Real-Time Predictability of Mutual Fund Performance Predictors

Abstract

Researchers have discovered abundant evidence that mutual fund performance is predictable in the cross-section *ex post*. This paper studies the *ex ante* predictability of 12 well-known predictors for fund performance from investors' perspective. Exploiting two types of fund picking strategies with either rule-based approach or machine learning methods, I find that utilizing machine learning can deliver superior real-time economic gains for investors with fund short-term performance being the primary driver underlying predictability. Moreover, using a novel approach to decomposing fund performance, I discover that investors' flow response to predictor-implied performance exhibits strong variations across predictors. These results suggest *ex ante* predictability as the compensation for employing costly algorithms to search for skilled managers.

JEL Classifications: G11; G14; G17; G23; G50

Keywords: Performance Evaluation; Personal Investing; Machine Learning; Investor Behavior; Market Efficiency

1 Introduction

The rapid growth of asset management industry over recent years has been accompanied with an increasing demand from households for diversified investment portfolios. As shown in Figure 1, the percentage of U.S. households owning mutual funds has grown from a negligible 5.7% in 1980 to almost 46% in 2020, and actively managed funds remain important accounting for 60% of the U.S. total net assets in 2020. Consequently, the request for investors such as households to distinguish mutual funds with superior performance has become an increasingly relevant and critical issue for their financial well-being.

At the same time, researchers have discovered a bunch of predictors suggesting that outperforming actively managed mutual funds can be identified with lagged information variables using full sample information[1]. A natural and relevant question henceforth arises: is it possible for investors to employ available predictors for better fund selection in real time, without knowing which predictor works *ex ante*? And a further question is: to what degree *de facto* do investors take advantage of any potential predictive information when choosing actively managed mutual funds? In this paper, I address these issues by conducting a comprehensive study of the economic benefits using 12 well-known fund performance predictors from the investors' perspective in real time.

Studies in existing literature mainly focus on discovering new predictors without accounting for the joint predictive power of existing predictors. This paper attempts to bridge the gap by utilizing two types of strategies to study investors' gains in using performance predictors: rule-based strategies and machine learning based strategies. The baseline rule-based strategies are modified from the approaches in Pesaran and Timmermann (1995) and Cooper et al. (2005) which have been used for predicting future stock returns. They are straightforward to understand but still require demanding computing

[1] See Hendricks et al. (1993), Carhart (1997), Chen et al. (2004), Kacperczyk et al. (2005, 2006), Kacperczyk and Seru (2007), Cremers and Petajisto (2009), Barras et al. (2010), Amihud and Goyenko (2013), Kacperczyk et al. (2014), Doshi et al. (2015), Cremers and Pareek (2016), Harvey and Liu (2018, 2019), Barras et al. (Forthcoming) for instance.

power to implement[2]. On the other hand, machine learning methodologies have been recently used by researchers to uncover patterns not detected by traditional OLS method. For instance, in the asset pricing literature, Gu et al. (2020) compares various machine learning methods for better measurement of equity risk premia, and Kozak et al. (2020) imposes economically-driven prior to identify characteristic-based principal components that can explain the cross-section of stock returns. For other asset classes, Bali et al. (2021) and Goyenko and Zhang (2021) use machine learning methods to study the cross-predictability between either corporate bonds and stocks, or options and stocks. One main advantage of using machine learning methods for predicting future fund performance is that they allow more flexible specifications for the relation between future fund performance and predictors, especially when we have limited knowledge on the specific sources of managerial skills[3].

Table 1 lists the 12 predictors studied in this paper (expense ratio, turnover, fund flow, fund size, one-year return, Carhart alpha, one-month return, return gap, active share, R-squared, active weight, and fund duration) classified into three categories[4]: characteristics, performance, and activeness. These predictors have been found to predict performance in their respective full sample in the original studies. The question in mind is whether an investor would have chosen those predictors for fund selection without *ex post* knowledge that those predictors would work. Is it possible for an investor in real-time to identify these predictors among a group of alternatives, or is the evidence that the outperforming funds can be screened out only due to the clarity of hindsight? This paper provides an answer to this question.

In my analysis, the investor may employ any of the 12 predictors individually or

[2] There have been plenty of studies such as Lo et al. (2000) on using technical rules for predicting stock returns.

[3] Kacperczyk et al. (2016) develops a theory of managers' optimal attention allocation over business cycles to identify skilled fund managers, and Kacperczyk et al. (2014) provides more evidence on the time-varying nature of skills. However, the exact functional form underlying the relation between performance and skills imperfectly captured by observed variables is not well-understood.

[4] Another category of predictors related to fund liquidity management found in Simutin (2014) and Boguth and Simutin (2018) has not been included in the current version of the paper due to limited number of funds in earlier periods.

a combination of them based on either rule-based strategies or machine learning strategies. One distinguishing feature of these strategies is that by examining combinations of predictors, specific fund skill embedded in one predictor can be isolated by controlling for other performance indicators. For instance, Amihud and Goyenko (2013) shows that among low R-squared funds, those with higher past Carhart four-factor alpha have better future performance. Another notable feature of these strategies is that I do not need to put additional *ex ante* restrictions on which of the 12 variables investors would like to use for fund selection. For instance, would it be a good decision to invest in a fund with high risk-adjusted alpha or a fund that is the most active, or choose neither and just invest in a passive market portfolio instead? For rule-based strategies, I identify the potential fund selection rules as cross-sectional sorts of all actively managed U.S. domestic equity funds based on the 12 predictors, while for machine learning I form strategies based on predictions from machine learning algorithms.

I test for investors' gains from performance predictability by analyzing whether a simulated real-time fund portfolio outperforms different benchmark stock portfolios after fees. For rule-based strategies, the real-time portfolio is constructed each year by choosing the fund selection rules that perform best during the prior in-sample period. I examine real-time simulations based on the mean monthly return criterion[5]. The results indicate that one version of the rule-based real-time portfolio is able to beat the market in real time but generates no alpha relative to Carhart four-factor model. In contrast, regression-based machine learning with variable selection feature (LASSO and elastic net) can also deliver outperformance not only relative to the market benchmark (with annualized alpha of 1.68%) but also compared to Carhart four factors (with annualized alpha of 1.32%). Across all methods, short-term performance (one-month return) is found to be the primary predictor for performance forecasting. Further inspecting the real-time machine learning portfolio, I find that through variable selection, elastic net or LASSO portfolios only take advantage of predictive information from predictors when

[5] Results for other criteria including buy-and-hold dollar return and Sharpe ratio will be incorporated in future version of the paper.

predictability is strong, and switch to passive market portfolio by ignoring all predictors when overall predictability is weak. This feature essentially trades off some positive gains for less volatility in real-time portfolio. However, other regression-based machine learning methods cannot beat the market. These results suggest that robo-advisors using machine learning algorithms with variable selection feature can add value to fund picking by investors.

Moreover, my paper further examines whether in reality investors attend to those well-known predictors constructed with publicly available information when evaluating mutual funds. I find that conditional on investors' usage of CAPM, investors react to the components of CAPM alpha implied by predictors in different ways, and investor attention to predictive information embedded in predictors is stronger among aggressive growth funds where those predictors are found to work well.

My findings help to resolve the ongoing debate with regards to what degree the asset management industry is informationally efficient. While Berk and Green (2004) argues that investors have perfect foresight for discovering skilled managers such that no real-time predictability exists *ex ante*, Gârleanu and Pedersen (2018) contends that there exist costs to acquire information for investors to identify skilled managers. My results suggest that real-time predictability exists not due to lack of investor attention to publicly available predictive information, instead the magnitude of any real-time excess gain found in this paper can be seen as a search cost an average investor needs to incur using intensive search algorithms to find skilled managers in the asset management industry.

A large body of previous research has been devoted to finding outperforming funds in the cross-section with full-sample *ex post* information (Chen et al., 2004, Kacperczyk et al., 2006, Cremers and Petajisto, 2009). My study contributes to this literature by assessing the real-time predictive power in an extensive and more flexible setup by evaluating multiple predictors simultaneously. Another strand of related literature is on mutual fund investors' flow response to returns (Ippolito, 1992, Chevalier and Ellison, 1997, Sirri and Tufano, 1998). This literature has found that fund flows tend to be a convex function of past performance. Barber et al. (2016) and Berk and van Binsbergen (2016) argue that investors are most likely to use the Capital Asset Pricing Model (CAPM) to risk-adjust fund performance. My paper is also related to the literature on investor learning and return predictability. Lewellen and Shanken (2002) argues that investor learning may distort empiricists' test for market efficiency and demonstrate how in-sample stock predictability emerges in absence of real-time predictability through investor learning. Martin and Nagel (Forthcoming) further shows that with many predictors, out-of-sample performance instead of in-sample performance is a more proper validation for asset pricing tests if investors learn about predictors. More closely related to my paper, Baks et al. (2001) and Avramov and Wermers (2006) show that skeptical prior beliefs of mean-variance investors can identify funds that predict alpha *ex ante* while Avramov and Wermers (2006) finds that if investors do not believe in fund return predictability, their optimal fund portfolios would not have positive out-of-sample performance. However, those papers do not examine any real-time predictability of specific predictors as part of investors' information set. Given my results that variable selection machine learning methods[6] are able to identify superior mutual funds *ex ante* while other approaches cannot, it would be interesting to recover investor beliefs in the asset management industry given the *ex ante* predictability I discover in this paper. Last but not least, my paper contributes to the household finance literature (see Campbell (2006)) by demonstrating investors' gains using either rule-based approaches or machine learning methods, given increasing popularity among households in diversified investment vehicles such as mutual funds.

My paper also complements recent examinations of the out-of-sample predictability of the cross-section of mutual fund performance. Jones and Mo (2021) finds that after the original sample periods, the predictive power of 27 mutual fund predictors have fallen by around a half. They find that increases in arbitrage activities and mutual fund competition tend to be the main reasons for the drop in predictability beyond the original

[6] In Bayesian setup, variable selection with L_1 regularization corresponds to the Laplace prior.

sample periods. Both Jones and Mo (2021) and my study highlight a marked difference between *ex ante* and *ex post* performance predictability. However, my paper differs in motivations and aims to answer to what degree investors can benefit from using fund predictors without knowing whether they would work, instead of comparing predictor performance before and after original sample periods. In essence, my empirical test is out-of-sample but at the same time incorporates an additional layer by considering selection for predictors or predictive information to be used by investors.

Contemporaneous works by Li and Rossi (2020), DeMiguel et al. (2021), and Kaniel et al. (2021) also examine fund performance using machine learning algorithms and find that machine learning helps to distinguish outperforming funds. Li and Rossi (2020) considers fund performance predictors based on fund stock holdings while DeMiguel et al. (2021) focuses on fund characteristics and performance measures. My paper shows that among three groups of predictors (fund characteristics, performance, and holding-based activeness measures), one-month short-term return is the primary driver that contributes to selecting outperforming funds in real time. This short-term fund momentum is further confirmed in Kaniel et al. (2021). However, beyond machine learning algorithms, a human-like rule-based portfolio approach is studied in my paper to see whether a relatively simple approach allowing for nonlinear interactions helps to find outperforming funds for investors in real time. I find that this simple approach can generate outperformance relative to the market via significant exposure to stock momentum factor. More importantly, my paper finds that investors tend to incorporate predictive information embedded in predictors to allocate capital across mutual funds, suggesting they may use those predictors to find skilled managers, which are new to the literature. These results together suggest that real-time return predictability exists in the competitive asset management industry not due to lack of attention from investors to use those predictors when choosing mutual funds but instead as a compensation for using complex algorithms which requires significant computing power to implement. In this regard, my paper provides empirical support for Gârleanu and Pedersen (2018) which argues that investors need to incur search costs to find skilled managers in an informationally efficient market.

The rest of the paper is organized as follows. Section 2 introduces the rule-based approach and machine learning methods used for predicting future fund performance in this paper. Section 3 describes the mutual fund data and the sample selection criteria. Section 4 illustrates the in-sample predictive power of each of the 12 predictors. Section 5 examines the performance of real-time portfolios constructed based on rule-based and machine learning strategies and evaluates investors' gains from using those predictors. Section 6 explores investors' flow response to the predictive information embedded in predictors. Section 7 concludes.

2 Methodology

Given the paper's objective is to examine investors' benefits in using various predictive information for fund selection, statistical tools that are adequately sophisticated to accommodate predictive variables in large scale are necessary to help investors obtain a comprehensive view on any predictive relation before making value-creating investment decisions. On the other hand, methods that are over-complicated may deliver results lack of robustness and credibility for fund investors, due to additional model risk[7]. Two types of methods stand out for achieving the trade-off between sophistication and robustness: rule-based portfolio sorting and regression-based machine learning. Rule-based portfolio sorting approach shares the same economic spirit as standard portfolio sorting approach but extends the standard one by incorporating interactions among many predictors. Regression-based machine learning methods are variants of standard least squares approach after accounting for correlations either among predictors or between predictors and the forecasting target (i.e., fund performance). In the following subsections, I describes each type of methods and their respective advantages in predicting fund

[7] This can be less an issue for more sophisticated institutional investors who have the capacity to understand and employ more complex methods in predicting fund performance. However, unsophisticated investors may be more concerned about potential model risk.

performance.

2.1 Rule-Based Portfolio Sorting Approach

For the rule-based portfolio sorting approach, I adapt the recursive two-way portfolio sorting procedure proposed in Cooper et al. (2005) to evaluate the real-time performance of combinations of 12 predictors from January 1995 to December 2016. Specifically, I form one-way and two-way dependent quintile sorts from those 12 predictors at the end of each month and select single best performing rule (i.e., a combination of predictors and quintiles) that is shown to perform the best in a given in-sample period for investors to form real-time portfolio in the following year. I adopt an expanding window[8] starting with a six-year in-sample period and then expand the in-sample window by one-year as the evaluation moves forward. The reason I use dependent sort is to control for correlations between different predictors such that for a pair of correlated predictors, one predictor does not drive out the predictive power of the other one. The one-way sorts yield $12 \times 5 = 60$ rules, and the two-way sorts add $A_{12}^2 \times 25 = 3,300$ more. In total, I assess $3,360$ fund selection rules.

Another variant of the portfolio sorting approach is to consider a fraction of rules instead of using one single rule. The advantage of using multiple combinations is to average out potential noises introduced with using only one rule[9]. This can be potentially helpful since even though mutual funds are diversified portfolios, distinguishing outperforming funds among alternative portfolios using multiple rules can be more informative to capture fund manager's skill in generating abnormal returns. In order to select the best fraction of rules, I split the in-sample period into two samples: a training sample and a one-year validation sample. The initial training sample is therefore five years out of the initial six-year in-sample period. The purpose of setting up a validation sample is to avoid over-fitting the in-sample period by selecting a fraction of rules only to perform

[8] Expanding window provides additional years for training models compared to rolling window.

[9] Recall a rule is either a single predictor quintile or a combination of quintiles of two predictors.

well in the sample but not out of the sample. Similar to the machine learning methods introduced in the following subsection, I treat the percent of rules to be selected as a hyperparameter which is determined in the validation period so that the selected rules based on the chosen fraction of rules would perform the best for the validation period. The range of percentage of rules is 0.1%, 0.2%, 0.5%, and 1%, which corresponds to 3, 7, 17, and 34 rules respectively[10].

2.2 Machine Learning Methods

Machine learning methodologies have been recently used by researchers to uncover patterns not detected by traditional methods. For instance, in the asset pricing literature, Gu et al. (2020) compares various machine learning methods for better measurement of equity risk premia, and Kozak et al. (2020) imposes economically-driven prior to motivate elastic net method and identifies characteristic-based principal components that can explain the cross-section of stock returns. In this section, I describe six regression-based machine learning methods that are relatively intuitive to understand and have been widely used for forecasting with many predictors.

The six machine learning methods can classified into two categories based on each method's specific purpose: penalized linear and dimension reduction. To fix idea, consider a simple performance generating process by fund manager's skill as follows:

$$r_{i,t+1} = \mathbb{E}_t[r_{i,t+1}] + \epsilon_{i,t+1}, \tag{1}$$

where

$$\mathbb{E}_t[r_{i,t+1}] = g^*(x_{i,t}; \theta). \tag{2}$$

$r_{i,t+1}$ is the net-of-fee return investors would realize by investing fund i in month $t+1$,

[10] Two alternative criteria (buy-and-hold dollar return and Sharpe ratio) to select rules to construct real-time fund portfolio will be included in future version of the paper as robustness check.

which can be decomposed into an expected performance component plus noise. My objective is to model the unknown expected component $\mathbb{E}_t[r_{i,t+1}]$ as a function of observable predictors that maximizes the expected performance for a mutual fund investor at $t+1$. I denote those predictors as a M-dimensional vector $x_{i,t}$, and assume the conditional expected return $g^*(\cdot)$ as a flexible function of these predictors. The following subsections present different methods and their advantages in estimating $\mathbb{E}_t[r_{i,t+1}]$.

2.2.1 Penalized Linear

The most familiar model I consider as a benchmark is the linear model for expected return $g^*(x_{i,t};\theta) = x'_{i,t}\theta$ with the following objective function:

$$\mathcal{L}(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(r_{i,t+1} - g^*(x_{i,t};\theta))^2 \tag{3}$$

For comparison, this loss function is firstly minimized to get the benchmark OLS estimator. Note that I assume θ is the same constant across all funds for a given in-sample estimation period T and predictor vector $x_{i,t}$ captures all skill heterogeneity across funds.

Penalized linear models still assume a linear form for expected performance but combine the original loss function with an additional penalty term:

$$\mathcal{L}(\theta;\cdot) = \mathcal{L}(\theta) + \phi(\theta;\cdot), \tag{4}$$

where I consider the general elastic net penalty which takes the following form:

$$\phi(\theta;\lambda,\rho) = \lambda(1-\rho)\sum_{m=1}^{M}|\theta_m| + \frac{1}{2}\lambda\rho\sum_{m=1}^{M}\theta_m^2. \tag{5}$$

The elastic net (EN) penalty involves two nonnegative hyperparameters, λ and ρ. Specifically, the case when $\rho = 0$ corresponds to the least absolute shrinkage and selection operator (LASSO) with only L_1 penalty. This penalty acts for variable selection where it allows coefficients on predictive variables to be exactly zero. In this sense, LASSO imposes sparsity so that only the most important variables are selected. On the other hand, the case when $\rho = 1$ corresponds to the ridge regression which only uses L_2 penalty. Although ridge regression does not impose sparsity as LASSO to push coefficients to be exactly zeros, it shrinks unduly large coefficients towards zero. This shrinkage feature is particular useful when predictors are correlated where standard OLS gives unstably large estimates with substantial estimation errors. The case in between when $0 < \rho < 1$ therefore incorporates both sparsity and shrinkage among predictors.

As shown in Table 1, fund performance predictors examined in this paper can be classified into three groups: characteristics-based, performance-based, and activeness measures. Given such group structure, it is desirable to have all coefficients within a group to be nonzero or zero simultaneously. On the other hand, I would like to incorporate sparsity within each group as well. Simon et al. (2013) proposes a penalty term that allows sparsity across groups and within each group. For J groups of predictors, the penalty term can be specified as

$$\phi(\theta; \alpha) = \lambda \sum_{j=1}^{J} \left[(1-\alpha) \|\theta_j\|_2 + \|\theta_j\|_1 \right], \tag{6}$$

where θ_j is a vector of coefficients corresponds to the j-th group of predictors.

2.2.2 Dimension Reduction

Although shrinkage helps deal with correlated predictors, a more direct and simple approach is to transform the predictor space such that the transformed predictors are orthogonal to each other. Principal component regression (PCR) and partial least squares (PLS) serve this purpose well.

PCR involves two steps. In the first step, it extracts principal components from existing predictors as a smaller set of linear combinations that best preserve the covariance structure among original predictors. In the second step, a few leading components are used in standard predictive regression as in OLS. The problem with PCR is that it does not incorporate any information on the covariance relation between predictors and the target performance measures or returns. PLS solves this issue by first estimating each predictor's contribution to predicting target performance and then forming linear combination of those predictors using each predictor's contribution as weight. A successful application of PLS to estimating overall equity market risk premia can be found in Kelly and Pruitt (2013).

Mathematically, rewrite the linear model $r_{i,t+1} = x'_{i,t}\theta + \epsilon_{i,t+1}$ as a vectorized version:

$$R = X\theta + E, \tag{7}$$

where R is the $NT \times 1$ vector of $r_{i,t+1}$, X is the $NT \times M$ matrix of stacked predictors $x_{i,t}$, and E is a $NT \times 1$ vector of residuals $\epsilon_{i,t+1}$.

Both PCR and PLS reduce the dimensionality of the predictor space by transforming the original predictor space into a smaller number of K linear combinations of predictors.

$$R = (X\Omega_K)\theta_K + \tilde{E}. \tag{8}$$

Ω_K is $M \times K$ matrix with columns $w_1, w_2, \ldots, w_K$. Each w_j is the set of linear combination weights used to create the jth predictive components, and θ_K is a $K \times 1$ vector.

PCR chooses the combination weights Ω_K recursively such that the jth linear combination solves

$$w_j = \arg\max_{w} \mathrm{Var}(Xw), \quad \text{s.t. } w'w = 1, \quad \mathrm{Cov}\,(Xw, Xw_l) = 0, l = 1, 2, \ldots, j-1. \tag{9}$$

On the other hand, PLS searches K linear combinations of predictors X such that the new combinations have maximal predictive relation with the performance measure.

Specifically, the chosen weight to construct the jth PLS component is found by solving

$$w_j = \arg\max_w \text{Cov}(R, Xw), \quad \text{s.t.} \quad w'w = 1, \quad \text{Cov}\,(Xw, Xw_l) = 0, \quad l = 1, 2, \ldots, j-1 \tag{10}$$

Eventually, after finding the solution for Ω_K, θ_K is estimated by OLS regression using R on $X\Omega_K$.

3 Data and Sample Selection

The mutual fund sample ranges from 1994 to 2016[11]. Fund monthly returns and characteristics are from Center for Research in Security Prices (CRSP) survivor-bias-free mutual fund database. Fund quarterly holdings are extracted from Thomson Reuters (former CDA/Spectrum) s12 file. I use MFLINKS constructed in Wermers (2000) to merge fund returns and holdings data. When a fund has multiple share classes, I construct the TNA-weighted average of CRSP net returns, expenses, turnover ratio, and other characteristics for each fund.

Since my analysis focuses on actively managed U.S. domestic equity funds, I exclude international, municipal bonds, bond and preferred, and balanced funds based on CDA/Spectrum investment objective code. I further classify actively managed funds using Lipper, Strategic Insight and Wiesenberger code. The final sample includes three fund styles (aggressive growth, growth, and growth and income) and the rest of funds are grouped as one style. Evans (2010) finds that mutual fund incubation introduces biases in fund performance. I therefore put three additional filters as to control for such biases: (1) only funds with total net asset no less than \$15 millions are included; (2) observations preceding a fund's first offer date as reported in CRSP are eliminated; (3) observations with missing fund names are not included. The Online Appendix provides further details

[11] Specifically, the predictor sample is from December 1994 to Novemeber 2016 and the corresponding return period is from January 1995 to December 2016. The sample ends in 2016 since I require complete information of all 12 predictors in my sample and two of the 12 predictors examined in this paper (active share and duration) is only available up to September 2015.

regarding the cleaning procedure for mutual fund data. The full sample period for fund characteristics and performance predictors are from December 1994 to November 2016. The 12 predictors assessed in this paper and their definitions are laid out in Table 1 and 2[12].

For performance evaluation, I obtain information variables measuring economic conditions including lagged values of one-month T-bill yield from Ken French's website, dividend yield of the CRSP value-weighted NYSE/AMEX stock index, term spread (measured by the difference between yields on 10-year treasuries and three-month T-bills), and default spread (measured by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds) from FRED.

Table 3 presents the summary statistics of fund characteristics at the end of each year from 1994 to 2016. I require that a fund with information on all 12 predictors to be included for any cross-section in my sample. There is a secular pattern that the average size of actively managed funds usually peaked before any economic downturn, and the number of funds do not increase significantly over the years. Moreover, in more recent years, actively managed equity funds have experienced declines in average turnover as their average size increases over time, suggesting that even actively managed funds have become increasingly passive throughout past few years. As actively managed funds become more passive, it would be more difficult to detect active outperforming funds in real time using the activeness measures discovered in previous literature.

Table 4 provides summary statistics of the 12 predictors from December 1994 to November 2016[13]. I consider these 12 predictors since their construction does not require hard-to-obtain fund information from investor's perspective. The descriptive statistics in Panel A are computed as time-series averages of monthly statistics in each cross-section, except the first-order autocorrelation coefficient. On average, funds earn a slightly negative net-fee one-year Carhart alpha as found in previous studies (Carhart, 1997, Fama

[12] See Appendix A for the construction details of some of the 12 performance predictors.

[13] Since holdings are reported to the SEC and a three-month delay is imposed for investors to use holding-based predictors including return gap, active share, active weight, and fund duration.

and French, 2010). It is worth mentioning that all predictors are highly persistent for a given fund, suggesting that they act as potential proxies for skills as argued in the original studies. Panel B shows the contemporaneous pairwise Pearson correlations between the 12 predictors. Consistent with the time-series pattern shown in Table 3, in the cross-section, larger funds are generally less active with lower turnover, lower active share and active weight, higher return R-squared, and longer equity holding duration. As expected, within either performance-based or activeness category, predictors are correlated with each other. For instance, two measures of managerial activeness (active share and active weight) are highly positively correlated as expected for actively managed equity funds. And R-squared, regarded as an opposite measure to activeness, has strongly negative correlations with both active share and active weight. Finally, fund duration has a negative correlation with active share, while a slightly positive correlation with active weight, which is in general consistent with the concept that funds with infrequent rebalancing (i.e., high duration funds) tend to be less active. Overall, the summary statistics of predictors are consistent with existing findings in previous studies.

4 In-Sample Performance of Individual Predictors

Before examining the real-time predictability of predictors, I first validate the in-sample performance of each individual predictor using full sample information from December 1994 to November 2016. I construct the in-sample Carhart four-factor alpha spread of each individual predictor. Specifically, at the end of each month, funds are grouped into quintiles based on the predictor value in current month. I compute the next-month return spread between funds within the highest quintile and funds within the lowest quintile for a given predictor. Portfolios are rebalanced at monthly frequency. Table 5 illustrates the full-sample unconditional performance of predictor-sorted fund portfolios using the standard Carhart four-factor (C4) model (Carhart, 1997) as the benchmark:

$$R^H_{P,t} - R^L_{P,t} = \alpha_P + \beta_P(R_{M,t} - R_{f,t}) + s_P R_{SMB,t} + h_P R_{HML,t} + m_P R_{MOM,t} + \epsilon_{P,t}, \quad (11)$$

where $R^H_{P,t} - R^L_{P,t}$ is the return spread between the highest quintile and the lowest quintile fund portfolio based on predictor P.

Consistent with previous studies, Panel A in Table 5 shows that with equal-weighting, fund size, one-year Carhart alpha, one-month return, active share, R-squared, and active weight are significant predictors for the following month fund performance in the full sample, and the predictive signs are consistent with original studies. Panel B with value-weighting shows a slightly different picture from Panel A. With value-weighting, low expense funds have significantly better future performance than high expense funds. And high turnover funds now perform significantly worse than low turnover funds. For other activeness measures, with value-weight schemes active share does not predict future fund performance by itself. Moreover, fund duration now becomes a significant predictor for performance[14]. In summary, for each weighting scheme, six out of 12 predictors generate economically significant Carhart alpha spread between highest and lowest quintiles fund portfolios within the full sample.

5 Real-Time Performance of Predictors

A drawback of evaluating each predictor separately is that it ignores covariance structure among multiple predictors. For instance, as shown in Table 5, weighting schemes matter for some of the predictors given that fund size is correlated with most of other predictors. Moreover, even if predictors are found to perform well to distinguish best performing funds relative to worst funds, it is not suitable for a typical mutual fund investor who can only long a fund portfolio instead of shorting. Moreover, we still know rela-

[14] Cremers and Petajisto (2009) find that active share lacks statistically significant predictive power for fund performance in the cross-section though a later study (Cremers and Pareek, 2016) find that conditional on fund duration, active share predicts performance significantly.

tively little on whether the best performing funds selected by predictors can outperform a passive market portfolio in real time.

This section assesses the *ex ante* real-time predictive power of mutual fund performance predictors to resolve these issues with rule-based approaches and machine learning methods outlined in Section 2. For the rule-based approach, I first characterize the optimal fund selection rules that are determined in-sample, and then evaluate a real-time fund portfolio with two standard evaluation frameworks.

5.1 Rule-Based Portfolio Sorting Approach

In this subsection, I implement two versions of the rule-based portfolio sorting approach described earlier. The first version (Rule 1 henceforth) only selects the single best-performing rule and involves no validation for how many rules to be selected within each in-sample period. The in-sample period is 1995-2000 and the last in-sample period is 1995-2015, with expanding window for each year. The corresponding out-of-sample year is from 2001 to 2016. The second version (Rule 2 henceforth) considers a one-year validation period within each in-sample period for tuning the hyperparameter (i.e., fraction of rules selected) to avoid potential over-fitting problems using in-sample information. More precisely, I split the in-sample period into a training period and a one-year validation period. The first in-sample evaluation uses 1995-1999 as the training period with 2000 as the validation period, and the last in-sample evalution uses 1995-2014 as the training period with 2015 as the last validation year. The corresponding OOS year is the same as the version without validation (2001-2016).

Table 6 shows the single best-performing rule selected using Rule 1 based on corresponding in-sample performance. Among all predictors, performance-based variables perform the best compared to either fund characteristics and activeness measures. Given rule-based portfolio sorts are dependent, the second variable in a two-way sort is the more relevant variable that contributes in-sample predictability. Using the rule-based portfolio sorting approach without validation shows that the one-year return after controlling for short-term (one-month) return performs the best for 15 out of 16 in-sample periods.

Compared to Rule 1, Rule 2 admits several rules in order to average out noises associated with picking only the single best-performing rule. Table 7 presents the top-3 best-performing rules from the best to the worst using Rule 2. The top performing rule is largely the same as using Rule 1. A more noticeable feature is that active measures such as turnover, R-squared, and active weight start to matter as either the second-best or third-best performing rules. For instance, R-squared appears to be either the first controlling variable or the second predictive variable among the top-3 rule in any OOS year from 2003 to 2016. Still, performance-based measures prevail as the second predictive variable (41 out of 48 rules), and R-squared as the only other predictive variable that matters (7 out of 48 rules).

Panel A in Table 8 shows the risk-adjusted OOS performance of the real-time portfolio formed based on rules selected using either Rule 1 and Rule 2. Surprisingly, the OOS performance of rule-based portfolio without validation outperforms the passive market portfolio by 21 basis points (or 2.52% per year) at 10% level of significance, with only the single best-performing rule used. In contrast, the OOS performance of rule-based portfolio with validation does not significantly outperform the market, possibly due to the fact that multiple rules dilute the real-time predictability. However, after controlling for additional risk factors, none of the real-time portfolios generate significant positive alpha.

I further examine risk exposures of these two real-time portfolios. Given the time-varying nature of performance predictability, I conduct the analysis using the conditional framework by Ferson and Schadt (1996). Specifically, I study whether low-frequency macroeconomic information can account for the time-varying performance of OOS portfolios:

$$\begin{aligned} R_t - R_{f,t} = \alpha + (\beta + B'z_{t-1})(R_{M,t} - R_{f,t}) + sR_{SMB,t} + hR_{HML,t} + mR_{MOM,t} \\ + \epsilon_t, \end{aligned} \tag{12}$$

where R_t is return for the OOS portfolios. The one-month lagged macroeconomic variables z_{t-1}[15] include one-month T-bill yield, dividend yield of the CRSP value-weighted NYSE/AMEX stock index, term spread (measured by the difference between yields on 10-year treasuries and three-month T-bills), and default spread (measured by the yield difference between Moody's Baa-rated and Aaa-rated corporate bonds). As shown in Panel A of Table 9, conditional macroeconomic information does not matter for explaining the performance of the OOS portfolio in either case, and the OOS portfolios share a strong positive loading on the size and momentum factor, which is expected given both Rule 1 and Rule 2 select performance-based predictors for the best-performing rules during the in-sample periods. In summary, although rule-based approach without validation outperforms the market during my OOS evaluation period, it cannot generate significant alpha after accounting for more risk factors.

5.2 Regression-Based Machine Learning Methods

In this subsection, I implement six regression-based machine learning methods described in Section 2. As mentioned earlier, all these six methods are variants of the standard least squares estimator either with different specifications on an additional penalty term or through transformation of the original predictor space. I also examine the performance of OLS as the benchmark when evaluating each of these methods in OOS tests.

To evaluate each predictor's marginal contribution to return predictability, I consider a notion of variable importance following Gu et al. (2020). Predictor P's importance is measured as the reduction in panel predictive R^2 from setting the coefficient estimate of predictor P to zero, while holding other model estimates fixed. As in the machine learning literature, I use the training sample for calculating variable importance. To make each method comparable to each other, I compute the relative importance of predictor P as the fraction of total R^2 reduction attributed to that predictor, which is bounded between 0 and 1.

[15] z_{t-1} is demeaned for more precise estimates of coefficients.

Figure 2 shows the relative variable importance of each predictor based on training sample estimation using each of the six machine learning methods. Across all methods, short-term performance (one-month return) is found to be the primary predictor for performance forecasting, accounting for more than 40% reduction in R^2 for 5 out of the 6 methods. And active share is found to be the second important variable in 5 of 6 methods, which is different from the predictor ranking uncovered using rule-based approach. R-squared appears in the top-3 important predictors in 4 of 6 methods. One thing worth mentioning is that since LASSO, elastic net, and sparse group LASSO (SGL) all involve variable selection in the estimation step, their respective variable importance ranking is close to each other, which turns out to be reflected in their real-time forecasting as well.

Panel B in Table 8 shows the risk-adjusted OOS performance of the real-time portfolio formed using the six machine learning methods. Out of the six methods, OOS portfolio formed based on predictions from LASSO and elastic net are found to have a monthly positive Carhart alpha of 11 basis points (or 1.32% per year) at 5% level of significance. It is prominent that these two methods yield almost identical results. Since LASSO is a special case of elastic net with only variable selection feature, this suggests that variable selection in the original predictor space is an essential feature to generate real-time return predictability. The other method that can generate significantly positive return is the sparse group LASSO which also involves variable selection. However, SGL fails to generate any significantly risk-adjusted return.

Panel B in Table 9 presents the conditional performance evaluation for machine learning OOS portfolios. For LASSO and elastic net, none of the macroeconomic information variables matter for explaining performance, while for other regression-based methods one-month short-term interest rate and term spread play some roles in explaining OOS portfolio performance. In contrast to rule-based methods, regression-based methods build OOS portfolios that are not exposed to the momentum factor even though short-term one-month return turns out to be the most important predictor in all setups.

I further check the real-time predictability of predictors using machine learning across different fund investment styles in Table 10. It turns out that LASSO and elastic net only enable predictors for forecasting performance among more growth-oriented funds (i.e., aggressive growth and growth funds), and using SGL can barely generate a marginally statistically significant conditional Carhart alpha among aggressive growth funds (though the economic magnitude is about 1.56% per annum). Another noticeable finding is that none of the six machine learning methods would deliver superior risk-adjusted performance for conservative investors who mainly invest in income-oriented funds (with significantly positive exposure to the value factor).

5.3 Time Variations in Real-Time Portfolios

Previous tests provide evidence that variable selection methods LASSO and elastic net can provide reliable OOS performance upon selecting among the 12 predictors, with short-term one-month return being the predictability driver. This subsection attempts to examine how rule-based approach and machine learning methods work over time. I only consider rule-based approach without validation and elastic net from machine learning since each of these two methods performs the best in respective methodology type. Figure 3 shows the market-adjusted performance of real-time portfolios constructed using rule-based approach and elastic net over different OOS periods. Plot A and B demonstrate that before 2011, rule-based portfolio can outperform the market in general but the outperformance starts to deteriorate from 2011. In contrast, elastic net portfolio navigates away from significant down times of performance predictability by investing in the passive market portfolio during these periods. However, this benefit is associated with costs by missing positive market-adjusted gains during the first few OOS evaluation periods, partly due to the relatively short initial in-sample window for estimation. In this sense, through variable selection, elastic net or LASSO portfolios only take advantage of predictive information from some of the 12 predictors when predictability is strong, and switch to passive market portfolio by ignoring all predictors when overall predictability is weak. This feature essentially trades off some positive gains for less volatility in real-time portfolio.

Figure 4 demonstrates the investment value from real-time portfolios. If an investor starts to invest at the beginning of 2001 in the elastic net portfolio, she would obtain 31% higher return than the market portfolio by the end of 2016. On the other hand, if she invests in the rule-based portfolio without validation, the outperformance relative to the market would be 45% higher. This is consistent with the results in Table 8 which shows that rule-based portfolio without validation has a higher CAPM alpha than elastic net portfolio.

6 Flow Response to Predictor-Implied Performance

Real-time tests in previous sections show that in a simulated or hypothetical environment, short-term performance (one-month return) plays the primary role in forecasting future fund performance in real time given an information set of 12 predictors. Beyond this hypothetical setting, it would be of theoretical interests to understand how in reality investors incorporate predictive information into their capital allocation decisions. In this section, I use variations in fund flows to study the investment impact of predictive information implied by six of the 12 predictors[16].

Following the prior literature on fund flows (Zheng, 1999, Frazzini and Lamont, 2008), I make the simplified assumption that investors invest and redeem money from funds only at the end of each month. Fund flows is then calculated as percentage changes in fund total net assets net of capital appreciation. A positive value represents net inflow and a negative value implies net outflow. The fund flow for fund i at the end of month $t+1$ is

$$F_{i,t+1} = \frac{TNA_{i,t+1}}{TNA_{i,t}} - (1 + R_{i,t+1}), \tag{13}$$

where $TNA_{i,t}$ is the total net asset of fund i at the end of month t, and $R_{i,t+1}$ is the

[16] Tests for all 12 predictors will be added in future version of the paper.

return to fund i in month $t+1$ net of fees and expenses. To mitigate impact of outliers, I winsorize flows at 1% in each cross-section.

A first thought in examining investors' attention to predictive information is to include standard predictors in a panel regression to test whether coefficient on predictors are significantly different from zero. However, this approach can be confounded by the fact that investors allocating capital may use those predictors for other non-performance related reasons. For instance, a high fee fund may not be attractive to investors but it does not mean that this fund would not have skill in generating net-of-fee abnormal returns for investors. To resolve this confounding effect in order to isolate predictive content of each predictor, I propose a novel approach by further extracting a return component that can be attributed to each performance predictor.

Specifically, I modify the return decomposition procedure in Barber et al. (2016) to extract the return component that can be attributed to each performance predictor. To achieve this, I first run time-series rolling-window regressions for each fund to estimate fund's exposure to the high-minus-low portfolio using the most recent 5-year performance[17]:

$$R_{i,\tau} - R_{f,\tau} = \alpha_{i,t} + \gamma_{i,t}^{P} R_{\tau}^{P} + \sum_{j} \beta_{i,j,t} f_{j,\tau} + \epsilon_{i,\tau} \tag{14}$$

for $\tau = t-1, \ldots, t-60$, where R_{τ}^{P} is the high-minus-low return spread in month τ between two fund portfolios that equally weights funds within the fifth quintile based on predictor P and the fund portfolio that equally weights funds within the first quintile based on predictor P, both of which are formed at the end of month $\tau - 1$. $f_{j,\tau}$ denotes return to factor j. The high-minus-low spread for a given predictor does not represent any specific risk factor as in the asset pricing literature. Instead, it represents the market price of a common managerial skill capture by the predictor. For instance, when a group of funds owning a common skill can be captured by fund size, a fund with positive loading γ on

[17] I restrict the sample by including only funds with a five-year history of fund returns in order to estimate factor loadings in flow analysis.

the return spread means that the fund behaves as if it has a similar skill as large funds[18]. The purpose of this step is to estimate month t fund loading ($\widehat{\gamma}_{i,t}^{P}$) to the factor-mimicking fund portfolio R_t^P and factor loadings ($\widehat{\beta}_{i,j,t}$'s).

In the second step, I decompose fund excess return in month t into three components (pure alpha, predictor-implied performance, and performance attributed to risk factors):

$$R_{i,t} - R_{f,t} = \underbrace{\widehat{\alpha}_{i,t}}_{\text{pure alpha}} + \underbrace{\widehat{\gamma}_{i,t}^{P} R_t^P}_{\text{predictor } P\text{-implied performance (PIP)}} + \underbrace{\sum_j \widehat{\beta}_{i,j,t} f_{j,t}}_{\text{risk premia}} . \tag{15}$$

This decomposition allows me to isolate the return component attributed to predictive content embedded in predictor P. Moreover, the realized pure alpha, $\widehat{\alpha}_{i,t}$, is computed as the residual term from the decomposition, which captures any abnormal components not absorbed by common risk factors and the predictor-implied performance (PIP henceforth).

Since flows tend to be responsive to the lagged performance as well (Chevalier and Ellison, 1997), I follow Barber et al. (2016) to estimate the exponential decay rate of the flow-performance sensitivity using the full sample, which is estimated through a market-adjusted return (MAR) model as follows:

$$F_{i,t+1} = a + b \sum_{s=0}^{17} e^{-\lambda s} MAR_{i,t-s} + c' X_{i,t} + \eta_{t+1} + \epsilon_{i,t+1}, \tag{16}$$

where $MAR_{i,t-s}$ is the marked-adjusted return for fund i in month $t-s$. The vector of control variables $X_{i,t}$ includes fund characteristics observable at the end of month t, including lagged monthly flows from $t-17$ to t, log of one-month lagged fund TNA and fund age, most recent available fund expense ratio[19], a fund dummy that indicate whether the fund has any load, and the total volatility of monthly fund net return in prior 12

[18] An alternative approach would be assigning funds into different groups based on a predictor and using the average return of that group to proxy predictor-implied performance.

[19] Expense ratio is reported at annual frequency.

months (from $t-11$ to t). The model is estimated using nonlinear least squares with month fixed effects. The estimated exponential decay rate is 0.28 at the 1% significance level.

To reduce the number of parameters in estimation when accounting for flow response to lagged performance, I weight past performance using the exponential decay function estimated from equation (16) and construct an index for each return component. Specifically,

$$\begin{aligned} Alpha_{i,t} &= \frac{\sum_{s=0}^{17} e^{-\widehat{\lambda} s}\widehat{\alpha}_{i,t-s}}{\sum_{s=0}^{17} e^{-\widehat{\lambda} s}}, \\ PIP_{i,t}^{P} &= \frac{\sum_{s=0}^{17} e^{-\widehat{\lambda} s}\widehat{\gamma}_{i,t-s}^{P} R_{t-s}^{P}}{\sum_{s=0}^{17} e^{-\widehat{\lambda} s}}, \\ FACTOR_{i,j,t} &= \frac{\sum_{s=0}^{17} e^{-\widehat{\lambda} s}\widehat{\beta}_{i,j,t-s} f_{j,t-s}}{\sum_{s=0}^{17} e^{-\widehat{\lambda} s}}, \end{aligned} \tag{17}$$

where $FACTOR_{i,j,t}$ varies depending on which model to use as the testing field.

To assess the impact of PIP on fund flows, I run the following panel regression for each predictor P separately:

$$F_{i,t+1} = b_0 + b_\alpha Alpha_{i,t} + b_P PIP_{i,t}^{P} + \sum_j b_j FACTOR_{i,j,t} + \theta' X_{i,t} + \eta_{t+1} + \epsilon_{i,t+1}, \tag{18}$$

where $F_{i,t+1}$ is the flow for fund i in month $t+1$. The parameter of interest is b_P, which measures the flow sensitivity to past predictive information implied by predictor P. The panel regression includes a vector of controls ($X_{i,t}$) and month fixed effects (η_{t+1}) as in equation (16). Most importantly, for different predictor-implied factor-mimicking portfolios, I include in $X_{i,t}$ the lagged predictor itself as a control for that characteristic[20]. This novel specification helps to isolate predictive information from characteristic preference by investors that are not motivated by performance predictability[21]. For a given factor

[20] Essentially all predictors are fund characteristics.

[21] An alternative approach is to use ranking functions for each predictor or standardize predictors so that the coefficient in front of each predictor is comparable, which is exploited in Jones and Mo (2021). The difference between this approach and mine resembles the difference between covariance-

model, I consider the magnitude b_P across different predictor P. The comparison across predictors is also conducted within alternative factor models. If investors incorporate the predictive information implied in predictor P, the coefficient b_P should be significant.

Table 11 shows monthly flow sensitivity to different performance components using CAPM as the benchmark model[22]. For comparison, I also estimate flow response to two performance components (performance attributed to risk factors and alpha) in the last column of each panel, where I re-estimate the equations (15) and (18) without extracting PIP.

The first column of Table 11 illustrates that an average 1% increase in size-implied return after adjusting for market risk and controlling for size characteristic itself corresponds to a 0.5% increase in monthly fund flows, comparable with a 0.6% increase in fund flows when there is a 1% increase in pure alpha. This suggests that investors do respond to predictive information implied by fund size, in an economically significant magnitude. Similarly, active weight and fund duration also have strong predictive information captured by flow variations. In contrast, estimates from the third to the fifth columns reject that investors respond to the return components implied by return gap, active share, and R-squared, after controlling for corresponding characteristics. Interestingly, in such cases, characteristics dominates over predictor-implied return components.

Table 12 exhibits additional tests of flow responses to PIP across three fund investment styles: aggressive growth, growth, and growth and income. For aggressive growth funds, investor flows respond more to PIP compared to flows to growth and income funds in terms of both economical and statistical significance, except for active weight-implied performance. In contrast, although none of the flow-PIP sensitivities for growth and income funds is statistically significant, the economic magnitude for size-implied perfor-

based and characteristics-based asset pricing tests.

[22] Barber et al. (2016) and Berk and van Binsbergen (2016) argue that investors are most likely to use CAPM for risk-adjusting performance. I also conduct the test using five different benchmark factor models (CAPM, Fama-French three-factor model (FF3) (Fama and French, 1993), Carhart four-factor model (C4) (Carhart, 1997), Fama-French six-factor model (FF6) (Fama and French, 2018), and q-factor (HXZ4) (Hou et al., 2015)). To save space, I only include the tests using CAPM as the benchmark in the main text.

mance is higher than that for aggressive growth funds. In overall, these results suggest that investors in more growth-oriented funds are more inclined to use predictors to select funds for performance concerns than investors in more income-oriented funds, suggesting that investor attention to predictors is stronger among funds where they usually work well.

7 Conclusion

How would a rational investor select mutual funds based on *ex ante* information? Can mutual fund performance predictors be effectively used in real-time for better capital allocation for investors? Researchers have found abundant evidence that mutual fund performance is predictable *ex post*. This paper examines whether investors can utilize predictors without knowing which one would work *ex ante*. Specifically, I assess if a real-time investor could have used 12 fund performance predictors (expense ratio, turnover, fund flow, fund size, one-year return, Carhart alpha, one-month return, return gap, active share, R-squared, active weight, and fund duration) to outperform different benchmark stock portfolios over the 2001-2016 period. Employing rule-based and machine learning methods, I find one version of the rule-based real-time portfolio is able to beat the market in real time but generates no alpha relative Carhart four-factor model. In contrast, regression-based machine learning with variable selection feature (LASSO and elastic net) can deliver outperformance not only relative to the market benchmark (with annualized market-adjusted alpha of 1.68%) but also relative to additional risk factors (with annualized Carhart four-factor alpha of 1.32%). Further inspection on the real-time machine learning portfolio reveals that through variable selection, either LASSO or elastic net portfolio only exploits predictive information from some of the predictors when predictability is strong, and switches to the passive market portfolio by ignoring all predictors when overall predictability is weak. This feature essentially trades off some positive expected returns for less volatility in the real-time portfolio. Short-term fund performance (one-month return) turns out to be the main driver underlying any real-time predictability discovered by LASSO or elastic net. These findings justify potential value added by robo-advisors which aim to assist unsophisticated households to pick outperforming funds.

My paper further shows that beyond investors' usage of CAPM, investors react to the components of CAPM alpha implied by predictors in different ways, and investor attention to predictors is stronger among aggressive growth funds where those predictors are found to work well. These results suggest that real-time predictability exists not due to lack to investor attention to publicly available predictive information, instead the magnitude of any real-time excess gain discovered in this paper can be seen as a proxy cost an average investor needs to incur using intensive search algorithms to find skilled managers in the asset management industry. More investigation of investors' time-varying attention allocation on predictors and investors' sophistication in using predictive information would be an interesting venue for future work.

References

Amihud, Y. and Goyenko, R. (2013), 'Mutual fund's r^2 as predictor of performance', *Review of Financial Studies* **26**(3), 667–694.

Avramov, D. and Wermers, R. (2006), 'Investing in mutual funds when returns are predictable', *Journal of Financial Economics* **81**(2), 339–377.

Baks, K. P., Metrick, A. and Wachter, J. (2001), 'Should investors avoid all actively managed mutual funds? a study in bayesian performance evaluation', *Journal of Finance* **56**(1), 45–85.

Bali, T., Goyal, A., Huang, D., Jiang, F. and Wen, Q. (2021), 'Different strokes: Return predictability across stocks and bonds with machine learning and big data', *Swiss Finance Institute, Research Paper Series* pp. 20–110.

Barber, B. M., Huang, X. and Odean, T. (2016), 'Which factors matter to investors? evidence from mutual fund flows', *Review of Financial Studies* **29**(10), 2600–2642.

Barras, L., Gagliardini, P. and Scaillet, O. (Forthcoming), 'Skill, scale, and value creation in the mutual fund industry', *Journal of Finance* .

Barras, L., Scaillet, O. and Wermers, R. (2010), 'False discoveries in mutual fund performance: Measuring luck in estimated alphas', *Journal of Finance* **65**(1), 179–216.

Berk, J. B. and Green, R. C. (2004), 'Mutual fund flows and performance in rational markets', *Journal of Political Economy* **112**(6), 1269–1295.

Berk, J. B. and van Binsbergen, J. H. (2016), 'Assessing asset pricing models using revealed preference', *Journal of Financial Economics* **119**(1), 1–23.

Boguth, O. and Simutin, M. (2018), 'Leverage constraints and asset prices: Insights from mutual fund risk taking', *Journal of Financial Economics* **127**(2), 325–341.

Bollen, N. P. and Busse, J. A. (2004), 'Short-term persistence in mutual fund performance', *Review of Financial Studies* **18**(2), 569–597.

Campbell, J. Y. (2006), 'Household finance', *The Journal of Finance* **61**(4), 1553–1604.

Carhart, M. M. (1997), 'On persistence in mutual fund performance', *Journal of Finance* **52**(1), 57–82.

Chen, J., Hong, H., Huang, M. and Kubik, J. D. (2004), 'Does fund size erode mutual fund performance? the role of liquidity and organization', *American Economic Review* **94**(5), 1276–1302.

Chevalier, J. and Ellison, G. (1997), 'Risk taking by mutual funds as a response to incentives', *Journal of Political Economy* **105**(6), 1167–1200.

Cooper, M., Gutierrez, Jr, R. C. and Marcum, B. (2005), 'On the predictability of stock returns in real time', *Journal of Business* **78**(2), 469–500.

Cremers, K. M. and Petajisto, A. (2009), 'How active is your fund manager? a new measure that predicts performance', *Review of Financial Studies* **22**(9), 3329–3365.

Cremers, M. and Pareek, A. (2016), 'Patient capital outperformance: The investment skill of high active share managers who trade infrequently', *Journal of Financial Economics* **122**(2), 288–306.

DeMiguel, V., Gil-Bazo, J., Nogales, F. J. and AP Santos, A. (2021), 'Can machine learning help to select portfolios of mutual funds?', *Can Machine Learning Help to Select Portfolios of Mutual Funds* .

Doshi, H., Elkamhi, R. and Simutin, M. (2015), 'Managerial activeness and mutual fund performance', *Review of Asset Pricing Studies* **5**(2), 156–184.

Elton, E. J., Gruber, M. J., Das, S. and Hlavka, M. (1993), 'Efficiency with costly information: A reinterpretation of evidence from managed portfolios', *Review of Financial Studies* **6**(1), 1–22.

Evans, R. B. (2010), 'Mutual fund incubation', *Journal of Finance* **65**(4), 1581–1611.

Fama, E. F. and French, K. R. (1993), 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics* **33**(1), 3–56.

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

Fama, E. F. and French, K. R. (2018), 'Choosing factors', *Journal of Financial Economics* **128**(2), 234–252.

Ferson, W. E. and Schadt, R. W. (1996), 'Measuring fund strategy and performance in changing economic conditions', *Journal of Finance* **51**(2), 425–461.

Frazzini, A. and Lamont, O. A. (2008), 'Dumb money: Mutual fund flows and the cross-section of stock returns', *Journal of Financial Economics* **88**(2), 299–322.

Gârleanu, N. and Pedersen, L. H. (2018), 'Efficiently inefficient markets for assets and asset management', *The Journal of Finance* **73**(4), 1663–1712.

Goyenko, R. and Zhang, C. (2021), 'The joint cross section of option and stock returns predictability with big data and machine learning', *Available at SSRN 3747238* .

Gu, S., Kelly, B. and Xiu, D. (2020), 'Empirical asset pricing via machine learning', *The Review of Financial Studies* **33**(5), 2223–2273.

Harvey, C. R. and Liu, Y. (2018), 'Detecting repeatable performance', *Review of Financial Studies* **31**(7), 2499–2552.

Harvey, C. R. and Liu, Y. (2019), 'Cross-sectional alpha dispersion and performance evaluation', *Journal of Financial Economics* **134**(2), 273–296.

Hendricks, D., Patel, J. and Zeckhauser, R. (1993), 'Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988', *Journal of Finance* **48**(1), 93–130.

Hou, K., Xue, C. and Zhang, L. (2015), 'Digesting anomalies: An investment approach', *Review of Financial Studies* **28**(3), 650–705.

Ippolito, R. A. (1992), 'Consumer reaction to measures of poor quality: Evidence from the mutual fund industry', *Journal of Law and Economics* **35**(1), 45–70.

Jones, C. S. and Mo, H. (2021), 'Out-of-sample performance of mutual fund predictors', *Review of Financial Studies* **34**(1), 149–193.

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

Kacperczyk, M. and Seru, A. (2007), 'Fund manager use of public information: New evidence on managerial skills', *Journal of Finance* **62**(2), 485–528.

Kacperczyk, M., Sialm, C. and Zheng, L. (2005), 'On the industry concentration of actively managed equity mutual funds', *Journal of Finance* **60**(4), 1983–2011.

Kacperczyk, M., Sialm, C. and Zheng, L. (2006), 'Unobserved actions of mutual funds', *Review of Financial Studies* **21**(6), 2379–2416.

Kacperczyk, M., Van Nieuwerburgh, S. and Veldkamp, L. (2016), 'A rational theory of mutual funds' attention allocation', *Econometrica* **84**(2), 571–626.

Kaniel, R., Lin, Z., Pelger, M. and Van Nieuwerburgh, S. (2021), 'Machine-learning the skill of mutual fund managers', *Available at SSRN 3977883* .

Kelly, B. and Pruitt, S. (2013), 'Market expectations in the cross-section of present values', *Journal of Finance* **68**(5), 1721–1756.

Kozak, S., Nagel, S. and Santosh, S. (2020), 'Shrinking the cross-section', *Journal of Financial Economics* **135**(2), 271–292.

Lewellen, J. and Shanken, J. (2002), 'Learning, asset-pricing tests, and market efficiency', *Journal of Finance* **57**(3), 1113–1145.

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

Lo, A. W., Mamaysky, H. and Wang, J. (2000), 'Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation', *Journal of Finance* **55**(4), 1705–1765.

Martin, I. and Nagel, S. (Forthcoming), 'Market efficiency in the age of big data', *Journal of Financial Economics* .

Pesaran, M. H. and Timmermann, A. (1995), 'Predictability of stock returns: Robustness and economic significance', *Journal of Finance* **50**(4), 1201–1228.

Simon, N., Friedman, J., Hastie, T. and Tibshirani, R. (2013), 'A sparse-group lasso', *Journal of Computational and Graphical Statistics* **22**(2), 231–245.

Simutin, M. (2014), 'Cash holdings and mutual fund performance', *Review of Finance* **18**(4), 1425–1464.

Sirri, E. R. and Tufano, P. (1998), 'Costly search and mutual fund flows', *Journal of Finance* **53**(5), 1589–1622.

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

Zheng, L. (1999), 'Is money smart? a study of mutual fund investors' fund selection ability', *Journal of Finance* **54**(3), 901–933.

Table 1: **List of Mutual Fund Performance Predictors**

Category	Predictor	Study
Characteristics-Based	Expense Ratio (ER)	Elton et al. (1993)
	Fund Flow (Flow)	Zheng (1999)
	Fund Size (Size)	Chen et al. (2004)
Performance-Based	One-Year Return (Ret1y)	Hendricks et al. (1993)
	Carhart Alpha (Car1y)	Carhart (1997)
	One-Month Return (Ret1m)	Bollen and Busse (2004)
	Return Gap (RG)	Kacperczyk et al. (2006)
Activeness	Turnover (TR)	Elton et al. (1993)
	Active Share (AS)	Cremers and Petajisto (2009)
	R-Squared (R^2)	Amihud and Goyenko (2013)
	Active Weight (AW)	Doshi et al. (2015)
	Fund Duration (Dur)	Cremers and Pareek (2016)

Table 2: Predictor Definition

Predictor	Definition
ER	Annual expense ratio in fraction of total net asset
Flow	Three-month dollar flow in millions
Size	Log of total net asset in million dollars
Ret1y	One-year cumulative return of a fund
Car1y	Monthly Carhart four-factor alpha using 12 monthly returns from last 12 months
Ret1m	Most recent one month return net of fees
RG	Difference between net fund return and the net return to most recent fund stock holdings
TR	Minimum of aggregated sales or aggregated purchases of securities divided by the average 12-month total net assets of a fund
AS	Deviation of a fund portfolio holdings from its benchmark index holdings
R^2	R-squared from a regression of fund net excess return on Carhart four factors using monthly returns from last 24 months
AW	Deviation of a fund portfolio holdings from its market-cap weighted holdings
Dur	Average time (in years) a fund rebalances its stock holdings

Table 3: Summary Statistics - Number of Funds.
This table reports the summary statistics for actively managed U.S. domestic equity funds at the end of each year in the sample from 1994 to 2016. The fund sample is constructed such that only observations where each predictor is available are kept. Additional filters include: (1) only funds with at least $15 millions of total net assets (TNA) are kept; (2) incubation bias is adjusted by eliminating fund observations preceding a fund's first offer date as reported in CRSP and observations with missing fund names. TNA, Expense Ratio, and Turnover Ratio are reported as the cross-sectional average at the end of each year and winsorized at 1% and 99% levels.

Year	Num. of Funds	TNA (in Millions)	Turnover Ratio (%)	Expense Ratio (%)
1994	269	1204.65	72.01	1.16
1995	209	1677.63	73.65	1.16
1996	214	1683.13	72.71	1.17
1997	471	2206.39	78.97	1.19
1998	526	2497.50	80.74	1.14
1999	585	2911.42	88.49	1.17
2000	670	2481.49	96.81	1.20
2001	719	2022.17	88.43	1.26
2002	806	1376.50	85.68	1.29
2003	900	1711.11	78.10	1.27
2004	987	1778.94	74.87	1.26
2005	1035	1765.40	76.38	1.24
2006	1084	1924.94	75.06	1.20
2007	1149	1925.00	81.91	1.18
2008	1146	1110.40	89.15	1.20
2009	1182	1386.76	74.97	1.18
2010	1268	1421.84	71.09	1.15
2011	1241	1386.31	65.10	1.13
2012	1191	1582.06	61.96	1.11
2013	1181	2138.59	59.03	1.09
2014	1163	2332.54	58.11	1.07
2015	1128	2246.54	57.50	1.06
2016	1072	2359.53	57.05	1.04

Table 4: Summary Statistics - Fund Performance Predictors.
Panel A exhibits descriptive statistics of the 12 predictors described in Table 1 and 2 from December 1994 to November 2016. All predictors are winsorized at 1% and 99% level. Obs. is the time-series average of number of funds in each cross-section in the sample. Mean is the time-series average of cross-sectional mean of a predictor. Median is the time-series average of cross-sectional median. SD is the time-series average of cross-sectional standard deviation. Min (max) is the time-series average of cross-sectional minimum(maximum). AR(1) is the cross-sectional median of first-order autocorrelation of a predictor for a fund. Panel B exhibits the contemporaneous pairwise Pearson correlations among predictors.

Panel A: Descriptive Statistics

Predictor	Obs.	Mean	Median	SD	Min	Max	AR(1)
ER	900	1.17%	1.14%	0.36%	0.27%	2.18%	0.95
Flow	900	-1.20	-1.72	108.04	-459.57	484.43	0.78
Size	900	6.12	6.08	1.67	2.89	10.22	0.97
Ret1y	900	10.81%	10.06%	12.41%	-18.34%	187.00%	0.92
Car1y	900	-0.05%	-0.07%	0.90%	-4.19%	13.43%	0.84
Ret1m	900	0.87%	0.83%	2.38%	-7.04%	20.37%	0.10
RG	900	-0.01%	-0.02%	1.26%	-7.15%	17.95%	0.13
TR	900	75.74%	59.47%	61.18%	2.98%	317.57%	0.93
AS	900	0.81	0.84	0.15	0.15	1.00	0.96
R^2	900	0.91	0.93	0.07	0.33	0.99	0.94
AW	900	0.79	0.77	0.21	0.12	1.58	0.93
Dur	900	5.64	4.86	3.49	0.01	17.69	0.96

Panel B: Pairwise Correlation

	ER	Flow	Size	Ret1y	Car1y	Ret1m	RG	TR	AS	R^2	AW	Dur
ER	1											
Flow	0.068	1										
Size	-0.372	-0.097	1									
Ret1y	0.024	0.113	0.03	1								
Car1y	0.002	0.033	0.006	0.431	1							
Ret1m	0.007	0.013	0.004	0.268	0.253	1						
RG	0.013	-0.001	-0.003	0.145	0.192	0.001	1					
TR	0.186	0.019	-0.148	-0.017	-0.027	-0.003	0.007	1				
AS	0.336	0.052	-0.195	0.062	0.008	0.017	0	0.023	1			
R^2	-0.196	-0.045	0.092	-0.113	-0.107	-0.02	-0.036	-0.061	-0.367	1		
AW	0.106	0.027	-0.031	0.016	0.006	0.006	-0.005	0.006	0.16	-0.206	1	
Dur	-0.244	-0.097	0.224	-0.006	0.014	0.001	-0.003	-0.592	-0.166	0.108	0.008	1

Table 5: In-Sample Performance of Mutual Fund Predictors.
This table exhibits the Carhart four-factor (C4) Carhart (1997) alpha spread across quintile fund portfolios. Fund portfolios are formed based on value of previous month-end predictors defined in Table 1 and 2. Portfolios are rebalanced at the end of each month. The Newey-West corrected standard error with six-month lag is shown in parentheses. Alpha spread is in monthly percentage. Absolute t-statistics are shown in parentheses. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively. The sample for all predictors is the same, with returns from January 1995 to December 2016.

Panel A: Equal-Weighted Fund Portfolio

Predictor	Portfolio	C4 Alpha	Abs. t-stat
ER	High - Low	-0.03	(0.5)
Flow	High - Low	0.08	(1.03)
Size	High - Low	-0.15***	(2.61)
Ret1y	High - Low	0.24	(1.36)
Car1y	High - Low	0.29***	(3.34)
Ret1m	High - Low	0.60***	(2.84)
RG	High - Low	0.01	(0.11)
TR	High - Low	-0.05	(0.53)
AS	High - Low	0.12*	(1.67)
R^2	High - Low	-0.18*	(1.78)
AW	High - Low	0.19***	(2.77)
Dur	High - Low	0.12	(1.65)

Panel B: Value-Weighted Fund Portfolio

Predictor	Portfolio	C4 Alpha	Abs. t-stat
ER	High - Low	-0.20***	(3.27)
Flow	High - Low	-0.02	(0.33)
Size	High - Low	-0.12**	(2.11)
Ret1y	High - Low	0.09	(0.47)
Car1y	High - Low	0.25**	(2.55)
Ret1m	High - Low	0.62***	(2.78)
RG	High - Low	-0.10	(1.41)
TR	High - Low	-0.18**	(2.38)
AS	High - Low	0.00	(0.06)
R^2	High - Low	-0.14	(1.18)
AW	High - Low	0.09	(1.04)
Dur	High - Low	0.16***	(2.81)

Table 6: Best-Performing Rule Selected Using Rule-Based Approach Without Validation.

This table exhibits best-performing rule selected using rule-based approach without validation based on corresponding in-sample performance. A rule is either a single predictor quintile or a combination of quintiles of two predictors. 12 predictors described in Table 1 and 2 are considered to form the fund selection rules.

2001	2002	2003	2004	2005	2006	2007	2008
Car1y, 5; Ret1m, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5
2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5

Table 7: Top-3 Best-Performing Rule Selected Using Rule-Based Approach With Validation.

This table exhibits predictor ranking based on the training sample performance of selected rules to pick funds. 12 predictors described in Table 1 and 2 are considered to form the fund selection rules.

Rank	2001	2002	2003	2004	2005	2006	2007	2008
1	TR, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5
2	Car1y, 5; Ret1m, 5	TR, 5; Ret1m, 5	R2, 1; Ret1y, 5	R2, 1; Ret1y, 5	R2, 1; Ret1y, 5	R2, 1; Ret1y, 5	R2, 1; Ret1y, 5	R2, 1; Ret1y, 5
3	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Car1y, 5; Ret1m, 5	Flow, 4; Ret1y, 5	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5

Rank	2009	2010	2011	2012	2013	2014	2015	2016
1	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5	Ret1m, 5; Ret1y, 5
2	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	AW, 2; R2, 1	AW, 2; R2, 1	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5
3	R2, 1; Ret1y, 5	AW, 2; R2, 1	Car1y, 5; Ret1m, 5	Car1y, 5; Ret1m, 5	AW, 2; R2, 1	AW, 2; R2, 1	AW, 2; R2, 1	AW, 2; R2, 1

Table 8: Real-Time Performance of Rule-Based and Machine Learning Portfolios.
This table presents the monthly returns for fund portfolios constructed using rule-based approach (without and with validation) and six regression-based machine learning methods (OLS as the benchmark). FF3 Alpha is from Fama-French three factor model Fama and French (1993). C4 alpha is from Carhart four-factor model Carhart (1997). All returns are in monthly percentage. Absolute t-statistics are reported in parentheses.

Panel A: Rule-Based Approaches				
Validation	Average Return	CAPM Alpha	FF3 Alpha	C4 Alpha
No	0.79**	0.21*	0.11	0.08
	(2.14)	(1.71)	(1.20)	(0.81)
Yes	0.70*	0.11	0.02	-0.01
	(1.89)	(1.17)	(0.28)	(0.17)
Panel B: Machine Learning Methods				
Method	Average Return	CAPM Alpha	FF3 Alpha	C4 Alpha
OLS (Benchmark)	0.56	-0.07	-0.14	-0.12
	(1.37)	(0.61)	(1.35)	(1.22)
Ridge	0.58	-0.04	-0.11	-0.11
	(1.46)	(0.38)	(1.15)	(1.07)
LASSO	0.74**	0.14**	0.11**	0.11**
	(1.98)	(2.18)	(2.25)	(2.16)
Elastic Net	0.74**	0.14**	0.11**	0.11**
	(1.98)	(2.18)	(2.25)	(2.17)
PCR	0.61	0.00	-0.08	-0.09
	(1.60)	(0.01)	(1.15)	(1.22)
PLS	0.55	-0.07	-0.14	-0.12
	(1.37)	(0.65)	(1.37)	(1.23)
SGL	0.68*	0.07	0.03	0.03
	(1.83)	(0.96)	(0.44)	(0.44)

Table 9: Conditional Performance of Real-Time Fund Portfolios.
This table exhibits the conditional performance attribution of real-time fund portfolios within the Ferson and Schadt (1996) (FS) framework. Real-time fund portfolios are selected using two types of approaches: rule-based and machine learning methods. Panel A shows the performance of rule-based approach and Panel B shows the results from machine learning methods. 12 fund predictors defined in Table 1 and 2 are used as inputs for prediction. All fund portfolios are formed through equal-weighting. The one-month lagged conditional variables include one-month T-Bill, dividend yield (DY), term spread (TS), and default spread (DS). All conditional variables are demeaned to have zero sample means. Absolute t-statistics based on the Newey-West corrected standard error using six-month lag are shown in square brackets. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively. The real-time portfolio performance is from January 2001 to December 2016.

Panel A: Rule-Based Approach

Validation	Alpha	Market	Market ×1m T-Bill	Market ×DY	Market ×TS	Market ×DS	SMB	HML	MOM
No	0.10	0.96***	0.35	0.04	0.02	0.01	0.39***	0.05	0.11**
	(1.05)	(23.39)	(0.97)	(0.48)	(0.47)	(0.18)	(8.70)	(0.67)	(2.11)
Yes	0.00	0.98***	0.19	0.05	0.03	-0.01	0.36***	0.03	0.10**
	(0.06)	(31.56)	(0.87)	(0.90)	(0.99)	(0.14)	(11.74)	(0.48)	(2.29)

Panel B: Machine Learning Methods

Method	Alpha	Market	Market ×1m T-Bill	Market ×DY	Market ×TS	Market ×DS	SMB	HML	MOM
OLS (Benchmark)	-0.05	1.01***	0.58**	-0.08	0.05**	0.07	0.30***	0.03	-0.02
	(0.64)	(34.63)	(2.39)	(1.31)	(2.24)	(1.42)	(6.62)	(1.09)	(0.56)
Ridge	-0.04	0.99***	0.45**	-0.08	0.05**	0.08	0.30***	0.04	-0.01
	(0.48)	(37.51)	(2.00)	(1.08)	(2.45)	(1.29)	(7.71)	(1.09)	(0.12)
LASSO	0.12**	1.00***	0.07	0.00	0.01	0.03	0.13***	-0.02	0.00

Table 9: Conditional Performance of Real Fund Portfolios - FS (Continued)

	(2.06)	(40.58)	(0.46)	(0.04)	(0.45)	(0.66)	(2.87)	(0.77)	(0.08)
Elastic Net	0.12**	1.00***	0.07	0.00	0.01	0.03	0.13***	-0.02	0.00
	(2.07)	(40.52)	(0.46)	(0.05)	(0.44)	(0.66)	(2.87)	(0.76)	(0.08)
PCR	-0.10	1.00***	0.37**	0.13**	0.02	-0.05	0.31***	0.04	0.04
	(1.25)	(36.58)	(2.21)	(2.28)	(0.80)	(0.68)	(5.75)	(0.89)	(0.93)
PLS	-0.05	1.00***	0.57**	-0.08	0.04**	0.06	0.29***	0.04	-0.02
	(0.68)	(34.28)	(2.29)	(1.24)	(2.08)	(1.35)	(6.70)	(1.14)	(0.56)
SGL	0.05	1.00***	0.34***	0.06	0.04***	-0.02	0.18***	0.01	0.00
	(0.70)	(41.77)	(3.05)	(1.36)	(2.68)	(-0.38)	(3.71)	(0.24)	(0.05)

Table 10: Conditional Performance of Real-Time Fund Portfolios by Fund Styles.
This table exhibits the conditional performance attribution of real-time fund portfolios within the Ferson and Schadt (1996) (FS) framework for three styles of funds: Aggressive Growth, Growth, and Growth and Income. Real-time fund portfolios are selected using two types of approaches: rule-based and machine learning methods. Panel A shows the performance of rule-based approach and Panel B shows the results from machine learning methods. 12 fund predictors defined in Table 1 and 2 are used as inputs for prediction. All fund portfolios are formed through equal-weighting. The one-month lagged conditional variables include one-month T-Bill, dividend yield (DY), term spread (TS), and default spread (DS). All conditional variables are demeaned to have zero sample means. Absolute t-statistics based on the Newey-West corrected standard error using six-month lag are shown in square brackets. ***, **, and * denote significance at 1%, 5%, and 10% levels, respectively. The real-time portfolio performance is from January 2001 to December 2016.

Panel A: Rule-Based Approach

Style	Validation	Alpha	Market	Market ×1m T-Bill	Market ×DY	Market ×TS	Market ×DS	SMB	HML	MOM
Aggressive Growth	No	0.17	1.03***	1.28**	0.15	0.05	-0.04	0.48***	-0.01	0.16**
		(1.15)	(23.33)	(2.43)	(1.37)	(1.15)	(0.42)	(7.03)	(0.05)	(2.52)
	Yes	0.04	1.03***	0.75*	0.08	0.05	0.03	0.43***	-0.05	0.15***
		(0.37)	(31.72)	(1.95)	(0.98)	(1.31)	(0.39)	(8.95)	(0.50)	(3.45)
Growth	No	0.11	0.99***	0.22	0.04	0.01	-0.01	0.39***	0.02	0.12**
		(1.24)	(22.41)	(0.74)	(0.5)	(0.33)	(0.12)	(7.23)	(0.32)	(2.34)
	Yes	-0.03	1.00***	0.41**	0.03	0.03	0.01	0.35***	-0.02	0.08**
		(0.41)	(37.31)	(2.21)	(0.58)	(1.38)	(0.10)	(11.34)	(0.48)	(2.42)
Growth and Income	No	0.06	0.88***	-0.57**	-0.02	0.00	0.00	0.06**	0.13***	0.04**
		(0.86)	(38.26)	(-2.17)	(0.31)	(0.10)	(0.01)	(2.20)	(2.96)	(2.12)
	Yes	-0.04	0.94***	-0.26*	-0.02	-0.02	0.00	0.05***	0.11***	0.05***
		(0.95)	(63.7)	(1.68)	(0.35)	(1.13)	(0.05)	(2.70)	(3.06)	(2.70)

Table 10: Conditional Performance of Real Fund Portfolios - FS (Continued)

Panel B: Machine Learning Methods

Style	Method	Alpha	Market	Market ×1m T-Bill	Market ×DY	Market ×TS	Market ×DS	SMB	HML	MOM
Aggressive Growth	OLS (Benchmark)	-0.03	1.05***	0.74*	-0.10	0.07**	0.06	0.33***	-0.01	-0.01
		(0.33)	(30.49)	(1.90)	(1.00)	(2.18)	(0.84)	(6.35)	(0.15)	(0.16)
	Ridge	-0.06	1.04***	0.59*	-0.07	0.09**	0.06	0.33***	-0.03	0.03
		(0.65)	(29.54)	(1.75)	(0.65)	(2.41)	(0.69)	(6.54)	(0.55)	(0.42)
	LASSO	0.18**	1.01***	0.40**	0.09	0.05**	-0.03	0.17***	-0.07	0.03
		(2.15)	(34.58)	(2.38)	(1.37)	(2.26)	(0.42)	(3.11)	(1.50)	(0.58)
	Elastic Net	0.18**	1.01***	0.40**	0.09	0.05**	-0.03	0.17***	-0.07	0.03
		(2.15)	(34.59)	(2.39)	(1.38)	(2.28)	(0.42)	(3.11)	(1.49)	(0.58)
	PCR	-0.12	1.06***	0.68*	-0.05	0.07**	0.07	0.34***	-0.06	0.02
		(1.52)	(27.62)	(1.68)	(0.68)	(2.29)	(1.15)	(5.85)	(1.54)	(0.42)
	PLS	-0.04	1.05***	0.77**	-0.10	0.08**	0.06	0.32***	-0.02	-0.01
		(0.40)	(30.88)	(2.03)	(1.05)	(2.37)	(0.90)	(6.17)	(0.56)	(0.12)
	SGL	0.13*	1.01***	0.42**	0.08	0.06**	-0.03	0.17***	-0.04	0.01
		(1.93)	(43.98)	(2.33)	(1.55)	(2.43)	(-0.67)	(3.32)	(-1.18)	(0.43)
Growth	OLS (Benchmark)	-0.07	1.02***	0.54**	-0.06	0.05***	0.05	0.27***	-0.03	-0.03
		(0.91)	(35.68)	(2.38)	(1.00)	(2.69)	(0.97)	(6.50)	(0.89)	(-0.89)
	Ridge	-0.06	1.01***	0.47**	-0.06	0.06***	0.05	0.29***	-0.01	-0.02
		(0.85)	(36.48)	(2.07)	(0.79)	(2.69)	(0.87)	(7.54)	(0.45)	(0.36)
	LASSO	0.15**	1.01***	0.37***	0.06	0.05***	-0.02	0.15***	-0.02	0.01
		(2.55)	(41.14)	(3.35)	(1.13)	(3.08)	(0.41)	(3.21)	(0.55)	(0.25)
	Elastic Net	0.15**	1.01***	0.37***	0.06	0.05***	-0.02	0.15***	-0.01	0.01

Table 10: Conditional Performance of Real Fund Portfolios - FS (Continued)

		(2.56)	(41.17)	(3.36)	(1.13)	(3.08)	(0.41)	(3.21)	(0.54)	(0.25)
	PCR	-0.10	1.03***	0.50***	0.10*	0.04*	-0.03	0.3***	-0.02	0.04
		(1.49)	(38.68)	(3.09)	(1.86)	(1.87)	(0.47)	(5.58)	(0.38)	(1.08)
	PLS	-0.08	1.02***	0.57**	-0.05	0.05***	0.04	0.26***	-0.02	-0.03
		(1.02)	(33.87)	(2.36)	(0.83)	(2.62)	(0.86)	(6.54)	(0.81)	(0.73)
	SGL	0.05	1.02***	0.37***	0.08*	0.05***	-0.05	0.18***	-0.01	0.00
		(0.78)	(40.01)	(2.85)	(1.70)	(2.86)	(-0.96)	(3.68)	(-0.29)	(0.10)
Growth and Income	OLS (Benchmark)	-0.03	0.96***	0.07	-0.03	0.00	0.02	0.03	0.06***	-0.03
		(0.60)	(63.07)	(0.45)	(0.79)	(0.03)	(0.69)	(1.01)	(4.33)	(1.23)
	Ridge	0.00	0.94***	-0.01	-0.05	-0.01	0.03	0.03	0.07***	-0.03
		(0.10)	(55.56)	(0.07)	(0.94)	(0.47)	(1.06)	(1.12)	(4.51)	(1.25)
	LASSO	0.00	0.96***	0.20*	0.03	0.02*	-0.01	0.03*	0.02*	-0.02
		(0.08)	(77.68)	(1.89)	(1.23)	(1.88)	(0.57)	(1.71)	(1.68)	(1.37)
	Elastic Net	0.00	0.97***	0.22**	0.03	0.02*	-0.01	0.02	0.02*	-0.02
		(0.15)	(80.20)	(1.98)	(1.27)	(1.67)	(0.66)	(1.48)	(1.94)	(1.29)
	PCR	-0.02	0.91***	0.22**	0.09***	0.00	-0.04	0.05*	0.08***	-0.01
		(0.53)	(59.07)	(2.17)	(2.69)	(0.10)	(1.52)	(1.83)	(2.94)	(0.63)
	PLS	-0.03	0.96***	0.07	-0.02	0.00	0.01	0.02	0.06***	-0.03
		(0.70)	(65.92)	(0.58)	(0.67)	(0.22)	(0.57)	(0.67)	(4.24)	(1.41)
	SGL	0.01	0.96***	-0.02	-0.02	-0.01	0.02	0.02	0.02	-0.02
		(0.44)	(63.82)	(-0.10)	(-0.44)	(-0.79)	(0.67)	(1.58)	(1.12)	(-0.99)

Table 11: Monthly Flow Sensitivity to Performance Components.
This table exhibits monthly flow sensitivity to different performance components using CAPM as the benchmark model. PIP^P denotes for the predictor-implied performance based on predictor P, which is one of the six predictors: fund size (Size), return gap (RG), active share (AS), R-squared (R^2), active weight (AW), fund duration (Dur). Control variables are other observables at the end of month t, including lagged monthly flows from $t-17$ to t, one-month lagged log of TNA (size), one-month lagged log of fund age, one-month lagged value of fund's expense ratio, a fund's dummy that indicate whether the fund has any load, and the total volatility of monthly fund net return in prior 12 months (from $t-11$ to t). Standard errors clustered by fund and month are shown in parentheses.

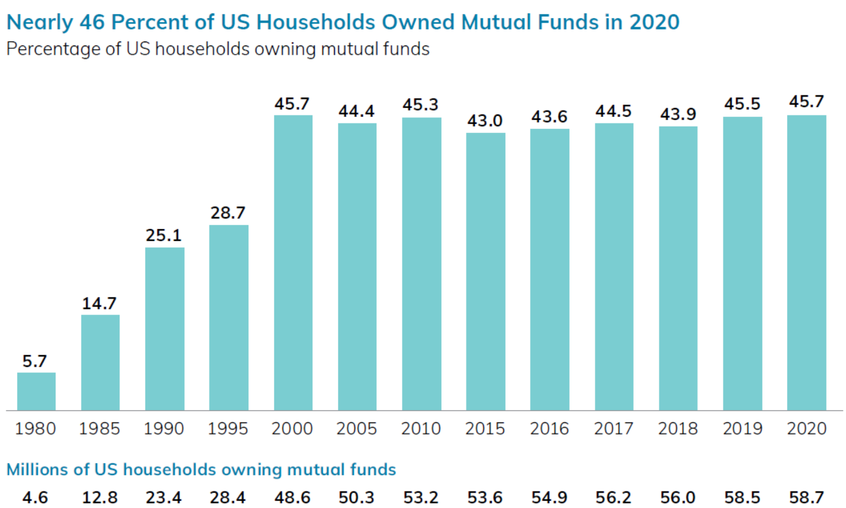
Benchmark Model: CAPM							
	Predictor P						
Monthly Flow	Size	RG	AS	R^2	AW	Dur	No PIP^P
Pure Alpha	0.632***	0.625***	0.647***	0.651***	0.625***	0.633***	0.552***
	(0.043)	(0.044)	(0.043)	(0.043)	(0.044)	(0.044)	(0.040)
PIP^P	0.520***	0.506	0.301	0.153	0.776***	0.466***	
	(0.163)	(0.339)	(0.202)	(0.187)	(0.227)	(0.156)	
Size	-0.166***	-0.166***	-0.176***	-0.172***	-0.165***	-0.168***	-0.163***
	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)	(0.022)	(0.021)
RG		21.591**					
		(9.010)					
AS			-0.692***				
			(0.192)				
R^2				1.347***			
				(0.448)			
AW					0.035		
					(0.112)		
Dur						0.007	
						(0.007)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	152,756	152,756	152,756	152,756	152,756	152,756	157,970
Adj. R^2	0.026	0.026	0.027	0.027	0.026	0.026	0.026

Table 12: Monthly Flow Sensitivity to Performance Components for Three Fund Styles.
This table exhibits monthly flow sensitivity to different performance components using CAPM as the benchmark model for three fund styles. PIP^P denotes for the predictor-implied performance based on predictor P, which is one of the six predictors: fund size (Size), return gap (RG), active share (AS), R-squared (R^2), active weight (AW), fund duration (Dur). Control variables include size, predictor P and other observables at the end of month t, including lagged monthly flows from $t-17$ to t, one-month lagged log of TNA (size), one-month lagged log of fund age, one-month lagged value of fund's expense ratio, a fund's dummy that indicate whether the fund has any load, and the total volatility of monthly fund net return in prior 12 months (from $t-11$ to t). Standard errors clustered by fund and month are shown in parentheses.

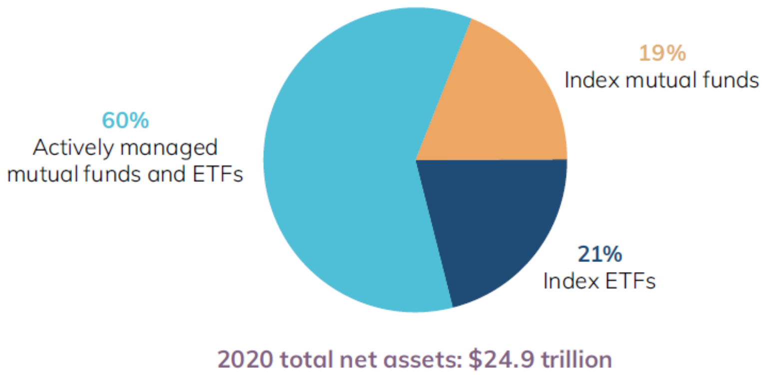
Benchmark Model: CAPM							
Panel A: Aggressive Growth							
	Predictor P						
Monthly Flow	Size	RG	AS	R^2	AW	Dur	No PIP^P
Pure Alpha	0.654***	0.654***	0.656***	0.662***	0.648***	0.655***	0.691***
	(0.065)	(0.063)	(0.065)	(0.065)	(0.064)	(0.065)	(0.065)
PIP^P	0.701***	0.899**	0.807***	0.643***	0.055	0.731***	
	(0.247)	(0.361)	(0.198)	(0.244)	(0.274)	(0.265)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	16,530	16,530	16,530	16,530	16,530	16,530	17,764
Adj. R^2	0.087	0.088	0.087	0.087	0.087	0.087	0.092
Panel B: Growth							
	Predictor P						
Monthly Flow	Size	RG	AS	R^2	AW	Dur	No PIP^P
Pure Alpha	0.750***	0.737***	0.760***	0.759***	0.737***	0.741***	1.003***
	(0.075)	(0.076)	(0.075)	(0.075)	(0.076)	(0.076)	(0.066)
PIP^P	0.355	0.364	0.422*	0.373	0.678***	0.361	
	(0.236)	(0.332)	(0.252)	(0.233)	(0.142)	(0.241)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	80,637	80,637	80,637	80,637	80,637	80,637	83,793
Adj. R^2	0.026	0.026	0.026	0.026	0.026	0.026	0.031
Panel C: Growth and Income							

Table 12: Monthly Flow Sensitivity to Performance Components (Continued)

	Predictor P						
Monthly Flow	Size	RG	AS	R^2	AW	Dur	No PIP^P
Pure Alpha	0.816***	0.806***	0.864***	0.853***	0.835***	0.815***	0.691***
	(0.119)	(0.113)	(0.110)	(0.113)	(0.117)	(0.116)	(0.065)
PIP^P	0.811	0.814	-0.880	-0.028	0.072	0.717	
	(0.518)	(0.594)	(0.814)	(0.501)	(0.601)	(0.457)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	36,859	36,859	36,859	36,859	36,859	36,859	17,764
Adj. R^2	0.014	0.014	0.014	0.014	0.014	0.014	0.092



(a) Percentage of U.S. Households Owning Mutual Funds over Time.



(b) U.S. Total Net Assets Managed by Three Types of Investment Vehicles. Note: Data for ETFs exclude non–1940 Act ETFs and data for mutual funds exclude money market funds.

Figure 1: Households Demand for Mutual Funds. Source: 2021 Investment Company Fact Book.

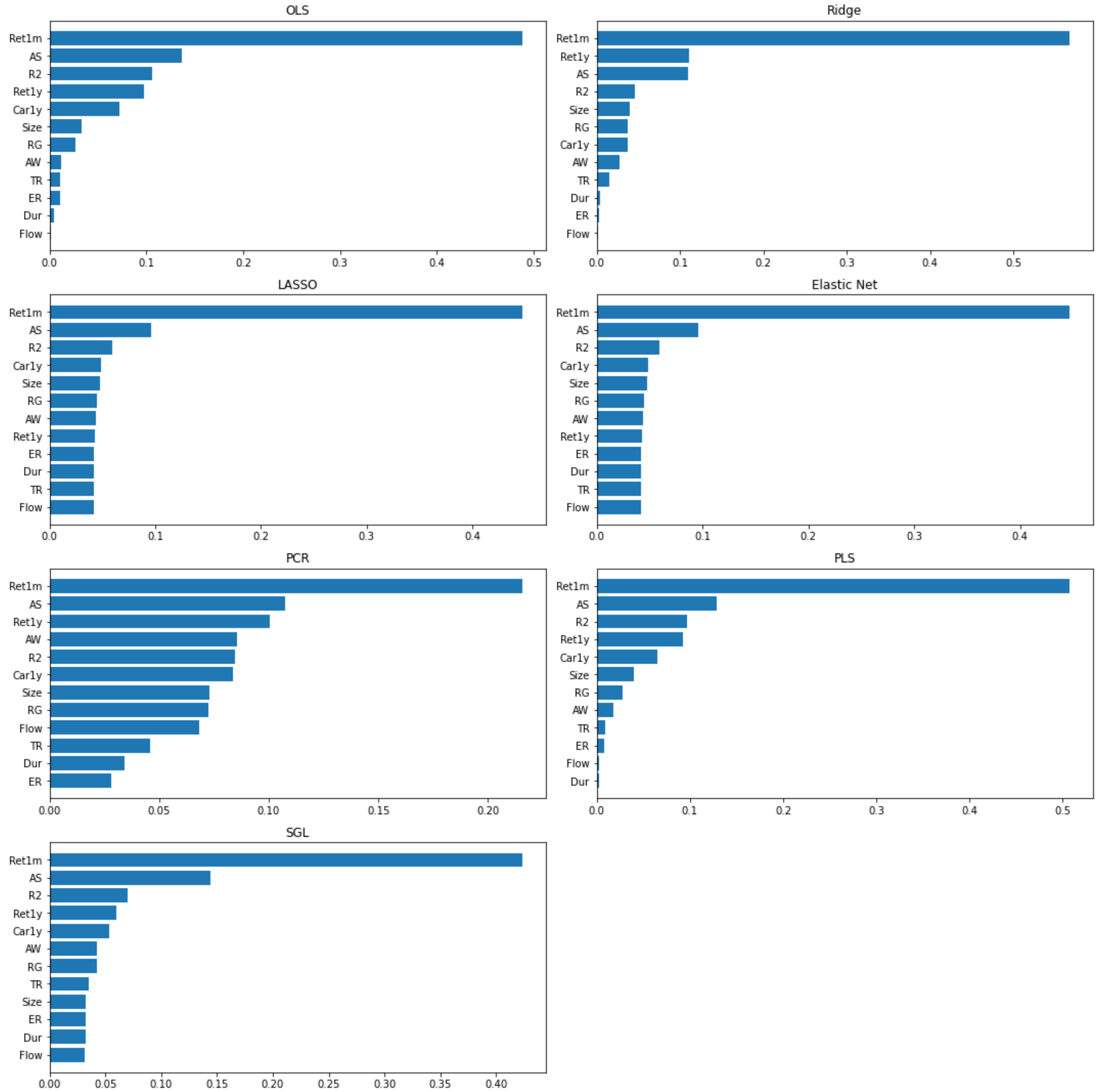


Figure 2: Relative variable importance by model: fund performance predictability, all funds

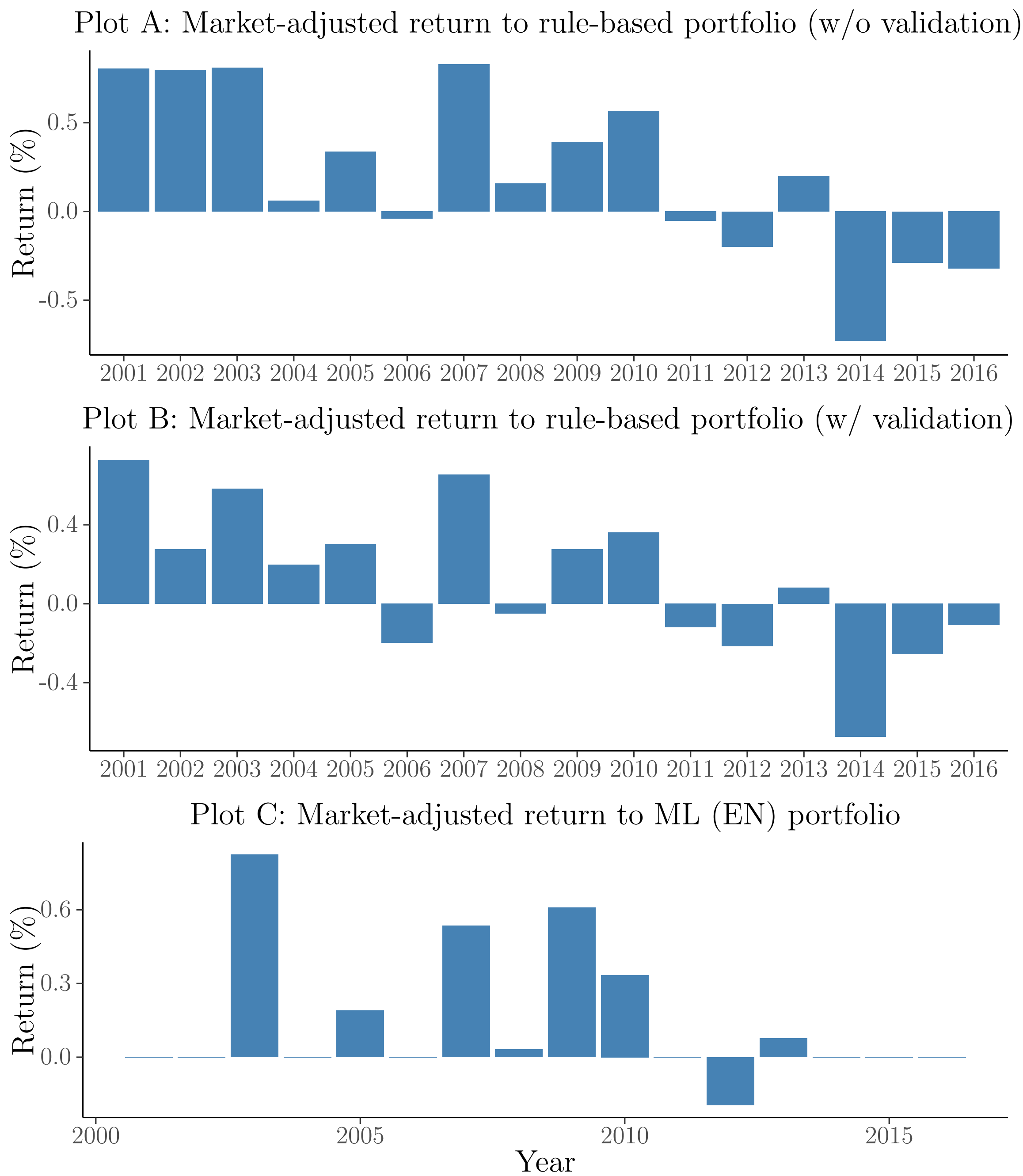


Figure 3: Mean monthly (%) return of real-time fund portfolio using rule-based approach and elastic net. Sample period: 2001-2016.

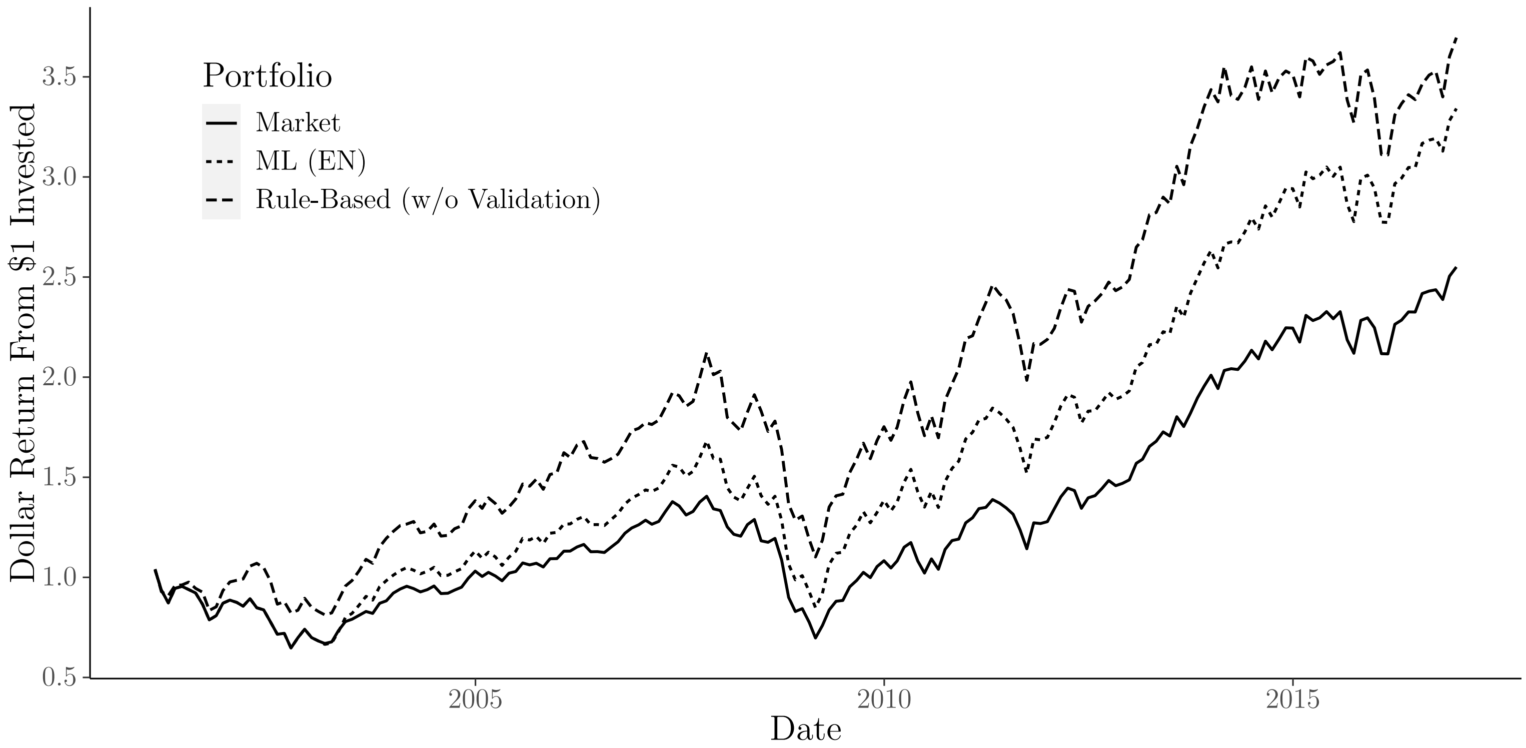


Figure 4: Dollar return from $1 invested using rule-based approach and elastic net. Sample period: 2001-2016.

Appendix A Details on Some Fund Predictors

A.1 Return Gap

Kacperczyk et al. (2006) define the return gap as the difference between net fund return and the net return to fund stock holdings, that is,

$$RG_{f,t} = RF_{f,t} - (RH_{f,t} - EXP_{f,t}), \tag{19}$$

where $RF_{f,t}$ is the net fund return in month t, $RH_{f,t}$ is the total return to a buy-and-hold portfolio based on the most recently disclosed fund stock holdings, and $EXP_{f,t}$ denotes expenses and fees. This measure is constructed at monthly frequency. I take the most recent return gap and lag it for three months if necessary to account for potential reporting delay.

A.2 Active Share

Cremers and Petajisto (2009) propose a measure for active portfolio management, which measures the deviation of a fund portfolio holdings from its benchmark index holdings. Specifically, active share for fund i at time t is defined as

$$AS_{i,t} = \frac{1}{2} \sum_{j=1}^{N_{i,t}} |w_{i,j,t} - w^B_{i,j,t}|, \tag{20}$$

where $w_{i,j,t}$ and $w^B_{i,j,t}$ are the portfolio weights of stock j in fund i and in its benchmark index respectively, and the sum is taken over stock positions only[23]. I obtain the active share data from Martijn Cremers' website `https://activeshare.nd.edu/data`, which originally ranges from 1984 to 2015. Given this paper's focus on real-time predictability, I use active share data computed from self-declared benchmarks instead of from minimum active share benchmarks which require full-sample information. Moreover, I assign to

[23] The investment universe here is defined as the joint union of a fund stock portfolio holding universe and its benchmark portfolio universe.

each fund the most recently available active share while restrict the maximal time lag to be 11 months between current date and the most recent date when active share is available. For instance, I keep fund observation in November 2000 if its most recent active share date is in December 1999, but drop fund observation if its most recent active share date is in November 1999. I take the most recent active share and lag it for three months if necessary to account for potential reporting delay.

A.3 R-squared

I compute R^2 following Amihud and Goyenko (2013) from a regression of mutual fund excess returns on Carhart four factors with based on monthly returns in the prior 24 months up to month t. Funds are required to have valid return in each of the prior 24 months.

A.4 Active Weight

Doshi et al. (2015) proposes an alternative measure for managerial activeness, i.e., active weight, which is defined as

$$AW_{i,t} = \frac{1}{2} \sum_{j=1}^{\widetilde{N}_{i,t}} |w_{i,j,t} - w_{i,j,t}^{M}|, \tag{21}$$

where $w_{i,j,t}^{M}$ is the market-cap weight of stock j within fund i's portfolio at time t, and $\widetilde{N}_{i,t}$ is the total number of stocks held long by the fund. The difference between active weight and active share is that active weight measures how funds allocate money across their long stock positions after determining their long-investment universe, while active share incorporates fund decisions to cover specific stocks. Therefore active weight exclusively captures managerial decisions for deviating from a simple benchmark on the intensive margin. I therefore compute quarterly active weight following Doshi et al. (2015) and require a fund to have at least 10 stocks. For each month, I keep the most recently available active weight. I take the most recent active weight and lag it for three months if necessary to account for potential reporting delay.

A.5 Fund Duration

Cremers and Pareek (2016) construct a fund duration measure to gauge how frequent a fund rebalances its stock holdings. They find that among high active share funds, only those with high fund duration are able to outperform. The fund duration data is available on Martijn Cremers' website `https://activeshare.nd.edu/data`. Since fund duration measures the rebalancing frequency of fund portfolio, it has a highly negative correlation with fund turnover measures. I take the most recent fund duration and lag it for three months if necessary to account for potential reporting delay.

Online Appendix: Data Cleaning

I modify the procedure in Kacperczyk et al. (2006) and Doshi et al. (2015) for cleaning mutual fund data.

Stock Holdings

I use three data files to create a dataset for mutual fund stock holdings: Thomson-Reuters (former CDA/Spectrum, or TFN for abbreviation) s12 type 1 file, type 3 file, mflink2 file in MFLINKS constructed by Wermers (2000) and provided by Wharton Research Data Services (WRDS). The mutual fund stock holdings data are from N-30D, N-30B-2, N-CSR, N-CSRS, N-Q. The cleaning procedure is outlined as follows:

- I exclude funds with CDA/Spectrum investment objective code (IOC) being 1, 5, 6, and 7, corresponding to international, municipal bonds, bond and preferred, and balanced funds. The left funds have investment objective code as aggressive growth, growth, growth & income, metals, unclassified, or missing.
- TFN s12 type 1 file reports two dates: RDATE (reported holding date) and FDATE

(vintage date for matching datasets). They generally do not coincide, and I screen out the first appearing FDATE for each FUNDNO-RDATE pair to avoid stale information. I also create a month-end date variable based on RDATE, which is useful to merge datasets.

- Some funds report more than once in a given month, and I keep only the last report of the month.
- After merging s12 type 1 file with mflink2 file, there are several cases when WFICN-RDATE is not unique (since there may be multiple FUNDNO's for one WFICN due to error or multiple WFICN's for one FUNDNO due to re-usage of fundno by TFN). In those cases, I keep only funds (identified by WFICN after eliminating observations with missing WFICN) with the largest total net assets (identified by the ASSETS variable in s12 type 1 file).
- I then merge the previous resulting file with s12 type 3 file which contains stock holding information.
- I link CUSIP from s12 type 3 file to NCUSIP from CRSP to get the PERMNO identifier.
- The last thing is to adjust back shares held by funds for stock splits and other events.
  - TFN has already adjusted stock splits according to FDATE. For instance, if a fund holds $1,000$ shares in stock A in March (RDATE) while stock A experiences $2:1$ stock splits in June which happens to be the vintage month (FDATE) for holdings reported in March. Then TFN would record $1,000 \times 2 = 2,000$ shares of stock A held by the fund in March (RDATE), based on stock splits in June (FDATE). I therefore need to adjust back the shares so that in March, the number of shares owned is indeed $1,000$.
  - To achieve this, I use CFACSHR from the CRSP MSF file. In the above example, the correct number of shares in March (RDATE) can be recovered as $shares\_rdate = \frac{shares\_fdate * CFACSHR\_fdate}{CFACSHR\_rdate}$.

– I then use the re-adjusted reported shares and ALTPRC from CRSP MSF file to calculate the dollar value of a security held by a fund as $shares\_rdate \times |ALTPRC|$.

– I only keep stock holdings with positive dollar values.

Equity Funds

I create a dataset that contains WFICN-CRSP_FUNDNO-DATE pairs for actively managed U.S. domestic equity funds[24]. I use fund style file to pre-select funds and later combine it with monthly fund information, fund names, and the newly created holdings to filter out index funds and classify funds into different styles at monthly frequency.

- I fill empty style of fund share class with the most recent available style.
- I pre-select fund styles based on style code in CRSP.
    1. I first exclude funds if the CRSP policy code is in 'Bal', 'Bonds', 'B & P', 'C & I', 'GS', 'MM', 'Pfd', 'TFM'.
    2. Then I keep funds if the Lipper classification code is in 'EIEI', 'G', 'LCCE', 'LCGE', 'LCVE', 'MCCE', 'MCGE', 'MCVE', 'MLCE', 'MLGE', 'MLVE', 'SCCE', 'SCGE', 'SCVE' or Lipper prospectus objective code is in 'CA', 'EI', 'G', 'GI', 'MC', 'MR', 'SG'.
    3. If Lipper code is not available, I keep funds if the Strategic Insight objective code is in 'AGG', 'GMC', 'GRI', 'GRO', 'ING', 'SCG', the fund is identified as domestic equity fund.
    4. If Strategic Insight code is not available either, I include funds if the Wiesenberger objective code is in 'G', 'GCI', 'IEQ', 'LTG', 'MCG', 'SCG'.
- I then merge fund styles with fund monthly returns, fund names, holdings (for IOC).
- Before style classification, I use CRSP index flag and fund names to identify index funds. Specifically, I first exclude fund share classes with non-missing CRSP index

[24] DATE is a month-end date variable for CALDT.

flag and then exclude funds if the name contains any of the following characters: 'index', 's&p', 'idx', 'dfa', 'program', 'etf', 'exchange traded', 'exchange-traded', 'target', '2000', '2005', '2010', '2015', '2020', '2025', '2030', '2035', '2040', '2045', '2050', '2055', '2060', '2065', '2070', '2075'.

- Finally, I classify funds into four styles (aggressive growth, growth, equity growth and income, and others) with a created STYLE variable.
  1. If Lipper objective code is 'CA', or Strategic Insight code is 'AGG', or Wiesenberge code is 'MCG', or IOC is 2, I classify the fund as aggressive growth fund.
  2. If Lipper objective code is 'G', or Strategic Insight code is 'GRO', or Wiesenberge code is in 'G' or 'LTG', or IOC is 3, I classify the fund as growth fund.
  3. If Lipper objective code is in 'GI' or 'EI', or Strategic Insight code is in 'GRI' or 'ING', or Wiesenberge code is in 'GCI' or 'IEQ', or IOC is 4, I classify the fund as equity growth and income fund.
  4. Other unclassified funds are denoted as 'Other' in variable STYLE.

More Filters

To be included in a cross-section, I require funds to have at least $15 million TNA in the portfolio formation month. I also adjust the incubation bias documented in Evans (2010) using two filters:

- Eliminate observations preceding the fund's first offer date as reported in CRSP, that is, observations with a missing value in the created AGE variable.
- Eliminate observations with missing fund names.