ESSAYS ON MUTUAL FUND PERFORMANCE AND PREDICTABILITY

Yu Xia

Desautels Faculty of Management

McGill University

Montreal, Quebec

June 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Abstract

This thesis consists of two essays on evaluating mutual fund performance and its predictability. In the first essay, I study the *ex ante* predictability of 12 well-known predictors for fund performance from investors' perspective. The 12 predictors cover three major categories: fund characteristics, fund performance, and holding-based activeness measures, which are constructed using real-time information. For performance evaluation, I exploit two types of fund picking strategies with either rule-based approach or machine learning methods and 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.

Specifically, using variable selection methods such as LASSO and elastic net at individual predictor level can generate annual 1.3%-1.7% real-time alphas after adjusting for standard risk factors. The essay further examines whether real-world investors react to those well-known predictors when evaluating mutual fund performance. Using a novel approach to decomposing fund returns, 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 reaction to predictive information embedded in predictors is stronger within aggressive growth funds. These results provide empirical support for Gârleanu and Pedersen (2018) and suggest *ex ante* predictability exists not due to lack of investor reaction but as the compensation for employing costly algorithms to identify skilled managers.

The second essay examines how decision-making hierarchy in team-managed U.S. equity mutual funds affects their performance and risk-taking behavior. Employing a unique hand-collected dataset, we find that vertically-managed funds with lead managers earn 75 bps per year lower Fama-French five-factor alpha than their horizontally-managed counterparts. Moreover,

vertically-managed funds hold less concentrated portfolios and are exposed to lower residual risk, thus showing signs of inferior security selection ability. Using mutual fund industry as a laboratory, the second essay provides evidence supporting a horizontal decision-making structure in organizations functioning in an uncertain expectation environment. These results echo similar mechanisms as in recent cross-country studies on the benefits of democratic form of government for country's economic growth.

Résumé

La présente thèse porte sur deux essais sur l'évaluation de la performance des fonds communs de placement et de sa prévisibilité. Dans le premier essai, j'étudie la prévisibilité *ex ante* de 12 indicateurs prévisionnels bien connus de la performance des fonds du point de vue des investisseurs. Les 12 indicateurs prévisionnels couvrent trois catégories principales: les caractéristiques du fonds, la performance du fonds et les mesures d'activité basées sur la détention, qui sont construites en utilisant des renseignements en temps réel. En ce qui concerne l'évaluation de la performance, j'exploite deux types de stratégies de sélection de fonds avec une approche basée sur des règles ou des méthodes d'apprentissage automatique et je fait le constat selon lequel l'utilisation de l'apprentissage automatique peut assurer des gains économiques supérieurs en temps réel pour les investisseurs, la performance à court terme des fonds étant le principal facteur sous-jacent à la prévisibilité.

Spécifiquement, l'utilisation de méthodes de sélection de variables, notamment LASSO et le filet élastique au niveau des indicateurs prévisionnels individuels peut générer des alphas annuels en temps réel de 1,3% à 1,7% après ajustement des facteurs de risque standard. L'essai examine ensuite si, en réalité, les investisseurs tiennent compte de ces indicateurs prévisionnels bien connus lorsqu'ils évaluent la performance de fonds communs de placement. En utilisant une nouvelle approche pour décomposer les rendements des fonds, je trouve que, dépendamment de l'utilisation du CAPM par les investisseurs, ceux-ci réagissent aux composantes de l'alpha du CAPM impliquées par les indicateurs prévisionnels de différentes manières, et que la réaction des investisseurs aux renseignements prédictifs intégrés dans les indicateurs prévisionnels est plus forte dans les fonds de croissance agressive. Ces résultats apportent un soutien empirique à Gârleanu et Pedersen (2018) et suggèrent que la prévisibilité *ex ante* existe non pas en raison du

manque de réaction des investisseurs, mais comme la compensation de l'emploi d'algorithmes coûteux pour identifier les gestionnaires compétents.

Le deuxième essai examine comment la hiérarchie de prise de décision dans les fonds communs de placement en actions américains gérés en équipe affecte leur performance et leur comportement de prise de risque. En utilisant un ensemble unique de données collectées manuellement, nous trouvons que les fonds gérés verticalement avec des gestionnaires principaux gagnent 75 points de base par an de moins du modèle d'alpha Fama-French à cinq facteurs que leurs homologues gérés horizontalement. De plus, les fonds gérés verticalement détiennent des portefeuilles moins concentrés et sont exposés à un risque résiduel plus faible, montrant ainsi des signes d'une capacité inférieure de sélection des titres. En utilisant l'industrie des fonds communs de placement en tant que laboratoire, le deuxième essai fournit des preuves à l'appui d'une structure décisionnelle horizontale dans les organisations fonctionnant dans un environnement d'attentes incertaines, reflétant des mécanismes similaires à ceux des récentes études transnationales sur les avantages d'une forme démocratique de gouvernement pour la croissance économique d'un pays.

Acknowledgement

I cannot express more gratitude towards my supervisor, Professor Sergei Sarkissian, for his mentorship throughout my doctoral studies. Without his guidance and support, I would not have come this far and overcome challenges along the road. His endless support, unwavering encouragement, and generous advice have proved to significant ingredients of my journey towards a Ph.D. In many senses, what I learned from Sergei as a researcher and more importantly as a person enlightens me to thrive to become one of the same caliber.

I want to also thank my thesis committee members Professors David Schumacher, Francesca Carrieri, Hongping Tan, Mikhail Simutin, and external advisor, Professor Lawrence Kryzanowski, for providing useful feedbacks on my research. Working with Professor Saurin Patel on the second essay of my thesis is a rewarding experience to see how good research work can be developed and improved. I also appreciate Professor Francesca Carrieri for her generous support during my job market.

The finance area at Desautels has helped me grow tremendously throughout the process. Professors Patrick Augustin and Daniel Andrei have always been offering me great advice with regards to research and career. Professor Thomas Rivera has given me genuine and unsolicited help during my job market. Also, special thanks to Stella Scalia and Yanyan Shu for taking care of the administrative part of my studies.

The journey cannot be described without many friends along the road. I enjoyed conversations with Evan, Byungjin, Ali, Linxiao, Chengyu, Yiliu, Pouya, Yaping, Haohua, Stephen, Changjie, and Zhu who have shared creative thoughts and kept me motivated during my Ph.D. I would like to express my special thanks to Weiyu Jiang and Nan Ma for their cheerful

encouragement and emotional support during difficult times of my Ph.D. which I would always remember. Their company as friends has made this journey brighter and more meaningful.

Finally, I want to thank my parents Qiujiao Wu and Jianian Xia for giving me the freedom and resources to explore who I am and providing endlessly support along the road. Their sacrifice and unwavering faith is something I could hardly pay back.

Contribution of Authors

This thesis is comprised of two essays (Chapters 2 and 3). The first essay is single-authored by Yu Xia. It is completed under the guidance of Yu's supervisor, Professor Sergei Sarkissian. The second essay is co-authored with Professor Saurin Patel and Professor Sergei Sarkissian. The three authors for the second essay share equal responsibilities where Yu is in charge of data collection and analysis, as well as research design and implementation.

The thesis contributes to the literature of mutual fund performance evaluation. Although academic researchers have found abundant evidence that actively-managed equity mutual fund performance is predictable *ex post* using full sample information, little is known whether investors can exploit *ex ante* information for achieving better investment outcomes (compared to passive benchmarks) in real time. The first essay titled, "Real-Time Predictability of Mutual Fund Performance Predictors", discovers superior fund performance after risk adjustments through computationally intensive algorithms and finds that investors react to predictive information for fund selection. These results suggest that real-time predictability can only be exploited with costly search algorithms instead of with traditional OLS method, and provide empirical support for the theoretical argument by Gârleanu and Pedersen (2018) that fund investors need to incur information costs to find skilled managers. The second essay titled, "The Leadership Effect: Evidence from the Fund Industry", uncovers a novel contributor to fund stock selection ability: managerial decision-making structure. Funds with vertical managerial decision-making structure perform significantly worse than funds with horizontal structure. Looking into details, we find that vertically-managed funds hold less concentrated portfolios and have lower residual risk, suggesting inferior security selection ability. Our findings support a positive impact of democratic decision-making structure on institutional performance in an uncertain expectation environment, which helps to resolve the debate on whether vertical (autocratic) or horizontal (democratic) policy making is better for institutional and country development.

Earlier versions of the first essay has been accepted/presented at the 2021 Annual Meeting of the Asian Finance Association, the 2021 Annual Meeting of the Northern Finance Association, the 2021 China Finance Review International Conference, the 2021 New Zealand Finance Meeting, the 2021 World Finance and Banking Symposium, the 2022 Annual Meeting of the American Finance Association, the 2021 Annual Meeting of the Southwestern Finance Association, McGill University, Wilfrid Laurier University, University of Manitoba, and Queens College – CUNY.

Contents

1. Introduction

This thesis evaluates mutual fund performance and predictability from the perspectives of different market participants. The first essay provides a systematic guidance for investors to choose actively-managed mutual funds in real time, using 12 well-known predictors discovered in the mutual fund literature. This is an important topic for personal and household finance, as nowadays about half of Canadian and U.S. households own mutual funds. However, it is unclear which features are the best predictors for future fund performance so that investors can make better real-time decisions. To evaluate investors' economic gains from using performance predictors, I adopt two approaches: regression-based machine learning and rule-based portfolio sorting. What I find is, after risk-adjustments, sparsity methods, which are a subset of machine learning algorithms, can generate alphas of 1.3%-1.7% per year in real time. I also find real-world investors indeed react to predictive information embedded in some of the predictors such as fund asset under management.

Existing studies mainly focus on discovering new predictors to capture managerial skills, which are examined using *ex post* full sample information[1]. Recent studies start to consider *ex ante* or out-of-sample predictability. For instance, Barras et al. (2010) controls for false discovery rate, and Jones and Mo (2021) studies the effect of academic publication on mutual fund performance predictability. This essay compliments and contributes to this literature by conducting real-time test which puts an additional layer beyond out-of-sample tests and examines predictors without knowing whether they will work from investors' perspective. Put it simply, the test is out-of-sample but at the same time incorporates predictor selection.

[1] For instance, Kacperczyk et al. (2006) discovers return gap as a predictor for future fund performance.

In terms of theoretical foundation, there is an ongoing debate with regards to what degree the asset management industry is efficient for investors to incorporate information when they allocate capital across mutual funds. The information can be public or private data, or complex information technology for data processing. While Berk and Green (2004) argues that investors have perfect foresight for discovering skilled managers without any costs and *ex ante* net-of-fee return should be zero, Gârleanu and Pedersen (2018) contend that there should exist costs for investors to acquire information to find skilled managers. In Gârleanu and Pedersen (2018), besides noisy traders who randomly supply underlying asset shares, additional noisy allocators exist to arbitrarily supply capital to either informed or uninformed managers. Noisy allocators are necessary in their model to not fully reveal which manager is skilled based on readily available public information such as fund asset under management. The first essay of my thesis provides empirical evidence for the second view, by showing that with computationally intensive algorithms, investors can detect skilled funds in real time even using publicly available information. And the magnitude of outperformance can be seen as a proxy for the searching cost an average investor needs to incur to find skilled funds in the asset management industry.

In order to make progress on this topic, I exploit two types of approaches to evaluating 12 well-known fund performance predictors for fund selection in real time. The 12 predictors can be categorized into 3 groups: fund characteristics, fund performance, and holding-based activeness measures. All these predictors are constructed using information publicly available to investors. The first approach is regression-based machine learning (e.g., LASSO, ridge, etc.) and the second approach is rule-based portfolio sorting. For each of these two approaches, I associate the best performing funds in the past with a particular predictor or a combination of predictors, and then choose funds based on selected predictors for the next period real-time investment.

First, I find that using sparsity methods with variable selection feature such as LASSO and elastic net can deliver superior real-time investment benefits for investors choosing mutual funds. The economic magnitude of outperformance after adjusting for common stock risk factors ranges from 1.3% to 1.7% per year, depending on specific risk adjustments. The rule-based portfolio sorting approach can generate a better outperformance of about 2.5% per year after adjusting the market factor. However, this outperformance diminishes after controlling for additional stock risk factors such as size and momentum. Second, it turns out short-term one-month fund return is the most important feature among the three categories of predictors I examine here. Other predictors commonly used in the literature have relatively small additional real-time predictive power compared to short-term one-month return. And this predictability can be only discovered with machine learning methods with variable selection features. Further inspecting the time variations of predictability, I find that elastic net and LASSO generate outperformance by selecting predictors for fund selection only when the overall predictability is strong. When the overall predictability is weak, elastic net and LASSO do not select any predictors and investors by default switch to the passive market portfolio. Lastly, the essay examines how investors incorporate predictive information embedded in predictors for choosing mutual funds. To answer this question, I develop a novel approach to decomposing fund performance into three components: a component due to fund exposure to common stock risk factors, a predictor-implied component (PIP) which captures how similar a fund performs relative to a benchmark portfolio of funds, and a residual component which is not captured by risk factors or PIP. Using mutual fund flow as a direct measure for investor reaction, I identify investors' flow reaction to the predictor-implied component for different predictors. The reason for employing this novel approach is that although investors may use predictors for fund selection, it is not clear whether predictors are used for performance related causes or performance irrelevant behavioral reasons. For instance, investors may mistakenly treat a high-fee fund as one that deliver low net-of-fee return even if the fee is justified as compensation for the superior skills provided by managers[2]. Through my approach, I find that investors in more actively managed funds such as aggressive growth funds are more inclined to use predictive information for fund selection than more income-oriented funds. Moreover, investors are found to consistently use fund asset under management (AUM) for performance evaluation across different asset pricing models.

These findings suggest that the real-time predictability for fund performance exists despite investors' awareness of fund performance predictors for performance evaluation. Henceforth the outperformance discovered by either machine learning or rule-based approach can be seen as the compensation for using intensive search algorithms to find informed or skilled managers.

The second essay is a joint work with my supervisor Professor Sergei Sarkissian from McGill University and Professor Saurin Patel from University of Western Ontario. We examine the driver of mutual fund performance from the perspective of fund management. Using a unique hand-collected data for fund decision-making structure from SEC EDGAR system, we find that among team-managed domestic active equity funds in U.S., vertically-managed funds with clear leaders that possess final authority over investment decisions compared to other portfolio managers perform significantly worse than horizontally-managed funds where each portfolio manager shares more equal responsibility.

Over the past three decades, team-managed funds have become very popular in the U.S. mutual fund industry. For instance, nowadays, more than 70% of U.S. domestic active equity mutual funds are managed by a team of portfolio managers. However, although existing studies

[2] See Sheng et al. (2022) which shows that net alphas are unrelated to fees after proper risk adjustments and confronts the common suggestion that investors should prefer low-fee funds over high-fee funds.

show that team-managed funds perform better than single-managed funds[3], little is known about how investment decision is made among managers. Do team-managed funds take in account all managers' investment opinions equally? Or do they behave more like single-manager funds with lead managers having the final say? The second essay aims to answer these questions.

Moreover, the second essay adds to the debate in the organizational behavior literature where there is no consensus on whether democratic or autocratic organizational structure is better for country and institutional development. The jury is still out there on what is the best decision-making structure. Existing studies approach this issue mainly by looking at country-level data which is limited in terms of sample size and characteristics that can be controlled for. A major contribution of the second essay is to approach this issue with granular mutual fund data which has the most information across all occupational databases, so that we can exactly pinpoint the effect of decision-making structure on institutional performance after controlling for various confounding factors.

Specifically, we find that among team-managed funds, horizontally-managed funds outperform vertically-managed funds by 50-75bps per year, after adjusting for common risk factors. Further inspecting the driver of outperformance, we find that horizontally-managed funds take on more residual risks and hold more concentrated portfolios, suggesting that a more democratic decision-making structure contributes to more informative investment ideas and better security selection ability. Furthermore, the performance gap between funds with different team structures exhibits a non-monotonic relation with team sizes. That is, the performance gap between horizontally-managed funds and vertically-managed funds is greater for funds with four managers than those with fewer or more managers.

[3] See Patel and Sarkissian (2017).

Our findings are consistent with the hypothesis that there exists a trade-off between team coordination costs and managerial incentives to engage in decision-making. Within a small group of two people, the lead portfolio manager may neglect a more optimal decision of only one group member, while for larger teams of three or four people, such policy could result in more severe suboptimal decisions. However, as the team size grows further, increasing coordination costs associated with larger teams start to play a more significant role, with less negative impact of vertical team management on performance for funds with larger team sizes, since such larger teams could be better coordinated and motivated.

2. Real-Time Predictability of Mutual Fund Performance Predictors

2.1. 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 merely 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[4]. 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 general investors' perspective in real time.

[4] 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. (2022) for instance.

Studies in existing literature mainly focus on discovering new predictors without accounting for the joint predictive power of existing predictors. This paper attempts to fill the gap by utilizing two types of strategies to study investors' gains in using available 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 though still require demanding computing power to implement[5]. 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[6]. 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[7].

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

[5] Studies such as Lo et al. (2000) have considered using technical rules for predicting stock returns.
[6] Giglio et al. (2021) provides a comprehensive survey on using machine learning in asset pricing.
[7] 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.

weight, and fund duration) classified into three categories[8]: 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, an investor may employ any of the 12 predictors individually or a combination of them with 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.

[8] 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.

I examine 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[9]. The results indicate that one version of the rule-based real-time portfolio can outperform 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 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 generate outperformance relative to the market. These results suggest that robo-advisors using machine learning algorithms with variable selection feature can add value to fund picking by general investors.

Moreover, my paper further examines whether real-world investors react 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 reaction to predictive information

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

embedded in predictors is stronger among aggressive growth funds where those predictors are usually 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 reaction 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 by 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 of 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). More recently, Barber et al. (2016) and Berk and van Binsbergen (2016) argue that investors are most likely to use the Capital Asset Pricing Model (CAPM) for fund performance risk adjustment. My paper further examines how investors react to the predictive component of abnormal return that is not explained by standard asset pricing models. Moreover, 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 (2021) 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[10] 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 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

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

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 is 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 reaction 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 2.2 introduces the rule-based approach and machine learning methods used for predicting future fund performance in this paper. Section 2.3 describes the mutual fund data and the sample selection criteria. Section 2.4 illustrates the in-sample predictive power of each of the 12 predictors. Section 2.5 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 2.6 explores investors' flow response to the predictive information embedded in predictors. Section 2.7 concludes.

2.2. 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[11]. 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

[11] 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.

and the forecasting target (i.e., fund performance). In the following subsections, I describe each type of methods and their respective advantages in predicting fund performance.

2.2.1. 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[12] 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 × 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[13]. 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

[12] Expanding window provides additional years for training models compared to rolling window.
[13] Recall a rule is either a single predictor quintile or a combination of quintiles of two predictors.

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 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[14].

2.2.2. 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:

$$r_{i,t+1} = E_t\left[r_{i,t+1}\right] + \epsilon_{i,t+1}, \tag{2.1}$$

[14] 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.

where

$$E_t[r_{i,t+1}] = g^*(x_{i,t};\theta). \tag{2.2}$$

$r_{i,t+1}$ is the net-of-fee return investors would realize by investing in fund i during month t + 1, which can be decomposed into an expected performance component plus noise. My objective is to model the unknown expected component $E_t[r_{i,t+1}]$ 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 $E_t[r_{i,t+1}]$.

2.2.2.1. 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:

$$L(\theta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{i=1}^{T}(r_{i,t+1} - g^*(x_{i,t};\theta))^2. \tag{2.3}$$

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:

$$L(\theta;\cdot) = L(\theta) + \phi(\theta;\cdot), \tag{2.4}$$

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

$$\phi(\theta;\lambda,\rho) = \lambda(1-\rho)\sum_{m=1}^{M}|\theta_m| + \frac{1}{2}\lambda\rho\sum_{m=1}^{M}\theta_m^2. \quad (2.5)$$

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 particularly 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, performance, 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

$$\phi(\theta;\alpha) = \lambda\sum_{j=1}^{J}[(1-\alpha)\left|\left|\theta_j\right|\right|_2 + \left|\left|\theta_j\right|\right|_1], \quad (2.6)$$

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

2.2.2.2. 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[15].

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

$$R = X\theta + \epsilon, \tag{2.7}$$

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.

$$R = (X\Omega_K)\theta_K + \bar{E}. \tag{2.8}$$

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 j-th predictive components, and θ_K is a $K \times 1$ vector.

[15] Kelly and Pruitt (2013) uses PLS to estimate overall equity market risk premia.

PCR chooses the combination weights Ω_K recursively such that the j-th linear combination solves

$$w_j = argmax_w Var(Xw), \quad s.t.\ w'w = 1, Cov(Xw, Xw_l) = 0, l = 1,2, \dots, j-1. \tag{2.9}$$

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 j-th PLS component is found by solving

$$w_j = argmax_w Cov(R, Xw), \quad s.t. w'w = 1, Cov(Xw, Xw_l) = 0, l = 1,2, \dots, j-1. \tag{2.10}$$

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

2.3. Data and Sample Selection

The mutual fund sample ranges from 1994 to 2016[16]. 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

[16] Specifically, the predictor sample is from December 1994 to November 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.

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 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. Appendix B provides further details 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[17].

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 a fund to have all available information of the 12 predictors to be included in 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

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

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[18]. I consider these 12 predictors since their construction involves only publicly available information that is commonly used by investors. 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 in my sample earn a slightly negative net-fee one-year Carhart alpha as found in previous studies (Carhart, 1997, Fama and French, 2010). It is worth mentioning that all predictors except one-month return and return gap are highly persistent for a given fund, suggesting that they act as skill measures 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 qualitatively consistent with existing findings in previous studies.

[18] 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.

2.4. 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:

$$R_{P,t}^{H} - R_{P,t}^{L} = \alpha_P + \beta_P\left(R_{M,t} - R_{f,t}\right) + s_P R_{SMB,t} + h_P R_{HML,t} + m_P R_{MOM,t} + \epsilon_{P,t}, \tag{2.11}$$

where $R_{P,t}^{H} - R_{P,t}^{L}$ 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-weighting schemes active share does not predict future fund performance by itself. Moreover, fund duration now becomes a significant predictor for performance[19]. 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.

2.5. 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 relatively little on whether the best performing funds selected by predictors can outperform a passive benchmark portfolio (e.g., 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 2.2.

2.5.1. Rule-Based Portfolio Sorting Approach

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 first in-sample period is 1995-2000 and the last in-sample period is 1995-2015, with expanding window for each year forward.

[19] 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.

The corresponding out-of-sample (OOS) 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 evaluation 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 for each OOS year based on previous 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 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 salient observation 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 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:

$$R_t - R_{f,t} = \alpha + (\beta + B' z_{t-1})\left(R_{M,t} - R_{f,t}\right) + sR_{SMB,t} + hR_{HML,t} + mR_{MOM,t} + \epsilon_t, \qquad (2.12)$$

where R_t is return for the OOS portfolios. The one-month lagged macroeconomic variables z_{t-1}[20] 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 explain much the performance of the OOS portfolio in either case, and the

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

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.

2.5.2. Regression-Based Machine Learning Methods

In this subsection, I implement six regression-based machine learning methods described in Section 2.2. 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.

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).

2.5.3. 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 main 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 the real-time elastic net or LASSO 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.

2.6. 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 real-world 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[21].

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

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

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

$$F_{i,t+1} = \frac{TNA_{i,t+1}}{TNA_{i,t}} - \left(1 + R_{i,t+1}\right) \tag{2.13}$$

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 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' reaction to predictive information is to include standard predictors in a panel regression to test whether coefficients 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 extend 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[22]:

$$R_{i,\tau} - R_{f,\tau} = \alpha_{i,t} + \gamma_{i,t}^{P} R_{\tau}^{P} + \sum_j \beta_{i,j,t} f_{j,\tau} + \epsilon_{i,\tau} \tag{2.14}$$

[22] 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.

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 in month τ. 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 captured by the predictor. For instance, a fund with positive loading γ on the return spread means that the fund behaves as if it has a similar skill as large funds[23]. The purpose of this step is to estimate month t fund loading ($\hat{\gamma}_{i,t}^P$) to the factor-mimicking fund portfolio R_t^P and factor loadings ($\hat{\beta}_{i,j,t}{}'s$).

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):

$$R_{i,t} - R_{f,t} = \hat{\alpha}_{i,t} + \hat{\gamma}_{i,t}^P R_t^P + \sum_j \hat{\beta}_{i,j,t} f_{j,t}. \tag{2.15}$$

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

[23] 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.

$$F_{i,t+1} = a + b\sum_{s=0}^{17} e^{-\lambda s} MAR_{i,t-s} + c' X_{i,t} + \eta_{t+1} + \epsilon_{i,t+1}, \tag{2.16}$$

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[24], a fund 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). 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 (2.16) and construct an index for each return component. Specifically,

$$Alpha_{i,t} = \frac{\sum_{s=0}^{17} e^{-\hat{\lambda} s}\, \hat{\alpha}_{i,t-s}}{\sum_{s=0}^{17} e^{-\hat{\lambda} s}},$$

$$PIP_{i,t}^{P} = \frac{\sum_{s=0}^{17} e^{-\hat{\lambda} s} \hat{\gamma}_{i,t-s}^{P} R_{t-s}^{P}}{\sum_{s=0}^{17} e^{-\hat{\lambda} s}}, \tag{2.17}$$

$$FACTOR_{i,j,t} = \frac{\sum_{s=0}^{17} e^{-\hat{\lambda} s} \hat{\beta}_{i,j,t-s} f_{j,t-s}}{\sum_{s=0}^{17} e^{-\hat{\lambda} s}},$$

where $FACTOR_{i,j,t}$ varies depending on which model to use as the testing field[25].

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

[24] Expense ratio is reported at annual frequency.
[25] For the main text, I only include results for CAPM.

$$F_{i,t+1} = b_0 + b_\alpha Alpha_{i,t} + b_P PIP_{i,t}^P + \sum_j b_j\, FACTOR_{i,j,t} + \theta' X_{i,t} + \eta_{t+1} + \epsilon_{i,t+1}, \tag{2.18}$$

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 (2.16). 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[26]. This novel specification helps to isolate predictive information from characteristic preference by investors that are not motivated by performance predictability[27]. For a given factor model, I consider the magnitude 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[28]. 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 (2.15) and (2.18) 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%

[26] Essentially all predictors are fund characteristics.
[27] 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-based and characteristics-based asset pricing tests.
[28] 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.

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 are statistically significant, the economic magnitude for size-implied performance 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 reaction to predictor-implied performance information is stronger among funds where they usually work well.

2.7. 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 reaction 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 reaction 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 investigations of investors' time-varying reaction to predictors and investors' sophistication in using predictive information would be interesting venues for future work.

2.8. Tables

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 aggregate sales or purchases of securities divided by the average 12-month fund TNA
AS	Deviation of a fund portfolio holdings from its benchmark index holdings
R^2	R^2 from a regression of fund net excess return on Carhart model using 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	1376.5	85.68	1.29
2003	900	1711.11	78.10	1.27
2004	987	1778.94	74.87	1.26
2005	1035	1765.4	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
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	-0.020	-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

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	(0.50)
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	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	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 regression is 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)	(1.2)	(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	FFC4 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	-0.08	-0.09
	(1.6)	(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. OLS regression is the benchmark. 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
	(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)	(6.7)	(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)	(0.5)	(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)	(0.1)	(11.34)	(0.48)	(2.42)
Growth and Income	No	0.06	0.88***	-0.57**	-0.02	0	0	0.06**	0.13***	0.04**
		(0.86)	(38.26)	(2.17)	(0.31)	(0.1)	(0.01)	(2.2)	(2.96)	(2.12)
	Yes	-0.04	0.94***	-0.26*	-0.02	-0.02	0	0.05***	0.11***	0.05***
		(0.95)	(63.7)	(1.68)	(0.35)	(1.13)	(0.05)	(2.7)	(3.06)	(2.7)

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)	(0.9)	(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)	(6.5)	(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
		(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)	(1.07)	(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)	(80.2)	(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)	(0.1)	(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.

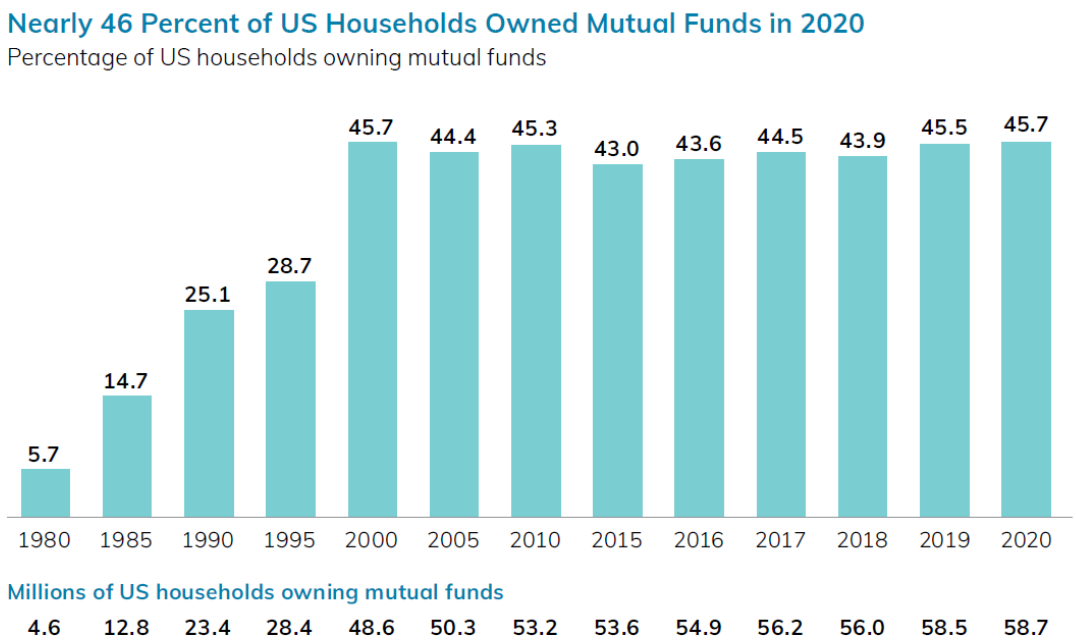
Benchmark Model: CAPM							
	Predictor P						
Monthly Flow	Size	RG	AS	R^2	AW	Dur	No PIP^P
Pure Alpha	0.632***	0.625***	0.647***	0.651***	0.625***	0.633***	0.552***
	(0.043)	(0.044)	(0.043)	(0.043)	(0.044)	(0.044)	(0.04)
PIP^P	0.520***	0.506	0.301	0.153	0.776***	0.466***	
	(0.163)	(0.339)	(0.202)	(0.187)	(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**					
		(9.01)					
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

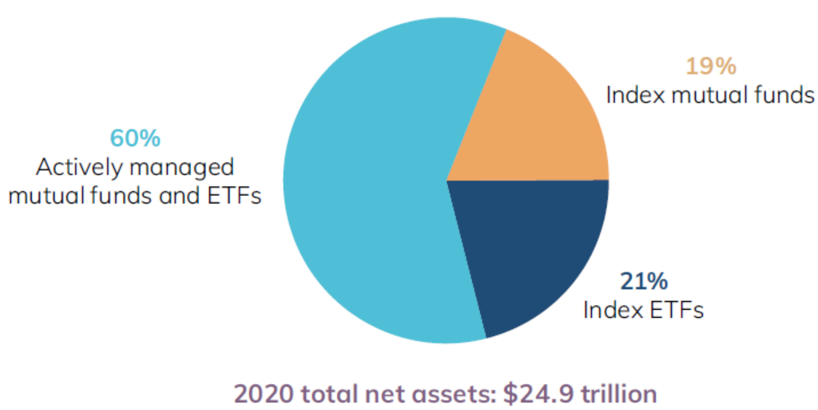
This table exhibits monthly flow sensitivity to different performance components using CAPM as the benchmark model for three fund styles. PIP^P denotes for the predictor-implied performance based on predictor P, which is one of the six predictors: fund size (Size), return gap (RG), active share (AS), 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*** (0.065)	0.654*** (0.063)	0.656*** (0.065)	0.662*** (0.065)	0.648*** (0.064)	0.655*** (0.065)	0.691*** (0.065)
PIP^P	0.701*** (0.247)	0.899** (0.361)	0.807*** (0.198)	0.643*** (0.244)	0.055 (0.274)	0.731*** (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*** (0.075)	0.737*** (0.076)	0.760*** (0.075)	0.759*** (0.075)	0.737*** (0.076)	0.741*** (0.076)	1.003*** (0.066)
PIP^P	0.355 (0.236)	0.364 (0.332)	0.422* (0.252)	0.373 (0.233)	0.678*** (0.142)	0.361 (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							
	Predictor P						
Monthly Flow	Size	RG	AS	R^2	AW	Dur	No PIP^P
Pure Alpha	0.816*** (0.119)	0.806*** (0.113)	0.864*** (0.11)	0.853*** (0.113)	0.835*** (0.117)	0.815*** (0.116)	0.691*** (0.065)
PIP^P	0.811 (0.518)	0.814 (0.594)	-0.88 (0.814)	-0.028 (0.501)	0.072 (0.601)	0.717 (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

2.9. Figures



(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.

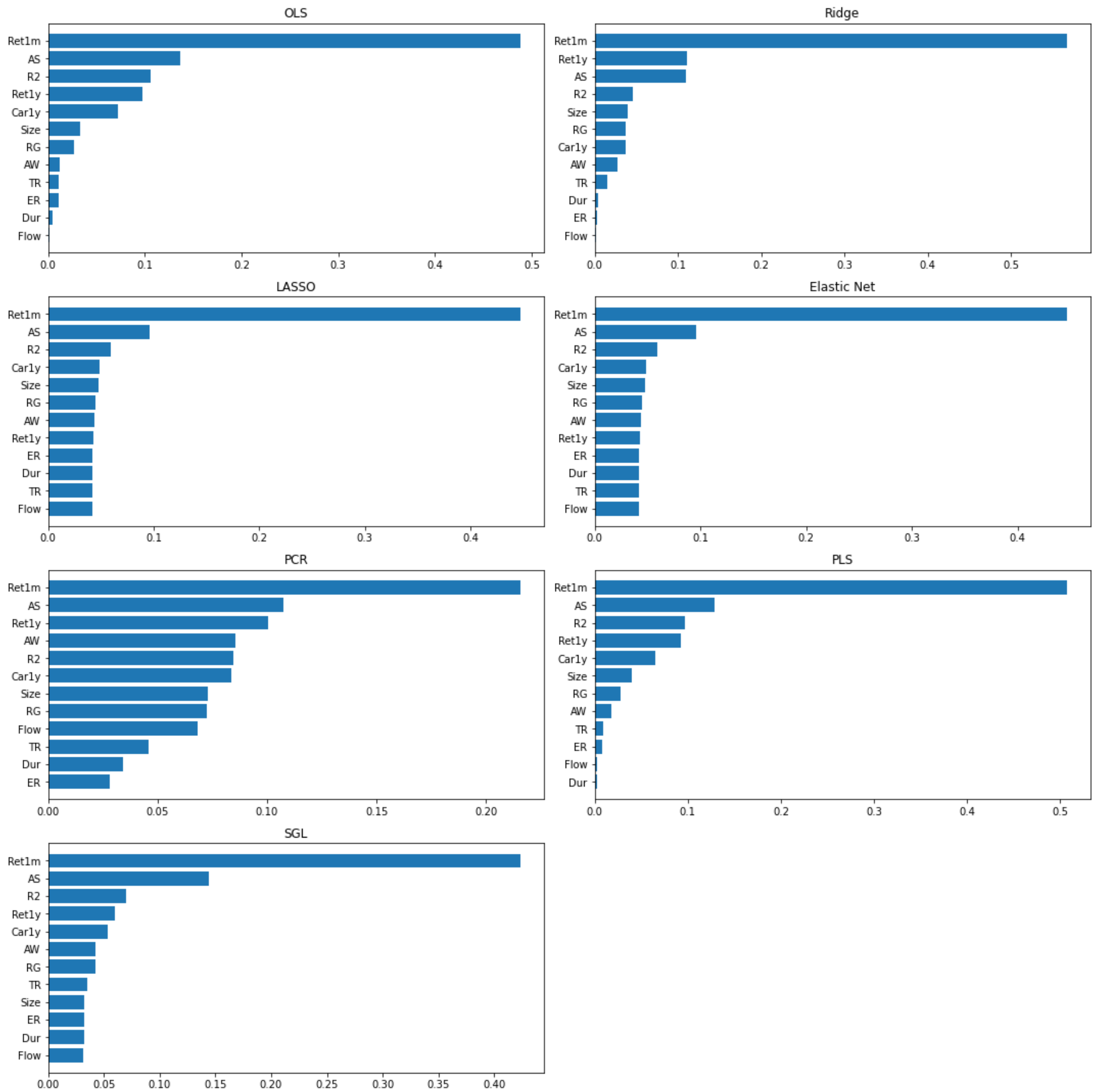


Figure 2: Relative variable importance by model: fund performance predictability

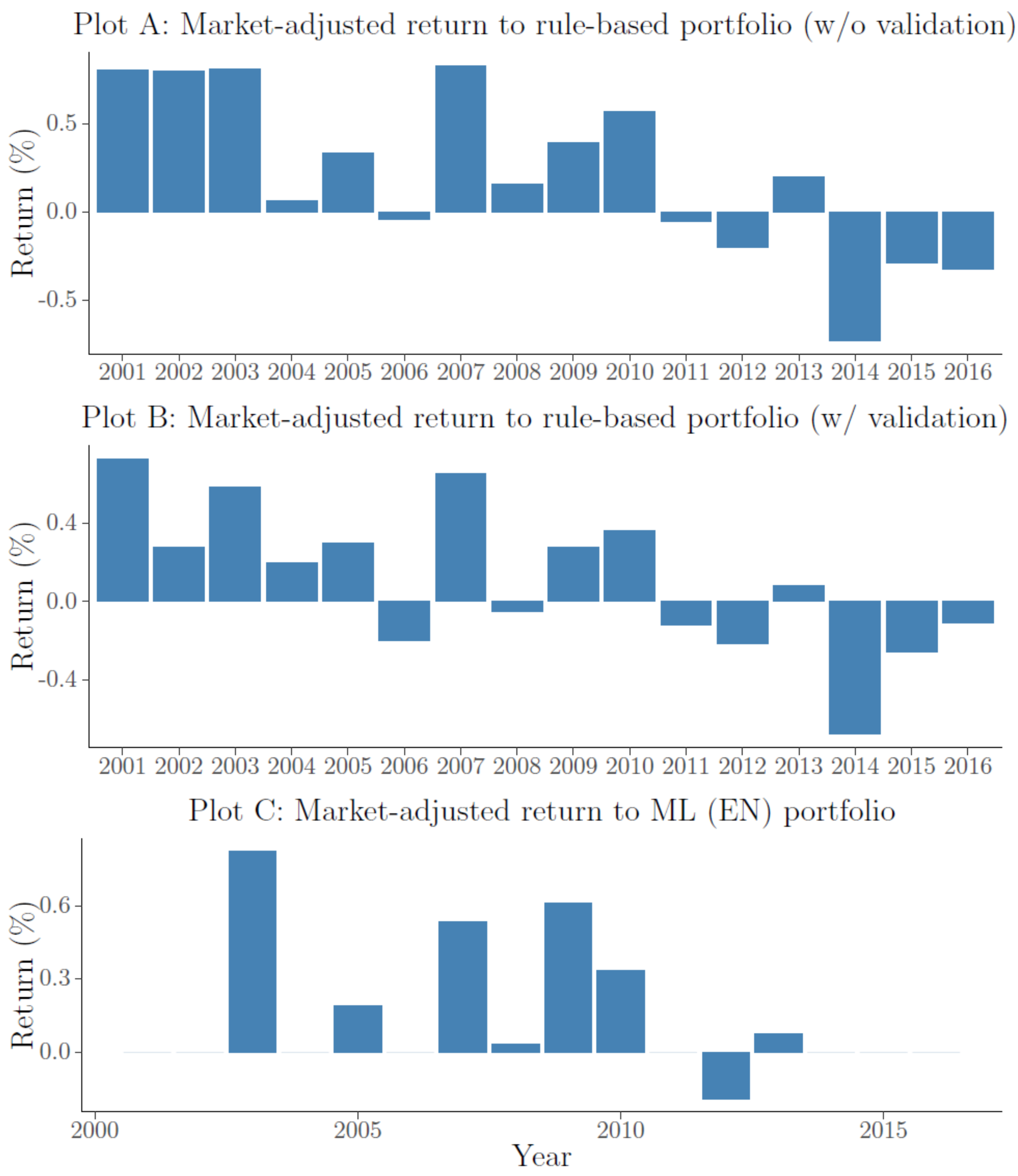


Figure 3: Mean monthly (%) return of real-time fund portfolio using rule-based approach and elastic net
Sample period: 2001-2016.

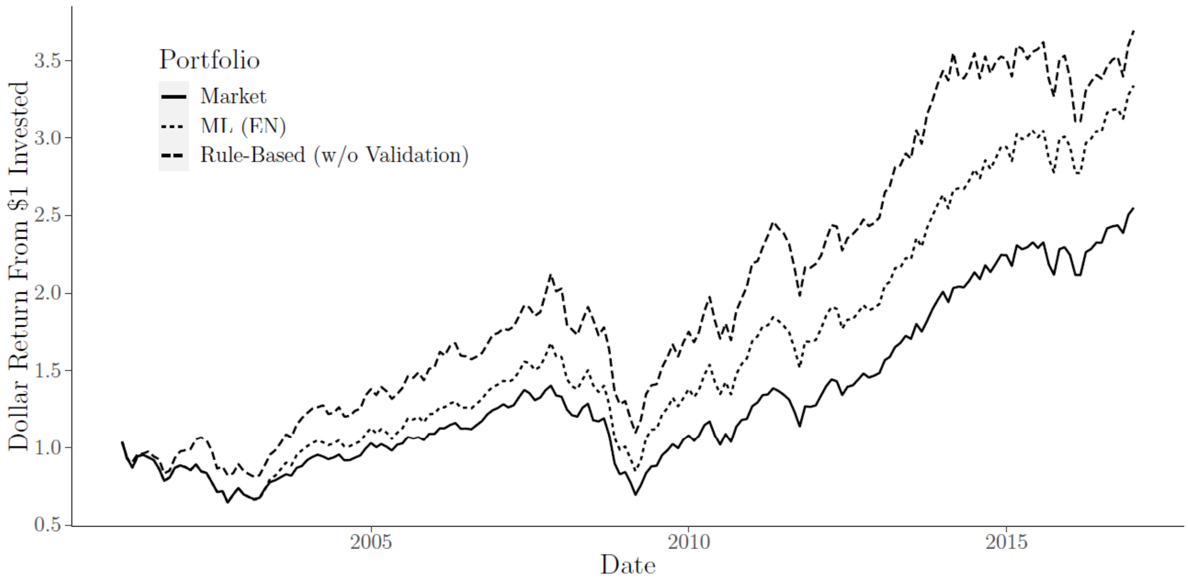


Figure 4: Dollar return from $1 invested using rule-based approach and elastic net
Sample period: 2001-2016.

2.10. Appendix

2.10.1. Details on Some Fund Predictors

2.10.1.1. Return Gap

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

$$RG_{f,t} = RF_{f,t} - (RH_{f,t} - EXP_{f,t}) \tag{2.19}$$

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.

2.10.1.2. Active Share

Cremers and Petajisto (2009) proposes 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

$$AS_{i,t} = \frac{1}{2}\sum_{j=1}^{N_{i,t}} |w_{i,j,t} - w_{i,j,t}^{B}|, \tag{2.20}$$

where 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[29]. 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

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

benchmarks instead of from minimum active share benchmarks which require full-sample information. Moreover, I assign to 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.

2.10.1.3. R-squared

I compute R-squared 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.

2.10.1.4. Active Weight

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

$$AW_{i,t} = \frac{1}{2}\sum_{j=1}^{\widetilde{N}_{i,t}} |w_{i,j,t} - w_{i,j,t}^{M}|, \tag{2.21}$$

where $w_{i,j,t}^{M}$ is the market-cap weight of stock j within fund i's portfolio at time t, and is the total number of stocks held long by fund i. 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.

2.10.1.5. Fund Duration

Cremers and Pareek (2016) constructs 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.

2.10.2. Data Cleaning

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

2.10.2.1. 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 × 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 $shares\_rdate = \frac{shares\_fdate \times CFACSHR\_fdate}{CFACSHR\_rdate}$.
    - 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.

2.10.2.2. Equity Funds

I create a dataset that contains WFICN-CRSP FUNDNO-DATE pairs for actively managed U.S. domestic equity funds[30]. 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.
    - I first exclude funds if the CRSP policy code is in 'Bal', 'Bonds', 'B & P', 'C & I', 'GS', 'MM', 'Pfd', 'TFM'.
    - 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'.
    - 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.
    - 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 flag and then exclude funds if the name contains any of the following characters: 'index', 's&p', 'idx',

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

‘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.
    - If Lipper objective code is ‘CA’, or Strategic Insight code is ‘AGG’, or Wiesenberger code is ‘MCG’, or IOC is 2, I classify the fund as aggressive growth fund.
    - If Lipper objective code is ‘G’, or Strategic Insight code is ‘GRO’, or Wiesenberger code is in ‘G’ or ‘LTG’, or IOC is 3, I classify the fund as growth fund.
    - If Lipper objective code is in ‘GI’ or ‘EI’, or Strategic Insight code is in ‘GRI’ or ‘ING’, or Wiesenberger code is in ‘GCI’ or ‘IEQ’, or IOC is 4, I classify the fund as equity growth and income fund.
    - Other unclassified funds are denoted as ‘Other’ in variable STYLE.

2.10.2.3. 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.

3. The Leadership Effect: Evidence from the Fund Industry

3.1. Introduction

The question on the form of government, that is, the decision-making structure that brings more political and economic benefits dates back to Plato and Aristotle's debate two millennia ago. Since then and even nowadays, there is no clear conclusion whether democratic governance spurs more development than its autocratic counterpart. On the one side, many studies argue that democracy with its largely *horizontal policy-making* structures that cooperate yet balance each other leads to economic growth and/or better overall socio-economic conditions (e.g., Wittman, 1989; Persson and Tabellini, 2006; Doucouliagos and Ulubasoglu, 2008; Papaioannou and Siourounis, 2008; Acemoglu et al., 2019). On the other side, an equally impressive list of papers argues the opposite, i.e., that an autocratic system with *vertical policy-making* that involves fewer people in the decision process is superior in many instances (e.g., Barro, 1996, Tavares and Wacziarg, 2001; Glaeser et al. 2004; Pozuelo et al., 2016). One of the main problems with all these previous studies is that they could only deal with a relatively small data sample – very limited set of countries with democratic and autocratic systems suitable for the analysis. The other serious problem is the difficulty in proper accounting for various cross-country differences. Thus, the jury is still out on the best form of the decision-making structure.

We approach this ongoing debate from a completely different angle. We examine how decision-making hierarchy (horizontal without lead managers vs. vertical with lead managers) in team-managed U.S. equity mutual funds affects their performance and risk taking. The U.S. mutual fund data is perfectly and uniquely suited for our goal, since it has the largest cross-sectional and time-series sample of all occupational databases, while the differences across funds can be easily and precisely controlled for. The analysis of the impact of specific decision-making structures at the country level and for the fund industry has two vital commonalities. First, similar to cross-country studies, which can clearly define the outcomes of different policy-making hierarchies in terms of objective economic indicators, such as GDP growth, the main result of different approaches to investment decisions in mutual funds – fund returns – can also be unambiguously measured. Second, the decisions of both policy makers for country's strong economic development and portfolio managers for achieving better fund performance do not rely on precisely defined and unambiguous information.

The importance of team-management in the fund industry, especially mutual funds, has received a significant attention in the past two decades because of both a significant increase in the proportion of funds with such managerial structure and the finding by Patel and Sarkissian (2017) of the outperformance of team-based funds.[31] Yet, the question on the importance of specific team hierarchy in decision making on fund returns and risk-taking behavior remains unanswered largely due to the difficulties in collecting and identifying information on lead managers in a given fund.[32]

In this paper, we manually collect fund managerial information from fund prospectuses (485BPOS filings) through SEC EDGAR system from 2012 to 2016. We first classify each fund manager as a lead manager or non-lead manager at the manager level. When all managers within the same fund share a title of "lead manager" or "co-lead manager", no manager in the fund is classified as a lead manager. When the manager cannot be clearly classified based on the title, we

[31] Patel and Sarkissian (2017) use Morningstar Direct data and show that it captures the managerial structure of funds much more correctly than CRSP or Morningstar Principia datasets used in the earlier studies. Such studies include Prather and Middleton (2002), Chen et al. (2004), Bliss, Porter, and Schwarz (2008), Massa, Reuter, and Zitzewitz (2010), Bar, Kempf, and Ruenzi (2011), and others. All of them find that team-managed funds provide no more or lower gains than single-managed funds.

[32] Other recent studies on team-based funds include among others Tan and Sen (2019), Chen, Xie, and Zhou (2020), Evans et al. (2020), Lu, Naik, and Teo (2021), all of whom examine team diversity and fund performance, Evans et al. (2021), who look at the performance of anonymous funds, Patel and Sarkissian (2021), who analyse the impact of team-management on illegal trading activities.

collect manager's biographical details from the "fund management" section. We then classify a fund with vertical team structure if a lead manager is identified for the fund with our approach. The details of the procedure are in Section 3.2 and the Appendix. Overall, we identify 2,866 funds with vertical team management. This constitutes 38.52% of the total number of funds in our sample. In this sample, 81% of funds (89% of observations) have one management team, while the remaining 19% of funds (11 % of observations) have multiple teams of advisors and subadvisors. Our tests account for these nuances.

We observe that the proportion of lead-manager funds increases with team size. To further understand the determinants of funds with team-leaders, we apply a probit model for identifying the drivers of vertical team funds as well as changes in the team structure. We use various fund and manager characteristics as our predictive variables. Besides the importance of the team size for the likelihood of vertical teams, these estimations reveal that decreasing fund flows increase the probability of team management structure changes in both directions. This is intuitive, since it implies that substantial drops in such important characteristic of mutual funds as their net flows forces the fund and/or fund family administrators to change the fund's leadership structure. Furthermore, low past returns increase the probability of a vertical team fund to be transformed into a horizontal one. This suggests that fund family administrators may associate low fund returns with a specific team leader and, therefore, be eager to change the fund's managerial structure into a horizontal relationship as a response to the fund's poor performance.

We begin our main empirical tests in a univariate setting by showing the existence of significant performance differences between funds with different team structure. We find that vertical team funds underperform horizontal ones based on risk-adjusted returns by 51-75 bps per year, depending on one of the four performance evaluation models: Fama and French (1993) three-factor, Carhart (1997) four-factor, Pastor and Stambaugh (2003) five-factor, or Fama and French (2015) five-factor. The subsequent multivariate panel regression tests that control for fund and manager characteristics reveal the same picture: funds with vertical team structure significantly underperform those with horizontal team structure. The underperformance of vertical team funds in these tests surpasses 75 bps per year when using an alpha computed from the Fama and French (2015) five-factor model.

We observe that negative value of lead team managers is the smallest in magnitude for two-member teams followed by large teams of five members and more. The vertical team structure of funds with three or four team members posts the lowest risk-adjusted returns. This pattern is present in both univariate and multivariate settings and potentially can be explained as follows. Within a small group of two people, the lead portfolio manager may neglect a more optimal decision of only one group member, while for larger teams of three or four people, such policy could result in more severe suboptimal decisions. However, as the team size grows further, increasing coordination costs associated with larger teams start playing a significant role as well.[33] Hence less negative impact of vertical team management for large team sizes, since such large teams could be better coordinated and motivated. We further document the underperformance of funds with vertical team structure across all investment objective categories with the largest values recorded for aggressive growth funds (almost 100 bps per year) and the smallest for the growth & income and equity income funds (30 bps per year) based on risk-adjusted returns from the Carhart (1997) model.

In addition, we find that vertical team funds hold less concentrated portfolios, which is consistent with Kacperczyk, Sialm, and Zheng (2005), who report that funds holding more

[33] For instance, Mueller (2012) shows that groups experience significant coordination costs and diminishing motivation when they are composed of four or more individuals.

concentrated portfolios perform better. Finally, we investigate the risk-taking behaviour of horizontal and vertical team funds using four factors from the Carhart (1997) model as well as comparing them in terms of total and residual risk. Our tests reveal that funds with lead managers load marginally more on the market risk than horizontal team funds. More significantly, vertical team funds have less residual risk, making this result consistent with these funds holding more diversified portfolios.

Thus, we view our paper as contributing not only to the extensive list of studies on mutual fund performance and group-decision making, but also to the ongoing debate on the benefits of specific forms of government structure (democratic versus autocratic) for economic and social well-being. Using an example of one specific industry, which has a large cross-section and time-series of data and clearly defined preferred outcomes, we are able to show that horizontal, i.e., more democratically and collaboratively managed teams are associated with unambiguously higher gains. It is also worth noting that our results, while questioning the role of dedicated leadership in the fund industry, where decisions are made in a vague and difficult to evaluate environment, are not novel in the literature. For example, as Thompson (1967) shows, deviant (inappropriate) discretion can be harmful in certain situations. Waldman et al. (2001) find many CEOs in industries with scarce opportunities whose actions negatively impact company performance. At any point in time, due to a noisy information environment the opportunities and resources for fund portfolio managers can be considered relatively limited for making good investment decisions.

The rest of the paper is organized as follows. Section 3.2 describes the mutual fund data, including the identification process of lead-manager funds, provides the summary statistics and examines the drivers of vertical team management. Section 3.3 presents the main performance difference tests between horizontal and vertical team structures in mutual funds based on various risk-adjusted measures. Section 3.4 examines the impact on portfolio holdings and risk-taking behavior of funds with lead managers. Section 3.5 concludes.

3.2. Data

3.2.1. Lead manager(*s*) (vertical team) identification

We use Morningstar Direct database covering team-managed U.S. domestic equity funds with the following four investment objective categories: aggressive growth, growth, growth & income, and equity income. All sector and index funds are excluded. When only one fund manager is named at the end of the calendar year, we classify that fund as sole-managed for that year and exclude it from the sample as well. Since all data is reported at the fund share–class level, and that includes the fund manager names, we aggregate mutual fund share–class level observations to one fund-level observation using a unique fund identifier.

Fund managerial information is manually collected from fund prospectuses (485BPOS filings) through SEC EDGAR system from 2012 to 2016 only for those funds that remained team-managed during the whole five-year time period. Funds from EDGAR are then manually matched with funds from Morningstar via Series ID and Ticker Symbol. If ticker is not available from EDGAR, fund name is used for matching. For a given year in Morningstar, we first check that year's earliest prospectus to see if the managers from Morningstar exactly match with the managers from that prospectus. If the data is matched, we use the managerial information from that prospectus for that year. Otherwise, we further check the next recent prospectus (either from the same year or the following year) to see if the managers from Morningstar exactly match with the managers from that prospectus. Then if the data is matched, we use the information from that prospectus, otherwise we combine information from both prospectuses, where information for managers that appears in both prospectuses and Morningstar comes from the earliest prospectus. For managers that appear in only one prospectus and Morningstar, we use managerial information from that prospectus. This applies to both advisor and sub-advisor levels for each fund. For certain funds, some managers may not appear in that year's prospectus, and we search for their names from all the prospectuses available for that fund for managerial information and use the managerial information from the most recent prospectus.

By default, we treat all managers as non-lead managers. We then further classify a fund manager as a lead manager or not based on keywords from the managerial descriptions in fund prospectus. However, before we use any keyword for manager classification, we impose the following restrictions on classification:

1. For a fund managed by a single advisor or subadvisor (i.e., single-team fund), when all managers within the same fund can be classified as a lead manager based on keywords, none of them are classified as a lead manager.
2. For funds with multiple advisors and subadvisors which make decisions independently from each other (i.e., multiple management teams), the previous restriction applies at the team level. Moreover, some of their advisors or subadvisors may have only one single manager: in such scenarios those single managers are not classified as a lead manager, regardless of their keyword descriptions. [34]

[34] The following keywords are used to indicate hierarchy (vertical structure) in fund management from fund prospectus: lead manager, co-lead manager, (team) leader, led by; assistant (portfolio) manager, assist/assisted, support/supported; final authority, ultimate responsibility, ultimate decision-making authority, ultimate veto, ultimate decision-making authority, final decision-making responsibility, final decision maker, jointly and primarily responsible for day-to-day management for the fund. Keywords used to indicate non-hierarchy (horizontal structure) in fund management include: share equal responsibility, (play) equal roles, etc. Those keywords are not comprehensive in terms of classifying fund managers but when we later used our own judgement for classification, the same spirit applies.

After classifying managers for each fund, we create a fund-level dummy variable indicating whether the fund is a vertical team fund in terms of its organizational structure that is the main variable of interest in our paper. A fund is classified as a vertical team fund if it has at least one manager that is classified as a lead manager. Due to a substantial time involved in data collection, we use a recursive data imputation procedure to collect fund managerial information with 2015 as the starting year. The details are in the Appendix.

3.2.2. Other fund and manager data

For each fund we also obtain from Morningstar Direct its standard characteristics to use as control variables in our tests. These characteristics are fund size, measured by the total net assets under management of the fund at the end of calendar year; fund age, defined as the difference between the fund's inception year and the current year; expenses, measured by the annual net expense ratio of the fund; fund family size, measured by the total net assets under management of the fund complex to which the fund belongs at the end of calendar year; fund return volatility, measured by standard deviation of gross monthly fund returns over the past 12 months within the calendar year; turnover, defined as the minimum of aggregated sales or purchases of securities of the year divided by the average 12-month total net assets of the fund over the past year; net fund flows, defined as the net growth in the total net assets of funds as a percentage of their total net assets adjusted for prior year returns. We winsorize expenses, turnover, and fund flows at the 1% and 99% levels to minimize the effect of outliers.

We follow Chevalier and Ellison (1999) and use an additional managerial variable – manager tenure with a fund. We define the manager tenure as the difference between the year when a fund manager started as a portfolio manager for a given fund and the current year. As in

Patel and Sarkissian (2017), the manager tenure within a team is the equally-weighted average of manager tenures of all fund managers in the team. Finally, we add a gender (female) indicator variable, which equals one when at least one fund manager in a team is a female and zero otherwise.

3.2.3. Summary statistics

In Figure 1, Plot A shows the proportion of funds with lead managers for team-managed funds with different number of managers within a fund. The largest team size in our sample consists of 36 members. The general pattern is positive as larger teams are more likely to have a clearly identified lead manager(s): while the percent of teams with leaders in a two-member team is about 30%, that percent for teams with more than ten members reaches almost 80%. Note that an increase in the proportion of funds with lead managers becomes particularly profound for funds with more than five team members.

In Table 1, Panel A provides more details on the distribution (numbers and percentages) of lead-manager funds across years for all funds and across funds with different team sizes (with two, three, four, and five or more managers). In total we were able to identify 2,866 funds with lead team managers. This represents around 38.52% of all team-managed funds in our sample. This percentage decreases slightly but monotonically from 41.22% in 2012 to 36.96% in 2016. This pattern, which shows a more than 4% drop in lead-team-member funds in five years, allows us to cautiously suggest that over time the fund industry is moving towards more collaborative and “democratic” decision-making. We also note that the proportion of lead-manager funds increases with the number of managers in all five years in our sample.

Often a fund is managed by multiple advisors and subadvisors, who make investing decisions independently from each other. In this case, some teams may have identifiable lead managers, while others do not. Plot B of Figure 1 depicts the proportion of funds with lead managers for different number of management teams within a fund with the corresponding percentages of data observations. The maximum number of individually-managed teams is 14. We can see that the vast majority of data in our sample (88%) belongs to funds with only one management team. The other 12% of observations is spread across funds with more than one team (up to 14). Expectedly, the proportion of funds with lead managers increases with the number of teams within the fund.

Panel B of Table 1 shows the distribution of vertical team funds managed by one versus multiple management teams for each year in our sample period. The vast majority of funds with lead managers have only one management team (2,329 out of 2,866 or about 81%). There are only 537 lead-manager funds with more than one team and their numbers drop monotonically from 134 in 2012 to 92 in 2016. Expectedly, the percent of vertical team funds is higher among funds managed by several teams in every year of our sample, since the increasing number of teams increases the probability that at least of them will have a lead manager. This mimics the overall pattern depicted in Figure 1, Plot B.

Table 2 provides the summary statistics of fund and manager characteristics for horizontal and vertical team structure funds and highlights their statistical differences. We observe that horizontal team funds are on average larger and older than vertical peers. The fund family size, between the two groups of funds is, however, almost similar to each other. Furthermore, horizontal team funds are much less volatile and much cheaper to investors than vertical ones, in spite of no differences in their turnover levels. Finally, the average tenure of managers within horizontal teams is longer than those within vertical teams.

3.2.4. Drivers of teams with lead manager(s)

In this section, we examine which fund and managerial characteristics may increase the likelihood of adoption of a vertical team management structure in mutual funds. We achieve this by using a probit model, where the dependent variable is a dummy $VTeam_{i,t}$, which equals 1 if the management team of fund *i* at time *t* is classified as vertical and equals 0 if classified as horizontal. The independent variables include fund and manager characteristics from Table 2, and all these characteristics are lagged. Our model, therefore, is:

$$Pr(VTeam_{it} = 1) = \delta_0 + \delta_1 NM_{it} + \delta_2' Fund\_Char_{it-1} + \delta_3' Mgr\_Char_{it-1} + FE_{it} + \epsilon_{it}, \quad (3.1)$$

where $Pr(VTeam_{it} = 1)$ is the probability that the team-management structure of fund *i* at time *t* is vertical, while $Fund\_Char$ and Mgr_Char are fund- and manager-specific characteristics, respectively, from Table 2 and the lagged by one year fund performance metric – the risk-adjusted return from the Carhart (1997) four-factor model, α_C4. This and other risk-adjusted performance measures that we use in our study are based on the 12-month return window with the yearly calendar basis. In addition, we account for the number of managers (NM) within the team, but we include this variable without lagging, since decisions on lead managers and team size could be taken concurrently at the fund or fund family levels. The fixed effects, FE, include the year times fund investment objectives.

Besides Model (1), we also consider its modification to examine the drivers for the changes in the leadership structure from horizontal to vertical and vice versa. In this case, we deal with the following specification:

$$Pr(\Delta Team\_Str_{it} = 1) = \delta_0 + \delta_1 NM_{it} + \delta_2' Fund\_Char_{it-1} + \delta_3' Mgr\_Char_{it-1} + FE_{it} + \epsilon_{it} \quad , \tag{3.2}$$

where $Pr(\Delta Team\_Str_{it} = 1)$ is the probability that the team-management structure of fund *i* at time *t* changes from horizontal to vertical or vertical to horizontal, one at a time. All independent variables are the same as in Model (1). In our sample, we have 96 instances of changes from horizontal to vertical team structure and 160 cases of the reverse direction.

Table 3 shows the estimation results of both probit models. It reports the coefficient estimates and their corresponding absolute *t*-statistics based on standard errors clustered by fund and year. Column 1 gives the outcome from estimating Model (1). Consistent with Figure 1 and Table 1 patterns, we find a strong positive relation between the team size and the probability of a fund having lead managers within the team. In addition, directly corroborating with Table 2 results, younger funds and funds with higher expenses are more likely to have vertical team structure. Finally, we observe that higher past performance lowers the probability of having a lead manager within the team. The regressions in Columns 2 and 3 provide a more informative picture. Now we see that the coefficient on number of managers is positive and significant only for changes from horizontal to vertical team funds. We also observe that a drop in fund flows significantly increases the likelihood of team structure change in both directions: horizontal to vertical and vice versa. This result is quite intuitive as it suggests that any drastic change in the team management is driven by substantial changes in one of the most important characteristics of mutual funds – their net flows. Even more interestingly to us is the differentiated importance of other predictor variables. In particular, we observe that low fund performance leads to a higher probability of abandoning a vertical team leadership structure in favour of horizontal. This implies that lead managers, who are likely to be personally associated with poor fund returns, are stripped from their leadership roles.

Once again, higher expenses seem to be linked to vertical team funds. Then, our results also indicate that as funds become older, they are more likely to adopt horizontal team structure. Yet, we find that a switch to horizontal teams has a higher probability as the average tenure of managers within the fund remains relatively low. Finally, there is some association between an increasing number of female managers within the fund and horizontal team structure preference.

3.3. Main Empirical Results

3.3.1. Univariate tests

We first examine performance differences between funds with horizontal and vertical team structure in the univariate settings. Besides computing the fund alpha from the Carhart's four-factor model, α_C4, as alternative measures of fund risk-adjusted returns we also consider fund alphas from the French (1993) three-factor model, the French (2015) five-factor model, and the Pastor and Stambaugh (2003) five-factor model, which adds a liquidity factor to the Carhart's model, α_FF3, α_FF5, and α_PS5, respectively.[35]

Table 4 shows our four fund alphas in the following year for team funds that are identified to have horizontal or vertical structure in the current year. It also provides the difference in the mean test of their performance metrics for the whole sample of teams and individually for team sizes of two, three, four, and five or more fund managers. We can see that funds with vertical team structure significantly underperform economically and statistically those with horizontal team structure, irrespective of the type of risk-adjusted return. In particular, in annual terms the underperformance ranges from 51 bps (12×0.0429 percent) for α_PS5 to 75 bps (12×0.0625

[35] All these metrics are computed in gross terms, i.e., without fund subtracting expenses from their gross returns.

percent) for α_FF5. The difference tests for alphas across funds with different team sizes reveal that the overall pattern is present for each team size: for any number of managers within the team, the horizontal team structure outperforms the vertical one. However, this evidence is the weakest among two-manager funds. In addition, the magnitude of the performance difference is also somewhat smaller for funds with very large teams of five or more people. The largest discrepancy in all risk-adjusted returns between vertical and horizontal team structures is observed for funds with four-member managerial teams.

The observed a U-shaped pattern in return differences, which is depicted on Figure 2 using α_C4 estimates in annual terms alongside with the 95% confidence bounds, can be explained based on the benefits and costs of lead ("autocratic") manager in small and large teams. In very small teams, where coordination costs are low, a lead manager may neglect more optimal decisions of the only one other group member. In larger teams of three to four people, a vertical hierarchy within portfolio managers could lead to more frequent and severe suboptimal investment decisions. Yet, as the team size grows even further, the coordination costs quickly increase, thus reducing the benefits of collective decision-making (e.g., see Mueller, 2012). In this case, the team leaders could play a more prominent and positive role in making timely portfolio allocation and trading decisions.

3.3.2. Multivariate tests

We now move to evaluating the differences in performance between vertical and horizontal team-based funds in a multivariate setting. The general model that we estimate is as follows:

$$Alpha_{it} = g_0 + g_1 VTeam_{it-1} + g_2' Fund\_Char_{it-1} + g_3' Mgr\_Char_{it-1} + FE_{it} + \epsilon_{it}, \qquad (3.3)$$

where $Alpha$ is one of our four risk-adjusted return measures. The independent variable of interest is the vertical team dummy, $VTeam$, which equals 1 if the fund has a vertical team management structure and equals 0 if it has a horizontal structure. Other control variables come from Table 2. Unlike Table 3, now both fund and manager characteristics are lagged by one year. As before, the fixed effects include the year times fund investment objectives.

Table 5 shows the impact of the team leadership structure on future fund performance. It reports the coefficient estimates, their absolute t-statistics based on standard errors clustered by fund and year, as well as the number of observations and the adjusted R–squared for each regression. In Panel A we use the $VTeam$ dummy following exactly Model (3) regression. In Columns 1, 3, 5, and 7 we estimate Model (3) without controls but with fixed effects. Across all estimations, the coefficient on $VTeam$ is negative and significant at 1% or 5% levels. After the inclusion of control variables in Columns 2, 4, 6, and 8, both the magnitude and statistical significance of $VTeam$ do not change materially. In economic terms the underperformance of funds with vertical teams ranges between about 58 bps based on α_PS5 and almost 76 bps per year based on α_FF5. These numbers are even higher than the corresponding estimates in univariate tests in Table 4. Most of control variables, including fund size, are insignificant, which may look surprising at first, considering the importance of team size in many other studies (e.g., Chen et al., 2004; Patel and Sarkissian, 2017; etc.). However, this result should not be surprising, since we deal only with team-based funds and relatively short sample period. The only consistently significant control is fund expense, which is negative and significant at the 1% level in all regressions. This shows that from the managerial structure perspective, even within a more homogeneous group of team-mutual funds expensive funds underperform, and these funds have lead managers much more often (Table 2) than comparable less expensive funds.

We note that our vertical team dummy, $VTeam$, may not be viewed as a fully accurate measure of the leadership structure if a fund has several teams and only one of them has a lead manager. To reflect this reality, in Panel B of Table 5 we repeat our estimation of Model (3) but instead of the $VTeam$ dummy we use a team-weighted vertical team measure, $WVTeam$, which is the fraction of management teams that have lead managers. The set of control variables is the same as in Panel A, but their estimates are not shown. We can see that the introduction of an alternative measure for the fund leadership structure effectively leads to the same estimation outcome for all four fund alphas – negative association between vertical teams and subsequent performance. Thus, Table 5 shows that on average mutual funds with lead managers within teams severely underperform other team-managed funds without such lead portfolio managers, even after accounting for various fund and manager characteristics.[36]

To understand the impact of multiple teams on the documented negative relation between vertical teams and fund performance, in Table 6 using Model (3) we evaluate the effect of leadership structure on fund returns for different numbers of management teams within the fund. We split our funds into two subsamples: one contains funds with only one managerial team, while the second has two or more teams. As mentioned earlier, the first subsample absorbs 89% of all data observations. Our dependent variable in these tests is the Carhart (1997) four-factor alpha. We estimate regressions using the $WVTeam$ measure, which for funds with only one managerial team is equivalent to $VTeam$ dummy. We detect a significant underperformance of vertical teams for one-team funds only. Note that the economic magnitude of this underperformance, 6.13 bps per month is larger than comparable estimates of α_C4 in Table 5. The coefficient of $WVTeam$ is very small not only statistically, which could be due to much smaller sample size, but also economically for funds with two or more teams. Thus, the impact of team leadership is more

[36] In Table A1 in the Online Appendix we repeat the estimation of Model (3) using both $VTeam$ and $WVTeam$ measures but make all independent variables contemporaneous with the fund alphas. We obtain test results qualitatively similar to those in Table 5.

pronounced in one-team funds and exactly these types of funds show poor returns relative to their peers managed by teams without lead managers. [37]

Out next step is to examine how team leadership affects fund performance for funds with different number of managers within the team. That is, we want to see if the U-shaped pattern of underperformance of vertical team funds documented in Table 3 remains after accounting for controls. Table 7 reports the effect of leadership structure on fund performance across different team sizes. The tests are based on Model (3) and all estimation specifications are similar to those in Table 5. As the dependent variable we use only the Carhart (1997) four-factor alpha. Columns, 1, 3, 5, and 7 report the regression results only with fixed effects, while Columns 2, 4, 6, and 8 are based on Model (3) in its full extent. Similar to results in Table 3, we observe that vertical team funds underperform horizontal across team sizes: the coefficient on $VTeam$ is negative in all estimations. As in the univariate setting, vertical teams underperform horizontal ones the least for the smallest team size of two managers (44 bps per year) and the most for four-member teams (113 bps per year). The underperformance of funds with lead managers is highly significant for all team sizes but those with two people, consistent with our earlier observation. Therefore, we again are able to document a U-shaped underperformance pattern of vertical team structure funds.

Finally, in this section it is imperative to understand whether our documented underperformance of lead-manager funds is present across funds with different investment objectives or is a phenomenon of a particular fund investment strategy. Table 8 reports the effect of leadership structure on fund performance across three investment objective categories: aggressive growth, growth, and income. Due to small sample sizes and similar income producing investing agenda, we combine growth & income and equity income funds into one category –

[37] Given the similarity of our test results in Panels A and B, thereafter in our tests we use only the vertical team dummy, $VTeam$. We show our other tests with a team-weighted vertical team measure, $WVTeam$, in the Online Appendix.

income funds. We again use the Carhart (1997) four-factor alpha as the sole dependent variable and our estimations are based on Model (3). Columns, 1, 3, and 5 report the test results without controls and Columns 2, 4, and 6 – with controls. We find an underperformance of vertical team funds in all estimations. In economic terms, the coefficient on $VTeam$ ranges from 31 bps per year for income funds to almost 100 bps per year for aggressive growth funds when regressions include control variables. The statistical significance is lower for aggressive growth funds and absent for income funds, but much smaller sample sizes of these fund categories relative to growth funds, which have the most significant $VTeam$ dummy, largely explain these discrepancies.[38]

3.4. Additional Tests

3.4.1. Portfolio concentration

In the previous section, using a variety of tests, we could show a consistent underperformance of team-managed funds with lead managers. Prior research shows that poorly performing funds also hold less concentrated portfolios than funds with superior performance (Kacperczyk, Sialm, and Zheng, 2005). Therefore, we should expect that vertical team funds are more diversified than those with the horizontal team structure.

We verify our expectations in Table 9. It shows the effect of leadership structure on fund industry portfolio concentration, which is constructed following Kacperczyk, Sialm, and Zheng (2005). Panel A reports the mean industry concentration differences between full samples of horizontal and vertical team funds and that for the different team sizes. We also provide the absolute *t*-statistics of these differences. We can see that indeed vertical team funds are less

[38] In Tables A2 and A3 in the Online Appendix we replicate the estimation of Tables 7 and 8 using $WVTeam$. The results resemble those in Tables 7 and 8.

concentrated than horizontal ones across all teams and for every team size individually. This difference for the whole sample of both groups is of highly statistical significance. To exclude the possibility that this difference in portfolio concentration is fully or partially due to other differences in fund and managerial characteristics, in Panel B we show the results of multivariate regressions of industry concentration on the vertical team dummy, $VTeam$, and all controls from Table 5. As before, all regressors, which are not shown, are lagged by one year, include time by investment objective fixed effects, and standard errors are clustered by fund and year. Similar to Panel A, we show the estimation outcomes for the whole sample and separately for each team size. We arrive to the same finding: the coefficient on $VTeam$ is negative in all regressions and is significant at the 1% level for the whole sample estimation. Moreover, even with much smaller sample sizes, we detect the statistical significance of less concentrated portfolio holdings in vertical team funds also for team sizes of three and four managers.

3.4.2. Team leadership and risk-taking

In this subsection, we analyse whether, besides differences in performance, there exist differences in risk taking between vertical and horizontal team funds. While the recent literature such as Sah and Stiglitz (1991), Sharpe (1981), Barry and Starks (1984), Adams and Ferreira (2010), Patel and Sarkissian (2017) supports theoretically and empirically that group-decision making reduces risk, the impact of leadership structure within the team remains unclear.

We examine the effect of vertical and horizontal team structure on fund's risk-taking using the following model:

$$Risk_{it} = g_0 + g_1 VTeam_{it-1} + g_2' Fund\_Char_{it-1} + g_3' Mgr\_Char_{it-1} + FE_{it} + \epsilon_{it}, \qquad (3.4)$$

where $Risk_{it}$ is one of fund's i risk measures at time t. Our risk measures, similar to Patel and Sarkissian (2017), include the fund's total volatility and five risk factor benchmarks from the Carhart (1997) model, namely: market beta, the exposure to size, book-to-market, and momentum portfolios, as well as the idiosyncratic residual volatility.

Table 10 reports the test results on the impact of vertical teams on various risk measures from the Carhart (1997) model and Model (4) estimation. All fund and manager controls from Table 2 (except fund family size and flows) that we used in performance evaluation tests are included in these regressions. The tests show that vertical team funds load marginally more on the market risk. This means that in spite of some extra systematic risk-taking, vertical team funds are unable to outperform their horizontal counterparts. However, our most unequivocal result is that team-managed funds with team leaders have significantly less residual risk than team-managed funds without lead managers. This corroborates well with the earlier finding that such funds hold less concentrated portfolios. Combined with their underperformance relative to horizontal funds documented earlier, this finding also indicates that vertical team funds are most likely making inferior stock selection choices.

3.5. Conclusion

In this paper, we examine the impact of leadership in team-managed U.S. domestic equity mutual funds by differentiating between vertical ("autocratic") and horizontal ("democratic") team structures. This task in the fund industry resembles the long-time debate on whether autocratic or democratic policy-making is better for institutional and country development because in both instances decision makers face with uncertain expectation environment, yet have clear ways to evaluate the performance of their decisions.

To accomplish our goal, we manually collect 485BPOS filings through SEC EDGAR system from 2012 to 2016 and identify funds with leader(s) within the teams of portfolio managers for each fund. We show in both univariate and multivariate settings that vertical team-managed funds that have clearly identified lead portfolio manager(s) significantly underperform funds with horizontal team structure. This underperformance is economically large reaching 50-70 bps per year depending on the type of risk-adjusted returns. The significant underperformance of vertical teams occurs for all team sizes except those with two managers, even though the proportion of funds with identified leaders increases with the number of managers in the fund.

We further show that the observed performance differences between horizontal and vertical team-managed funds are present across funds with all investment objectives, with the most significant difference being recorded for aggressive growth funds. Moreover, consistent with performance differences, we also find that vertical team-managed funds hold more diversified portfolios, load marginally more on the market risk and have less residual risk. These findings point out that vertical team funds possess low security selection ability.

Thus, our findings add not only to our understanding on the determinants of value-creation in the fund industry but also to the extensive debate in the economics and political science literature on the superiority of a specific form of governance (democratic or autocratic) for economic and social well-being. Using the fund industry as a laboratory, our evidence showing a clear performance dominance of horizontal team-management structure effectively reflects the same mechanisms as those in recent cross-country studies like Acemoglu et al. (2019) on the benefits of democratic form of government for country's economic growth.

3.6. Tables

Table 1. Summary statistics of fund leadership structure

This table reports the number and percentage of actively managed U.S. domestic equity mutual funds with leaders (i.e., vertical team funds) from 2012 to 2016. A fund is classified as a vertical team fund if it has at least one manager that is classified as a lead manager, otherwise, it is classified as a horizontal team fund. Panel A shows these statistics for all team-managed funds and for funds with different team sizes: two, three, four, and five or more managers (2 FM, 3 FM, 4 FM, and 5+ FM, respectively). Panel B shows these statistics for multiple management teams, i.e., when a fund is managed by multiple advisors and subadvisors, who make decisions independently from each other.

Panel A: Lead manager (vertical team) funds for all team-managed funds

	All		2 FM		3 FM		4 FM		5+ FM	
Year	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
2012	608	41.22	178	31.5	139	37.67	87	42.86	204	60.36
2013	571	39.3	166	30.18	137	36.34	83	40.29	185	57.81
2014	560	37.94	173	30.4	136	36.86	95	44.6	156	48
2015	573	37.28	187	30.41	139	36.29	93	39.41	154	50.83
2016	554	36.96	192	30.72	128	33.77	92	41.63	142	51.82
Total	2,866	38.52	896	30.64	679	36.17	450	41.71	841	53.91

Panel B: Lead manager (vertical team) funds for different number of management teams

	1 Management Team		2+ Management Teams	
Year	Number	Percent	Number	Percent
2012	474	36.72	134	72.83
2013	453	35.25	118	70.24
2014	461	35.24	99	58.93
2015	479	34.86	94	57.67
2016	462	34.32	92	60.13
Total	2,329	35.27	537	64.23

Table 2: Fund and manager characteristics

This table gives the summary statistics of team-managed U.S. domestic equity mutual funds based on their leadership structure. The sample period is from 2012 to 2016. A fund is classified as a vertical team (V) fund if it has at least one manager that is classified as a lead manager; otherwise it is classified as a horizontal team (H) fund. Panel A reports fund characteristics. Fund Size ($ bln) is total net assets under management of a fund in a given year. Fund Age (years) is the difference between a fund's inception year and the current year. Family Size ($ bln) is measured by the total net assets under management of the fund complex to which the fund belongs at the end of the calendar year. Volatility (%) is the S.D. of monthly fund returns over the past 12 months. Expenses (%) is the annual total expense ratio of the fund. Turnover (%) is the minimum of aggregated sales or aggregated purchases of securities of the year divided by the average 12-month total net assets of the fund. Flows (%) is defined as the net growth in the total net assets of funds, as a percentage of their total net assets, adjusted for prior-year returns. Expenses, Turnover, and Flows are winsorized at the 1% and 99% levels. Panel B reports manager characteristics. Fund Tenure (years) is the number of years the fund manager remains with the fund. Female (Fraction) is defined as the proportion of female managers in a fund. The last column Diff (V-H) shows differences in fund and manager characteristics between vertical (V) and horizontal (H) team funds. The absolute t-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Team Structure	Obs.	Mean	S.D.	Min	Max	Diff (V-H)
Fund Characteristics							
Fund Size ($ bln)	V	2,863	1.33	3.03	0	47.83	-0.5793***
	H	4,557	1.91	7.4	0	145.55	(3.98)
Fund Age (years)	V	2,866	14.44	11.99	0	86.00	-0.9738***
	H	4,574	15.40	12.98	0	92.00	(3.21)
Family Size ($ bln)	V	2,866	33.85	89.53	0	586.46	-3.8057*
	H	4,574	37.65	89.01	0	586.46	(1.79)
Volatility (%)	V	2,751	4.00	1.26	0	9.72	0.0892***
	H	4,384	3.91	1.21	0	9.91	(2.99)
Expenses (%)	V	2,780	1.08	0.32	0.19	2.20	0.0313***
	H	4,421	1.05	0.34	0.19	2.20	(3.92)
Turnover (%)	V	2,810	61.50	49.25	2.00	303.00	0.3868
	H	4,443	61.89	49.12	2.00	303.00	(0.33)
Flows (%)	V	2,749	0.19	1.55	-0.89	17.80	0.0147
	H	4,380	0.21	1.48	-0.89	17.80	(0.40)
Manager Characteristics							
Tenure (years)	V	2,866	5.5	4.15	0	25.0	-0.2647***
	H	4,574	5.77	4.32	0	25.5	(2.59)
Female (proportion)	V	2,866	0.09	0.16	0	1	-0.0045
	H	4,574	0.09	0.17	0	1	(1.15)

Table 3: Determinants of the fund leadership structure

This table reports probit tests of the determinants of leadership structure using team-managed U.S. domestic equity mutual funds from 2013 to 2016. In Column (1) the dependent variable is the probability that the team structure of fund *i* at time *t* is vertical. In Column 3, the dependent variable is the probability that the team structure of fund *i* at time *t* switches from horizontal to vertical. In Column 4, the dependent variable is the probability that the team structure of fund *i* at time *t* switches from vertical to horizontal. In the sample 13.72% funds have ever changed their team leadership structure at least once. There are 107 cases where a team-managed fund changes its structure from horizontal to vertical. There are 182 cases where a team-managed fund changes its structure from vertical to horizontal. All fund and manager variables are defined in Table 2, but Fund Size, Fund Age, and Family Size are taken in the log form. The performance measure, α_C4, is the fund alpha from the Carhart (1997) model. All regressors but Team Size are lagged by one year. All regressions include time by investment objective fixed effects (FE), and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

		Team Leadership Structure Changes	
	Vertical Team	Horizontal to Vertical	Vertical to Horizontal
Team Size $_{i,t}$	0.1121***	0.0379***	0.0028
	(8.47)	(5.46)	(0.16)
α_C4 $_{i,t-1}$	-0.1242**	-0.0886	-0.2217**
	(2.27)	(0.93)	(2.27)
Fund Size $_{i,t-1}$	0.0136	-0.0410	-0.0472
	(0.67)	(0.90)	(0.91)
Fund Age $_{i,t-1}$	-0.0727*	0.0375	0.1263**
	(1.87)	(0.92)	(2.41)
Family Size $_{i,t-1}$	-0.0028	-0.0146	-0.0067
	(0.19)	(0.64)	(0.32)
Expenses $_{i,t-1}$	0.3151***	-0.0920***	-0.1868
	(2.79)	(2.75)	(0.73)
Turnover $_{i,t-1}$	-0.0003	0.0014**	0.0008*
	(0.58)	(2.17)	(1.78)
Volatility $_{i,t-1}$	0.0575*	0.1094	-0.0721
	(1.84)	(1.41)	(1.62)
Flows $_{i,t-1}$	-0.0076	-0.1080**	-0.0778*
	(0.81)	(2.54)	(1.92)
Tenure $_{i,t-1}$	0.0011	0.0155	-0.0259***
	(0.13)	(1.23)	(3.57)
Female $_{i,t-1}$	-0.0325	0.2790	0.3748*
	(0.18)	(0.83)	(1.86)
Constant	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes
Obs.	6,433	6,275	5,088

Table 4: Performance of funds with horizontal and vertical team management structure

This table presents the summary statistics of the next-year performance based on the leadership structure of team-managed U.S. domestic equity mutual funds from 2013 to 2016. It reports the mean and standard deviation (S.D.) of three fund performance measures: α_FFC3 is the monthly (percentage) risk-adjusted gross fund returns computed each year over 12 monthly observations using Fama and French (1993) 3-factor model, and α_C4 is the similarly computed risk-adjusted return from the Carhart (1997) 4-factor model, α_FF5 is the similarly computed risk-adjusted return from the Fama and French (2015) 5-factor model, and α_PS5 is the similarly computed risk-adjusted return from the 5-factor model, which includes the liquidity factor of Pastor and Stambaugh (2003) added to the Carhart (1997) model. The table also reports the difference Diff (V-H) in performance test results between the groups of horizontal funds and vertical funds for the full sample and across funds with different numbers of fund managers. 2FM, 3FM, 4FM, and 5+FM denote various number of managers as defined in Table 1. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

					Number of Managers: Diff (V-H)			
	All	Vertical	Horizontal	Diff (V-H)	2 FM	3 FM	4 FM	5+ FM
α_FF3								
Mean	-0.1586	-0.1912	-0.1378	-0.0535*** (4.67)	-0.0368* (1.75)	-0.0606*** (2.72)	-0.1008*** (3.39)	-0.0585*** (2.67)
S.D.	0.4197	0.4059	0.4270					
Obs.	5,495	2,143	3,352		2,039	1,408	810	1,238
α_C4								
Mean	-0.1659	-0.1989	-0.1449	-0.0540*** (4.67)	-0.0359* (1.69)	-0.0582*** (2.60)	-0.1088*** (3.65)	-0.0598*** (2.70)
S.D.	0.4226	0.4105	0.4289					
Obs.	5,495	2,143	3,352		2,039	1,408	810	1,238
α_FF5								
Mean	-0.1995	-0.2375	-0.1752	-0.0625*** (4.96)	-0.0376 (1.62)	-0.0597** (2.45)	-0.1295*** (4.11)	-0.0788*** (3.22)
S.D.	0.4627	0.4444	0.4723					
Obs.	5,495	2,143	3,352		2,039	1,408	810	1,238
α_PS5								
Mean	-0.1506	-0.1769	-0.1338	-0.0429*** (3.56)	-0.0354 (1.57)	-0.0354 (1.50)	-0.1046*** (3.37)	-0.0369* (1.64)
S.D.	0.4397	0.4280	0.4463					
Obs.	5,495	2,143	3,352		2,039	1,408	810	1,238

Table 5: The effect of leadership structure on fund performance

This table reports the effect of leadership structure on fund performance using team-managed U.S. domestic equity mutual funds from 2013 to 2016. The dependent variable contains four performance measures, defined in Table 3. Independent variables are fund and manager controls as defined in Table 2. The independent variable of interest in Panel A is the vertical team dummy, VTeam, which equals 1 if the fund has a vertical team management structure and equals 0 if it has a horizontal structure. The independent variable of interest in Panel B is team-weighted vertical team measure, WVTeam, which is used to account for funds with multiple management teams. WVTeam is the fraction of management teams that have lead managers. All regressors are lagged by one year. All regressions include time by investment objective fixed effects (FE) and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Vertical team dummy

	α_FF3		α_FFC4		α_FF5		α_PS5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$VTeam_{i,t-1}$	-0.0599*** (3.48)	-0.0556*** (2.88)	-0.0615*** (3.32)	-0.0564*** (2.79)	-0.0702*** (3.90)	-0.0632*** (3.27)	-0.0498** (2.34)	-0.0484** (2.16)
$Fund\ Size_{i,t-1}$		-0.0019 (0.46)		-0.0027 (0.63)		-0.0036 (0.39)		0.0026 (0.88)
$Fund\ Age_{i,t-1}$		0.0070 (0.28)		0.0066 (0.28)		0.0159 (0.89)		-0.0017 (0.08)
$Family\ Size_{i,t-1}$		0.0010 (0.24)		0.0010 (0.22)		0.0023 (0.63)		-0.0011 (0.40)
$Expenses_{i,t-1}$		-0.1216*** (4.56)		-0.1258*** (5.08)		-0.1517*** (3.33)		-0.1063*** (5.98)
$Turnover_{i,t-1}$		-0.0000 (0.07)		-0.0000 (0.11)		-0.0002 (0.56)		0.0002 (0.33)
$Volatility_{i,t-1}$		-0.0404 (1.54)		-0.0495* (1.75)		-0.0743** (2.16)		-0.0112 (0.52)
$Flows_{i,t-1}$		0.0028 (0.41)		0.0033 (0.56)		0.0046 (1.30)		-0.0008 (0.12)
$Tenure_{i,t-1}$		-0.0040 (1.11)		-0.0045 (1.32)		-0.0051 (1.23)		-0.0022 (0.62)
$Female_{i,t-1}$		0.0353 (1.52)		0.0372* (1.83)		0.0260 (0.89)		0.0473 (1.49)
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	5,495	5,179	5,495	5,179	5,495	5,179	5,495	5,179
Adj. R^2	0.035	0.051	0.038	0.059	0.022	0.053	0.045	0.051

Table 5 (continued)

Panel B: Team-weighted vertical team measure

	α_FF3		α_FFC4		α_FF5		α_PS5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$WVTeam_{i,t-1}$	-0.0681*** (4.15)	-0.0575*** (2.81)	-0.0698*** (3.82)	-0.0582*** (2.71)	-0.0832*** (4.15)	-0.0682*** (3.07)	-0.0539** (2.55)	-0.0471* (1.95)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	5,495	5,179	5,495	5,179	5,495	5,179	5,495	5,179
Adj. R^2	0.035	0.051	0.039	0.058	0.023	0.053	0.045	0.050

Table 6: The effect of leadership structure on fund performance for different numbers of teams in the fund

This table reports the effect of leadership structure on fund performance across funds with different numbers of teams (multiple advisors and subadvisors) using team-managed U.S. domestic equity mutual funds from 2013 to 2016. The dependent variable is the Carhart (1997) four-factor alpha. The independent variable of interest is the team-weighted vertical team measure, WVTeam, which accounts for funds with multiple management teams. It is the fraction of management teams that have lead managers. In funds with one management team WVTeam is equivalent to the vertical team dummy, VTeam, which equals 1 if the fund has a vertical team management structure and equals 0 if it has a horizontal structure. All regressors are lagged by one year. All regressions include time by investment objective fixed effects (FE) and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	1 Management Team		2+ Management Teams	
Variable	(1)	(2)	(3)	(4)
$WVTeam_{i,t-1}$	-0.0720***	-0.0613***	-0.0057	0.0047
	(3.75)	(2.71)	(0.12)	(0.13)
Controls	No	Yes	No	Yes
Constant	Yes	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes	Yes
Obs.	4,843	4,561	652	618
Adj. R^2	0.043	0.061	0.041	0.074

Table 7: The effect of leadership structure on fund performance across different team sizes

This table reports the effect of leadership structure on fund performance across different number of fund managers (FM) using team-managed U.S. domestic equity mutual funds from 2013 to 2016. The dependent variable is the Carhart (1997) four-factor alpha. Independent variables are fund and manager controls as defined in Table 2. The independent variable of interest is the vertical team dummy, VTeam, which equals 1 if the fund has a vertical team management structure and equals 0 if it has a horizontal structure. All regressors are lagged by one year. All regressions include time by investment objective fixed effects (FE) and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	2 FM		3 FM		4 FM		5+ FM	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VTeam $_{i,t-1}$	-0.0431	-0.0367	-0.0741***	-0.0665***	-0.1105***	-0.0943***	-0.0697***	-0.0784***
	(1.19)	(1.14)	(4.81)	(3.18)	(4.65)	(3.01)	(2.64)	(2.74)
Fund Size $_{i,t-1}$		-0.0023		-0.0268***		-0.0034		0.0227***
		(0.30)		(6.78)		(0.29)		(2.64)
Fund Age $_{i,t-1}$		-0.0020		0.0529***		-0.0151		-0.0001
		(0.07)		(3.60)		(0.67)		(0.00)
Family Size $_{i,t-1}$		-0.0026		0.0078**		0.0018		-0.0042
		(0.30)		(2.44)		(0.18)		(0.90)
Expenses $_{i,t-1}$		-0.1498**		-0.1642***		-0.0933**		-0.0515
		(2.36)		(2.75)		(2.10)		(1.18)
Turnover $_{i,t-1}$		0.0000		-0.0001		0.0003		-0.0006
		(0.08)		(0.16)		(0.69)		(1.20)
Volatility $_{i,t-1}$		-0.0452***		-0.0395		-0.0683**		-0.0577
		(3.31)		(1.08)		(2.48)		(1.05)
Flows $_{i,t-1}$		0.0026		0.0085*		-0.0095***		0.0069
		(0.27)		(1.84)		(3.21)		(0.75)
Tenure $_{i,t-1}$		-0.0044		-0.0040		0.0006		-0.0125***
		(1.17)		(0.93)		(0.10)		(6.88)
Female $_{i,t-1}$		-0.0029		0.1255		-0.1033		0.0621
		(0.17)		(1.56)		(0.73)		(0.96)
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	2,039	1,914	1,408	1,332	810	765	1,238	1,168
Adj. R^2	0.032	0.050	0.060	0.085	0.032	0.040	0.038	0.074

Table 8: The effect of leadership structure on fund performance across different investment objectives

This table reports the effect of leadership structure on fund performance using team-managed U.S. domestic equity mutual funds from 2013 to 2016 across three investment objective categories: aggressive growth, growth, and income (a combination of growth & income and equity income funds). The dependent variable is the Carhart (1997) four-factor alpha. Independent variables are fund and manager controls as defined in Table 2. The independent variable of interest is the vertical team dummy, VTeam, which equals 1 if the fund has a vertical team management structure and equals 0 if it has a horizontal structure. All regressors are lagged by one year. All regressions include year fixed effects (FE) and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Aggressive Growth		Growth		Income	
	(1)	(2)	(3)	(4)	(5)	(6)
VTeam $_{i,t-1}$	-0.0972*	-0.0829**	-0.0558***	-0.0523***	-0.0349	-0.0251
	(1.88)	(2.00)	(3.78)	(3.33)	(1.63)	(1.25)
Fund Size $_{i,t-1}$		0.0081		-0.0049		-0.0048
		(0.77)		(0.80)		(0.21)
Fund Age $_{i,t-1}$		0.0065		0.0115		-0.0094
		(0.24)		(0.63)		(0.24)
Family Size $_{i,t-1}$		-0.0197**		0.0060**		0.0014
		(2.09)		(2.37)		(0.12)
Expenses $_{i,t-1}$		-0.0512		-0.1379***		-0.1726*
		(0.77)		(3.02)		(1.87)
Turnover $_{i,t-1}$		-0.0003		0.0001		-0.0003***
		(0.37)		(0.26)		(10.06)
Volatility $_{i,t-1}$		-0.0362		-0.0515*		-0.0522
		(0.79)		(1.95)		(1.23)
Flows $_{i,t-1}$		0.0067		0.0024		-0.0053
		(0.50)		(0.49)		(0.31)
Tenure $_{i,t-1}$		-0.0105**		-0.0043		0.0009
		(2.25)		(1.03)		(0.22)
Female $_{i,t-1}$		-0.0519		0.0628		0.0469*
		(0.83)		(1.25)		(1.73)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1,215	1,143	3,330	3,154	950	882
Adj. R^2	0.028	0.037	0.037	0.066	0.031	0.059

Table 9: Portfolio concentration and leadership structure

This table reports the effect of leadership structure on fund industry portfolio concentration (proportion) using the team-managed U.S. domestic equity mutual funds from 2013 to 2016. Industry portfolio concentration is constructed following Kacperczyk et al. (2005). Panel A reports the mean industry concentration for different number of fund managers: 2 FM, 3 FM, 4 FM, and 5+ FM denote the number of managers in the team as defined in Table 1 and the absolute *t*-statistics for the difference in industry concentrations between vertical and horizontal team-managed funds. Panel B reports multiivariate regression estimations of industry concentration on the vertical team dummy, VTeam, and fund controls as in Table 5. All regressors are lagged by one year, include time by investment objective fixed effects (FE), and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Industry concentration across horizontal and vertical team funds

	All	2 FM	3 FM	4 FM	5+ FM
Vertical	0.0318	0.0350	0.0325	0.0318	0.0283
Horizontal	0.0350	0.0363	0.0352	0.0339	0.0324
Diff (V-H)	-0.0032***	-0.0013	-0.0026	-0.0021	-0.0040*
	(3.46)	(0.78)	(1.47)	(1.03)	(1.80)
Obs.	3,952	1,445	1,017	572	918

Panel B: Multivariate analysis for industry concentration across horizontal and vertical team funds

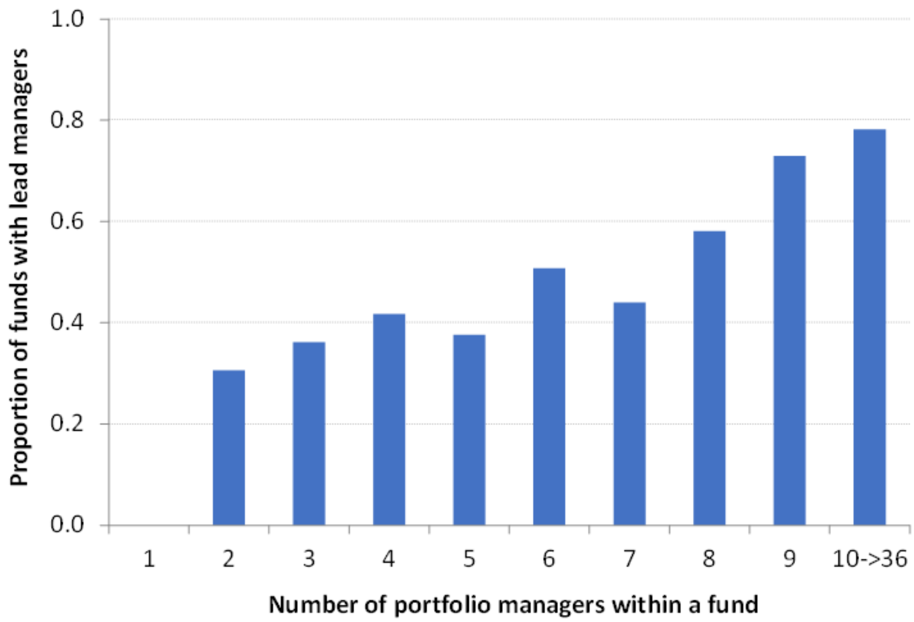
	All	2 FM	3 FM	4 FM	5+ FM
$VTeam_{i,t-1}$	-0.0043***	-0.0027	-0.0049**	-0.0036*	-0.0060
	(3.05)	(1.41)	(2.01)	(1.66)	(1.12)
Controls	Yes	Yes	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes	Yes	Yes
Obs.	3,728	1,355	960	546	867
Adj. R^2	0.113	0.124	0.172	0.111	0.087

Table 10: The effect of leadership structure on risk-taking behavior

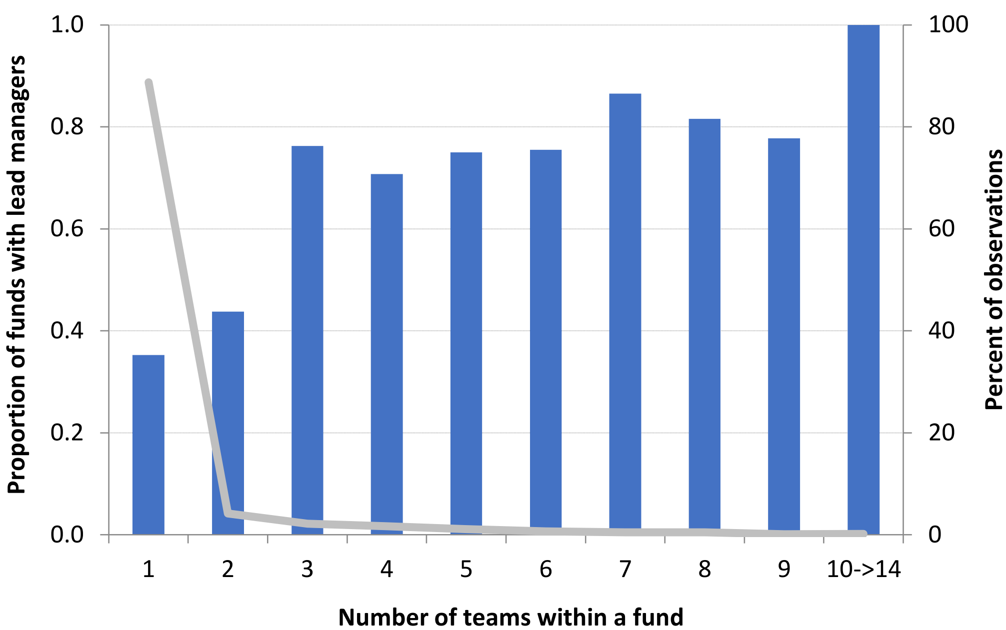
This table reports the effect of leadership structure on the risk-taking behavior using the team-managed U.S. domestic equity mutual funds from 2013 to 2016. It shows the estimates from panel regressions of fund risk taking on the Vertical Team (VTeam) dummy and other controls. TOTAL is defined as the standard deviation of monthly gross fund returns over the past 12 months. MKT, SMB, HML, and MOM are the coefficients on market, size, book-to-market ratio, and momentum portfolios based on the Carhart (1997) four-factor model. RESIDUAL is the standard deviation of the fund's residual return from the Carhart model. The independent variable of interest is the vertical team dummy, VTeam, which equals 1 if the fund has a vertical team management structure and equals 0 if it has a horizontal structure. Other independent variables are fund and manager characteristics as controls as defined in Table 6. All regression specifications include time by investment objective fixed effects (FE), and standard errors are clustered by fund and year. The absolute *t*-statistics are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

		Carhart (1997) Model				
	TOTAL	MKT	SMB	HML	MOM	RESIDUAL
VTeam $_{i,t-1}$	0.0248	0.0120*	0.0189	-0.0093	-0.0064	-0.0332***
	(1.22)	(1.78)	(1.63)	(0.75)	(1.29)	(2.68)
Fund Size $_{i,t-1}$	-0.0034	-0.0039**	-0.0082**	-0.0172***	-0.0003	0.0161**
	(0.48)	(2.38)	(2.27)	(3.10)	(0.08)	(2.50)
Fund Age $_{i,t-1}$	0.0173	0.0176***	-0.0142	-0.0002	0.0125	-0.0690***
	(0.92)	(2.62)	(1.42)	(0.03)	(1.38)	(7.06)
Family Size $_{i,t-1}$	0.0047	0.0051	0.0052	-0.0033	0.0017	-0.0169***
	(0.62)	(1.44)	(1.50)	(0.87)	(0.94)	(4.19)
Expenses $_{i,t-1}$	0.2337***	-0.0095	0.1133***	-0.1165***	-0.0292	0.3516***
	(3.14)	(0.51)	(3.94)	(3.32)	(0.80)	(11.38)
Turnover $_{i,t-1}$	0.0009**	0.0001	0.0006***	-0.0006***	0.0004	0.0002
	(2.25)	(0.98)	(3.44)	(3.43)	(1.10)	(1.26)
Tenure $_{i,t-1}$	0.0022	-0.0015	0.0019	-0.0008	-0.0030*	0.0118***
	(0.65)	(1.18)	(0.74)	(0.46)	(1.90)	(4.93)
Female $_{i,t-1}$	-0.0748	-0.0174	-0.0347	0.0480	0.0271***	0.0198
	(1.01)	(1.60)	(0.94)	(1.21)	(14.90)	(0.28)
Constant	Yes	Yes	Yes	Yes	Yes	Yes
Year × Obj. FE	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	5,191	5,187	5,187	5,187	5,187	5,187
Adj. R^2	0.486	0.073	0.472	0.120	0.116	0.193

3.7. Figures



Plot A



Plot B

Figure 1: Proportion of funds with lead managers across different team sizes
The figure shows the proportion of funds with lead managers for all diversified and actively managed team-based U.S. domestic equity mutual funds from 2012 to 2016. Plot A shows this proportion for different team sizes within a fund, while Plot B – for different number of teams within a fund (columns) with the corresponding percentages of data observations (line).

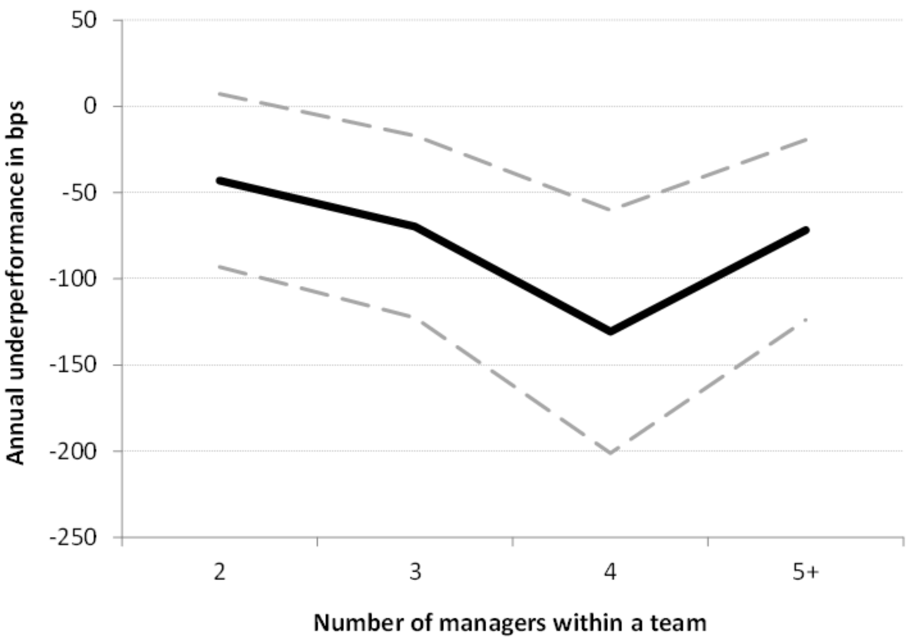


Figure 2: The annual underperformance of vertical team funds across different team sizes
The figure shows the annual difference in basis points between vertical and horizontal team funds based on the Carhart (1997) alpha, α_C4, across funds with different number of managers. The sample consists of all diversified and actively managed team-based U.S. domestic equity mutual funds from 2012 to 2016. The corresponding mean point estimates from Table 4 are multiplied by 1,200 (solid line). For each estimate, the plot also shows the upper and lower 95% confidence bounds (dashed lines).

3.8. Appendix

Due to a very significant time involved in data collection, we adopt a recursive imputation procedure to collect fund managerial information. Specifically, we assume that *if a fund maintains the exact same managers from year to year, its leadership structure does not change*.[39] Based on this assumption, we first manually collect data for funds in 2015 and then impute leadership structure for funds in 2012, 2013, 2014, and 2016 with 2015 managerial information. After the first-round imputation, we then manually collect managerial information for un-imputed funds in 2014 and 2016. The imputation procedure is recursive since we use the manually collected information in 2014 to further impute the un-imputed funds in 2013. For funds in 2013 that are left not imputed using either 2015 or 2014 managerial information, we manually collect their managerial information. Similarly, we impute 2012 funds with 2015, 2014, 2013 managerial information, in such order, and conduct manual collection for un-imputed funds.

The following table shows the current imputation rate at the manager level. For instance, in 2014, 3294 out of 5876 fund managers are the same as in 2015 once we match the funds that exist in both 2014 and 2015.

Number of Managers	Target Year				
	2012	2013	2014	2015	2016
Manual Collection	2,804	2,733	2,582	6,052	2,470
Imputation using 2015 info	1,475	2,177	3,294		3,361
Imputation using 2014 info	504	901			
Imputation using 2013 info	1,227				
Total Number of Managers	6,010	5,811	5,876	6,052	5,831
Imputation Rate (%)	53.34%	52.97%	56.06%	NA	57.64%

The following table shows the imputation rate at the fund level.

[39] There are cases where manager names present in the Morningstar Direct database were not found in the fund prospectus for a given year or nearby years. A fund in such case will have some of its managers with available managerial information filled, while other managers from the same fund do not. In such scenarios, we do not impute such fund in other years and manually checked its managerial information in other years.

Number of Funds	Target Year				
	2012	2013	2014	2015	2016
Manual Collection	568	528	505	1,583	478
Imputation using 2015 info	521	729	1,015		1,055
Imputation using 2014 info	143	239			
Imputation using 2013 info	292				
Total Number of Funds	1,524	1,496	1,520	1,583	1,533
Imputation Rate (%)	62.73%	64.71%	66.78%	NA	68.82%

To justify our assumption for imputation and evaluate the accuracy of our imputation procedure, we randomly select 25 imputed funds from each of the years (2012, 2013, 2014, and 2016) and manually collect managerial information for those imputed funds from the SEC EDGAR system. By comparing the manually collected information with the imputed information, we find our procedure has an overall accuracy of more than 95% for the 100 randomly selected funds. The following table shows the match rates in terms of leadership classification at the manager level for the 100 randomly selected funds.

Number of Managers	Target Year				
	2012	2013	2014	2015	2016
Manual Collection	74	70	80	87	311
Total Number of Matches	71	68	80	86	305
Total Number of Non-Matches	3	2	0	1	6
Match Rate (%)	95.95%	97.14%	100.00%	98.85%	98.07%

The following table shows the match rates in terms of leadership classification at the fund level for the 100 randomly selected funds.

Number of Funds	Target Year				
	2012	2013	2014	2015	2016
Manual Collection	25	25	25	25	100
Total Number of Matches	25	23	25	24	97
Total Number of Non-Matches	0	2	0	1	3
Match Rate (%)	100.00%	92.00%	100.00%	96.00%	97.00%

4. Conclusion

This thesis evaluates mutual fund performance from both the investors’ and managerial perspectives. To evaluate investors’ economic gains in real time, I adopt two approaches: rule-based portfolio sorting and regression-based machine learning algorithms. Both approaches can deliver additional abnormal returns to investors, depending on specific risk adjustments. I further show that average investors seem to react to predictive information embedded in predictors. These results suggest that real-time benefits of using fund performance predictors can only be recovered by complex algorithms which are computationally costly to implement for average investors, and the benefits should proxy for the search costs investors need to incur to find skilled managers if the asset management industry is informationally efficient. From the managerial perspective, using a novel dataset, we find that horizontally-managed equity funds perform better than vertically-managed equity funds in U.S., supporting that horizontal decision-making structure in organizations functioning in an uncertain expectation environment adds extra values in overall.

All these results provide new evidence that actively managed equity mutual funds can provide additional values to investors beyond passive portfolios. Meanwhile, there exists substantial heterogeneity in values added discovered either through new methodologies or new datasets. It would be interesting to distinguish these benefits across investors with different preferences and evaluation benchmarks which I leave for future work.

5. Bibliography

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J., 2019, Democracy does cause growth, *Journal of Political Economy* 127, 47-100.

Adams, R. and Ferreira, D., 2010, Moderation in groups: Evidence from betting on ice break-ups in Alaska, *Review of Economic Studies* 77, 882–913.

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.

Bar, M., Kempf, A., and Ruenzi, S., 2011, Is a team different from the sum of its parts? Evidence from mutual fund managers, *Review of Finance* 15, 359-396.

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., 2022, Skill, scale, and value creation in the mutual fund industry, *Journal of Finance* 77(1), 601-638.

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.

Barro, R.J., 1996: Democracy and growth, *Journal of Economic Growth* 1, 1–27.

Barry, C. and Starks, L., 1984, Investment management and risk sharing with multiple managers, *Journal of Finance* 39, 477–491.

Bliss, R., Porter, M., and Schwarz, C., 2008, Performance characteristics of individually-managed versus team-managed mutual funds, *Journal of Portfolio Management* 34, 110-119.

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, *Journal of Finance* 61(4), 1553–1604.

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

Carhart, M., Kaniel, R., Musto, D. and Reed, A., 2002, Leaning for the tape: Evidence of gaming behavior in equity mutual funds, *Journal of Finance* 57, 661-693.

Chen, J., Hong, H., Huang, M., and Kubik, J., 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* 94, 1276-1302.

Chen, J., Xie, L., and Zhou, S. 2020, Managerial multi-tasking, team diversity and mutual fund performance, *Journal of Corporate Finance* 65, 1017-1066.

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

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

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?, *Working Paper*.

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

Doucouliagos, H. and Ulubasoglu, M., 2008, Democracy and economic growth: A meta-analysis, *American Journal of Political Science* 52, 61–83.

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.

Evans, R., Prado, M., Rizzo, A., and Zambrana, R. 2020, Identity, diversity, and team performance: Evidence from U.S. mutual funds, CEPR Discussion Paper No. DP14305.

Evans, R. Ferreira, M., Matos, P., and Young, M. 2021, Hiding in plain sight: The global implications of manager disclosure, Working paper, University of Virginia.

Fama, E. and French, K., 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 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. and French, K., 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1-22.

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, *Journal of Finance* 73(4), 1663–1712.

Glaeser, E.L., La Porta, R., Lopez-de-Silanes, F., and Shleifer, A., 2004, Do institutions cause growth?, *Journal of Economic Growth* 9, 271-303.

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, *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* 59, 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.

Lu, Y., Naik, N., and Teo, M., 2021, Diverse hedge funds, Working paper, London Business School.

Martin, I. and Nagel, S., 2021, Market efficiency in the age of big data, *Journal of Financial Economics* 145(1), 154-177.

Massa, M., Reuter, J., and Zitzewitz, E., 2010, When should firms share credit with employees? Evidence from anonymously managed mutual funds, *Journal of Financial Economics* 95, 400-424.

Mueller, J., 2012, Why individuals in larger teams perform worse, *Organizational Behavior and Human Decision Processes* 117, 111-124.

Papaioannou, E. and Siourounis, G., 2008, Democratization and growth, *Economic Journal* 118, 1520-1551.

Pastor, L. and Stambaugh. R., 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Patel, S. and Sarkissian, S., 2017, To group or not to group? Evidence from mutual fund databases, *Journal of Financial and Quantitative Analysis* 52, 1989-2021.

Patel, S. and Sarkissian, S., 2021, Portfolio pumping and managerial structure, *Review of Financial Studies* 52, 1989-2021.

Persson, T. and Tabellini, G., 2006, Democracy and development: The devil in the details, *American Economic Review* 96, 319–24.

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

Pozuelo, J., Slipowitz, A., and Vuletin, G., 2016, Democracy does not cause growth: The importance of endogeneity arguments, *IDB Working paper No: IDB-WP-694*.

Prather, L. and K. Middleton, 2002, Are N+1 heads better than one? The case of mutual fund managers, *Journal of Economic Behavior and Organization* 47, 103-120.

Sah, R. and J. Stiglitz, 1991, The quality of managers in centralized versus decentralized organizations, *Quarterly Journal of Economics* 106, 289–295.

Sharpe, W., 1982, Decentralized investment management, *Journal of Finance* 36, 217–234.

Sheng, J., Simutin, M., and Zhang, T., 2022, Cheaper is not better: On the ‘superior’ performance of high-fee mutual funds, *Rotman School of Management Working Paper No. 2912511*.

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.

Tan, E. and Sen, A., 2019, Does educational diversity of managers matter for the performance of team-managed funds? *Accounting & Finance* 59, 801-830.

Tavares, J. and Wacziarg, R., 2001, How democracy affects growth, *European Economic Review* 45, 1341-1378.

Thompson, J.D., 1967, *Organizations in action*, New York: McGraw-Hill.

Waldman, D., Nohria, N., and Anand, B., 2001, When does leadership matter? The contingent opportunities view of CEO leadership, Working paper, Harvard University.

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.

Wittman, D., 1989, Why democracies produce efficient results, *Journal of Political Economy* 97, 1395-1424.

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