1.591)	(1.664)	(2.876)	(2.574)	(1.231)	(1.298)	(1.366)	(1.974)	(1.662)	(1.125)
P2	0.577	0.577	0.621	0.552	0.620	0.229	0.229	0.278	0.201	0.144	0.214	0.214	0.112	0.117	0.146
						(2.341)	(2.449)	(3.139)	(2.258)	(1.458)	(2.536)	(2.635)	(1.497)	(1.536)	(1.718)
P3	0.586	0.585	0.624	0.556	0.609	0.225	0.225	0.339	0.277	0.249	0.188	0.188	0.241	0.253	0.124
						(2.999)	(3.137)	(4.748)	(3.995)	(3.314)	(2.568)	(2.669)	(3.780)	(3.883)	(1.633)
P4	0.421	0.420	0.464	0.392	0.476	0.342	0.342	0.526	0.462	0.347	0.266	0.266	0.279	0.306	0.265
						(4.771)	(4.990)	(7.921)	(7.038)	(4.914)	(4.405)	(4.578)	(5.148)	(5.544)	(4.292)
High	0.071	0.057	0.086	0.092	0.117	0.353	0.362	0.380	0.388	0.346	0.364	0.364	0.386	0.452	0.351
						(6.524)	(6.581)	(7.498)	(7.579)	(6.402)	(6.993)	(7.272)	(7.031)	(9.608)	(6.729)
High_Low						0.180	0.189	0.095	0.126	0.217	0.232	0.231	0.201	0.294	0.240
						(1.783)	(1.610)	(0.852)	(1.108)	(1.835)	(2.035)	(2.101)	(1.841)	(2.760)	(2.150)

Panel B – Nonzero R^2 Funds

	R^2					6 months					24 months				
	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART
Low	0.736	0.736	0.766	0.697	0.776	0.056	0.059	0.135	0.134	0.118	0.085	0.088	0.035	-0.004	0.129
						(0.580)	(0.633)	(1.574)	(1.490)	(1.240)	(0.954)	(1.019)	(0.422)	(-0.051)	(1.481)
P2	0.792	0.792	0.813	0.759	0.817	0.111	0.111	0.015	-0.002	0.207	0.090	0.090	0.009	-0.024	0.177
						(1.213)	(1.269)	(0.177)	(-0.024)	(2.194)	(1.030)	(1.071)	(0.114)	(-0.296)	(1.987)
P3	0.797	0.796	0.823	0.777	0.821	0.134	0.134	0.072	0.014	0.211	0.116	0.116	0.019	-0.073	0.187
						(1.509)	(1.578)	(0.891)	(0.177)	(2.270)	(1.439)	(1.496)	(0.256)	(-1.005)	(2.253)
P4	0.803	0.803	0.818	0.775	0.814	0.187	0.187	0.130	0.067	0.272	0.125	0.125	0.044	-0.005	0.208
						(2.530)	(2.646)	(1.927)	(1.032)	(3.259)	(1.806)	(1.877)	(0.694)	(-0.087)	(2.724)
High	0.496	0.495	0.593	0.473	0.575	0.228	0.230	0.310	0.271	0.256	0.232	0.233	0.148	0.197	0.258
						(3.190)	(3.351)	(5.062)	(4.445)	(3.483)	(3.610)	(3.768)	(2.662)	(3.476)	(3.953)
High_Low						0.172	0.171	0.175	0.137	0.139	0.147	0.145	0.113	0.201	0.129
						(1.426)	(1.479)	(1.662)	(1.261)	(1.154)	(1.332)	(1.371)	(1.128)	(2.000)	(1.182)

Note: This table reports the Fung-Hsieh seven-factor alphas and machine learning model-based alphas, with t-statistics in parentheses, for the quintile portfolios sorted on the Strategy Distinctiveness Index (SDI) and held for the short term (6 months) or the long term (24 months). Each fund's SDI is estimated as (1 – *corr*), where *corr* is the correlation between a fund's return and the average return for funds belonging to the same style measured over the prior 24 months. The quintile portfolios are equal-weighted and updated every three months. Also reported are the average R^2 values for funds in each SDI quintile portfolio.



Note: This figure presents the growth of $1 investment in the top decile portfolios of hedge funds. The portfolios are equally-weighted and contain hedge funds ranked in the top decile on the basis of estimated fund alphas at the start of each year. The portfolios are updated annually. The alphas are based on the Fung and Hsieh (2004) model, or three versions of the BART model. At the start of each year we sort funds based on the alpha t-statistics computed using data over the prior 24 months.

Figure 1. Growth of $1 Investment in the Top Decile Portfolios

A. Appendix: Tables

Table A1
Simulation Results

Panel A – R^2

	Zero R^2 Funds					Nonzero R^2 Funds				
	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART
α=0	0.165	0.138	0.415	0.299	0.421	0.396	0.380	0.588	0.482	0.591
α=0.22%	0.161	0.136	0.411	0.297	0.413	0.391	0.386	0.590	0.481	0.588
α=0.99%	0.168	0.139	0.416	0.297	0.418	0.389	0.379	0.589	0.484	0.591

Panel B – Estimated α

	Zero R^2 Funds					Nonzero R^2 Funds				
	FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART
α=0	0.011	0.034	0.033	0.027	0.038	0.019	0.027	0.104	0.114	0.104
α=0.22%	0.227	0.234	0.248	0.246	0.243	0.216	0.230	0.301	0.312	0.300
α=0.99%	1.011	1.019	1.020	1.029	1.016	0.997	1.009	1.097	1.109	1.092

Panel C – Proportion of iterations with significant estimated α

		Zero R^2 Funds					Nonzero R^2 Funds				
		FH	Enet	Enet*	RF	BART	FH	Enet	Enet*	RF	BART
α=0	1st. Q	14%	9%	17%	19%	18%	13%	9%	21%	24%	22%
	Median	6%	5%	12%	13%	13%	6%	5%	15%	15%	15%
	Mean	10%	8%	15%	15%	15%	10%	9%	17%	18%	18%
	3rd. Q	4%	3%	9%	9%	9%	4%	3%	10%	10%	10%
α=0.22%	1st. Q	5%	3%	8%	6%	9%	4%	4%	9%	9%	9%
	Median	14%	11%	20%	20%	20%	11%	11%	23%	25%	23%
	Mean	24%	23%	28%	28%	28%	20%	19%	28%	30%	28%
	3rd. Q	31%	32%	41%	44%	43%	31%	30%	44%	47%	44%
α=0.99%	1st. Q	58%	57%	65%	62%	66%	59%	64%	64%	61%	65%
	Median	94%	96%	94%	92%	95%	97%	98%	93%	95%	93%
	Mean	77%	76%	78%	77%	79%	78%	78%	77%	78%	78%
	3rd. Q	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Note: This table presents the performance of the Fung-Hsieh 7-factor model and machine learning models based on a simulation exercise. We randomly select 300 funds (without replacement) from each of the zero-R^2 and nonzero-R^2 fund groups. Fund returns are generated as the sum of pre-specified alpha values and the product of estimated factor loadings and factor realizations, plus the re-sampled residuals from a benchmark model. The benchmark model comprises 9 factors, including the 7 factors from Fung and Hsieh (2004) model, the squared market excess return, and the interaction term between the market excess return and the traded liquidity factor (Pástor and Stambaugh, 2003). Three pre-specified alpha values are considered: zero, the median value ($\alpha \approx 0.22\%$), and the 95th percentile ($\alpha \approx 0.99\%$) value of the cross-sectional distribution of Fung-Hsieh estimated alphas across all funds. We estimate the Fung and Hsieh (2004) model and machine learning models using the simulated fund returns. We obtain the model R^2 values, estimated alphas, and their corresponding t-statistics. The machine learning models include Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest (*RF*), and Bayesian Additive Regression Tree (*BART*). In the case of the *RF* and *BART* models fund alphas are estimated based on proxy linear models described in the text. For each model, Panel A reports the median R^2 value across funds, computed using the median R^2 value for each fund across 500 simulations. Similarly, we compute the median alpha across the 500 simulations for each fund, and report the cross-sectional median alpha (in percent) in Panel B. In Panel C, summary statistics are provided for the proportion of iterations (out of 500) in which the estimated alpha is significantly different from zero at the 5% level (for the case when the true $\alpha = 0$), or the estimated alpha is positive and significantly different from zero at the 5% level (for the cases where true $\alpha \approx 0.22$ and true $\alpha \approx 0.99$).

Table A2
Diebold-Mariano Test – Machine Learning vs. Fung-Hsieh Model

Group	Panel A – Zero R^2 Funds					Panel B – Nonzero R^2 Funds				
	Obs.	Enet	Enet*	RF	BART	Obs.	Enet	Enet*	RF	BART
All	593	0%	89%	0%	90%	3609	0%	97%	0%	96%
Dedicated Short Bias	2	0%	100%	0%	100%	21	0%	100%	0%	95%
Emerging Markets	39	0%	82%	0%	90%	289	0%	98%	0%	95%
Global Macro	40	0%	93%	0%	93%	129	0%	98%	2%	95%
Managed Futures						4	0%	100%	0%	100%
Directional	81	0%	88%	0%	91%	443	0%	98%	0%	95%
Event Driven	41	0%	90%	0%	88%	282	0%	95%	0%	98%
Long/Short Equity Hedge	153	0%	96%	1%	97%	1041	0%	98%	0%	97%
Multi-Strategy	60	0%	82%	0%	85%	268	0%	96%	0%	97%
Semidirectional	254	0%	92%	0%	93%	1591	0%	97%	0%	97%
Convertible Arbitrage	15	0%	87%	0%	87%	82	0%	95%	0%	95%
Equity Market Neutral	61	0%	93%	2%	92%	79	0%	94%	0%	91%
Fixed Income Arbitrage	33	0%	88%	0%	91%	81	0%	83%	0%	86%
Nondirectional	109	0%	91%	1%	91%	242	0%	90%	0%	91%
Fund of Funds	66	0%	82%	0%	83%	1173	0%	97%	0%	96%
Options Strategy	8	0%	88%	0%	100%	16	0%	81%	0%	81%
Other	69	0%	81%	0%	86%	133	0%	98%	0%	98%
Undefined	6	0%	100%	0%	100%	11	0%	100%	0%	100%
Others	83	0%	83%	0%	88%	160	0%	96%	0%	97%

Note: This table presents the performance of the Diebold-Mariano test for hedge funds grouped by category. The Diebold-Mariano test compares the performance of machine learning models with the Fung and Hsieh (2004) 7-factor model. The table reports the proportion of funds with a significant Diebold-Mariano statistic. The DM statistic is defined to be significant if the time-series estimation residuals from the machine learning models are significantly lower than the estimation residuals from the Fung-Hsieh 7-factor model. The machine learning models analyzed in the test include Elastic Net with the same set of Fung-Hsieh 7 factors (*Enet*), Elastic Net with an extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). Additional details about these models can be found in the model description section. Panel A and Panel B present the results for zero and nonzero R^2 funds, respectively. The proportion of funds with a significant Diebold-Mariano statistic is reported for each category. This statistic indicates the superiority of machine learning models over the Fung-Hsieh 7-factor model in terms of lower estimation residuals.

Table A3
Average R^2 of Zero- vs. Nonzero-R^2 Funds — Up to 3 factors

	Panel A – Zero R2 Funds						Panel B – Nonzero R2 Funds					
Group	Obs.	FH	Enet	Enet*	RF	BART	Obs.	FH	Enet	Enet*	RF	BART
All	301	0.411	0.427	0.447	0.439	0.512	3901	0.199	0.231	0.305	0.304	0.169
Dedicated Short Bias	0	-0.081	-0.118	-0.019	-0.246	-0.286	23	0.359	0.291	0.246	0.018	0.193
Emerging Markets	14	0.244	0.318	0.442	0.159	0.700	314	0.233	0.277	0.434	0.216	-0.005
Global Macro	18	0.492	0.488	0.450	0.351	0.380	151	0.314	0.347	0.318	0.313	0.224
Managed Futures	0						4	0.048	0.100	0.369	0.246	0.682
Directional	32	0.359	0.391	0.435	0.243	0.518	492	0.260	0.296	0.391	0.236	0.077
Event Driven	22	0.351	0.387	0.429	0.464	0.588	301	0.253	0.276	0.384	0.359	0.235
Long/Short Equity Hedge	79	0.534	0.549	0.581	0.542	0.660	1115	0.239	0.294	0.348	0.371	0.186
Multi-Strategy	30	0.350	0.375	0.418	0.547	0.538	298	0.262	0.275	0.340	0.318	0.231
Semidirectional	131	0.461	0.481	0.518	0.531	0.620	1714	0.245	0.288	0.353	0.360	0.203
Convertible Arbitrage	8	0.470	0.466	0.463	0.484	0.340	89	0.299	0.315	0.290	0.159	0.036
Equity Market Neutral	35	0.345	0.351	0.319	0.393	0.416	105	0.229	0.267	0.335	0.276	0.217
Fixed Income Arbitrage	19	0.363	0.376	0.353	0.331	0.254	95	0.346	0.353	0.362	0.318	0.241
Nondirectional	62	0.368	0.375	0.349	0.387	0.356	289	0.292	0.312	0.329	0.250	0.164
Fund of Funds	29	0.204	0.194	0.230	0.265	0.277	1210	0.087	0.104	0.197	0.270	0.168
Options Strategy	3	0.333	0.318	0.306	0.312	0.381	21	0.051	0.122	0.302	0.263	0.029
Other	43	0.606	0.623	0.629	0.617	0.635	159	0.281	0.322	0.352	0.265	0.085
Undefined	1	-0.107	-0.022	-0.133	0.225	0.005	16	0.204	0.271	0.330	0.457	0.446
Others	47	0.528	0.547	0.543	0.560	0.565	196	0.253	0.298	0.346	0.278	0.104

Note: This table provides the average R^2 values based on two types of benchmark models: the optimal fund-specific linear benchmark model determined using the method proposed by Bollen (2013), and the machine learning benchmark models. For each fund, the optimal linear benchmark model is determined through a step-wise regression process. This process selects up to 3 out of the 7 factors from the Fung and Hsieh (2004) model that yield the highest R^2 value. The average R^2 values are then computed based on these optimal linear benchmark models. Additionally, for each fund, a fund-specific benchmark model is estimated using the machine learning framework described in the text. The machine learning models considered in this analysis include Elastic Net with the same subset of factor(s) as the optimal model (*Enet*), Elastic Net with an extended set of factors (the original subset of factor(s) from the optimal model plus all two-way interactions of the factor(s)) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). The table presents the results for both zero and nonzero R^2 funds in Panel A and Panel B, respectively. To determine whether a fund is classified as a zero-R^2 fund, each fund's optimal linear model R^2 value is compared to a null distribution of R^2 values using a simulation exercise. Funds with R^2 values below the 95th percentile of the null distribution are classified as zero-R^2 funds according to the method proposed by Bollen (2013). The remaining funds are classified as nonzero-R^2 funds. The reported results are organized by primary category.

Table A4
Summary Statistics of Zero- vs. Nonzero-R^2 Funds by Number of Observations — Up to 3 factors

Group	Number		FH		Enet		Enet*		RF		BART	
	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero	Zero	Nonzero
$t \leq 60$	87	586	0.071	0.353	0.068	0.350	0.104	0.394	0.109	0.394	0.069	0.314
$60 < t \leq 72$	39	489	0.053	0.354	0.047	0.352	0.078	0.393	0.085	0.391	0.055	0.323
$72 < t \leq 84$	45	375	0.055	0.345	0.052	0.343	0.080	0.380	0.087	0.381	0.055	0.320
$84 < t \leq 108$	54	682	0.046	0.321	0.044	0.319	0.074	0.354	0.077	0.356	0.048	0.303
$108 < t \leq 132$	48	581	0.031	0.332	0.030	0.330	0.057	0.368	0.055	0.369	0.038	0.315
$132 < t \leq 168$	18	593	0.030	0.337	0.027	0.335	0.051	0.365	0.058	0.367	0.035	0.327
$t > 168$	10	595	0.018	0.337	0.018	0.336	0.023	0.364	0.028	0.365	0.022	0.329

Note: This table provides the average R^2 values based on two types of benchmark models: the optimal fund-specific linear benchmark model determined using the method proposed by Bollen (2013), and the machine learning benchmark models. For each fund, the optimal linear benchmark model is determined through a step-wise regression process. This process selects up to 3 out of the 7 factors from the Fung and Hsieh (2004) model that yield the highest R^2 value. The average R^2 values are then computed based on these optimal linear benchmark models. Additionally, for each fund, a fund-specific benchmark model is estimated using the machine learning framework described in the text. The machine learning models considered in this analysis include Elastic Net with the same subset of factor(s) as the optimal model (*Enet*), Elastic Net with an extended set of factors (the original subset of factor(s) from the optimal model plus all two-way interactions of the factor(s)) (*Enet**), Random Forest with a proxy model (*RF*), and Bayesian Additive Regression Tree with a proxy model (*BART*). The table presents the results for both zero and nonzero R^2 funds in Panel A and Panel B, respectively. To determine whether a fund is classified as a zero-R^2 fund, each fund's optimal linear model R^2 value is compared to a null distribution of R^2 values using a simulation exercise. Funds with R^2 values below the 95th percentile of the null distribution are classified as zero-R^2 funds according to the method proposed by Bollen (2013). The remaining funds are classified as nonzero-R^2 funds. The reported results are organized by the number of observations in our sample.

Table A5
Portfolios Formed on Past Performance Measures by Category

Panel A – Directional

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	4.372	-5.725	9.288	1.925	4.042	-1.203	4.008	-3.318	2.263	-6.196
Std. Dev	9.427	14.642	12.814	15.818	11.306	16.795	12.006	17.062	13.961	18.725
SharpeRatio	0.464	-0.391	0.725	0.122	0.358	-0.072	0.334	-0.194	0.162	-0.331
SortinoRatio	0.737	-0.518	1.324	0.232	0.585	-0.108	0.477	-0.280	0.239	-0.468
MaxDrawdown	25.591	81.433	25.764	39.059	44.143	62.302	54.562	74.052	56.117	85.132
Max1MLoss	-14.709	-25.354	-16.678	-11.519	-14.302	-27.690	-24.346	-27.831	-22.948	-36.336
Alpha	0.274	-0.270	0.802	0.375	0.193	0.023	0.155	-0.345	0.120	-0.449
t value	(1.888)	(-1.036)	(3.616)	(1.472)	(1.004)	(0.075)	(0.829)	(-1.173)	(0.515)	(-1.376)
IR	0.428	-0.235	0.819	0.334	0.227	0.017	0.188	-0.266	0.117	-0.312
MPPM	1.286	-11.101	5.092	-3.388	0.388	-7.465	-0.101	-9.812	-2.555	-13.950

Panel B – Semidirectional

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	5.712	1.093	6.285	0.590	6.573	1.347	6.421	1.007	6.426	0.638
Std. Dev	7.046	4.801	5.618	7.604	5.882	7.274	5.784	6.732	6.401	7.424
SharpeRatio	0.811	0.228	1.119	0.078	1.118	0.185	1.110	0.150	1.004	0.086
SortinoRatio	1.380	0.396	1.979	0.131	1.937	0.324	1.821	0.242	1.613	0.143
MaxDrawdown	19.803	13.527	13.494	29.303	15.984	29.378	15.722	26.635	14.490	31.815
Max1MLoss	-9.323	-7.025	-6.770	-8.232	-6.837	-8.023	-7.825	-8.408	-9.505	-8.483
Alpha	0.252	0.179	0.354	0.111	0.351	0.191	0.355	0.111	0.344	0.120
t value	(4.017)	(2.187)	(5.496)	(0.921)	(5.365)	(1.795)	(5.333)	(0.948)	(4.658)	(1.007)
IR	0.910	0.495	1.245	0.209	1.215	0.407	1.208	0.215	1.055	0.228
MPPM	3.243	-0.955	4.094	-1.980	4.332	-1.151	4.194	-1.384	4.083	-1.892

Table A5
Portfolios Formed on Past Performance Measures by Category

Panel C – Nondirectional

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	3.545	0.912	4.060	3.396	4.283	4.113	3.693	4.771	5.004	2.869
Std. Dev	4.813	6.975	4.931	7.744	4.797	7.233	4.919	8.340	4.714	7.381
SharpeRatio	0.737	0.131	0.823	0.439	0.893	0.569	0.751	0.572	1.061	0.389
SortinoRatio	0.950	0.180	1.068	0.719	1.163	0.987	0.910	0.957	1.407	0.606
MaxDrawdown	15.247	36.098	14.432	19.307	14.666	25.054	14.432	22.544	13.423	28.634
Max1MLoss	-15.140	-16.943	-14.430	-14.078	-14.430	-13.327	-14.430	-13.931	-13.423	-13.411
Alpha	0.189	0.088	0.206	0.337	0.246	0.352	0.242	0.510	0.313	0.311
t value	(2.456)	(0.711)	(2.637)	(2.534)	(3.202)	(2.730)	(3.084)	(3.642)	(4.316)	(2.476)
IR	0.556	0.161	0.597	0.574	0.725	0.618	0.699	0.825	0.978	0.561
MPPM	1.451	-1.575	1.949	0.787	2.190	1.626	1.577	2.024	2.929	0.341

Panel D – Fund of Funds

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	2.680	0.204	4.779	2.843	4.534	3.454	4.939	3.331	4.640	3.377
Std. Dev	4.852	4.140	4.871	4.314	4.845	4.022	4.905	4.646	4.792	4.464
SharpeRatio	0.552	0.049	0.981	0.659	0.936	0.859	1.007	0.717	0.968	0.756
SortinoRatio	0.859	0.087	1.639	1.206	1.610	1.649	1.767	1.415	1.676	1.357
MaxDrawdown	18.851	22.700	16.036	8.356	16.100	12.009	15.102	10.795	16.385	8.440
Max1MLoss	-7.002	-3.618	-8.087	-4.310	-7.158	-3.871	-7.255	-3.984	-6.922	-4.640
Alpha	0.102	0.056	0.290	0.247	0.263	0.325	0.308	0.277	0.264	0.285
t value	(1.643)	(0.765)	(4.374)	(3.088)	(3.972)	(4.546)	(4.462)	(3.208)	(3.887)	(3.512)
IR	0.372	0.173	0.991	0.699	0.900	1.030	1.011	0.727	0.880	0.796
MPPM	0.616	-1.758	2.715	0.864	2.477	1.516	2.874	1.312	2.589	1.379

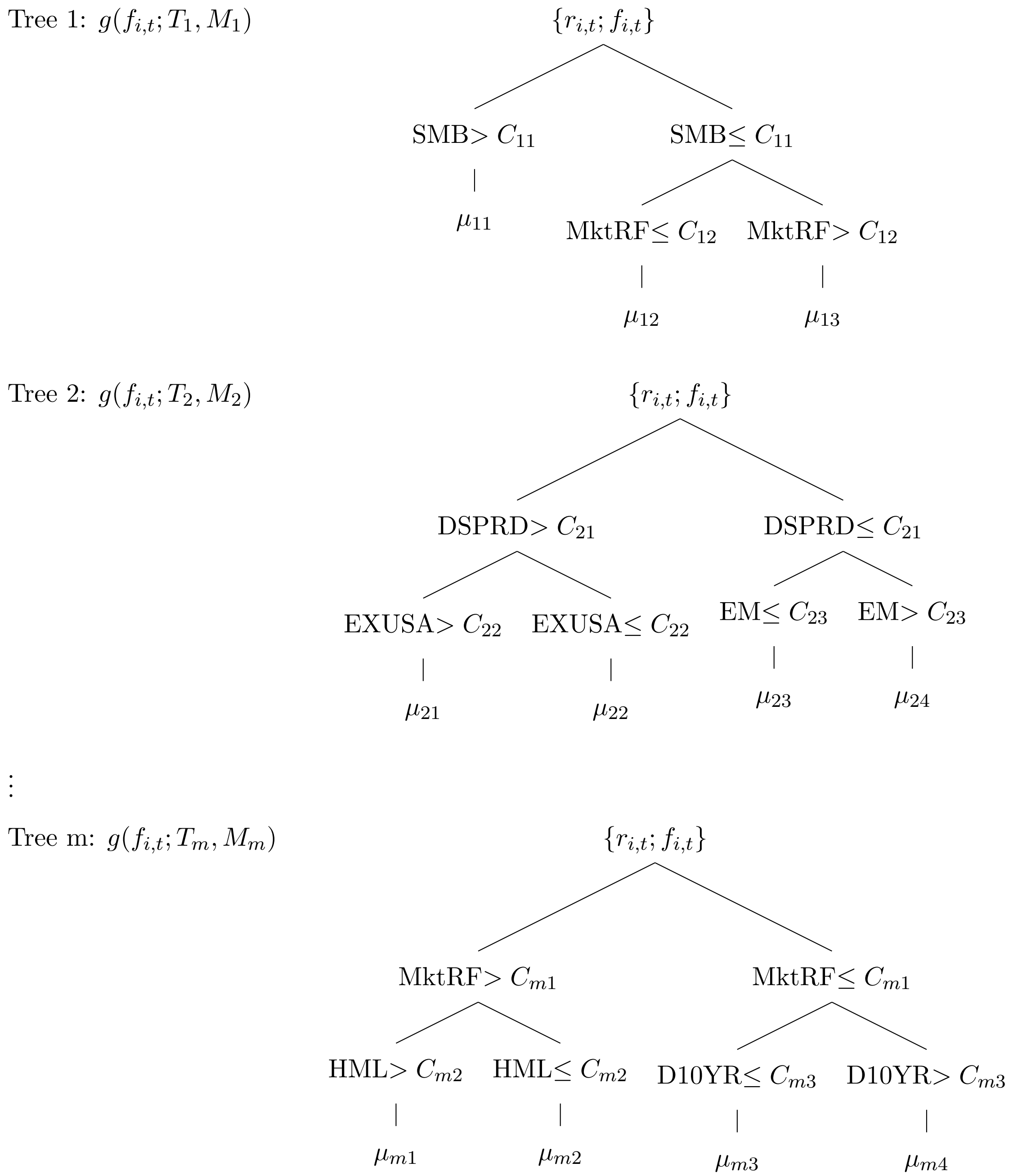
Table A5
Portfolios Formed on Past Performance Measures by Category

Panel E – Others

	FH		Enet		Enet*		RF		BART	
	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1	P10	P10-P1
MeanReturn	5.057	4.045	2.387	-1.848	1.992	-3.161	3.418	1.359	3.742	1.271
Std. Dev	4.491	11.168	4.435	11.913	3.752	11.143	4.037	12.964	4.727	12.404
SharpeRatio	1.126	0.362	0.538	-0.155	0.531	-0.284	0.847	0.105	0.792	0.102
SortinoRatio	1.895	0.651	0.678	-0.263	0.667	-0.474	1.118	0.194	1.090	0.190
MaxDrawdown	9.697	34.591	21.613	60.791	22.630	57.700	17.484	50.564	18.498	48.388
Max1MLoss	-9.413	-17.172	-8.714	-16.890	-8.532	-16.537	-8.714	-17.893	-8.714	-17.032
Alpha	0.426	0.334	0.195	-0.160	0.158	-0.255	0.276	0.149	0.313	0.098
t value	(5.093)	(1.599)	(2.318)	(-0.711)	(2.211)	(-1.226)	(3.595)	(0.613)	(3.469)	(0.420)
IR	1.172	0.368	0.533	-0.164	0.509	-0.282	0.827	0.141	0.798	0.097
MPPM	2.984	0.457	0.315	-5.680	0.005	-6.719	1.395	-2.788	1.622	-2.666

Note: The table reports the performance of portfolios within 5 fund categories: directional (Panel A), semidirectional (Panel B), nondirectional (Panel C), fund of funds (Panel D), and others (Panel E). Within each category, funds are grouped into decile portfolios based on the respective alphas derived from the Fung and Hsieh (2004) model, or machine learning models, including Elastic Net with the same set of 7 factors (*Enet*), Elastic Net with extended set of factors (the original 7 factors plus all two-way interactions of the factors) (*Enet**), Random Forest with proxy model (*RF*), and Bayesian Additive Regression Tree with proxy model (*BART*). The panels report the performance of the top decile portfolio (*P10*) and the high-minus-low hedging portfolio (*P10-P1*) constructed based on the t-statistic of estimated alpha over the prior 24 months.The performance measures include the annualized mean excess returns and standard deviations (expressed in percent). Also reported are the annualized Sharpe ratios, Sortino ratios, maximum drawdown, and maximum 1-month loss (all in percent), the annualized information ratios, the annualized Fung and Hsieh (2004) 7-factor alphas (expressed in percent, with t-statistics in parentheses), and the manipulation-proof performance measures (MPPM) of the portfolios, following the methodology outlined by Goetzmann et al. (2007), with a risk aversion level of 3.

B. Appendix: Figures



The return of fund i in month t:

$$\hat{r}_{i,t} = \sum_{j=1}^{m} g(f_{i,t}; T_j, M_j)$$

Figure A1. Example of Regression Trees: $g(f_{i,t}; T_j, M_j)$