Journal of Financial Economics 150 (2023) 103737

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Machine learning and fund characteristics help to select mutual funds with positive alpha ☆

Victor DeMiguel [a,*], Javier Gil-Bazo [b], Francisco J. Nogales [c], André A.P. Santos [d]

[a] Management Science and Operations, London Business School, United Kingdom
[b] Department of Economics and Business, Universitat Pompeu Fabra, Barcelona School of Economics, and UPF Barcelona School of Management, Spain
[c] Department of Statistics, Universidad Carlos III de Madrid, Spain
[d] CUNEF Universidad, Spain

ARTICLE INFO

Dataset link: https://doi.org/10.17632/rpgb99m5zy.3

JEL classification:
G11
G17
G23

Keywords:
Active asset management
Mutual-fund performance
Mutual-fund misallocation
Machine learning
Tradable strategies
Nonlinearities and interactions

ABSTRACT

Machine-learning methods exploit fund characteristics to select tradable long-only portfolios of mutual funds that earn significant out-of-sample annual alphas of 2.4% net of all costs. The methods unveil interactions in the relation between fund characteristics and future performance. For instance, past performance is a particularly strong predictor of future performance for more active funds. Machine learning identifies managers whose skill is not sufficiently offset by diseconomies of scale, consistent with informational frictions preventing investors from identifying the outperforming funds. Our findings demonstrate that investors can benefit from active management, but only if they have access to sophisticated prediction methods.

1. Introduction

Mutual-fund research consistently shows that the average active fund earns negative risk-adjusted returns (alpha) after transaction costs, fees, and other expenses (Sharpe, 1966; Jensen, 1968; Gruber, 1996; Ferreira et al., 2013). Moreover, although several studies document the existence of a subset of managers that outperform their benchmarks (Wermers, 2000; Barras et al., 2010; Fama and French, 2010; Kacper-

☆ Nikolai Roussanov was the editor for this article. A previous version of this manuscript was circulated under the title “Can Machine Learning Help to Select Portfolios of Mutual Funds?” We are grateful for detailed and constructive feedback from an anonymous referee and the editor. We are also grateful for comments from Eddie Anderson, Fahiz Baba-Yara, Paul Borochin, Andrea Buraschi, Joao Cocco, Pasquale Della Corte, Francisco Gomes, Jin Guo, Martin Haugh, Juan Imbet, Marcin Kacperczyk, Howard Kung, Narayan Naik, Jean Pauphilet, Anna Pavlova, Markus Pelger, Zhen Qi, Alexandre Rubesam, Stephen Schaefer, Henri Servaes, Raman Uppal, Michael Young, and Paolo Zaffaroni, as well as seminar participants at CUNEF, Imperial College Business School, London Business School, Universidad Autónoma de Madrid, Universidad de Zaragoza, University of Bath, University of Bristol and Université Paris Dauphine and conference participants at the 2021 conference of the French Finance Association, 2021 Finance Forum (Spanish Finance Association), 2021 EFMA annual meeting, 2021 conference of the Brazilian Finance Society, 2021 FMA annual meeting, 2021 INFORMS Annual Meeting, 2021 Paris December Finance Meeting, 2022 Funcas (Jornadas de Análisis Financiero y Big Data), 2022 Global Finance Conference, 2022 UF Research Conference on Machine Learning in Finance, 2023 Oxford-Man Institute Machine Learning in Quantitative Finance Conference, and 2023 SoFiE conference. Javier Gil-Bazo acknowledges financial support from the Spanish Government, Ministry of Science and Innovation grant PID2020-118541GB-I00, and Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S). Francisco J. Nogales acknowledges the financial support from the Spanish Government through project PID2020-116694GB-I00. André A. P. Santos acknowledges the financial support from the Comunidad de Madrid Government through project 2022-T1/SOC-24167.
* Corresponding author.
E-mail addresses: avmiguel@london.edu (V. DeMiguel), javier.gil-bazo@upf.edu (J. Gil-Bazo), fjnm@est-econ.uc3m.es (F.J. Nogales), andre.santos@cunef.edu (A.A.P. Santos).

https://doi.org/10.1016/j.jfineco.2023.103737
Received 19 September 2022; Received in revised form 12 October 2023; Accepted 12 October 2023
Available online 26 October 2023

czyk et al., 2014; Berk and Van Binsbergen, 2015), it is notoriously difficult to identify the outperforming funds ex ante. We show that machine-learning methods that exploit nonlinearities and interactions in the relation between fund characteristics and performance can help to construct tradable long-only portfolios of mutual funds that earn significant out-of-sample alphas net of all costs. Our results imply that investors can earn economically significant alpha by investing in active mutual funds, but only if they have access to sophisticated prediction methods that capture the complexity in the relation between fund characteristics and performance.

To understand the economic mechanism behind our results, we study whether the performance of our portfolios can be explained by capital misallocation in the mutual-fund market (Roussanov et al., 2021), and indeed find that nonlinear machine-learning methods select funds that are "too small" relative to their managers' skill. Thus, machine learning helps to select outperforming funds not only because it can identify skilled managers, but also because it can identify managers whose skill is not sufficiently offset by diseconomies of scale. This is consistent with informational frictions preventing investors from identifying some of the funds whose managers have the highest skill, and thus, these funds remaining small relative to their manager's skill. Our work implies that there is scope for pension-plan administrators and financial advisors to integrate machine learning with other tools in order to help investors select active mutual funds with positive alpha.

Passive funds have recently surpassed active funds in terms of assets under management in U.S. domestic equity mutual funds. Many interpret this victory of passive management as a result of the persistent inability of the average active manager to outperform cheaper passive alternatives (Gittelsohn, 2019). To determine whether at least some active managers outperform, researchers have investigated if future fund performance can be predicted using past returns. The consensus that emerges from this literature is that positive net alpha does not persist, particularly after accounting for the exposure of mutual-fund returns to the momentum factor (Carhart, 1997).[1]

Lack of persistence in fund net alpha is consistent with the model of Berk and Green (2004), in which investors supply capital with infinite elasticity to funds they expect to outperform, based on past returns. If there are diseconomies of scale in portfolio management, in equilibrium funds with positive past alpha attract more assets, and thus, earn the same expected net alpha as any other active fund: that of the alternative passive benchmark (zero). However, informational frictions may prevent investor flows from driving fund performance to zero (Dumitrescu and Gil-Bazo, 2018; Roussanov et al., 2021). Consequently, whether mutual-fund performance is predictable is ultimately an empirical question that has received considerable attention in the literature. Several studies have shown that mutual-fund characteristics can be used to predict fund performance; see Jones and Mo (2020) for a review. Typically, these studies rank funds every month or quarter on the basis of a mutual-fund characteristic. They then allocate funds to quintile or decile portfolios and evaluate the performance of *long-short* portfolios of funds. However, only a small subset of the mutual-fund characteristics considered in the literature can be used to select *long-only* portfolios of funds with positive alpha after transaction costs, fees, and other expenses. This is crucial because open-end funds cannot be easily shorted, and thus, investors can only benefit from active management via long-only portfolios of funds that deliver positive net alpha.

Our goal is to study whether investors can benefit from active management, and thus, we take on the challenge of identifying long-only portfolios of mutual funds with positive future alpha net of all costs. Our approach departs from the existing literature along three dimensions. First, we jointly exploit 17 mutual-fund characteristics to predict fund performance, which allows us to account for the complex nature of the problem. Fund performance is determined by a host of different factors including the manager multifaceted ability, portfolio constraints, manager incentives and agency problems, as well as fund trading costs, fees, and other expenses. Thus, it seems unlikely that using a single variable to predict performance would be as efficient as exploiting a large set of characteristics.

Second, we use three machine-learning methods to forecast fund performance: elastic net, gradient boosting, and random forests. These methods can accommodate irrelevant or highly correlated predictors, and thus, they allow us to consider multiple characteristics with lower risk of overfitting than Ordinary Least Squares (OLS). In addition, the two decision-tree based methods (gradient boosting and random forests) can exploit nonlinearities and interactions, and thus, they may uncover predictability that would be missed by linear methods such as elastic net or OLS. As a robustness test, Section IA.6 of the Internet Appendix considers also neural networks.

Third, we focus on identifying *tradable* portfolios of funds. In particular, we consider long-only portfolios of mutual funds, we construct the portfolios using exclusively past data, and we evaluate their future (out-of-sample) performance in terms of alpha net of fees, transaction costs, and other expenses. Finally, we employ a dynamic approach—the decision whether to exploit a fund characteristic is taken every time we rebalance the portfolio. By allowing for variation over time in the relation between characteristics and performance, our method can accommodate changes in the determinants of fund performance due to investor learning or shifts in market conditions.

We compare the out-of-sample and net-of-costs performance of the portfolios of funds constructed using the three machine-learning methods, OLS, and two naive strategies (equally weighted and asset-weighted portfolios of all funds). We use monthly data on the returns and 17 characteristics of no-load actively managed U.S. domestic equity mutual funds spanning the 1980 to 2020 period. We consider only no-load funds to ensure that our alphas are net of all costs. We use the first 10 years of data to train the three machine-learning methods and OLS to predict future annual net alpha, estimated using the five-factor model of Fama and French (2015) augmented with momentum. As predictors, we use lagged values of the 17 fund characteristics. We then form a long-only equally weighted portfolio of the funds in the top decile of predicted net alpha, and compute the net return of the portfolio in the following 12 months. For every remaining year, we expand the training sample forward by one year, construct a new top-decile portfolio, and track its net return for the next 12 months. This way, we construct a time series of monthly out-of-sample net returns of the top-decile portfolio spanning the period from 1990 to 2020. Finally, we evaluate the net alpha of the portfolio over the whole out-of-sample period with respect to four models: Carhart (1997) four-factor model; Fama and French (2015) five-factor model (FF5); FF5 augmented with momentum; and FF5 augmented with momentum and the liquidity factor of Pástor and Stambaugh (2003).

We highlight five findings. First, the two machine-learning methods that exploit nonlinearities and interactions (gradient boosting and random forests) select long-only portfolios of funds that earn statistically significant alphas net of all costs of 2.36% and 2.69% per year, respectively, relative to the FF5 model augmented with momentum. These alphas are also economically significant—for instance, they are more than double the average expense ratio in our sample (1.11%). In contrast, the portfolios based on the linear methods (elastic net and OLS) deliver annual net alphas of 1.09% and 1.21%, respectively, which are statistically indistinguishable from zero. The equally weighted and asset-weighted portfolios earn negative annual net alphas of −0.22% and −0.44%, respectively, consistent with existing evidence that the average active fund underperforms passive benchmarks after costs. Our findings are similar when we evaluate out-of-sample alpha using other factor models. In summary, while portfolios that exploit predictability in the data help investors to avoid underperforming funds, only the machine-learning methods that exploit nonlinearities and interactions—gradient

[1] A notable exception is the study of Bollen and Busse (2005), who find evidence of short-term (quarterly) persistence among top-performing funds.

boosting and random forests—allow them to earn significantly positive net alpha by investing in active funds.

Second, machine learning unveils nonlinearities and interactions in the relation between fund characteristics and future performance. The most important characteristics for the nonlinear machine-learning methods include various measures of past performance and fund activeness. We find that the relation between fund activeness and future performance is highly nonlinear, with the relation being strongly positive for the most active funds, but flat for the rest of the funds. The nonlinear methods also unveil important interactions between past-performance and fund-activeness measures. In particular, we find that, although investors may generally achieve higher net alpha by holding funds with good past performance, past performance is a particularly strong predictor of future performance for more active funds.

Third, given the importance of the interactions between past performance and fund activeness for the nonlinear machine-learning portfolios, we explore whether it is possible to achieve positive net alpha by double sorting funds across one measure of past performance and one measure of fund activeness. We find that, although it is possible to achieve positive net alpha by double sorting mutual funds, the performance of such double-sorted portfolios is quite sensitive to the particular measures of past performance and fund activeness considered. Moreover, we find that the relative predicting ability of the measures of past performance and fund activeness varies substantially over time, and thus, to achieve superior out-of-sample performance, investors should use machine learning dynamically to identify the characteristics and interactions that are important at each point in time using only past data.

Fourth, we build on the work by Roussanov et al. (2021) to study whether capital misallocation in the mutual-fund market explains the performance of the nonlinear machine-learning portfolios. Roussanov et al. (2021) estimate managerial skill using a Bayesian approach and find that funds in the top decile of the skill distribution are "too small" for diseconomies of scale to offset the skill of their managers. We compute the average net skill and fund size of the decile portfolios of funds generated by the four prediction methods and, consistent with Roussanov et al. (2021), we find that the top decile of funds are "too small" given the skill of their managers, with funds in the top decile of the two nonlinear machine-learning methods being particularly small. These findings provide an economic interpretation of our results: Machine learning helps to select mutual funds not only because it can identify skilled managers, but also because it can identify managers whose skill is not sufficiently offset by diseconomies of scale. This is consistent with a competition framework à la Berk and Green (2004) in which informational frictions prevent a substantial fraction of the investor population from identifying some of the funds whose managers have the highest skill, and thus, these funds remaining small relative to their manager's skill.

Fifth, Jones and Mo (2020) show that the ability of fund characteristics to predict performance has declined over time due to increased arbitrage activity and mutual-fund competition. Motivated by their work, we study how the alpha of the different portfolios varies from 1991 to 2020. We find that the three prediction-based portfolios (gradient boosting, random forests, and OLS) outperform the two naive portfolios (equally weighted and asset weighted) from 1991 to 2011. Consistent with Jones and Mo (2020), however, the performance of the prediction-based portfolios is similar to that of the naive portfolios from 2012 until 2018. Interestingly, all three prediction-based portfolios outperform the two naive portfolios in the last two years of our sample (2019 and 2020). We also find that the difference in the performance of the nonlinear machine-learning portfolios across different business-cycle and sentiment regimes is not statistically significant.

We check the robustness of our findings to considering various alternative methodological choices in the Internet Appendix. First, we show that our results are robust to considering the post-publication decay in predictability documented by McLean and Pontiff (2016). Second, our results continue to hold if we use other performance measures, such as alphas based on the factor models of Cremers et al. (2013), Hou et al. (2015), and Stambaugh and Yuan (2017). Third, the performance of the top-decile portfolio is just as good or even better if we exclude from our sample institutional share classes, which implies that our results are not driven by the presence of share classes targeted to sophisticated investors. Fourth, performance is only slightly weaker if we construct portfolios consisting of funds in the top 5% or 20% of the predicted alpha distribution. Fifth, if we extend the holding period to 24 months instead of 12 months, the performance of the top-decile portfolios selected by gradient boosting and random forests improves substantially. For instance, the annual net alpha for the random-forest portfolio is 4%. Sixth, we find that although neural networks can deliver portfolios with positive alphas, their alphas are systematically smaller and less significant than those obtained with gradient boosting and random forests. Seventh, the performance of the machine-learning portfolios is similar if we use a cross-validation method that accounts for time-series properties of the data. Eighth, the performance of the machine-learning methods does not decline if we invest in at most one share class per fund. Ninth, the performance of the machine-learning methods is similar if we use as a predictor the "value-added" characteristic proposed by Berk and Van Binsbergen (2015) estimated over a 36-month window instead of a 12-month window. Finally, the performance of the machine-learning methods is similar if we use alternative methods to impute missing observations of fund characteristics.

We emphasize two implications of our work for investment managers and regulators. First, the economically large positive net alphas that we document show that investors can benefit from active management in the mutual-fund industry, but only if they have access to the predictions of sophisticated nonlinear methods. Thus, our findings suggest that there is scope for managers of funds of funds, pension-plan administrators, financial advisors, and independent analysts to integrate machine learning with other tools in order to help investors select active mutual funds with positive alpha. This may help to improve the efficiency of capital allocation in the mutual-fund market. Second, we show that mutual-fund characteristics that do not require information on fund portfolio holdings are enough to predict positive alpha. This is particularly relevant given the recent debate on the SEC proposal to raise the asset threshold for mandatory portfolio disclosure (Form 13F) from US$ 100 million to US$ 3.5 billion (Aliaj, 2020). While information on portfolio holdings is potentially valuable to investors, it can also reveal portfolio strategies and reduce active managers' incentives to identify mispriced assets, which can be detrimental for market efficiency (Aragon et al., 2013; Shi, 2017). Our results imply that even if no information on portfolio holdings had been available during our sample period, our methods would have identified funds with positive net alpha on average.

Our work is related to the literature that documents associations between a single mutual-fund characteristic and fund performance (Jones and Mo, 2020). A strong association between a fund characteristic and performance does not guarantee that long-only portfolios of funds based on that characteristic earn positive net alphas. For instance, higher expense ratios are negatively associated with net fund alphas (in our sample, funds in the bottom decile of the expense-ratio distribution outperform funds in the top decile by 1% per year relative to the FF5 model augmented with momentum), but a portfolio that invests only in the cheapest funds does not outperform passive benchmarks in net terms. Thus, expense ratios help investors to avoid expensive underperforming funds, but not to select outperforming funds with positive net alphas. In fact, only seven of the 27 studies identified by Jones and Mo (2020) report positive and statistically significant in-sample Carhart (1997) alphas after fees and transaction costs for long-only portfolios of mutual funds (Chan et al., 2002; Busse and Irvine, 2006; Mamaysky et al., 2008; Cremers and Petajisto, 2009; Elton et al., 2011; Amihud and Goyenko, 2013; Gupta-Mukherjee, 2014). We contribute to this literature by showing that it is possible to select long-only portfolios of mutual funds with significant positive net alpha by exploiting multiple characteristics and using machine learning.

Our paper is related to an emerging literature that uses machine learning to predict fund performance. Wu et al. (2021) predict future *hedge-fund returns* by exploiting characteristics constructed from fund historical returns. Instead, we predict future *mutual-fund alphas* by exploiting both fund historical returns as well as other fund characteristics. Like us, Li and Rossi (2020) use machine learning to select portfolios of mutual funds, but a fundamental difference between the two papers is that they use disjoint sets of predictors: while Li and Rossi (2020) exploit data on *fund holdings* and *stock characteristics*, we exploit data on *fund characteristics*. Our findings complement theirs by showing that investors can select portfolios of mutual funds with positive net alpha by exploiting *solely* the information contained in fund characteristics. Kaniel et al. (2023) use neural networks to predict mutual-fund alpha using a comprehensive set of predictors that includes stock characteristics, fund characteristics, and macroeconomic variables. They not only corroborate our finding that fund characteristics predict performance, but also show that when fund characteristics are included as predictors, stock characteristics no longer help to predict alpha. A key distinguishing feature of our work is the focus on tradable portfolios of mutual funds, which allows us to study whether investors can actually benefit from active management. In particular, we identify long-only portfolios of mutual funds using exclusively past data, and evaluate their future (out-of-sample) performance net of all costs (including loads). Kaniel et al. (2023) focus on *long-short* portfolios of mutual funds, forecast performance using three-fold cross validation over the entire sample, and do not account for fund loads. Moreover, most of the predictability in *after-fee* alpha documented by Kaniel et al. (2023, Figure 6b) comes from the short leg of their long-short portfolios of funds.

Our paper is also related to studies that use Bayesian methods to construct optimal portfolios of mutual funds (Baks et al., 2001; Pástor and Stambaugh, 2002; Jones and Shanken, 2005; Avramov and Wermers, 2006; Banegas et al., 2013). Unlike these papers, we do not study how investors should allocate their wealth across funds given their preferences and priors about managerial skill and predictability. Instead, our goal is to identify active funds with positive alpha that investors can combine with passive funds to achieve better risk-return tradeoffs.

Finally, our paper is related to the growing literature that employs machine learning to address empirical problems in finance such as predicting global equity-market returns (Rapach et al., 2013); predicting consumer credit-card defaults (Butaru et al., 2016); measuring equity-risk premia (Gu et al., 2020; Chen et al., 2020a); detecting predictability in bond risk premia (Bianchi et al., 2021); building test assets that capture nonlinearities and interactions in asset pricing (Feng et al., 2020; Bryzgalova et al., 2019); forecasting inflation (Garcia et al., 2017; Medeiros et al., 2021), and studying the relation between investor characteristics and portfolio allocations (Rossi and Utkus, 2020). In the context of mutual funds, Pattarin et al. (2004), Moreno et al. (2006), and Mehta et al. (2020) employ machine learning to classify mutual funds by investment category, but they do not study fund performance. Chiang et al. (1996) and Indro et al. (1999) use neural networks to predict mutual-fund net asset value and return, respectively. While these authors focus on forecasting accuracy, our goal is to identify funds with superior performance.

2. Data

In this section, we describe the data we use in our analysis. Section 2.1 describes the sample data. Section 2.2 defines the 17 monthly mutual-fund characteristics that we consider. Section 2.3 explains how we transform these monthly characteristics to generate the annual target and predicting variables for the machine-learning methods.

2.1. CRSP sample data

We collect monthly information on U.S. domestic-equity mutual funds from the CRSP Survivor-Bias-Free US Mutual Fund database. To keep our analysis as close as possible to the actual selection problem faced by investors, we perform the analysis at the share-class level.[2] Moreover, we restrict our analysis to share classes that charge no front-end or back-end loads, and thus rebalancing our portfolios of mutual funds do not incur any costs. Our sample includes both institutional and retail share classes and spans from January 1980 to December 2020.[3]

We apply a few filters that are common in the mutual-fund literature. First, we include only share classes of actively managed funds, therefore excluding ETFs and passive mutual funds.[4] Second, we include only share classes of funds with more than 70% of their portfolios invested in equities. Third, to avoid previously documented biases in the CRSP database, we exclude observations of a share class before it reaches 36 months of age and before the first observation with at least US$ 5 million of Total Net Assets (TNA), see Elton et al. (2001) and Evans (2010). Our final sample contains 8,767 unique share classes, of which 7,921 correspond to diversified equity funds (representing 95% of aggregate TNA in the sample) and 846 to sector funds.

2.2. Mutual-fund characteristics

We construct a dataset of 17 share-class characteristics using readily available information on fund characteristics and historical returns. None of our characteristics requires information about portfolio holdings, and thus, our set of predictors is disjoint from that used by Li and Rossi (2020).

For the ith share class in the mth month, we obtain data on its *return* in excess of the risk-free rate net of expenses and transaction costs ($r_{i,m}$), *total net assets* ($TNA_{i,m}$), *expense ratio* ($ER_{i,m}$), and portfolio *turnover* ratio.[5] In addition, we compute the class *age* as the number of months since its inception; we estimate the monthly *flows* as the relative growth in the class TNA adjusted for returns net of expenses

$$flow_{i,m} = \frac{TNA_{i,m} - TNA_{i,m-1}\left(1 + r_{i,m}\right)}{TNA_{i,m-1}}; \tag{1}$$

we estimate the *volatility of flows* as the standard deviation of flows in the calendar year; and we compute the *manager tenure* in years.[6] All of these characteristics have been identified as predictors of mutual-fund performance (Chen et al., 2004; Rakowski, 2010; Jones and Mo, 2020).

Moreover, we obtain several characteristics associated with the time-series regression of share-class returns on the five Fama and French (2015) and momentum factors (hereafter, FF5+MOM). In particular, for each share class and month in our sample, we run a "rolling-window" regression of the share-class returns on the FF5+MOM factor returns for the previous 36 months.[7] We then compute *alpha t-stat* (the inter-

[2] Section IA.8 of the Internet Appendix shows that our findings are robust to investing in at most one share class per fund.

[3] Section IA.3 of the Internet Appendix shows that our results are robust to considering only retail classes and it also studies how the differences between retail and institutional classes affect the different prediction methods.

[4] We use the index-fund identifier from CRSP, index_fund_flag, to identify funds that aim to replicate an index. When the identifier is missing, we use the fund name to infer whether it is passively managed.

[5] We proxy for the risk-free rate using the one-month T-bill rate downloaded from Ken French's website.

[6] We cross-sectionally winsorize flows at the 1st and 99th percentiles; that is, each month we replace extreme observations that are below the 1st percentile or above the 99th percentile with the value of those percentiles. The computation of the standard deviation of flows is based on winsorized flows. For each calendar year, we require at least ten monthly flow observations to compute volatility of flows.

[7] To run each regression, we require at least 30 months of non-missing returns in the 36-month window.

Table 1

Share-class characteristics: Definitions. This table lists the 17 monthly mutual-fund share-class characteristics that we consider. The first column gives the name of each characteristic and the second column provides its definition.

Variable	Definition
realized alpha	Monthly realized alpha calculated using Equation (2)
flows	Monthly flows calculated using Equation (1)
value added	Monthly dollar value extracted by the fund's manager from asset market calculated using Equation (3)
volatility of flows	Standard deviation of monthly flows in calendar year
total net assets (TNA)	Total assets minus total liabilities at end of month
expense ratio	Annual expenses as percentage of assets under management
age (months)	Number of months since share-class's inception date
manager tenure (years)	Number of years since beginning of manager's mandate
turnover ratio	Minimum of annual aggregate sales and annual aggregate purchases divided by total net assets
alpha t-stat	Alpha t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
market beta t-stat	Market beta t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
profitability beta t-stat	Profitability beta t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
investment beta t-stat	Investment beta t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
size beta t-stat	Size beta t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
value beta t-stat	Value beta t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
momentum beta t-stat	Momentum beta t-stat from rolling-window regression on FF5+MOM factors for previous 36 months
R^2	R-squared from rolling-window regression on FF5+MOM factors for previous 36 months

cept scaled by its standard error) and *beta t-stats*. We use t-stats instead of raw alphas and betas as predictors to account for estimation error (Hunter et al., 2014). In addition, we use the R^2 from the FF5+MOM rolling-window regression as a predictor of fund performance, as proposed by Amihud and Goyenko (2013), who explain that R^2 is a measure of fund activeness because low-R^2 funds track the benchmark less closely.[8] We also compute the monthly realized alpha for the ith share class in the mth month ($\alpha_{i,m}$) as:

$$\alpha_{i,m} = r_{i,m} - \hat{\beta}_{MKT,i,m}\ MKT_m - \hat{\beta}_{SMB,i,m}\ SMB_m - \hat{\beta}_{HML,i,m}\ HML_m - \hat{\beta}_{RMW,i,m}\ RMW_m - \hat{\beta}_{CMW,i,m}\ CMW_m - \hat{\beta}_{MOM,i,m}\ MOM_m, \tag{2}$$

where MKT_m, SMB_m, HML_m, RMW_m, CMW_m, and MOM_m are the returns in month m of the five Fama-French and momentum factors, and $\hat{\beta}_{MKT,i,m}$, $\hat{\beta}_{SMB,i,m}$, $\hat{\beta}_{HML,i,m}$, $\hat{\beta}_{RMW,i,m}$, $\hat{\beta}_{CMW,i,m}$, $\hat{\beta}_{MOM,i,m}$ are the factor loadings of the ith share class excess return with respect to the FF5+MOM factors estimated using the 36-month estimation window ending in month $m-1$.

Finally, we use the realized alpha defined in Equation (2) to compute the *value added* for each class and month, which we define as in Berk and Van Binsbergen (2015):

$$value\ added_{i,m} = (\alpha_{i,m} + ER_{i,m}/12) \times TNA_{i,m-1}. \tag{3}$$

This variable captures the dollar value extracted by the fund's manager from the asset market.[9]

Table 1 lists the 17 share-class characteristics and their definitions, and Table 2 reports the mean, median, standard deviation, and number

Table 2

Share-class characteristics: Descriptive statistics. This table reports monthly descriptive statistics (mean, median, standard deviation, and number of class-month observations) for the mutual-fund share-class characteristics we consider. All variables are measured at the share-class level and correspond to U.S. domestic equity funds in the 1980 to 2020 period.

	Mean	Median	Standard deviation	Class-month observations
monthly return	0.86%	1.25%	5.23%	718,928
monthly realized alpha	-0.14%	-0.13%	2.22%	676,147
alpha t-stat	-0.431	-0.430	1.209	676,475
TNA (USD mill.)	679.9	97.4	2,593	719,398
expense ratio	1.11%	1.04%	0.52%	712,564
age (months)	145.7	117.0	109.8	719,398
flows	0.002	-0.004	0.094	718,734
manager tenure (years)	8.219	7.005	5.352	656,418
turnover ratio	0.790	0.550	1.141	711,568
volatility of flows	0.173	0.091	0.240	704,945
value added	-0.295	-0.016	37.233	669,727
market beta t-stat	16.667	15.064	10.591	676,475
profitability beta t-stat	-0.125	-0.125	1.463	676,475
investment beta t-stat	-0.444	-0.495	1.544	676,475
size beta t-stat	1.460	0.617	3.801	676,475
value beta t-stat	0.022	-0.081	2.195	676,475
momentum beta t-stat	0.009	0.026	1.878	676,475
R^2	0.907	0.944	0.122	676,475

of class-month observations for each characteristic. Consistent with the mutual-fund literature, we observe that the average share class in our sample has negative alpha and loads positively on the market factor. The average R^2 is 90.7%, which suggests that the FF5+MOM factors explain most of the time-series variation in equity mutual-fund returns. The total number of class-month observations varies across variables from 656,418 to 719,398.

2.3. Target and predicting variables

We now explain how we transform the 17 mutual-fund characteristics to generate the target and predicting variables for machine learning.

[8] Another popular measure of fund activeness is the active share of Cremers and Petajisto (2009). We do not use this measure because we rely only on fund characteristics that do not require information on mutual-fund holdings.

[9] Berk and Van Binsbergen (2015) estimate before-fee alpha by regressing fund gross returns on the gross returns of passive mutual funds tracking different indexes. In unreported analysis, we follow their approach and obtain similar results to those based on the FF5+MOM model.

First, we convert our sample from monthly to annual frequency because some of the characteristics are available only at the quarterly or annual frequency, and even some of the characteristics available at the monthly frequency are very persistent. For each calendar year, we compute annual realized alpha, value added, and flows as the average of their monthly values multiplied by twelve.[10] Flow volatility is already defined for each calendar year and we multiply it by square root of 12 to annualize it. For all other characteristics, we use their values in December of each year.

Second, like Green et al. (2017) we standardize each characteristic so that it has a cross-sectional mean of zero and a standard deviation of one. This ensures the estimation process of the machine-learning methods is scale invariant. We set missing observations of each standardized characteristic equal to its cross-sectional mean (zero). Section IA.10 of the Internet Appendix shows that our findings are robust to using an alternative imputation method for missing observations that exploits cross-sectional and time-series dependence in the data.

Third, we build our final dataset consisting of the target variable and the characteristics that we use as predictors when training the prediction methods. Our target variable is the share-class realized alpha in the calendar year. This choice is consistent with our goal to exploit share-class characteristics to generate positive alpha. In contrast, Li and Rossi (2020) use fund excess returns as their target variable, which allows them to study whether the returns of mutual funds can be predicted from the characteristics of the stocks they hold. The 17 characteristics we use as predictors are the following one-year-lagged standardized variables: realized alpha, alpha t-stat, TNA, expense ratio, age, flows, volatility of flows, manager tenure, value added, R^2, and the t-stats of the market, profitability, investment, size, value, and momentum betas.[11] Fig. 1 shows the correlation matrix of the target and predicting variables. The target variable has low correlation with lagged predictors. However, some predictors exhibit substantial correlations, with the highest absolute correlation being that between lagged flows and volatility of flows (61%).

3. Machine-learning methods

We use well-known software packages to implement the machine-learning methods—the interested reader can refer to their documentation for a detailed description of the methods.[12] Gu et al. (2020) also provide an extensive description of various machine-learning methods in the context of asset pricing. In the remainder of this section, we briefly describe the methods we consider and the five-fold cross-validation procedure we use to tune their hyper parameters.

We organize our data in panel structure, with years indexed as $t = 1, 2, \ldots, T$ and share classes as $i = 1, 2, \ldots, N_t$. As a benchmark, we use the ordinary least squares (OLS) method:

$$\min_{\theta} \sum_{t=1}^{T-1} \sum_{i=1}^{N_t} (\alpha_{i,t+1} - z'_{i,t}\theta)^2,$$

where $\alpha_{i,t+1}$ is the realized alpha of the ith share class in year $t+1$, $z_{i,t}$ is a K-dimensional vector of standardized characteristics for the ith share class in year t, and θ is the K-dimensional parameter vector. The OLS estimator of realized alpha, $z'_{i,t}\theta$, is a *linear* function of the share-class characteristics. Although OLS provides an unbiased and interpretable prediction, machine-learning methods often outperform OLS for data that exhibit high variance, nonlinearities, and interactions.

We consider three machine-learning methods: elastic net, random forests, and gradient boosting. *Elastic net* is a linear method, like OLS, but uses regularization to alleviate overfitting. To capture nonlinearities and interactions, we consider two types of ensembles of decision trees (*random forests* and *gradient boosting*), which often outperform the linear methods on structured (tabular) data like our mutual-fund database; see, for instance, Medeiros et al. (2021).

Another popular machine-learning method is neural networks, which tend to perform well on non-structured data or highly nonlinear structured data. To capture these nonlinearities, neural networks employ a large number of parameters, and hence, they require a large number of observations to deliver accurate estimates. Consequently, neural networks are not as well suited to our setting as ensembles of trees. Nonetheless, as a robustness check we evaluate the performance of feed-forward neural networks with up to three hidden layers in Section IA.6 of the Internet Appendix.[13]

3.1. Elastic net

Regularization is often employed to alleviate overfitting in datasets with a large number of predicting variables. The elastic net of Zou and Hastie (2005) uses both 1-norm and 2-norm regularization terms to *shrink* the size of the estimated parameters. The objective function for the elastic net, with two regularization terms, is:

$$\min_{\theta} \sum_{t=1}^{T-1} \sum_{i=1}^{N_t} (\alpha_{i,t+1} - z'_{i,t}\theta)^2 + \lambda\rho\, ||\theta||_1 + \lambda(1-\rho)\, ||\theta||_2^2, \tag{4}$$

where $||\theta||_1 = \sum_{k=1}^{K} |\theta_k|$ and $||\theta||_2 = (\sum_{k=1}^{K} \theta_k^2)^{1/2}$ are the 1-norm and 2-norm of the parameter vector θ, and λ and ρ are hyper parameters. The 1-norm term ($\lambda\rho\, ||\theta||_1$) can be used to control the sparsity of the estimated parameter vector θ and the 2-norm term ($\lambda(1-\rho)\, ||\theta||_2^2$) to increase its stability. For the case with $\rho = 0$, the objective function in (4) includes only the 2-norm term, and thus, elastic net is equivalent to ridge regression, which provides a dense estimator of the parameter vector θ. If, on the other hand, $\rho = 1$, the objective function includes only the 1-norm term, and a Least Absolute Sum of Squares Operator (LASSO) regression is performed, which provides a sparse estimator. We explain in Section 3.4 how we calibrate the two hyper parameters ρ and λ.

3.2. Random forests

Random forests are ensembles of decision trees formed by bootstrap aggregation (Breiman, 2001). Decision trees split a sample recursively into homogeneous and non-overlapping regions shaped like high-dimensional boxes. The procedure to generate these boxes is often represented as a tree, in which the sample is split at each node based on the characteristic that is most relevant at that particular node. The tree grows from the root node to the leaf nodes, and the prediction is the average value of the target variable for the observations in each leaf node.

Decision trees are highly interpretable, but their performance can be poor because of the high variance of their predictions. Random forests reduce the prediction variance by averaging across the predictions of numerous decision trees in a *forest*. The reduction in prediction variance

[10] We require at least ten monthly observations in a calendar year to compute annual realized alpha, value added, and flows in that year. Section IA.9 of the Internet Appendix shows that using a 36-month window to estimate value added instead of a 12-month window does not help to improve the performance of the different portfolios.

[11] The target variable and some predictors are not observable and must be estimated from the data. While this may pose a problem for inference, our goal is to predict future performance rather than conduct inference.

[12] Specifically, we use `glmnet`, `randomForest`, `xgboost`, and `h2o` packages to implement elastic net, random forests, gradient boosting, and neural networks, respectively. The documentation for these four packages can be found in Friedman et al. (2010), Liaw and Wiener (2002), Chen et al. (2020b), and LeDell et al. (2020), respectively.

[13] We have not considered other classes of machine-learning methods such as principal-component regression or partial least squares because they are typically outperformed by elastic net; see Elliott et al. (2013).

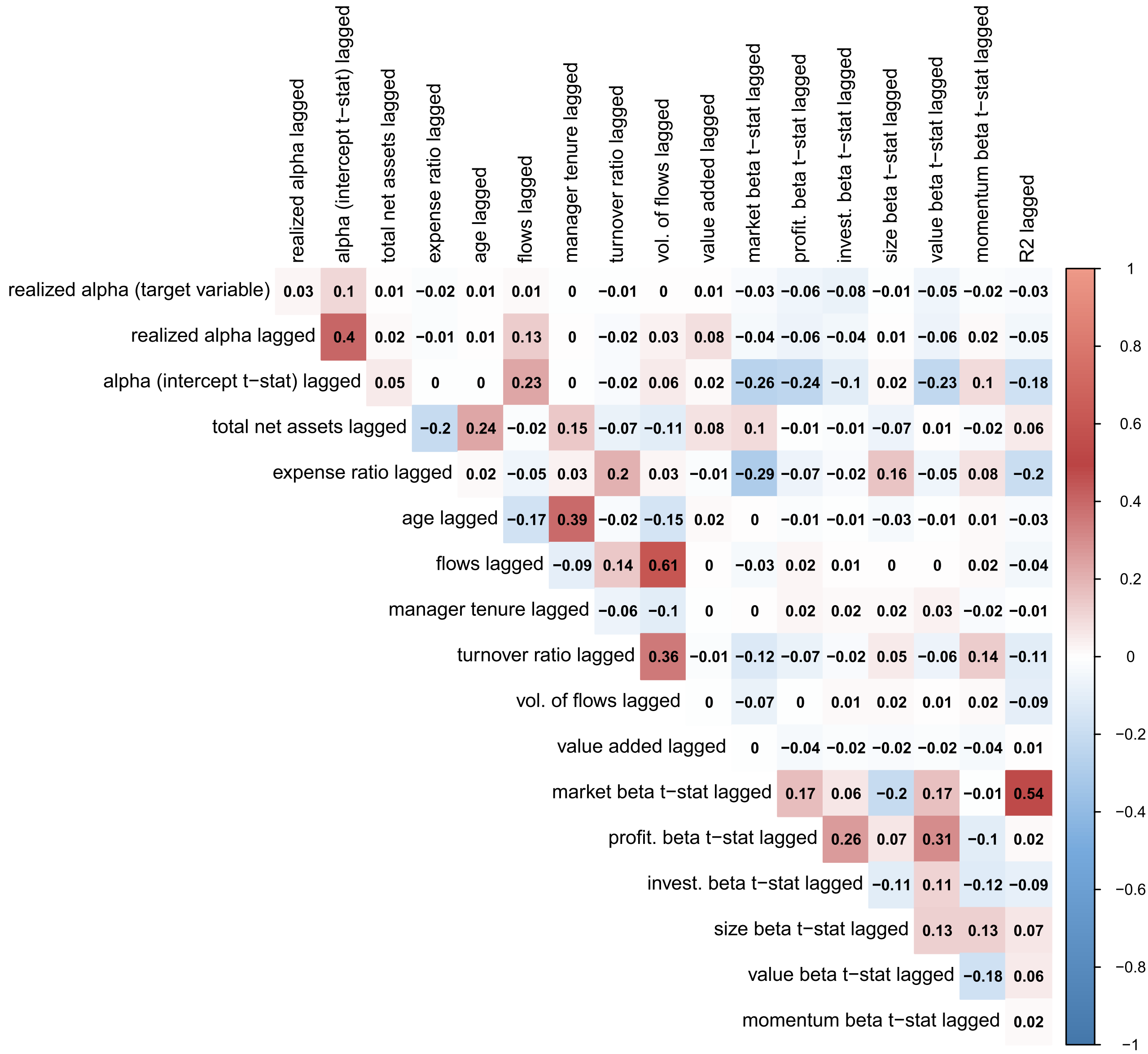

Fig. 1. Correlation matrix between the target variable and fund characteristics. This figure reports correlation coefficients between the target variable (annual realized alpha) and the 17 fund characteristics used as predictors. Predictors are lagged one year with respect to the target variable. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

is inversely related to the correlation between trees, and thus, ideally the trees should be uncorrelated. To accomplish this, random forests use bootstrap to select the observations for each tree, and consider a random subset of characteristics for each node.

Our random-forest method uses bootstrap with replacement to generate $B = 1{,}000$ samples from the original data. For each bootstrap sample, the method grows a decision tree by choosing a random subset of $m < K$ characteristics at each node, and choosing the best out of these m characteristics to split the sample. Section 3.4 discusses how we tune the hyper parameter m. The existing literature shows that random forests achieve good prediction performance, specially when there are many prediction variables and their relation to the target variable is nonlinear and contains interactions (Medeiros et al., 2021; Coulombe et al., 2020).

3.3. Gradient boosting

Gradient boosting uses ensembles of decision tress, but instead of aggregating independent decision trees like random forests, gradient boosting aggregates decision trees *sequentially* to give more influence to those observations that are poorly predicted by previous trees. As a result, the gradient-boosting method starts from weak decision trees (those with prediction performance only slightly better than random guessing) and converges to strong trees (better performance). In this fashion, boosting achieves improved predictions by reducing not only the prediction variance, but also the prediction bias (Schapire and Freund, 2012).

At each iteration of gradient boosting, a new decision tree is used to fit the *residuals* of the current ensemble of decision trees. Thus, this new decision tree gives more weight to those observations that are poorly predicted by the current ensemble. Then, gradient boosting updates the ensemble using the new decision tree. A key hyper parameter in gradient boosting is the learning rate, which determines the weight the ensemble gives to the most recent decision tree.

Unlike random forests, gradient boosting tends to overfit the data. To avoid overfitting, gradient boosting employs several regularization techniques that require tuning additional hyper parameters. For instance, gradient boosting often imposes constraints on the number of decision trees aggregated, the depth and number of nodes of each tree, and the minimum number of observations in a leaf node.

3.4. Cross validation of hyper parameters

For each estimation window, we tune the hyper parameters of the elastic net, random forests, and gradient boosting using five-fold cross-validation; see Hastie et al. (2009, Chapter 7). Specifically, we select a grid of possible values for the hyper parameters. We divide the sample into five equal intervals or "folds." For j from 1 to 5, we remove the jth fold and use the remaining four folds to obtain the predictions corresponding to the different values of the hyper parameters. We then evaluate the prediction error (or cross-validation error) of the prediction associated with each value of the hyper parameters on the jth fold. After completing this process for each of the five folds, we select the value of the hyper parameters that minimizes the average cross-validation error.

An alternative to k-fold cross validation that accounts for the time-series properties of the data is *time-series cross validation*, which reserves a section at the end of the training sample for evaluation. Section IA.7 of the Internet Appendix reports the results of a robustness check where we use time-series cross validation. We find that five-fold cross validation performs slightly better, consistent with Bergmeir et al. (2018) and Coulombe et al. (2020).

4. Performance of machine-learning portfolios

In this section, we first describe our performance-evaluation methodology and then compare the out-of-sample performance of the various portfolios.

4.1. Performance-evaluation methodology

We now describe the procedure we use to select share classes and evaluate the performance of the resulting portfolios. Although the analysis is carried out at the share-class level, for simplicity herein we refer to share classes as funds.

We use the first 10 years of data on one-year ahead realized alphas (from 1981 until 1990) and one-year-lagged fund characteristics (from 1980 until 1989) to train each machine-learning method and OLS. We then use the values of fund characteristics in December of 1990, which are not employed in the training process, to predict fund performance in 1991. We form an equally weighted portfolio of the funds in the top decile of the predicted-performance distribution and track its return (net of expenses, fees, loads, and transaction costs) in the 12 months of 1991. If, during that period, a fund that belongs to the portfolio disappears from the sample, the amount invested in that fund is equally distributed across the remaining funds. For every successive year, we expand the training sample forward one year, train the algorithm again on the expanded sample, make new predictions for the following year, construct a new top-decile fund portfolio and track its net return in the next 12 months. This way, we construct a time series of monthly out-of-sample net returns of the top-decile fund portfolio that spans from January 1991 to December 2020 (360 months). The average number of funds selected into the top-decile portfolios is 159 with a minimum of 11 and a maximum of 326.

To evaluate the out-of-sample performance of the top-decile fund portfolio, we run a time-series regression of the 360 out-of-sample monthly portfolio excess returns on contemporaneous risk-factor returns. The portfolio alpha is the intercept of the time-series regression. We consider four risk-factor models to evaluate portfolio performance: the Fama and French (1993) three-factor model augmented with momentum (FF3+MOM) proposed by Carhart (1997); the Fama and French (2015) five-factor model (FF5); the FF5 model augmented with momentum (FF5+MOM); and the FF5 model augmented with momentum and the aggregate liquidity factor of Pástor and Stambaugh (2003) (FF5+MOM+LIQ). Note however, that in all cases, fund selection is based on performance predicted according to the FF5+MOM model.

Table 3

Out-of-sample alpha of fund portfolios. This table reports the monthly out-of-sample alphas (in %) net of all costs of the top-decile fund portfolios obtained with three machine-learning methods (gradient boosting, random forests, and elastic net), with Ordinary Least Squares (OLS), and with two naive strategies (equally weighted and asset-weighted portfolios of all available funds). Alphas are computed by regressing the out-of-sample excess monthly portfolio returns net of all costs against the Fama and French (1993) three-factor model augmented with momentum (FF3+MOM), the Fama and French (2015) five factors (FF5), and the FF5 model augmented with momentum (FF5+MOM) and with the liquidity risk factor of Pástor and Stambaugh (2003) (FF5+MOM+LIQ). The out-of-sample period spans from January 1991 to December 2020. We report standard errors with Newey-West adjustment for 12 lags in parentheses. One, two, and three asterisks indicate that the alpha is significant at the 10%, 5%, and 1% level, respectively.

	FF3+MOM	FF5	FF5+MOM	FF5+MOM + LIQ
Gradient boosting	0.178**	0.222***	0.197**	0.198**
	(0.077)	(0.085)	(0.080)	(0.081)
Random forest	0.210**	0.263***	0.224**	0.226**
	(0.086)	(0.097)	(0.087)	(0.089)
Elastic net	0.044	0.075	0.091	0.098
	(0.065)	(0.067)	(0.069)	(0.068)
OLS	0.056	0.085	0.101	0.109*
	(0.063)	(0.065)	(0.066)	(0.066)
Equally weighted	-0.018	-0.007	-0.018	-0.017
	(0.045)	(0.045)	(0.044)	(0.045)
Asset weighted	-0.043	-0.033	-0.037	-0.036
	(0.036)	(0.035)	(0.035)	(0.036)

4.2. Out-of-sample and net-of-costs performance

Table 3 reports the out-of-sample alpha net of all costs of the top-decile fund portfolios selected by the three machine-learning methods—gradient boosting, random forests, and elastic net—and by OLS. For comparison purposes, we also report the alpha of two naive fund portfolios: an equally weighted and an asset-weighted portfolio of all share classes, both rebalanced annually.

Our main finding is that the two machine-learning methods that exploit nonlinearities and interactions (gradient boosting and random forests) select long-only portfolios of funds that deliver statistically significant net alphas of 19.7 bp and 22.4 bp per month (2.36% and 2.69% per year), respectively, relative to the FF5+MOM model. In contrast, the portfolios based on linear methods (elastic net and OLS) deliver net alphas of 9.1 bp and 10.1 bp per month (1.09% and 1.21% per year), respectively, which are statistically indistinguishable from zero. The equally weighted and asset-weighted portfolios earn negative net alphas of −1.8 bp and −3.7 bp per month (−0.22% and −0.44% per year), respectively. Interestingly, the asset-weighted portfolio underperforms the equally weighted portfolio, which implies that the average dollar invested in active funds earns lower risk-adjusted after-cost returns than the average fund. In summary, while portfolios that exploit predictability in the data help investors to avoid underperforming funds, only the machine-learning methods that exploit nonlinearities and interactions (gradient boosting and random forests) allow them to significantly benefit from investing in actively managed funds. Table 3 shows that these findings are remarkably stable when we evaluate out-of-sample alpha using the other three factor models we consider, with the only excep-

Table 4

Out-of-sample alpha with respect to OLS. This table reports the monthly out-of-sample alphas (in %) net of all costs of the portfolio that goes long in the funds selected by one of the methods we consider (gradient boosting, random forests, elastic net, equally weighted, asset weighted) and short in the funds selected by OLS. For instance, "gradient boosting minus OLS" refers to a long-short portfolio that is long on the prediction-based top-decile portfolio obtained with the gradient-boosting method and short on the top-decile portfolio obtained with the OLS method. Alphas are computed by regressing the out-of-sample excess monthly long-short portfolio returns net of all costs against the Fama and French (1993) three-factor model augmented with momentum (FF3+MOM), the Fama and French (2015) five factors (FF5), and the FF5 model augmented with the momentum factor (FF5+MOM) and with the liquidity risk factor of Pástor and Stambaugh (2003) (FF5+MOM+LIQ). The out-of-sample period spans from January 1991 to December 2020. We report standard errors with Newey-West adjustment for 12 lags in parentheses. One, two, and three asterisks indicate that the alpha is significant at the 10%, 5%, and 1% level, respectively.

	FF3+MOM	FF5	FF5+MOM	FF5+MOM + LIQ
Gradient boosting minus OLS	0.122***	0.136**	0.096**	0.089**
	(0.046)	(0.056)	(0.043)	(0.044)
Random forest minus OLS	0.154***	0.178**	0.123**	0.117**
	(0.053)	(0.069)	(0.050)	(0.051)
Elastic net minus OLS	-0.012	-0.010	-0.010	-0.010
	(0.011)	(0.011)	(0.011)	(0.010)
Equally weighted minus OLS	-0.074	-0.092*	-0.119**	-0.126***
	(0.048)	(0.052)	(0.048)	(0.048)
Asset weighted minus OLS	-0.100**	-0.118**	-0.137***	-0.145***
	(0.050)	(0.054)	(0.052)	(0.051)

tion being that OLS is statistically significant at the 10% level for the FF5+MOM+LIQ factor model.[14]

The positive net alphas achieved by the long-only portfolios of funds selected by gradient-boosting and random forests are also economically significant. For instance, the median of the *in-sample* alpha *spreads* between the top and bottom quintile portfolios of funds sorted by the predictors considered by Jones and Mo (2020, Table 2) is 21.91 bp per month (2.62% per year). Gradient-boosting and random forests achieve a similar net alpha *for long-only portfolios and out of sample*. Note also that the out-of-sample net alphas achieved by the portfolios of funds selected by gradient boosting and random forests are more than double the average expense ratio in our sample of active funds (1.11%). This means that if the average fund decided to cut down all fees and expenses to zero, it would only boost its net performance by less than half the size of the alpha we find for our best portfolios.

Our best method, random forests, selects a portfolio of mutual funds that earns a net alpha of 21 bp per month (2.52% per year) with respect to the FF3+MOM model, which is very similar to that of the best top-decile portfolio of Li and Rossi (2020, Table 4), 2.88% per year. This is somewhat surprising given that the two studies use disjoint sets of predictors: fund characteristics in our case, and stock characteristics combined with fund holdings in Li and Rossi (2020). Thus, our empirical findings complement those of Li and Rossi (2020) by showing that just like manager portfolio holdings, fund traits contain information that can be used to construct portfolios of funds with large positive alpha.[15] Moreover, our findings demonstrate that it is possible to select mutual funds with positive net alpha even in the absence of information on portfolio holdings, which is relevant for the debate on the costs and benefits of mandatory portfolio disclosure (Aliaj, 2020).

Although the alphas of the nonlinear machine-learning portfolios are significantly different from zero, it is unclear whether they are also significantly different from that of the OLS portfolio. To answer this question, we evaluate the performance of a self-financed portfolio that goes long in each machine-learning portfolio and short in the OLS portfolio. Table 4 shows that the difference in performance between the gradient-boosting and OLS portfolios is positive and significant, ranging from 8.9 bp to 13.6 bp per month (1.1% to 1.6% per year) with respect to the four factor models we consider. A similar conclusion holds for the random-forest portfolio, whose outperformance of the OLS portfolio ranges between 11.7 bp and 17.8 bp per month (1.4% and 2.1% per year) depending on the model. In contrast, the performance of the elastic-net portfolio is statistically indistinguishable from that of the OLS portfolio. Finally, both the equally weighted and asset-weighted portfolios underperform OLS, with the difference being generally statistically significant.

Our main goal is to identify funds with positive net alpha. The alpha of a fund measures its ability to improve the Sharpe ratio of an investor who already has access to the factors in the model (Gibbons et al., 1989). However, investors may choose to invest only in mutual funds instead of combining them with benchmark portfolios. Thus, it is interesting to study how the various portfolios of active funds perform in terms of mean return and risk. To answer this question, Table 5 reports the following measures for each portfolio of funds: mean excess net returns; standard deviation of net returns; Sharpe ratio (mean excess net return divided by standard deviation); Sortino ratio (mean excess net return divided by semi-deviation); information ratio (alpha net of all costs with respect to FF5+MOM model divided by idiosyncratic volatility); maximum drawdown; and value-at-risk (VaR) based on the historical simulation method with 99% confidence. The ranking of mean excess net returns closely mirrors the ranking in alphas. This result is far from obvious because the target variable we use to train the methods is fund alpha, and not fund excess returns, unlike the studies of Wu et al. (2021) and Li and Rossi (2020). Higher mean excess net returns for the prediction-based portfolios are at least par-

[14] Section IA.2 of the Internet Appendix shows that our findings are also robust to evaluating performance with respect to the factor models proposed by Cremers et al. (2013), Hou et al. (2015), and Stambaugh and Yuan (2017).

[15] Li and Rossi (2020, Sections 5.3 and 6.3) show that a linear combination of fund characteristics cannot improve the information contained in fund holdings and stock characteristics about future fund returns. Nonetheless, we show that using only fund characteristics with machine learning, one can construct portfolios of mutual funds with alphas similar to those obtained by exploiting fund holdings and stock characteristics.

Table 5

Out-of-sample mean excess return and risk. For each fund portfolio, this table reports the following monthly out-of-sample performance metrics: mean excess returns net of all costs; standard deviation; Sharpe ratio (mean excess return divided by the standard deviation); Sortino ratio (mean excess return divided by the semi-deviation); information ratio (alpha net of all costs with respect to FF5+MOM model divided by idiosyncratic volatility); maximum drawdown; and value-at-risk (VaR) based on the historical simulation method with 99% confidence. The last column reports the average annual portfolio turnover.

	Mean	Standard deviation	Sharpe ratio	Sortino ratio	Information ratio	Maximum drawdown	VaR 99%	Turnover
Gradient boosting	0.90%	4.71%	0.192	0.292	0.174	50.3%	12.0%	1.476
Random forest	0.93%	4.96%	0.188	0.290	0.163	55.4%	13.4%	1.410
Elastic net	0.81%	4.81%	0.168	0.249	0.075	58.3%	12.4%	1.219
OLS	0.82%	4.80%	0.170	0.253	0.083	58.5%	12.3%	1.218
Equally weighted	0.78%	4.39%	0.178	0.263	-0.029	51.4%	10.2%	0.414
Asset weighted	0.73%	4.42%	0.166	0.243	-0.069	52.8%	10.7%	0.369

tially explained by higher standard deviation. However, the two best methods in terms of alpha (gradient boosting and random forests) also deliver portfolios with the highest Sharpe ratios. Our conclusions do not change if we consider downside risk: gradient boosting and random forests select portfolios of funds with the highest Sortino ratio. In terms of maximum drawdown, the portfolios selected by elastic net and OLS appear to be the riskiest, and in terms of VaR, the equally weighted and asset-weighted portfolios are the safest. Finally, the relative performance of the different portfolios in terms of information ratio closely parallels that based on net alpha reported in Table 3.[16]

Although our measures of performance are net of all costs, it is useful to know how much trading the top-decile portfolios require. The last column of Table 5 reports the average annual turnover of the top-decile portfolios. Annual turnover is calculated at the beginning of each calendar year, when the portfolio is rebalanced, as the sum of the absolute values of changes in portfolio weights with respect to the last month of the previous year across all funds in the sample. For instance, a turnover value of one means that 50% of the wealth in the portfolio is reallocated across funds each year. As expected, the naive portfolios have very low turnover. Approximately, only 20% of the portfolio is reallocated from year to year due to changes in the pool of available funds and (for the equally weighted portfolio) also to changes in fund values. In contrast, managing a portfolio based on the performance predictions of elastic net and OLS involves trading roughly 60% of the portfolio value each year, whereas investing based on gradient boosting and random forests requires trading 70% of the portfolio value. These findings suggest that to achieve superior performance investing in actively managed funds, portfolio managers must also actively trade their wealth across these funds, and thus, it is important to account for fund loads when we evaluate portfolio performance.

Taken together, the results in this section suggest that it is possible to exploit readily available fund characteristics to select portfolios of mutual funds that significantly outperform (in terms of net alpha) the equally weighted or asset-weighted average mutual fund. This is true even if investors use the worst-performing forecasting methods, elastic net and OLS, to predict performance. In other words, elastic net and OLS help investors to avoid underperforming funds. However, neither elastic net nor OLS allow investors to identify funds with significant positive net alpha ex-ante. Only methods that allow for nonlinearities and interactions in the relation between fund characteristics and subsequent performance, namely gradient boosting and random forests, can detect funds with large and significant alphas. Moreover, the resulting portfolios also have the highest Sharpe, Sortino, and information ratios of all the portfolios considered.

5. Which characteristics and interactions matter?

We now study the *importance* of characteristics and their interactions for the performance of gradient boosting and random forests. We also analyze the *nature* of the nonlinearities and interactions exploited by these nonlinear machine-learning methods. Finally, we investigate whether it is possible to replicate the performance of the machine-learning portfolios by using a simple strategy based on double sorting funds across two of the most important characteristics.

To study the importance of characteristics, we estimate SHAP values (Lundberg and Lee, 2017). SHapley Additive exPlanations (SHAP) is a method based on cooperative game theory and used to estimate the contribution of each characteristic to each individual prediction. SHAP is an additive method because aggregating SHAP values across characteristics, one recovers the difference between the prediction for an individual observation and the average prediction across all observations.[17] Fig. 2 reports characteristic importance for OLS, elastic net, gradient-boosting, and random forests. To quantify the importance of a characteristic, we compute the mean across all observations of the absolute SHAP value for the characteristic. We evaluate importance within the last estimation window, which spans the 1980 to 2019 period.

We highlight two main findings from Fig. 2. First, value added, alpha intercept t-stat, market beta t-stat, and R^2 are among the top five most important characteristics for both nonlinear methods (gradient boosting and random forests). This demonstrates that the nonlinear machine-learning methods can exploit at least two different measures of past performance (alpha intercept t-stat and value added) to predict future alpha.[18] The nonlinear methods also exploit measures of fund activeness to predict future performance. To see this, note that market beta t-stat can be interpreted as a measure of fund activeness because one

[16] Note that there is a close relation between information ratio and alpha t-stat. In particular, equation (4) in Gibbons et al. (1989) implies that the alpha t-stat of a portfolio is proportional to its information ratio, with the proportionality constant depending on the number of observations and the maximum Sharpe ratio of the factors in the model. This explains why the relative performance of the different portfolios in terms of out-of-sample information ratio is similar to that in terms of alpha.

[17] The SHAP method is model-agnostic, applicable to any type of data, and provides additive interpretation (contribution of each characteristic to the prediction) of machine-learning models, including feature importance, feature dependence, interactions, clustering and summary plots. Moreover, the tree-based versions take into account the dependencies between characteristics (Lundberg et al., 2020). For these reasons, SHAP has recently become the method of choice to visualize feature importance and interactions. For a general discussion see Molnar (2019) and for applications in finance see Pedersen (2022) and Bali et al. (2023).

[18] Note that the other measure of past performance we consider (realized alpha) is only the eighth most important characteristic for gradient boosting and the twelfth for random forests, which demonstrates that alpha intercept t-stat and value added are much more important measures of past performance for our nonlinear methods. This finding contrasts with that of Kaniel et al. (2023), who find that their 12-month fund-momentum characteristic, which is closely related to our annual realized alpha, is the second most important predictor for their neural networks.

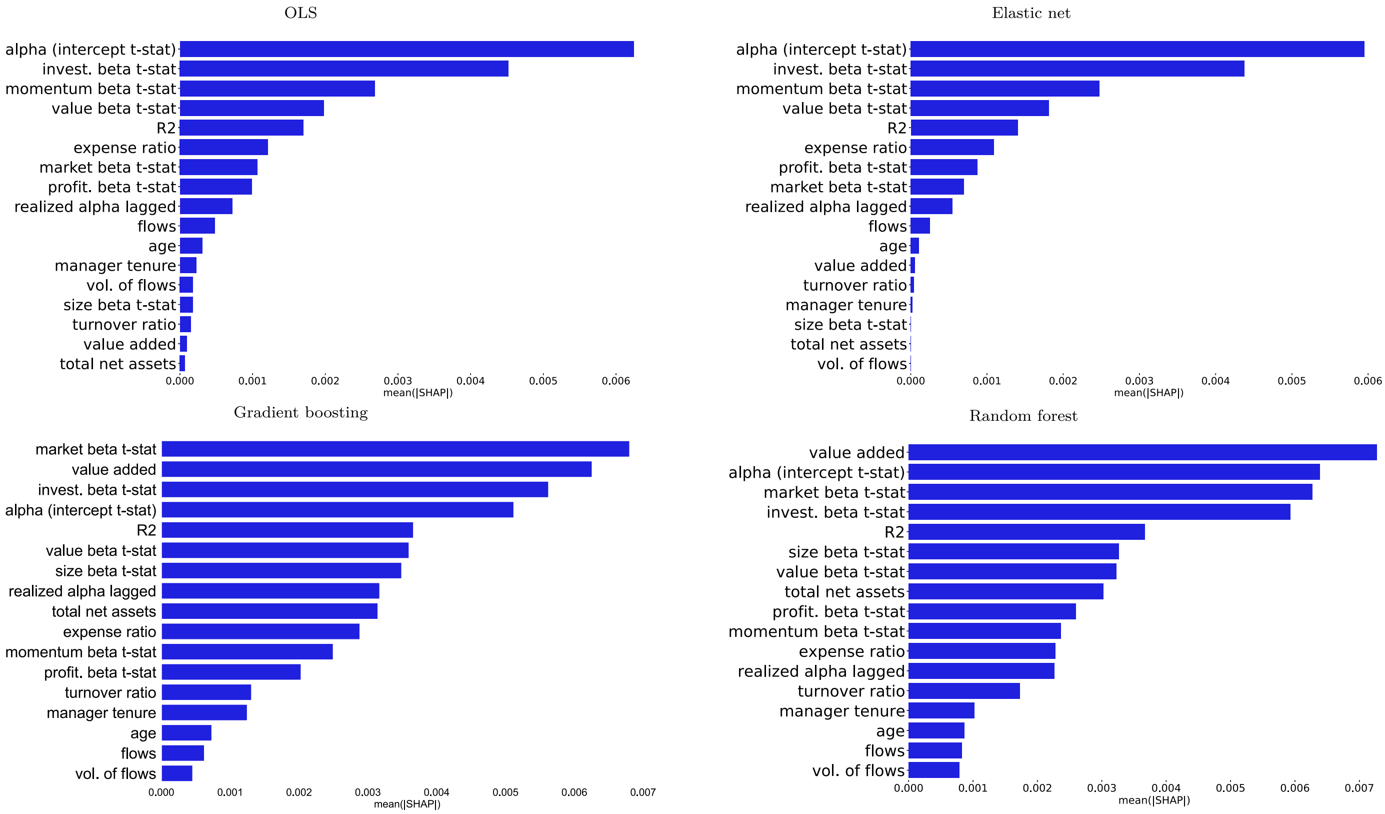

Fig. 2. Characteristic importance. This figure reports the importance of each characteristic measured as the average across all observations of the absolute SHAP value of the characteristic for ordinary least squares (OLS), elastic net, gradient boosting, and random forests. We compute characteristic importance for the last estimation window, which spans the period from 1980 to 2019.

would expect less active funds to have highly statistically significant betas on the market. Indeed, Fig. 1 shows that market beta t-stat has a high correlation of 54% with R^2, which Amihud and Goyenko (2013) consider as a measure of fund activeness.

Our second finding is that nonlinear and linear methods differ in characteristic importance. For example, for the two linear methods characteristic importance declines sharply beyond the two most important characteristics, but it declines much more gradually for the two nonlinear methods, for which around seven characteristics are similarly important. Another difference is that value added, which is one of the two most important characteristics for the nonlinear methods, is not very important for the linear methods. Finally, fund expense ratio is the sixth most important characteristic for the linear methods, but it is less important for the nonlinear methods.

The differences between nonlinear and linear methods in terms of both performance and characteristic importance suggest that there exist nonlinearities and interactions in the relation between characteristics and performance that investors can exploit to select actively managed equity funds. To explore the nature of these nonlinear relations, Figs. 3 and 4 display SHAP plots for four of the most important characteristics for gradient boosting and random forests: alpha intercept t-stat, value added, market beta t-stat, and R^2. For each SHAP plot, the horizontal axis shows the cross-sectionally standardized characteristic and the vertical axis the characteristic SHAP value for each observation (green dots) and the mean SHAP value conditional on the value of the characteristic (solid dark green line).[19]

Comparing Figs. 3 and 4, we find that the nonlinear patterns identified by the two machine-learning methods are very similar. In particular, the solid lines depicting the conditional mean SHAP value for each characteristic are quite similar across the two nonlinear methods.[20] Interestingly, we find that there is an approximately linear relation between alpha intercept t-stat and its conditional mean SHAP value. This may explain why alpha intercept t-stat is the most important characteristic for both linear methods, OLS and elastic net.[21] However, there is a substantial degree of nonlinearity in the relation between the other three characteristics, which are important mainly for the nonlinear methods, and predicted performance. For instance, we find that the relation between fund activeness and future performance is highly nonlinear, with the relation being strongly positive for the most active funds, but flat for the rest of the funds. In particular, we observe that very low standardized market beta t-stats predict superior performance, but the relation between market beta t-stat and future performance is flat for larger market beta t-stats. Similarly, consistent with Amihud and Goyenko (2013) there is an inverse relation between R^2 and performance for values of R^2 between −2.75 and −2, but the relation is roughly flat for values of standardized R^2 above −2. Finally, the relation between value added and its conditional mean SHAP value is flat for standardized value added below −0.06, u-shaped for intermediate value added, monotonically increasing for standardized value added between zero and 0.15, and decreasing above 0.15.

[19] To estimate the conditional mean SHAP value, we split the horizontal axis into a set of bins and compute the average SHAP value for each bin.

[20] Comparing Figs. 3 and 4, we also find that one difference between the two nonlinear methods is that the SHAP values for random forests are much more dispersed that those for gradient boosting. This is because, as explained in Section 3, while random forests employ ensembles of uncorrelated regression trees, gradient boosting employs a sequence of regression trees that build on each other, and thus, are potentially correlated.

[21] In unreported results, we also find that there is a linear relation between expense ratio and predicted alpha. This is not surprising as the expense ratio is linearly subtracted from gross alpha to obtain net alpha.

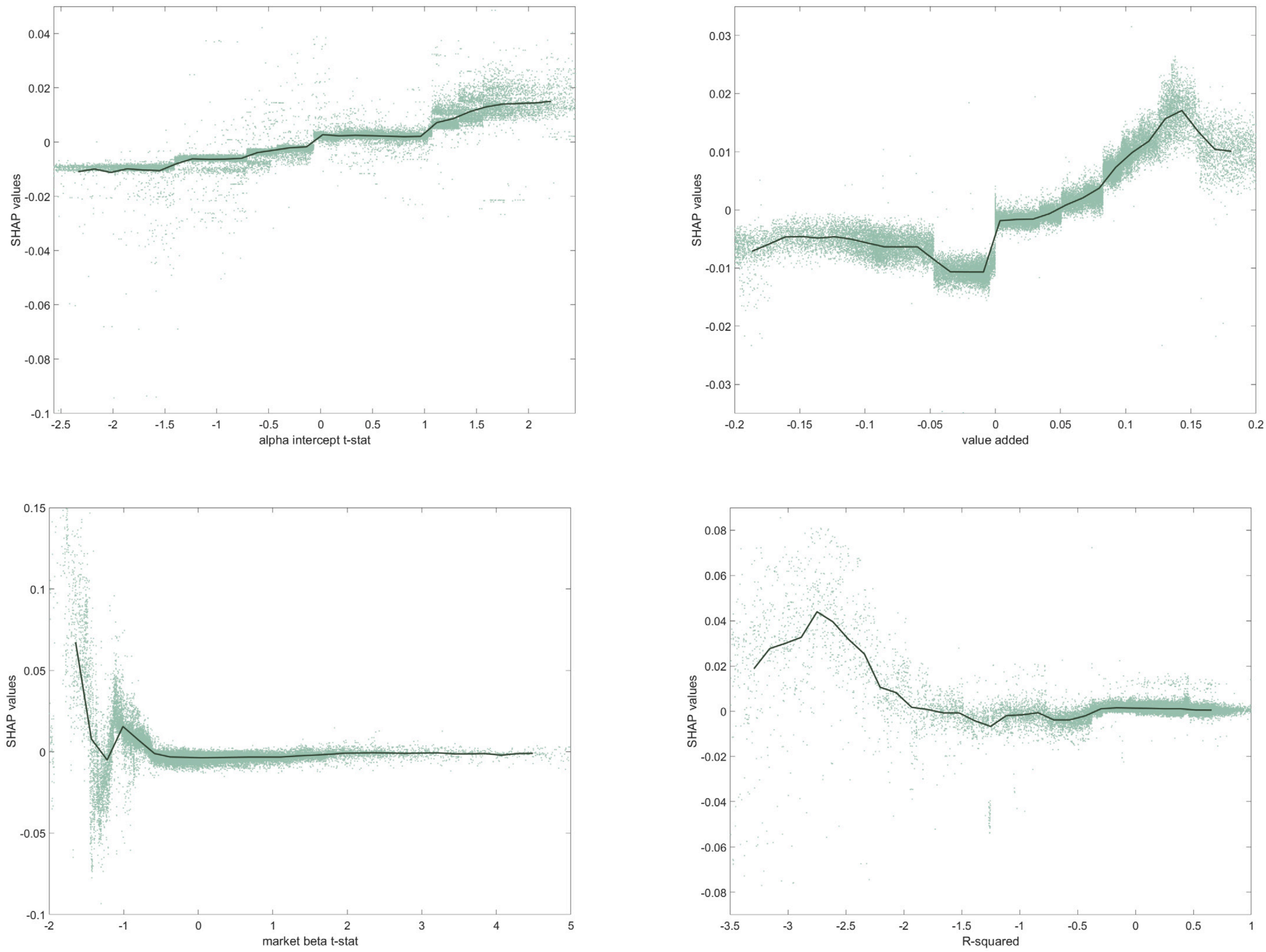

Fig. 3. Nonlinearity in the relation between fund characteristics and performance for gradient boosting. This figure displays SHAP plots for the gradient-boosting method corresponding to four characteristics: alpha intercept t-stat (top left graph), value added (top right graph), market beta t-stat (bottom left graph), and R^2 (bottom right). For each SHAP plot, the horizontal axis shows the cross-sectionally standardized characteristic and the vertical axis the characteristic's SHAP value for each observation (green dots) and the mean SHAP value conditional on the value of the characteristic (solid dark green line). Estimates are for the last estimation window spanning the period from 1980 to 2019.

We now turn our attention to interaction importance. Fig. 5 depicts the strength of the 30 most important interactions of characteristics for gradient boosting and random forests.[22] The figure reveals that past performance measures such as alpha intercept t-stat and value added are not only important as standalone predictors as shown in Fig. 2, but are also crucial through their interactions with measures of fund activeness such as market beta t-stat and R^2. For instance, the most important interaction for random forests is alpha intercept t-stat with market beta t-stat. Also, all four possible interactions between the two aforementioned measures of past performance and fund activeness are among the top 30 most important interactions.[23] Similarly, for gradient boosting three of the four possible interactions between the aforementioned measures of past performance and fund activeness are among the top 30. This suggests that the ability of fund past performance to predict future performance may depend on the activeness of the fund.

To further explore this conjecture, Figs. 6 and 7 illustrate the interaction between measures of past performance (alpha intercept t-stat or value added) and measures of fund activeness (market beta t-stat or R^2) for gradient boosting and random forests. For each interaction, we split all observations into deciles of the fund-activeness characteristic and depict, for each decile, the conditional mean SHAP value of the past-performance characteristic. For instance, the top-left graph in Fig. 6 illustrates the interaction between alpha intercept t-stat and market beta t-stat for gradient boosting. As expected, the SHAP values increase with alpha intercept t-stat for every decile of market beta t-stat, but the increase is much steeper for lower deciles of market beta t-stat (blue solid lines). That is, alpha intercept t-stat is a particularly strong predic-

[22] As mentioned before, SHAP values are additive across characteristics: aggregating SHAP values for each observation across the characteristics, we recover the difference between the prediction for each observation and the average prediction across all observations. Moreover, the SHAP value for each characteristic can also be decomposed into the pure effect of the characteristic and the SHAP *interaction* value of the characteristic with each of the other characteristics; see Molnar (2019, Section 9.6.8). Thus, the SHAP method estimates interaction strength by computing the mean across all observations of the absolute SHAP interaction value for each pair of characteristics.

[23] Note that there is a total of 136 pairwise interactions between the 17 characteristics in our dataset, and thus, all interactions among the top 30 are at the top quartile of importance.

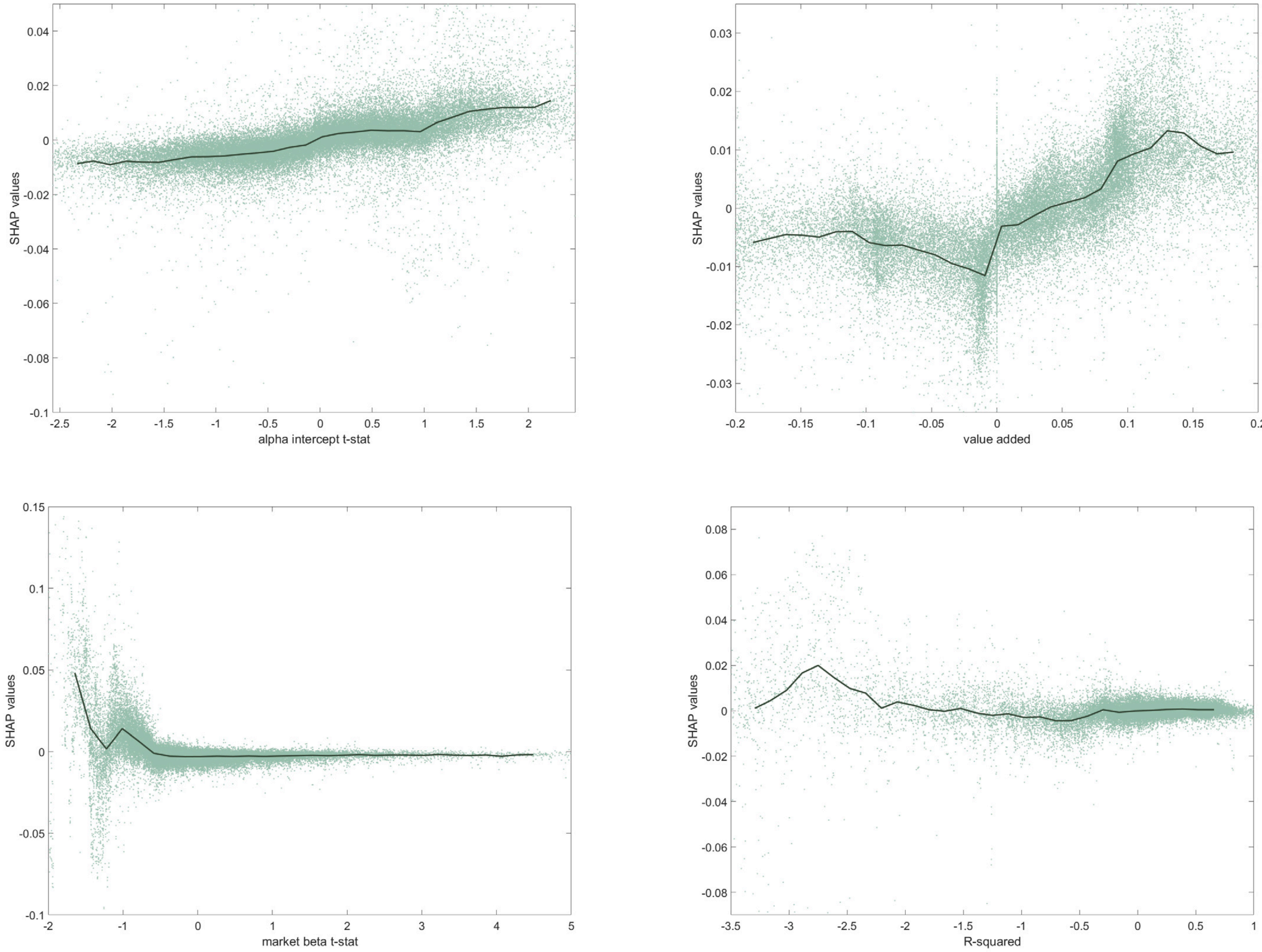

Fig. 4. Nonlinearity in the relation between fund characteristics and performance for random forests. This figure displays SHAP plots for the random-forest method corresponding to four characteristics: alpha intercept t-stat (top left graph), value added (top right graph), market beta t-stat (bottom left graph), and R^2 (bottom right). For each SHAP plot, the horizontal axis shows the cross-sectionally standardized characteristic and the vertical axis the characteristic's SHAP value for each observation (green dots) and the mean SHAP value conditional on the value of the characteristic (solid dark green line). Estimates are for the last estimation window spanning the period from 1980 to 2019.

tor of future performance for more active mutual funds. In other words, although investors may generally achieve higher net alpha by holding funds with good past performance, the effect is much stronger for more active funds. Similarly, the top-right graph in Fig. 6 shows that alpha intercept t-stat is particularly helpful to predict the future performance of funds with low R^2, that is, funds whose returns are not explained by common risk factors. The bottom-left and bottom-right graphs in Fig. 6 show that the effect of the interactions between value added and the two measures of fund activeness is similar, albeit weaker. Fig. 7 shows very similar effects for random forests.[24]

Given the importance of the measures of past performance and fund activeness and their interactions for the nonlinear machine-learning portfolios, it is interesting to study whether it is possible to earn positive net alpha by using a simple strategy based on double sorting funds across one measure of past performance and one measure of fund activeness. To do this, at the beginning of each year in our out-of-sample period, we first sort all funds in terms of the performance measure for the previous year and select funds that are above the top-$\sqrt{10}$th percentile. Second, we sort the selected funds in terms of the activeness measure at the end of the previous year and select funds below the bottom-$\sqrt{10}$th percentile.[25] This procedure results in a portfolio that contains 10% of the funds. Table 6 reports the monthly out-of-sample alphas net of all costs of the resulting long-only portfolios of funds obtained by combining one of two past-performance measures (alpha t-stat and value added) with one of two fund-activeness measures (R^2 and market beta t-stat).

[24] To understand the impact on portfolio composition of the nonlinearities and interactions exploited by machine learning, we compute the fund overlap for the portfolios of the four prediction methods averaged over the out-of-sample period. We find that while the fund portfolios selected by the two linear methods (OLS and elastic net) are very similar, with an average 94% fund overlap, the overlap between the portfolios of the two nonlinear methods and OLS is much smaller, around 45%. This shows that while the shrinkage of elastic net has negligible impact on portfolio composition, the nonlinearities and interactions exploited by gradient boosting and random forests lead to portfolios of funds that differ substantially from the OLS portfolios.

[25] Note that R^2 and market beta t-stat are *inverse* measures of fund activeness, and thus, we select funds below the bottom-$\sqrt{10}$th percentile of their distribution.

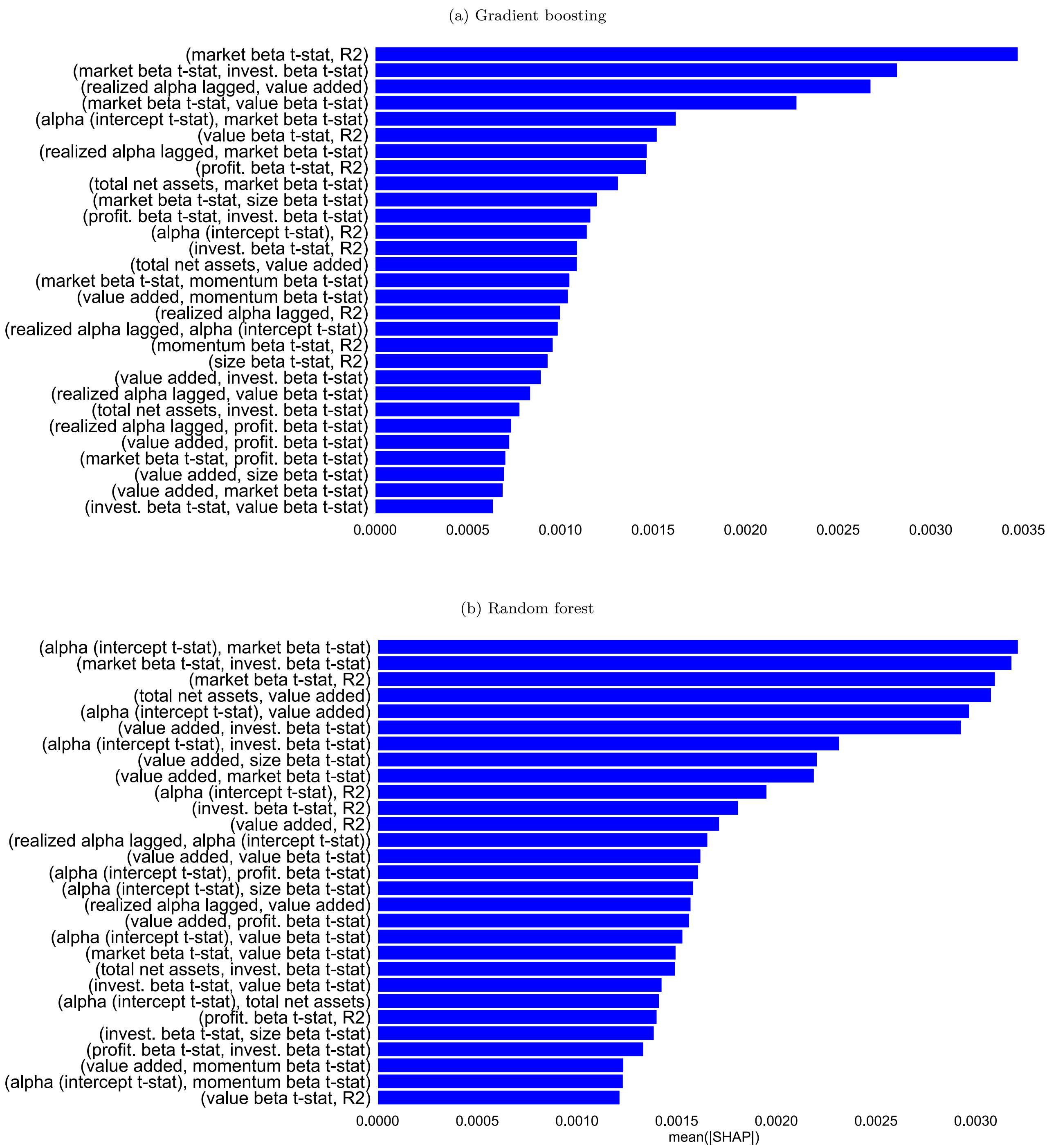

Fig. 5. Interaction importance. This figure reports the interaction strength of the 30 most important interactions for the gradient-boosting and random-forest methods. We compute interaction strength as the average across all observations of the absolute SHAP interaction value for each pairwise combination of characteristics. We compute interaction importance for the last estimation window, which spans the period from 1980 to 2019.

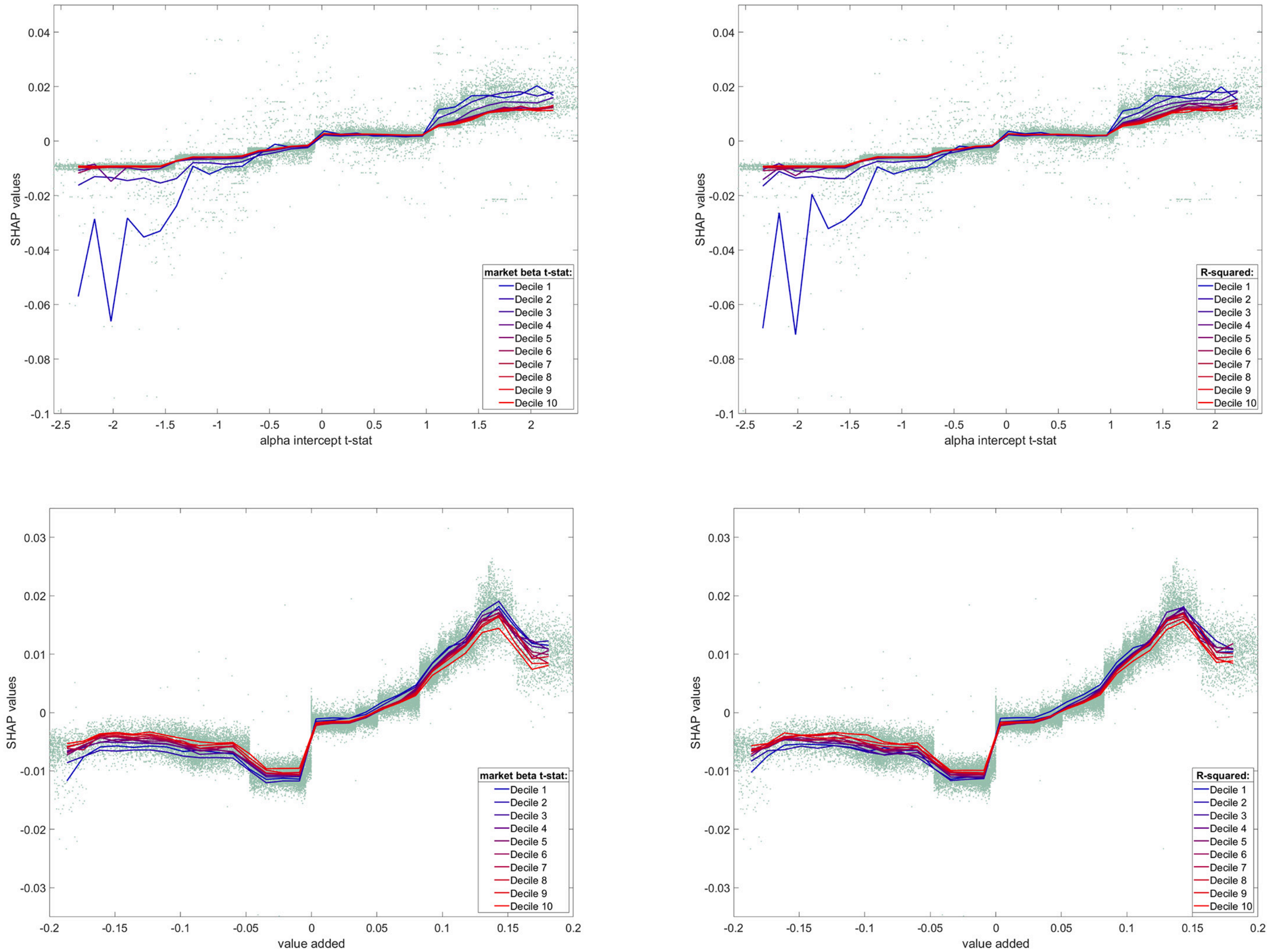

Fig. 6. Interactions between past performance and activeness measures for gradient boosting. Each graph illustrates the interaction between one past-performance characteristic (alpha intercept t-stat or value added) and one fund-activeness characteristic (market beta t-stat or R^2) for gradient boosting. For each graph, the horizontal axis depicts the cross-sectionally standardized past-performance characteristic and the vertical axis the characteristic's SHAP value for each observation (green dots). To visualize the interaction, we split all observations into deciles of the fund-activeness characteristic and depict, for each decile, the conditional mean SHAP value of the past-performance characteristic (solid lines). Estimates are for the last estimation window spanning the period from 1980 to 2019.

Table 6 shows that it is indeed possible to achieve positive net alpha by double sorting mutual funds based on past performance and fund activeness. For instance, the portfolios of funds based on a double sort of alpha t-stat and R^2 achieve alphas that are statistically significant at the 10% level, albeit slightly smaller than those attained by the nonlinear machine-learning methods in Table 3. Interestingly, the portfolios of funds based on a double sort of alpha t-stat and market beta t-stat achieve even higher alphas that are generally statistically significant at the 5% level and comparable in magnitude to those attained by the nonlinear machine-learning methods. This confirms the importance of the interaction of R^2 with measures of past performance as documented by Amihud and Goyenko (2013), but also reveals market beta t-stat as an alternative measure of fund activeness whose interaction with past performance helps to identify outperforming funds. However, Table 6 also shows that the performance of the portfolios of funds based on the double sorts is quite heterogeneous across different pairs of characteristics. For instance, the out-of-sample net alphas of the double-sorted portfolios based on *value added* and either market beta t-stat or R^2 are not significantly different from zero, and their magnitude is substantially smaller than those of the nonlinear machine-learning portfolios. Moreover, it is important to note that the results in Table 6 suffer from look-ahead bias because the pairs of characteristics for the double sort have been selected based on characteristic and interaction importance computed using the entire sample. The results in Table 6 demonstrate that although the portfolios obtained from a simple double sort can achieve good out-of-sample performance, investors should resort to nonlinear machine-learning methods in order to identify the relevant characteristics and interactions at each point in time (based only on past data) and achieve good performance in real time.

To investigate whether the predictive ability of some characteristics changes over time, Figs. 8 and 9 depict the importance of each predictor in each year of the out-of-sample period for gradient boosting and random forests, respectively. Figs. 8 and 9 exhibit some remarkable similarities, which suggests that the two methods identify similar patterns in the data. More importantly, the figures show that the importance of characteristics such as alpha t-stat, value added, and R^2 varies substantially over time.

Overall, our findings suggest that various measures of past performance and fund activeness and their interactions are important for the ability of the nonlinear machine-learning portfolios to achieve significant positive net alphas. We also find that, although it is possible to achieve positive net alpha by double sorting mutual funds based on

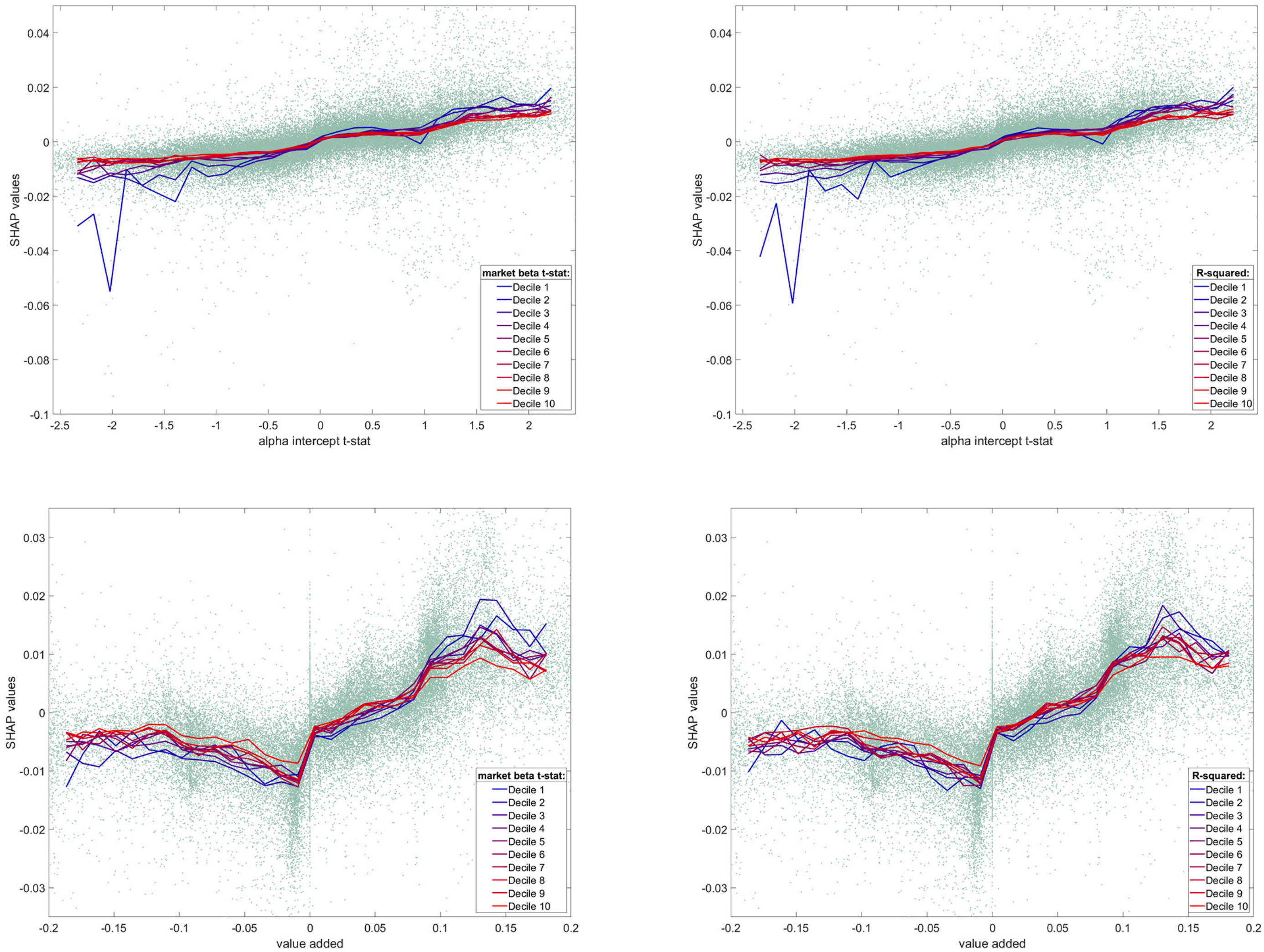

Fig. 7. Interactions between past-performance and activeness measures for random forests. Each graph illustrates the interaction between one past-performance characteristic (alpha intercept t-stat or value added) and one fund-activeness characteristic (market beta t-stat or R^2) for random forests. For each graph, the horizontal axis depicts the cross-sectionally standardized past-performance characteristic and the vertical axis the characteristic's SHAP value for each observation (green dots). To visualize the interaction, we split all observations into deciles of the fund-activeness characteristic and depict, for each decile, the conditional mean SHAP value of the past-performance characteristic (solid lines). Estimates are for the last estimation window spanning the period from 1980 to 2019.

past performance and fund activeness, the performance of such double-sorted portfolios is heterogeneous across different pairs of characteristics. Moreover, the relative predicting ability of the measures of past performance and fund activeness varies substantially over time, and thus, to achieve superior out-of-sample performance, investors should use machine learning dynamically to identify the characteristics and interactions that are important at each point in time.

6. Capital misallocation and machine learning

To investigate the economic mechanism behind our results, we now build on the work by Roussanov et al. (2021) and study whether capital misallocation in the mutual-fund market can explain the performance of the nonlinear machine-learning portfolios. To do this, we compute the average net skill and size of funds in the decile portfolios generated by the four prediction methods. Our main finding is that funds in the top decile are "too small" for diseconomies of scale to completely offset the skill of their managers, with funds in the top decile generated by the nonlinear methods being particularly small. This provides an economic interpretation of our results: Nonlinear machine-learning methods help to select outperforming mutual funds, not only because they can identify skilled managers, but also because they can identify managers whose skill is not offset by diseconomies of scale.

In the perfectly competitive equilibrium of Berk and Green (2004), fund size is such that diseconomies of scale and fees completely offset the manager's ability to generate gross alpha, and thus, expected net alpha is zero for every fund. However, Roussanov et al. (2021) show that, in a structural model where investors face informational frictions, funds do not necessarily reach their Berk and Green (2004) equilibrium size. Consequently, in expectation a subset of funds may earn positive net alpha while others may earn negative net alpha. Using data on U.S. active domestic equity funds from 1964 to 2015, Roussanov et al. (2021) employ a Bayesian approach to estimate managerial skill and find that about 80% of funds manage assets above their efficient size, while funds in the top decile of skill are "too small" relative to their manager's skill.

Following Roussanov et al. (2021), we assume that the net alpha of a fund can be decomposed into skill, diseconomies of scale, expense ratio, and a zero-mean idiosyncratic shock. Thus, the expected net alpha of fund i can be written as:

$$E\left(\alpha_{i,t+1} | \mathcal{F}_t\right) = \hat{a}_{i,t+1} - D(Q_{i,t}) - p_{i,t}, \tag{5}$$

Table 6

Out-of-sample alpha of double-sorted portfolios. This table reports the monthly out-of-sample alphas (in %) net of all costs of the portfolio of funds obtained by double sorting the funds in terms of past performance and fund activeness. Specifically, at the beginning of each year in the out-of-sample period, we sort all funds in terms of the performance measure for the previous year and select funds that are above the top-$\sqrt{10}$th percentile. Second, we sort the remaining funds in terms of the activeness measure at the end of the previous year and select funds below the bottom-$\sqrt{10}$th percentile. This procedure results in a portfolio that contains 10% of the funds. We consider two past-performance measures (alpha t-stat and value added) and two fund-activeness measures (R^2 and market beta t-stat). The portfolio alphas reported in the table are computed by regressing the out-of-sample excess monthly portfolio returns net of all costs against the Fama and French (1993) three-factor model augmented with momentum (FF3+MOM), the Fama and French (2015) five factors (FF5), and the FF5 model augmented with the momentum factor (FF5+MOM) and with the liquidity risk factor of Pástor and Stambaugh (2003) (FF5+MOM+LIQ). The out-of-sample period spans from January 1991 to December 2020. We report standard errors with Newey-West adjustment for 12 lags in parentheses. One, two, and three asterisks indicate that the alpha is significant at the 10%, 5%, and 1% level, respectively.

Double sort on	FF3+MOM	FF5	FF5+MOM	FF5+MOM +LIQ
Alpha t-stat and R^2	0.179**	0.195**	0.177*	0.179*
	(0.088)	(0.097)	(0.090)	(0.092)
Alpha t-stat and market beta t-stat	0.181*	0.235**	0.207**	0.211**
	(0.096)	(0.108)	(0.099)	(0.100)
Value added and R^2	0.109	0.154	0.113	0.111
	(0.091)	(0.102)	(0.095)	(0.096)
Value added and market beta t-stat	0.110	0.181	0.137	0.136
	(0.098)	(0.113)	(0.102)	(0.104)

where $\hat{a}_{i,t+1} = E\left(a_{i,t+1}|\mathcal{F}_t\right)$ is the expected skill of fund i conditional on the information set $\mathcal{F}_t$, $D(Q_{i,t})$ is the impact of diseconomies of scale given the size of fund i at time t, $Q_{i,t}$, and $p_{i,t}$ is the expense ratio of fund i at time t, which, given the persistence of fund expense ratios, is a reliable predictor of the expense ratio at time $t+1$. Roussanov et al. (2021) further assume that the diseconomies of scale are logarithmic, $D(Q_{i,t}) = \eta \log\left(Q_{i,t}\right)$. Thus, in the perfectly competitive equilibrium of Berk and Green (2004), the efficient size of fund i should satisfy $\log\left(Q_{i,t}^{BG}\right) = (\hat{a}_{i,t+1} - p_{i,t})/\eta$, where $\hat{a}_{i,t+1} - p_{i,t}$ is the net skill of fund i at time $t+1$.

To estimate the expected skill for fund i in year t, $\hat{a}_{i,t+1}$, we follow Zhu (2018) and average the fund's (annual) realized alphas before fees and diseconomies of scale from the fund's inception. We compute the diseconomies of scale as $D(Q_{i,t}) = \eta \log\left(Q_{i,t}\right)$ where $\eta = 0.0048$, as estimated by Roussanov et al. (2021), and $Q_{i,t}$ equals the assets under management of all of the fund's share classes at the end of year t, expressed in 2015 dollars.[26]

Fig. 10 illustrates capital misallocation for the decile portfolios generated by the four prediction methods. For the jth decile portfolio of funds ranked by predicted alpha, the horizontal axis gives the mean net skill, $E(\hat{a}_{i,t+1} - p_{i,t}|i \in D_j)$, where D_j is the set of funds in the jth decile, and the vertical axis the mean log size, $E(\log\left(Q_{i,t}\right)|i \in D_j)$. The colored lines plot the mean log size for each decile portfolio generated by OLS (orange stars), elastic net (yellow squares), gradient boosting (purple crosses), and random forests (green diamonds). For every method, the first decile portfolio has the lowest net skill and mean log size. We also plot the efficient (Berk-Green) log size, $\log\left(Q_{i,t}^{BG}\right)$, for each level of net skill (straight black line).

Fig. 10 shows that mean net skill increases monotonically for the decile portfolios of all four prediction methods; that is, the four prediction methods identify managers with higher net skill. The figure also shows that fund size also increases monotonically for the bottom nine decile portfolios, consistent with investors being generally able to identify funds with higher net skill. However, we observe that funds in the top decile of alpha predicted by all four methods manage on average substantially smaller portfolios than funds in the second-best decile. This pattern is particularly striking for funds in the top decile of alpha predicted by the two nonlinear machine-learning methods (gradient boosting and random forests), which are surprisingly small with size similar to that of funds in the bottom fourth decile of the predicted alpha distribution.

These findings suggest that informational frictions prevent investors from identifying some of the funds whose managers have the highest net skill, and thus, these funds remain small relative to their manager's skill. Comparing the mean log size of the decile portfolios of the four prediction methods to the straight black line that depicts the efficient (Berk and Green) log size, we observe that our findings are largely consistent with those of Roussanov et al. (2021) despite the different methodologies employed in the two papers. Funds in the bottom 80% of the predicted net alpha distribution are "too large" for their estimated skill while funds in the top 10% of the distribution are below their efficient size.

Overall, the findings in this section suggest that the conclusions of Roussanov et al. (2021) regarding capital misallocation in the U.S. mutual-fund industry are robust to the method of finding misallocated funds. Moreover, the findings provide an economic interpretation of our results. Nonlinear machine-learning methods help to select mutual funds not only because they can identify skilled managers, but also because they can identify managers whose skill is not sufficiently offset by diseconomies of scale. Our findings are consistent with a competition framework à la Berk and Green (2004) in which frictions prevent a substantial fraction of the investor population from identifying some of the funds whose managers have the highest skill, and thus, these funds remaining small relative to their manager's skill.

[26] To adjust assets under management for inflation, we follow Roussanov et al. (2021) and multiply assets in year t by the Consumer Price Index (CPI) at the end of 2015 divided by the CPI at the end of year t. We download data for CPI using the FRED series "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, Index 1982-1984 = 100, Monthly, Seasonally Adjusted.".

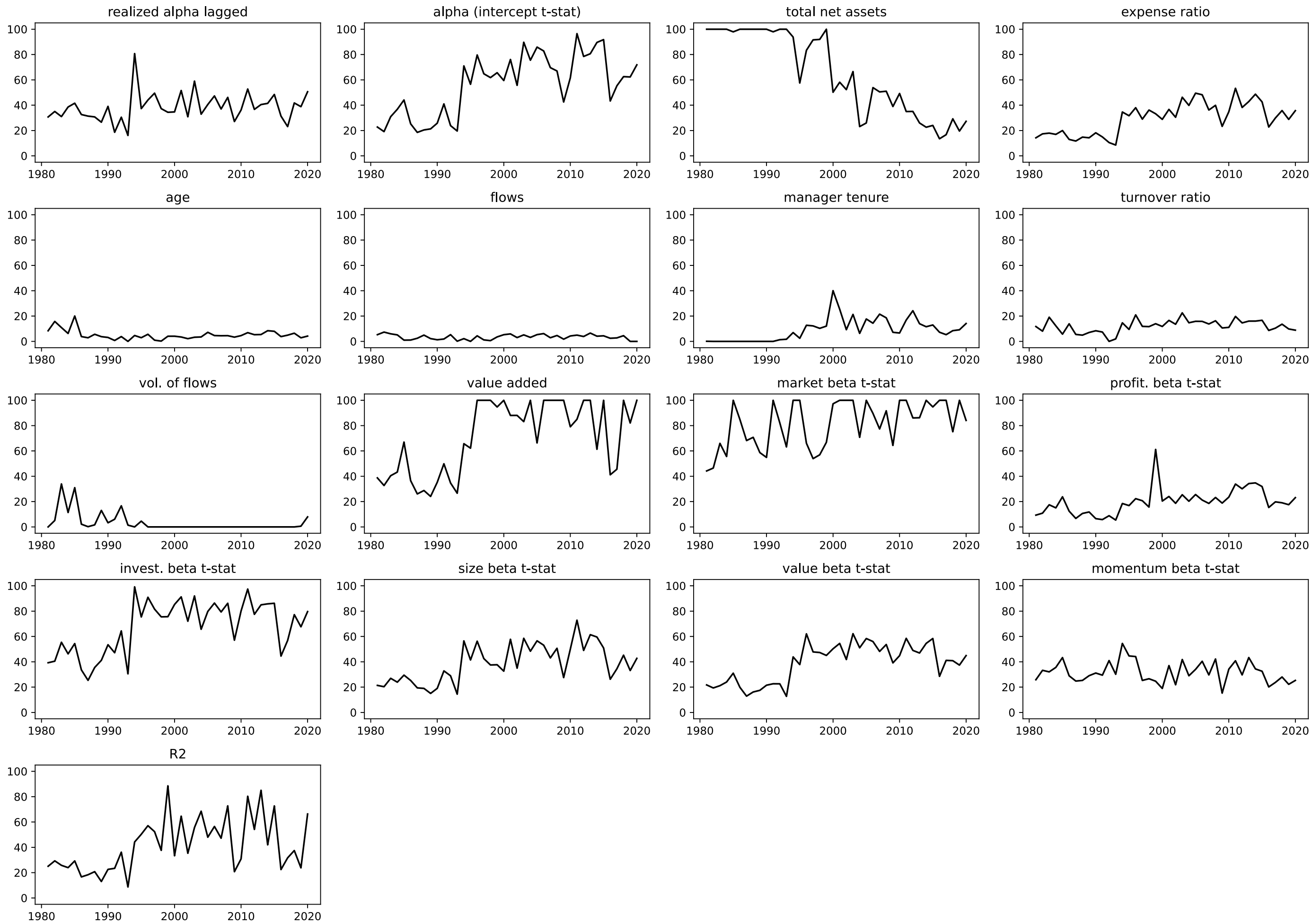

Fig. 8. Time evolution of characteristic importance for gradient boosting. This figure plots the time evolution of the importance of each characteristic for gradient boosting. We measure the importance of each characteristic as the average across all observations of the absolute SHAP value of the characteristic. We scale characteristic importance so that it ranges between zero for the least important characteristic and 100 for the most important characteristic and report relative importance for each year from 1980 to 2019.

7. Performance over time and across market conditions

Jones and Mo (2020) show that the ability of fund characteristics to predict performance has declined over time due to increased arbitrage activity and mutual-fund competition. Motivated by their work, we study how the alpha of the different portfolios varies over time. To do this, we compute the cumulative net alpha of the top-decile portfolio for gradient boosting, random forests, and OLS in each month of the out-of-sample period from 1991 to 2020 as well as those of the equally weighted and asset-weighted portfolios.[27] Fig. 11 shows the time-series of cumulative abnormal returns. The three prediction-based portfolios (gradient boosting, random forests, and OLS) outperform the two naive portfolios (equally weighted and asset weighted) over the whole 30-year out-of-sample period. In particular, while the gradient-boosting, random-forests, and OLS portfolios achieve cumulative net alphas of 69%, 78%, and 34%, respectively, the equally weighted and asset-weighted portfolios earn negative cumulative net alphas of −7% and −13%, respectively. Consistent with Jones and Mo (2020), however, the performance of the prediction-based portfolios is similar to that of the naive portfolios from 2012 until 2018. Nevertheless, all three prediction-based portfolios outperform the two naive portfolios in the last two years of our sample (2019 and 2020). In particular, while the gradient-boosting, random-forests, and OLS portfolios achieve cumulative (2019–2020) net alphas of 4.7%, 2.2%, and −0.1%, respectively, the equally weighted and asset-weighted portfolios earn negative cumulative net alphas of −2.8% and −3.9%, respectively.

Li and Rossi (2020) study whether the ability of *mutual-fund holdings and stock characteristics* to predict fund performance varies across market conditions. Inspired by their work, we now investigate whether the ability of *fund characteristics* to select funds with positive alpha changes across market conditions. Like Li and Rossi (2020), we condition estimates of performance on expansions and recessions, as well as on high and low investor sentiment. Specifically, we regress the out-of-sample monthly excess returns of the top decile portfolios selected by gradient boosting and random forests on the Fama and French (2015) five factors and momentum as well as indicator variables for expansions and recessions, and high and low investor sentiment. Expansions and recessions are defined following the NBER convention. The high (low) investor sentiment indicator equals one if investor sentiment, as defined in Baker and Wurgler (2006, 2007), is above (below) the median of the July 1965 to December 2020 period. Specifically, we download from Jeffrey Wurgler's website the version of investor sentiment based on the first principal component of five sentiment proxies, where each of the proxies has first been orthogonalized with respect to six macroeco-

[27] We compute monthly net alphas as the portfolio excess returns net of all costs each month minus the product of the factor realization in that month and the portfolio betas estimated over the whole out-of-sample sample period using the FF5 model augmented with momentum.

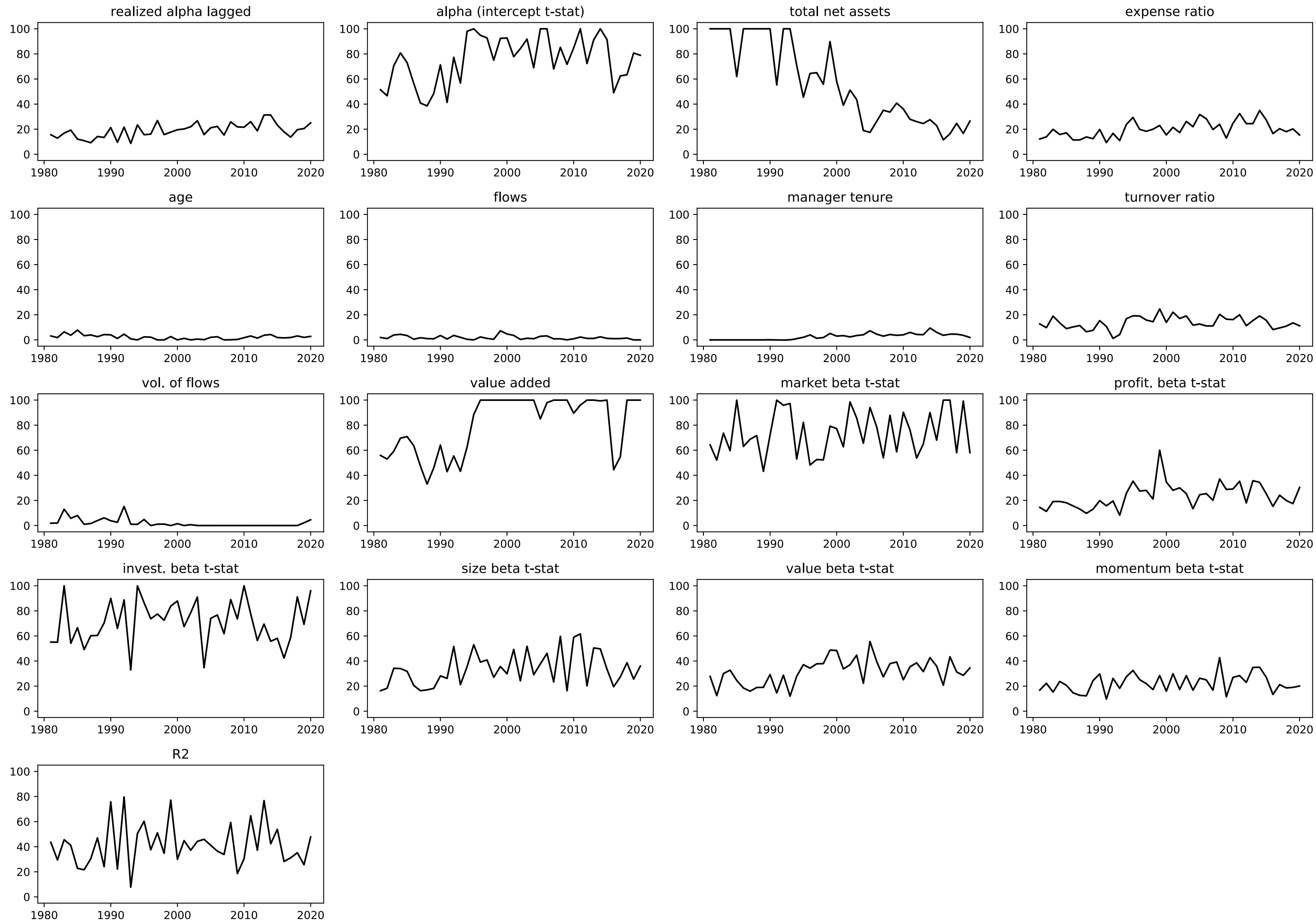

Fig. 9. Time evolution of characteristic importance for random forests. This figure plots the time evolution of the importance of each characteristic for random forests. We measure the importance of each characteristic as the average across all observations of the absolute SHAP value of the characteristic. We scale characteristic importance so that it ranges between zero for the least important characteristic and 100 for the most important characteristic and report relative importance for each year from 1980 to 2019.

Table 7

Out-of-sample alpha of fund portfolios under different market conditions. This table reports the monthly out-of-sample alphas (in %) net of all costs for the top-decile fund portfolios obtained with gradient boosting and random forests under different market conditions. Alphas are computed by regressing the out-of-sample excess monthly portfolio returns net of all costs against the Fama and French (2015) five factors and momentum as well as indicator variables for expansions and recessions (Panel A), and high and low investor sentiment (Panel B). Expansions and recessions are defined following the NBER convention. The high (low) investor sentiment indicator equals one if investor sentiment, as defined in Baker and Wurgler (2006, 2007), is above (below) the median of the July 1965 to December 2020 period. The out-of-sample period spans from January 1991 to December 2020. We report standard errors with Newey-West adjustment for 12 lags in parentheses. One, two, and three asterisks indicate that the alpha is significant at the 10%, 5%, and 1% level, respectively.

	Panel A. Business Cycle			Panel B. Investor Sentiment		
	Expansion	Recession	Exp.– Rec.	High	Low	High – Low
Gradient boosting	0.179**	0.375	-0.196	0.233***	0.150	0.083
	(0.082)	(0.228)	(0.226)	(0.085)	(0.109)	(0.106)
Random forests	0.202**	0.445*	-0.243	0.266****	0.169	0.097
	(0.087)	(0.248)	(0.236)	(0.102)	(0.118)	(0.131)

nomic indicators. Table 7 reports estimated alphas for different market conditions and their standard errors with Newey-West adjustment for 12 lags. We also report differences in alphas across market conditions. Our main finding is that the gradient-boosting and random-forest portfolios achieve positive alphas across all market conditions, and although they perform better in recessions and times of high investor sentiment, the differences in alpha across different market conditions are not statistically significant.

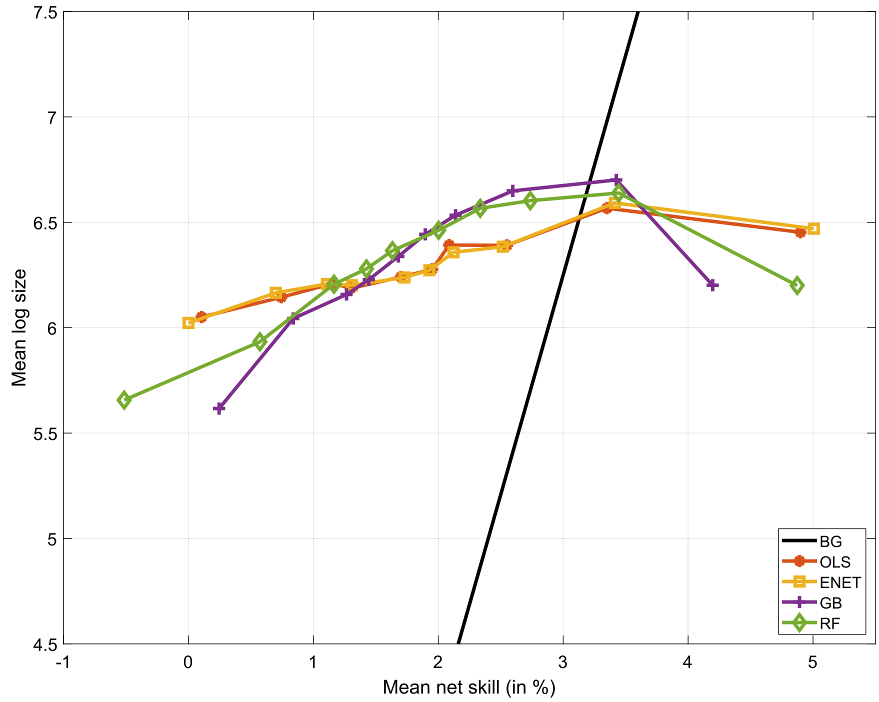

Fig. 10. Capital misallocation and machine learning. This figure illustrates capital misallocation for the decile portfolios generated by the four prediction methods we consider. For the jth decile portfolio of funds ranked by predicted alpha, the horizontal axis gives the mean net skill, $E(\hat{a}_{i,t+1} - p_{i,t}|i \in D_j)$, where D_j is the set of funds in the jth decile, and the vertical axis the mean log size, $E(\log(Q_{i,t})|i \in D_j)$. The colored lines plot the mean log size for each decile portfolio generated by OLS (orange stars), elastic net (yellow squares), gradient boosting (purple crosses), and random forests (green diamonds). For every method, the first decile portfolio has the lowest mean net skill and mean log size. We also plot the efficient (Berk-Green) log size, $\log\left(Q_{i,t}^{BG}\right)$, for each level of net skill (straight black line). Net skill is the average of past realized alpha before fees and diseconomies of scale estimated using the approach of Zhu (2018) minus the current expense ratio. Diseconomies of scale are computed based on Roussanov et al. (2021) as the log of fund size multiplied by the diseconomies of scale parameter, $\eta = 0.0048$ as estimated by Roussanov et al. (2021). The efficient (Berk and Green) fund sizes for each level of skill are computed by dividing net skill by the diseconomies of scale parameter, η.

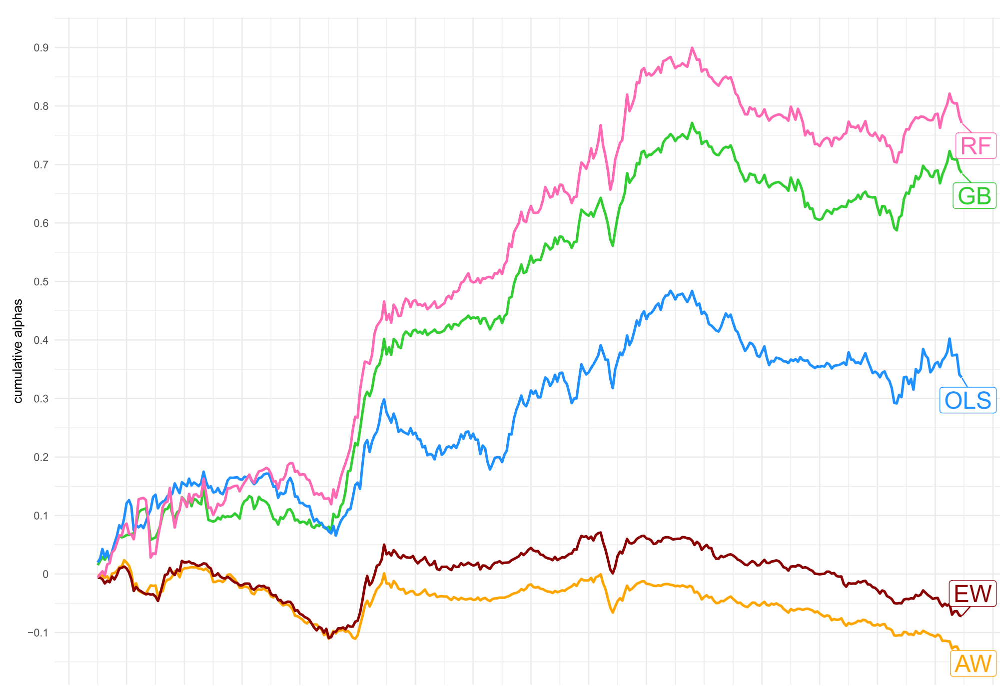

Fig. 11. Cumulative portfolio alpha. This figure plots the time series of cumulative out-of-sample portfolio realized alphas of the excess returns net of all costs of the top-decile fund portfolios. Realized portfolio alphas are based on the regressions on the five Fama-French factors augmented with momentum (FF5+MOM). Portfolios are obtained with gradient boosting (GB), random forests (RF), OLS, and with two naive strategies (equally weighted (EW) and asset-weighted (AW) portfolios of all available funds).

8. Conclusions

The question of whether mutual-fund investors can earn positive net alpha by investing in active mutual funds has received much attention from academics, practitioners, and regulators. We posit that the pessimistic results that dominate the literature could be a consequence of the methods employed to exploit predictability in fund performance. In particular, we show that machine-learning methods can dynamically identify and exploit nonlinearities and interactions in the relation between fund characteristics and performance and help investors to select funds that earn significant and positive alphas net of fees and transaction costs. The machine-learning methods reveal that the interactions between measures of past performance and fund activeness help to predict future fund performance. Our results demonstrate that investors can benefit from actively managed mutual funds, but only if they have access to sophisticated predictions that allow flexibility in the relation between fund characteristics and performance.

To understand the economic mechanism behind our results, we study whether the performance of our portfolios can be explained by capital misallocation in the mutual-fund market, and find that indeed machine learning selects funds that are small relative to their managers' skill, consistent with informational frictions preventing some investors from identifying the outperforming funds. Our work implies that there is scope for pension-plan administrators and financial advisors to integrate machine learning with other tools in order to help investors select active mutual funds with positive alpha.

Finally, our finding that mutual-fund characteristics that do not require information on fund portfolio holdings are enough to predict positive alpha implies that even if no information on portfolio holdings had been available during our sample period, our methods would have identified funds with positive net alpha on average. This is relevant to the debate around the recent SEC proposal to raise the asset threshold for mandatory portfolio disclosure.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code and data packet for this article can be found at https://doi.org/10.17632/rpgb99m5zy.3:

Machine Learning and Fund Characteristics Help to Select Mutual Funds with Positive Alpha (Original Data) (Mendeley Data).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfineco.2023.103737.

References

Aliaj, O., 2020. Most hedge funds to be allowed to keep equity holdings secret. Financial Times, July 11, https://www.ft.com/content/c68ca89c-3f9b-45f9-8205-6dbea70ed859.

Amihud, Y., Goyenko, R., 2013. Mutual fund's R^2 as predictor of performance. Rev. Financ. Stud. 26 (3), 667–694.

Aragon, G.O., Hertzel, M., Shi, Z., 2013. Why do hedge funds avoid disclosure? Evidence from confidential 13F filings. J. Financ. Quant. Anal. 48 (5), 1499–1518.

Avramov, D., Wermers, R., 2006. Investing in mutual funds when returns are predictable. J. Financ. Econ. 81 (2), 339–377.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. J. Finance 61 (4), 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. J. Econ. Perspect. 21 (2), 129–152.

Baks, K.P., Metrick, A., Wachter, J., 2001. Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation. J. Finance 56 (1), 45–85.

Bali, T.G., Beckmeyer, H., Moerke, M., Weigert, F., 2023. Option return predictability with machine learning and big data. Rev. Financ. Stud. 36 (9), 3548–3602.

Banegas, A., Gillen, B., Timmermann, A., Wermers, R., 2013. The cross section of conditional mutual fund performance in European stock markets. J. Financ. Econ. 108 (3), 699–726.

Barras, L., Scaillet, O., Wermers, R., 2010. False discoveries in mutual fund performance: measuring luck in estimated alphas. J. Finance 65 (1), 179–216.

Bergmeir, C., Hyndman, R.J., Koo, B., 2018. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Comput. Stat. Data Anal. 120, 70–83.

Berk, J., Green, R., 2004. Mutual fund flows and performance in rational markets. J. Polit. Econ. 112 (6), 1269–1295.

Berk, J.B., Van Binsbergen, J.H., 2015. Measuring skill in the mutual fund industry. J. Financ. Econ. 118 (1), 1–20.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Rev. Financ. Stud. 34 (2), 1046–1089.

Bollen, N.P., Busse, J.A., 2005. Short-term persistence in mutual fund performance. Rev. Financ. Stud. 18 (2), 569–597.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Bryzgalova, S., Pelger, M., Zhu, J., 2019. Forest Through the Trees: Building Cross-Sections of Stock Returns. Available at SSRN 3493458.

Busse, J.A., Irvine, P.J., 2006. Bayesian alphas and mutual fund persistence. J. Finance 61 (5), 2251–2288.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A.W., Siddique, A., 2016. Risk and risk management in the credit card industry. J. Bank. Finance 72, 218–239.

Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52 (1), 57–82.

Chan, L.K., Chen, H.-L., Lakonishok, J., 2002. On mutual fund investment styles. Rev. Financ. Stud. 15 (5), 1407–1437.

Chen, J., Hong, H., Huang, M., Kubik, J.D., 2004. Does fund size erode mutual fund performance? The role of liquidity and organization. Am. Econ. Rev. 94 (5), 1276–1302.

Chen, L., Pelger, M., Zhu, J., 2020a. Deep learning in asset pricing. Manag. Sci. https://doi.org/10.1287/mnsc.2023.4695. Forthcoming.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., 2020b. xgboost: Extreme gradient boosting. R package version 1.2.0.1.

Chiang, W.-C., Urban, T.L., Baldridge, G.W., 1996. A neural network approach to mutual fund net asset value forecasting. Omega 24 (2), 205–215.

Coulombe, P.G., Leroux, M., Stevanovic, D., Surprenant, S., 2020. How is machine learning useful for macroeconomic forecasting? Available at arXiv:2008.12477 [https://.org/abs].

Cremers, K.M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Rev. Financ. Stud. 22 (9), 3329–3365.

Cremers, M., Petajisto, A., Zitzewitz, E., 2013. Should benchmark indices have alpha? Revisiting performance evaluation. Crit. Finance Rev. 2 (1), 001.

Dumitrescu, A., Gil-Bazo, J., 2018. Market frictions, investor sophistication, and persistence in mutual fund performance. J. Financ. Mark. 40, 40–59.

Elliott, G., Gargano, A., Timmermann, A., 2013. Complete subset regressions. J. Econom. 177 (2), 357–373.

Elton, E.J., Gruber, M.J., Blake, C.R., 2001. A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and morningstar mutual fund databases. J. Finance 56 (6), 2415–2430.

Elton, E.J., Gruber, M.J., Blake, C.R., 2011. Holdings data, security returns, and the selection of superior mutual funds. J. Financ. Quant. Anal. 46 (2), 341–367.

Evans, R.B., 2010. Mutual fund incubation. J. Finance 65 (4), 1581–1611.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.

Fama, E.F., French, K.R., 2010. Luck versus skill in the cross-section of mutual fund returns. J. Finance 65 (5), 1915–1947.

Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116 (1), 1–22.

Feng, G., Polson, N.G., Xu, J., 2020. Deep Learning in Characteristics-Sorted Factor Models. Available at SSRN 3243683.

Ferreira, M.A., Keswani, A., Miguel, A.F., Ramos, S.B., 2013. The determinants of mutual fund performance: a cross-country study. Rev. Finance 17 (2), 483–525.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22.

Garcia, M.G., Medeiros, M.C., Vasconcelos, G.F., 2017. Real-time inflation forecasting with high-dimensional models: the case of Brazil. Int. J. Forecast. 33 (3), 679–693.

Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Gittelsohn, J., 2019. End of era: passive equity funds surpass active in epic shift. Bloomberg, September 11. https://www.bloomberg.com/news/articles/2019-09-11/passive-u-s-equity-funds-eclipse-active-in-epic-industry-shift.

Green, J., Hand, J.R., Zhang, X.F., 2017. The characteristics that provide independent information about average us monthly stock returns. Rev. Financ. Stud. 30 (12), 4389–4436.

Gruber, M.J., 1996. Another puzzle: the growth in actively managed mutual funds. J. Finance 51 (3), 783–810.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. Rev. Financ. Stud. 33 (5), 2223–2273.

Gupta-Mukherjee, S., 2014. Investing in the “new economy”: mutual fund performance and the nature of the firm. J. Financ. Quant. Anal. 49 (1), 165–191.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28 (3), 650–705.

Hunter, D., Kandel, E., Kandel, S., Wermers, R., 2014. Mutual fund performance evaluation with active peer benchmarks. J. Financ. Econ. 112 (1), 1–29.

Indro, D.C., Jiang, C., Patuwo, B., Zhang, G., 1999. Predicting mutual fund performance using artificial neural networks. Omega 27 (3), 373–380.

Jensen, M.C., 1968. The performance of mutual funds in the period 1945–1964. J. Finance 23 (2), 389–416.

Jones, C.S., Mo, H., 2020. Out-of-sample performance of mutual fund predictors. Rev. Financ. Stud. 34 (1), 149–193.

Jones, C.S., Shanken, J., 2005. Mutual fund performance with learning across funds. J. Financ. Econ. 78 (3), 507–552.

Kacperczyk, M., Nieuwerburgh, S.V., Veldkamp, L., 2014. Time-varying fund manager skill. J. Finance 69 (4), 1455–1484.

Kaniel, R., Lin, Z., Pelger, M., Van Nieuwerburgh, S., 2023. Machine-learning the skill of mutual fund managers. J. Financ. Econ. 150 (1), 94–138.

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., Malohlava, M., 2020. H_2O: R interface for the ‘H2O’ scalable machine learning platform. R package version 3.30.1.3.

Li, B., Rossi, A.G., 2020. Selecting mutual funds from the stocks they hold: A machine learning approach. Available at SSRN 3737667.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2 (3), 18–22.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30, 1–10.

Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. Rev. Financ. Stud. 21 (1), 233–264.

McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? J. Finance 71 (1), 5–32.

Medeiros, M.C., Vasconcelos, G.F., Veiga, Á., Zilberman, E., 2021. Forecasting inflation in a data-rich environment: the benefits of machine learning methods. J. Bus. Econ. Stat. 39 (1), 1–22.

Mehta, D., Desai, D., Pradeep, J., 2020. Machine learning fund categorizations. Available at arXiv:2006.00123 [https://.org/abs].

Molnar, C., 2019. Interpretable Machine Learning. Lulu.com. https://christophm.github.io/interpretable-ml-book/.

Moreno, D., Marco, P., Olmeda, I., 2006. Self-organizing maps could improve the classification of Spanish mutual funds. Eur. J. Oper. Res. 174 (2), 1039–1054.

Pástor, L., Stambaugh, R.F., 2002. Investing in equity mutual funds. J. Financ. Econ. 63 (3), 351–380.

Pástor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. J. Polit. Econ. 111 (3), 642–685.

Pattarin, F., Paterlini, S., Minerva, T., 2004. Clustering financial time series: an application to mutual funds style analysis. Comput. Stat. Data Anal. 47 (2), 353–372.

Pedersen, L.H., 2022. Big data asset pricing 5: Machine learning in asset pricing. Available at SSRN 4068797.

Rakowski, D., 2010. Fund flow volatility and performance. J. Financ. Quant. Anal. 45 (1), 223–237.

Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: what is the role of the United States? J. Finance 68 (4), 1633–1662.

Rossi, A.G., Utkus, S.P., 2020. Who Benefits from Robo-Advising? Evidence from Machine Learning. Available at SSRN 3552671.

Roussanov, N., Ruan, H., Wei, Y., 2021. Marketing mutual funds. Rev. Financ. Stud. 34 (6), 3045–3094.

Schapire, R.E., Freund, Y., 2012. Boosting: Foundations and Algorithms. MIT Press.

Sharpe, W.F., 1966. Mutual fund performance. J. Bus. 39 (1), 119–138.

Shi, Z., 2017. The impact of portfolio disclosure on hedge fund performance. J. Financ. Econ. 126 (1), 36–53.

Stambaugh, R.F., Yuan, Y., 2017. Mispricing factors. Rev. Financ. Stud. 30 (4), 1270–1315.

Wermers, R., 2000. Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses. J. Finance 55 (4), 1655–1695.

Wu, W., Chen, J., Yang, Z., Tindall, M.L., 2021. A cross-sectional machine learning approach for hedge fund return prediction and selection. Manag. Sci. 67 (7), 4577–4601.

Zhu, M., 2018. Informative fund size, managerial skill, and investor rationality. J. Financ. Econ. 130 (1), 114–134.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc., Ser. B, Stat. Methodol. 67 (2), 301–320.