CFR Working Paper No. 26-01

Forecasting Mutual Fund Performance – Combining Return-Based with Portfolio Holdings-Based Predictors

S. Müller • N. Pugachyov • F. Weigert

centre for financial research
cologne

Forecasting Mutual Fund Performance – Combining Return-Based with Portfolio Holdings-Based Predictors

Sebastian Müller

Technical University of Munich

Nikolay Pugachyov

University of Neuchâtel

Florian Weigert

University of Neuchâtel

September 16, 2024

Sebastian Müller is at the Technical University of Munich (TUM) School of Management, TUM Campus Heilbronn, Bildungscampus 9, 74076 Heilbronn, Germany. Email: sebastian.mueller.hn@tum.de. Nikolay Pugachyov and Florian Weigert are at the University of Neuchâtel, Institute of Financial Analysis, Rue Abram-Louis-Breguet 2, 2000 Neuchâtel, Switzerland. Emails: nikolay.pugachyov@unine.ch and florian.weigert@unine.ch. We thank Lieven Baele, Kris Boudt, Stefan Ruenzi, and seminar participants at the Belgian Financial Research Forum 2024, and University of Mannheim for their valuable comments. An older draft of the paper was named "The winner should take it all: How academic research helps to separate winners from losers in the market for actively managed funds". The coauthors of this paper were Sebastian Müller, Ekaterina Serikova, and Florian Weigert. Florian Weigert is also affiliated with the Centre of Financial Research (CFR) Cologne and thankful for their continuous support. Any errors are our own.

Abstract

We introduce a simple yet powerful method for enhancing mutual fund performance prediction by combining individual predictors into a composite predictor. This composite approach integrates information from 19 well-established return-based and portfolio holdings-based predictors from the literature. It effectively identifies top decile funds that outperform bottom decile funds by a risk-adjusted 4.56% per annum. Furthermore, it achieves statistically significant outperformance for long-only fund investments against the average active and passive fund. Both return-based predictors (e.g., fund alpha and the t-statistic of alpha) and holdings-based predictors (e.g., skill index and active weight) contribute equally to the composite predictor's success.

JEL classification: G11, G12, G20, G23

Keywords: Mutual funds, performance prediction, composite predictor

1 Introduction

Much of the academic literature on actively-managed mutual funds centers around their ability to generate positive risk-adjusted returns and whether such performance persists over time. Seminal studies such as Sharpe (1966), Jensen (1968), Gruber (1996), and Carhart (1997) establish that, on average, mutual funds are not able to generate superior risk-adjusted returns and find weak to no evidence supporting performance persistence.[1] More recent studies of Busse, Goyal, and Wahal (2010) and Ferreira et al. (2013) confirm these earlier findings.

Although active mutual funds, on average, do not outperform passive benchmarks, this does not rule out that a subset of skilled fund managers is able to generate superior performance. Consequently, numerous studies have attempted to identify variables associated with future outperformance of actively-managed funds. Generally, such predictor variables can be classified into return-based predictors (RBPs, i.e., variables based on funds' past returns as well as fund characteristics such as size, age, turnover, and expense ratios) and holdings-based predictors (HBPs, i.e., variables derived from funds' disclosed portfolio holdings).[2] While earlier studies have predominantly investigated the impact of RBPs on future performance (see, e.g. Grinblatt and Titman (1989), Ippolito (1989), Malkiel (1995), Chevalier and Ellison (1997), Sirri and Tufano (1998), Zheng (1999)), more recent papers also find value in applying HBPs to forecast fund returns and alphas (see, e.g., Cohen, Coval, and Pástor (2005), Kacperczyk and Seru (2007), and Hoberg, Kumar, and Prabhala (2018)) as detailed information on fund managers' trading styles cannot be inferred from common RBPs.

[1] Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Brown and Goetzmann (1995), Elton, Gruber, and Blake (1996), and Wermers (1999) provide early evidence that mutual funds exhibit performance persistence. However, more recent findings imply that, on average, actively managed U.S. equity mutual funds earn zero abnormal returns before expenses, and after accounting for fees, these returns become negative. Moreover, as noted by Fama and French (2010) and Barras, Scaillet, and Wermers (2010), mutual funds only show weak evidence of persistence before 1996, but not afterwards.

[2] We do not utilize fund manager characteristics in this study as coverage of those in commercial databases (such as CRSP Mutual Funds and Morningstar) is limited.

The goal of our paper is to suggest a new and distinct approach to forecast mutual fund performance by aggregating individual RBPs and HBPs into a composite score. Our work is inspired by recent meta-studies on return predictors in the stock market (see, e.g., Hou, Xue, and Zhang (2015), Harvey, Liu, and Zhu (2016), Green, Hand, and Zhang (2017), and Jacobs and Müller (2020)) and on influential calls to address the "factor zoo", i.e., the number of assets to explain the cross-section of a particular asset class (see, e.g., Cochrane (2011), and Harvey (2017)). While the "factor zoo" for stocks has now been tamed by a number of studies (Hou, Xue, and Zhang (2020), Kozak, Nagel, and Santosh (2020), Chen and Zimmermann (2021), and Jensen, Kelly, and Pedersen (2023)), we believe that there is still room to contribute to temper the "zoo" of variables used to predict mutual fund performance.

Our basic setup is straightforward. We replicate 19 prominent predictors of future fund performance from the literature (9 RBPs and 10 HBPs) and aggregate them in a consistent framework.[3] Precisely, we first perform rolling univariate cross-sectional regressions of funds' net-of-fees returns in month t on the individual predictor variables in month $t-1$ over an expanding window of at least 36 months to obtain predictor betas. Second, we generate fund return forecasts in month $t+1$ for each individual predictor by multiplying the predictor beta with the predictor value in month t. Third, we equal-weight the individual sign-adjusted forecasts to arrive at a composite return prediction for each fund in month $t+1$. Finally, in each month $t+1$, we sort the cross-section of funds based on this composite score from high to low and invest in the top decile of funds. Applying the same logic, we also divest from the bottom decile of funds.[4] In essence, the method is similar to the construction of the mispricing score of Stambaugh and Yuan (2017) and the quality score of Asness, Frazzini, and Pedersen (2019). By combining information across predictors, our aim is to obtain a diversified and less noisy proxy of (hidden) manager skill, which

[3] As RBPs, we select a fund's age, size, flows, expense ratio, turnover, volatility, alpha, t-statistic of alpha and R2. We utilize the Grinblatt and Titman (1993) performance measure, characteristic selectivity, characteristic timing, return gap, active share, active weight, industry concentration index, risk shifting, skill index, and fund duration as HBPs.

[4] Such a trading strategy is based on theoretical considerations, as it is not possible to short mutual funds. However, we also report results when we solely invest into funds with the highest predicted returns. Such a strategy is also practically feasible.

drives future fund performance compared to utilizing information of only one or a few individual predictors. Our estimation approach is entirely data-driven and allows to capture time-varying relationships between past individual predictors and future fund returns.[5]

We analyze data on 4,416 unique U.S. actively managed all-equity funds over the period 1985 to 2022. Fund returns and characteristics are from the Center for Research in Security Prices Mutual Fund (CRSP MF) database, while quarterly fund portfolio holdings data are taken from the Thomson Reuters Mutual Fund Holdings database. The merge of both databases is performed via the MFLinks merging table. To measure risk-adjusted performance of funds, our baseline models are the Carhart (1997) four-factor model, the Hou et al. (2021) five-factor q^5 model, and the Fama and French (2018) six-factor model.

Our main result is that combining individual predictors into a composite score improves the prediction of risk-adjusted fund performance considerably. While the average individual predictor achieves a six-factor alpha spread between top decile and bottom decile funds of insignificant 0.95% p.a., composite scores based on solely RBPs and HBPs predictors yield spreads of 2.76% ($t = 2.96$) and 3.48% ($t = 2.21$) p.a., respectively. Importantly, our results reveal that the best result is obtained when combining RBPs *and* HBPs predictors in a composite score. This score is able to select funds that yield a six-factor alpha spread of 4.56% ($t = 3.23$) p.a. Hence, consistent with our idea of mitigating noise in detecting manager skill, we find that averaging across all 19 predictors delivers superior predictive performance compared to utilizing individual predictors or averaging predictors separately within RBPs and HBPs categories.

We decompose the obtained alpha spreads by comparing the short and long legs of the strategies against the average active and average passive fund in our sample. Our results reveal that the composite score delivers improved performance predictability for *both* the long and the short sides. In particular, the top (bottom) decile of funds selected by the composite predictor outperforms (underperforms) the average active fund in the sample by statistically significant 1.92% (−2.64%) p.a. Noteworthy, it also selects funds that

[5] To enable reproducibility of our empirical findings, we provide detailed R code on GitHub (https://github.com/nikolaypugachyov).

significantly beat (lose against) the average *passive* fund on the long- and the short sides. Consequently, our results are also important for practitioners as fund selection strategies are primarily concerned with the long side of investing.

Finally, we (i) inspect the best fund selection strategy of a performance-chasing investor and (ii) evaluate the optimal combination across all RBP and HBP performance predictors. We empirically show that the best strategy for a performance-chasing investor who considers the relation between predictors and performance over the past 36 months is to invest into the composite score of *all* 19 predictors. Also, there is no other combination of individual predictors (which can now be constructed from both the RBP and HBP categories) that, on average, beats the performance of the 19-variable composite score. These results again document the importance of predictor diversification for investors when they intend to actively invest into mutual funds.

Our paper is closely related to DeMiguel et al. (2023) as well as Kaniel et al. (2023). Both papers investigate the use of machine learning (ML) algorithms for fund performance prediction and show that ML is able to consistently differentiate between high- and low-performing mutual funds, both before and after fees. We differentiate in two important aspects. First, both DeMiguel et al. (2023) and Kaniel et al. (2023) solely use RBPs as predictor variables to forecast fund performance. In contrast, we incorporate HBPs and demonstrate that they add sizeable value when predicting future fund returns and alphas. Second, our paper relies on a simple methodology to combine individual predictors without relying on vast computational power which is necessary to run certain regression trees or neural networks. Thus, practitioners can easily apply our methodology to select funds, whereas the use of ML techniques is supposedly reserved for mainly institutional investors with large computational facilities.

The remainder of the paper is as follows: Section 2 introduces the individual mutual fund predictors. In Sections 3 and 4, we describe our sample and the methodology to construct composite predictors. Sections 5 and 6 report the main empirical results on the combination of predictor variables and future fund performance. Finally, Section 7 reports robustness checks of our main findings, and Section 8 concludes.

2 Predictors of Mutual Fund Performance

From a theoretical point of view, Sharpe (1991) argues that after costs, the aggregate active returns must be less than the market return.[6] Furthermore, the efficient market hypothesis of Fama (1970) implies that achieving consistent risk-adjusted outperformance is extremely challenging, if not impossible. Combining these two insights, it is not only difficult to consistently beat the benchmark, but it also seems that the average active manager is doomed to underperform the market. Nevertheless, numerous studies have identified variables associated with the future performance of actively managed equity mutual funds, since the underperformance of the average mutual fund does not exclude the existence of a small subset of outperforming mutual funds (see, e.g. Pedersen (2018)). Table 1 lists alpha predictors covered in our study.

Table 1: Mutual Fund Performance Predictors Considered in the Study

	Original Study	Expected Sign
Fund- and Return-Based Predictors (RBPs)		
Age	Pástor, Stambaugh, and Taylor (2015, *JFE*)	(-)
Size	Chen, Hong, Huang, and Kubik (2004, *AER*)	(-)
Flows	Zheng (1999, *JF*)	(+)
Exp	Gil-Bazo and Ruiz-Verdú (2009, *JF*)	(-)
STO	Pástor, Stambaugh, and Taylor (2017, *JF*)	(+)
Vola	Jordan and Riley (2015, *JFE*)	(-)
CA	Carhart (1997, *JF*)	(+)
Tstat	Elton, Gruber, and Blake (1996, *JB*)	(+)
R2	Amihud and Goyenko (2013, *RFS*)	(-)
Portfolio Holding-Based Predictors (HBPs)		
GT	Grinblatt and Titman (1993, *JB*)	(+)
CS	Daniel, Grinblatt, Titman, and Wermers (1997, *JF*)	(+)
CT	Daniel, Grinblatt, Titman, and Wermers (1997, *JF*)	(+)
RG	Kacperczyk, Sialm, and Zheng (2008, *RFS*)	(+)
AS	Cremers and Petajisto (2009, *RFS*)	(+)
AW	Doshi, Elkamhi, and Simutin (2015, *RAPS*)	(+)
ICI	Kacperczyk, Sialm, and Zheng (2005, *JF*)	(+)
RS	Huang, Sialm, and Zhang (2011, *RFS*)	(-)
SI	Kacperczyk, Nieuwerburgh, and Veldkamp (2014, *JF*)	(+)
FD	Cremers and Pareek (2016, *JFE*)	(+)

This table lists nine RBPs and ten HBPs documented in the mutual fund literature. Columns 2 and 3 refer to the source and the sign of the relationship reported in the corresponding study.

[6] Fama and French (2010) report that "the aggregate portfolio of actively managed U.S. equity mutual funds is close to the market portfolio."

2.1 Fund- and Return-Based Predictors (RBPs)

Fund characteristics such as age (henceforth *Age*), size (henceforth *Size*), flows (henceforth *Flows*), expenses (henceforth *Exp*), and turnover have been frequently studied in the literature on their ability to predict performance of active mutual funds with mixed findings (Morey (2015)).

Theoretically, the effect of *Age* on a fund's performance can go in both directions. On the one hand, young mutual funds face pressure to perform better to avoid early termination. Alternatively, it is possible that mature fund performance improves due to fund managers' learning. Pástor, Stambaugh, and Taylor (2015) find that, on average, as a fund ages, its performance deteriorates.[7] When analyzing age-based investment strategies, the authors conclude that younger funds tend to outperform their older peers, which suggests that new entrants tend to be better incentivized or more skilled. Interestingly, once the authors control for mutual fund industry size, this negative relationship disappears and the average effect becomes positive. The latter suggests that fund managers' learning could be responsible for a positive fund age-performance relation.

As with *Age*, the effect of *Size* on funds' performance could be hypothesized to be either positive or negative. Growing larger may yield benefits such as the ability to negotiate better commissions due to larger position sizes, distribute fixed costs over a larger asset base, invest in opportunities unavailable to small funds, and have greater research capabilities. On the other hand, these benefits may be exhausted after a certain size level because it becomes more difficult to exploit good investment opportunities due to diseconomies of scale (Berk and Green, 2004). Chen et al. (2004) find support for the diseconomies of scale phenomenon and argue that it is closely related to liquidity and price impact costs.[8] Large funds are forced to trade in large quantities and, hence, are likely to take larger than optimal positions that erode performance due to lack of liquidity. Pollet and Wilson

[7] Most of the time the literature on US mutual funds finds no significant relationship between age and performance. For instance, Chen et al. (2004) and Yan (2008) find no significant coefficient on fund age when it is used as a control variable in their performance regressions; while Kacperczyk, Sialm, and Zheng (2005) report negative and statistically significant coefficient.

[8] While early literature reports mixed results (Grinblatt and Titman, 1989, 1994), the recent literature is in line with the study of Chen et al. (2004).

(2008) and Yan (2008) find further support that liquidity constraints limit the scalability of good investment ideas, confirming that fund size is negatively related to future fund performance. Further, trading in large volumes may lead to high price impact costs because of increased attention and trading from other market participants.

Similarly, evidence on the impact of *Flows*, *Exp* and turnover is not universal. Gruber (1996) and Zheng (1999) document support for the smart money hypothesis which assumes that fund investors can identify skilled managers, and hence, net inflows should distinguish better performing funds. On the other hand, Edelen (1999) finds a negative flow-performance relationship, while Ferreira et al. (2013) find no significant relationship between flows and future performance. One would naturally expect that an investor would be willing to pay a higher fee in return for a greater value generated by a fund. Nonetheless, Carhart (1997) and Gil-Bazo and Ruiz-Verdú (2009) observe a negative effect of expenses on performance, while Chen et al. (2004) and Ferreira et al. (2013) do not find a significant relationship. Unlike the previous literature, Pástor, Stambaugh, and Taylor (2017) document a strong positive time-series relationship between turnover and future performance.[9] The authors argue that profit opportunities are time-varying. In consequence, more skilled funds would trade more in periods when such opportunities are numerous. The cross-sectional turnover-performance is thus negatively impacted by trading costs as it takes time until gains are realized (while increased trading costs are contemporaneously measured).

The remaining set of RBPs is based on funds' past returns such as realized Carhart's alpha (henceforth *CA*), the *t*-statistic of Carhart's alpha (henceforth *Tstat*), R-squared (henceforth *R2*), and the one-year volatility of fund returns (henceforth *Vola*). Most of these predictors are derived from the Carhart (1997) four-factor model, which continues to be a prominent factor model within mutual fund research. While Hendricks, Patel, and Zeckhauser (1993) were the first to provide evidence that lagged past returns are associated with future outperformance of actively-managed funds, Carhart (1997) shows that past four-factor

[9] Elton et al. (1993) and Carhart (1997) document a negative turnover-performance relationship; Chen, Jegadeesh, and Wermers (2000) document a positive relationship; Wermers (2000) and Kacperczyk, Sialm, and Zheng (2005) document no significant relation.

alphas improve forecasting of future risk-adjusted performance. Similarly, Elton, Gruber, and Blake (1996) form portfolios of funds based on the *t*-statistic of the three-year alpha. The authors document that high *t*-statistic funds significantly outperform funds with a low alpha *t*-statistic. Amihud and Goyenko (2013) propose *R2* as a fund performance predictor. They document that a lower R^2, derived from the Carhart (1997) four-factor model, is indicative of great security selectivity of the fund manager, and positively and significantly predicts future fund performance. Lastly, Jordan and Riley (2015) find a significant negative relationship between *Vola* and subsequent performance, which is remarkably sizeable in a standard four-factor framework.

2.2 Portfolio Holdings-Based Predictors (HBPs)

The Grinblatt and Titman (1993) performance measure (henceforth *GT*) is one of the first holdings-based measures introduced in the mutual fund literature and the first measure to overcome the benchmark-choice problem of Roll (1978). Later work demonstrates that regression-based performance measurement is sensitive to the benchmark choice. Thus, ranking funds by the regression alpha can be unreliable. The *GT* measure overcomes such issues by proposing the self-benchmarking method, which uses the composition of the fund's portfolio instead of standard factor-based or peer-group benchmarks. Grinblatt and Titman (1993) document evidence of performance persistence, suggesting that *GT*-based past performance is a valid indicator of future performance.

Daniel et al. (1997) incorporate factors that drive stock returns into a holdings-based performance evaluation framework. The authors decompose fund performance using benchmark portfolios matched to funds' holdings based on size, book-to-market, and momentum characteristics and argue that this approach is superior to factor-based regressions. The DGTW holdings-based measures are characteristic selectivity (henceforth *CS*), characteristic timing (henceforth *CT*), and average style. *CS* and *CT* measures can help to detect skilled managers with stock-picking and timing abilities. High *CS* managers successfully pick stocks that outperform their characteristic-based benchmarks, while high *CT* managers successfully capture time-varying expected returns of the corresponding characteristic-based portfolios.

A related measure that captures both market timing and stock picking abilities is the skill index (henceforth *SI*) proposed by Kacperczyk, Nieuwerburgh, and Veldkamp (2014). *SI* can be viewed as a variant of such measures as *GT*, *CS*, and *CT* that measures a fund's ability to capture the systematic and idiosyncratic components of stock returns with respect to the market. Kacperczyk, Nieuwerburgh, and Veldkamp (2014) show that *SI* is a strong and persistent predictor of future fund returns.

Other HBPs are based on measures of portfolio concentration and deviation from a benchmark portfolio such as the industry concentration index (henceforth *ICI*), active share (henceforth *AS*), and active weight (henceforth *AW*). Kacperczyk, Sialm, and Zheng (2005) suggest that manager skills are industry-specific and find that mutual funds that hold concentrated positions in a few industries outperform funds with more diversified holdings. Similarly, *AS* and *AW* measure how much fund holdings deviate from their benchmark or a hypothetical cap-weighted portfolio, respectively. In both cases, higher values are associated with future outperformance, which suggests that there is a subset of skilled "high conviction" managers. Lastly, Cremers and Pareek (2015) develop a fund duration (henceforth *FD*) measure that quantifies the weighted average time a fund is holding its current portfolio, which they later reconsider as a fund performance predictor (Cremers and Pareek (2016)). The authors focus on high active share funds and find that high active share funds with long holdings durations outperform funds that frequently change the portfolio.

Kacperczyk, Sialm, and Zheng (2008) propose the return gap (henceforth *RG*), which combines holdings-based and returns-based performance evaluation (Fischer and Wermers, 2012). *RG* captures the impact of unobserved actions on future fund performance. The authors show that high *RG* funds outperform low *RG* funds.[10] Finally, Huang, Sialm, and Zhang (2011) infer the stability of a fund's risk profile by measuring risk shifting (henceforth *RS*). The authors document that funds that exhibit high levels of risk-shifting behavior underperform funds with a stable risk profile.

[10] Similar evidence is documented by Agarwal, Ruenzi, and Weigert (2023) in the hedge fund literature.

3 Data and Sample Selection

We utilize several datasets to construct our sample. Data on fund returns and fund characteristics come from the Center for Research in Security Prices Mutual Fund (CRSP MF) database. This database provides information on monthly net fund returns and Total Net Assets (TNA), as well as annual expense and turnover ratios. Data on quarterly fund portfolio holdings is retrieved from the Thomson Reuters Mutual Fund Holdings database. We merge holdings data with CRSP MF data using MFLinks files from the Wharton Research Data Services. Finally, we link fund holdings data with the CRSP and the Compustat North America database to obtain additional information for stock returns and accounting variables.

Due to the large scale of our analysis (in which we reconstruct 19 fund predictors), we need to follow a unified sample construction and use it to reconstruct each predictor. Hence, we apply data filters that are common among prominent studies that use holdings data (such as Daniel et al. (1997), Wermers (2000), Kacperczyk, Sialm, and Zheng (2005, 2008), and Doshi, Elkamhi, and Simutin (2015)).

We first merge CRSP MF data with MFLinks and apply common filters from the literature at this stage. We exclude funds that have no fund name and funds that have a non-unique *WFICN* (i.e., when a single *CRSP_FUNDNO* is matched with multiple *WFICN* in the MFLinks file).[11] We also perform a fund name search to identify index and target date funds based on a fund's name. The fund name search is performed following Amihud and Goyenko (2013). On top of the name search, we rely on the CRSP flags for index, target, and exchange-traded funds. We locate these funds and drop them. Following Barber, Odean, and Zheng (2005), funds with zero expense ratios are excluded as this is likely indicative of missing information. We delete incubated funds as in Evans (2010), i.e., we drop the fund if it has observations before the first offer date.[12] CRSP MF provides data on fund styles that total 26 styles (with sub-styles). We keep four major fund styles: *ED*, *EDS*, *EDC*,

[11] *WFICN* is the unique fund identifier in MFLinks, while *CRSP_FUNDNO* is the unique share class identifier. A fund can have multiple share classes, i.e., a single *WFICN* can belong to multiple *CRSP_FUNDNO*, but the reverse is not possible.

[12] A fund's (or a share class's) first offering date is denoted as *first_offer_dt* in CRSP MF.

and *EDY* (where $E = Equity$, $D = Domestic$, $S = Sector$, $C = Cap\text{-}based$, and $Y = Style$). We drop hedged (*EDY-H*) and short (*EDY-S*) style funds. In the last step, we locate and drop share classes that flipped styles following Doshi, Elkamhi, and Simutin (2015). The above list of filters is applied on a fund share class level.

Most funds in our sample have multiple share classes. Once the above-listed filters are applied, we begin to merge share-class level data to arrive at fund-level variables. First, we aggregate qualitative variables (such as the fund name or fund style) by utilizing information of the largest share class in terms of TNA. Next, the first set of quantitative variables (such as the net return, expense, and turnover ratios) are summed and weighted by the lagged TNA of the corresponding share classes. The next set of quantitative variables (such as TNA and NAV) are simply summed up. Lastly, fund age is simply the age of its oldest share class. Once we arrive at the fund level, we drop funds that are below US$ 15 million of TNA as returns on small funds are biased upwards in the CRSP MF database (see, Elton, Gruber, and Blake (2001)). Our final filter at this stage is to delete funds if the expense ratio (on a fund level) is below 0.1% as it is unrealistically low for an actively managed fund and, hence, signifies an index fund.

Once returns and fund characteristic data from the CRSP MF database are aggregated at the fund level, we begin merging MFLinks and the *s12 master file* from Thomson Reuters that contains data on fund stock holdings, and additional fund and stock characteristics.[13] We follow Kacperczyk, Sialm, and Zheng (2005, 2008) and Doshi, Elkamhi, and Simutin (2015), to exclude funds with investment objective codes set to (1) - International, (5) - Municipal Bonds, (6) - Bond and Preferred, (7) - Balanced. We also drop funds with missing *FUNDNO*, delete rows where a stock price or number of shares held by a fund is missing, and erase funds that have less than ten identifiable equity holdings in their portfolio. We take shares held by a fund and multiply them by the corresponding stock price to identify the percentage of the fund's TNA that is invested in equities. Funds are kept in our sample if the 12-month average of this percentage is between 80% and 110%.

[13]We thank Mike Simutin for providing detailed SAS code to replicate the results of Doshi, Elkamhi, and Simutin (2015) on his website (https://www-2.rotman.utoronto.ca/simutin/). We followed some of the steps in their code when constructing our data sample.

To reconcile HBPs from quarterly to monthly frequency, PERMNOs (stock identifiers) are rolled forward up until the next disclosure date or up to six months into the future (whichever comes sooner), and adjusted stock prices are merged from the CRSP-Compustat dataset. This method allows portfolio weights to vary between two consecutive disclosure dates (in a buy-and-hold manner). Since fund holdings are (in 80% of the cases) disclosed quarterly (in March, June, September, and December), this procedure assumes that a typical fund rebalances its portfolio at the end of each quarter, and holds it in between, i.e., buy-and-hold weights.

Our final dataset comprises 4,416 unique U.S. actively managed all-equity funds over the period from 1985 to 2022 and contains 564,082 fund-month observations. We provide detailed summary statistics of our sample in Appendix B.

4 Empirical Methodology

Our procedure for selecting predictors is similar to Jones and Mo (2021), who study the out-of-sample predictability of mutual fund alpha predictors. We review the literature for evidence of a predictive relationship between a variable and future fund performance, which can be measured as gross, net, or risk-adjusted returns. Our main requirement for a predictor to be included is that the predictor must be constructible using standard mutual fund databases such as the CRSP MF and Thomson Reuters Mutual Fund Holdings database, and standard databases for stock returns and accounting data such as the CRSP and the Compustat North America database. In comparison to Jones and Mo (2021), our data requirements are more restrictive. We also do not consider analyst data from the Thomson Reuters IBES global database and benchmark data from Standard and Poor's, FTSE Russell, and Barclays. Table 1 lists alpha predictors covered in our study.

Compared to Jones and Mo (2021), who cover 27 predictors, our study includes 19 predictors, of which 11 are overlapping. Additionally, we incorporate *Age*, *STO*, *Vola*, *Tstat*, *GT*, *CS*, *CT*, and *FD*.[14] We refer to Appendix A for details on the predictors' construction.

We provide summary statistics of the RBPs and HBPs in this study in Table 2. We display, among others, that the average fund charges an annual expense ratio of 1.2%, and a monthly Carhart four-factor alpha of −0.13% (−1.56% p.a.). The average age of a fund is 137 months or 11.4 years; the average size (TNA under management) amounts to 304.9 million US$. Furthermore, the average fund has a mean return gap close to zero, an active share of 44.37%, as well as a fund duration of 14.71. Note that the summary statistics of the RBPs and HBPs closely match the summary statistics of the original papers.

[14] Other predictors covered in Jones and Mo (2021), but not in our study include: Abnormal cash holdings (Simutin, 2014), direct-sold (Bergstresser, Chalmers, and Tufano, 2008), flow-induced trading (Lou, 2012), holdings-based alpha (Elton, Gruber, and Blake, 2011), intangibles (Gupta-Mukherjee, 2014), inverse of diversification (Pollet and Wilson, 2008), one-year return (Hendricks, Patel, and Zeckhauser, 1993), momentum (Grinblatt, Titman, and Wermers, 1995), public info (Kacperczyk and Seru, 2007), growth style (Chan, Chen, and Lakonishok, 2002), one-year Carhart alpha (Carhart, 1997), back-tested alpha (Mamaysky, Spiegel, and Zhang, 2007), active peer benchmarks (Hunter et al., 2014), Pastor-Stambaugh alpha (Busse and Irvine, 2006), and success overlap (Cohen, Coval, and Pástor, 2005).

Table 2: Summary Statistics of RBPs and HBPs, 1985-2022

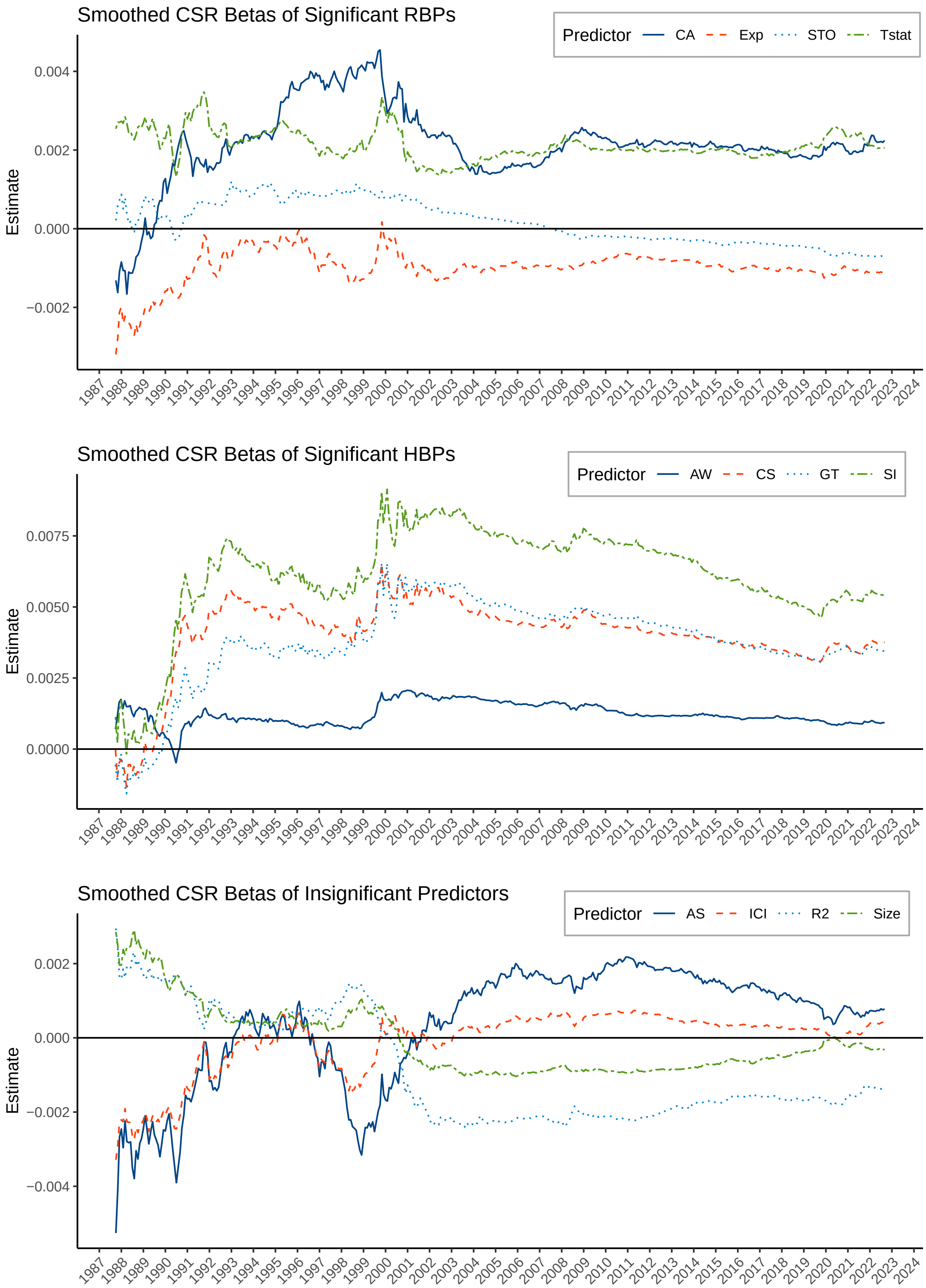
	Mean	SD	Skew	Kurt	Min	5%	25%	Median	75%	95%	Max	N obs.
Fund- and Return-Based Predictors (RBPs)												
Age	4.92	0.91	-0.52	3.28	1.10	3.26	4.36	5.02	5.55	6.27	6.98	549 509
Size	5.72	1.63	0.36	2.62	2.71	3.23	4.46	5.61	6.84	8.61	11.16	549 509
Flows	0.15	5.79	3.49	51.54	-67.50	-5.41	-1.48	-0.44	0.92	7.56	213.05	548 783
Exp	0.10	0.03	0.72	4.18	0.01	0.05	0.08	0.10	0.12	0.16	0.31	501 752
STO	-0.09	1.29	0.58	3.29	-5.41	-1.85	-1.02	-0.33	0.78	2.18	7.18	316 262
Vola	4.85	2.28	1.48	6.94	0.74	2.13	3.20	4.40	5.97	9.10	23.22	465 097
CA	-0.13	2.09	0.26	16.64	-22.19	-3.24	-1.05	-0.12	0.79	2.91	45.04	404 802
Tstat	-0.49	1.52	-0.24	3.80	-7.26	-3.08	-1.39	-0.43	0.46	1.89	8.12	408 957
R2	0.91	0.08	-2.04	7.83	0.51	0.74	0.89	0.94	0.96	0.99	1.00	434 316
Portfolio Holding-Based Predictors (HBPs)												
GT	0.42	3.45	0.04	17.91	-41.04	-4.50	-0.72	0.30	1.60	5.61	64.25	549 509
CS	0.05	1.96	0.86	28.09	-26.86	-2.69	-0.75	0.01	0.82	2.81	46.33	548 396
CT	0.35	2.49	-0.27	10.70	-26.17	-3.36	-0.61	0.29	1.32	4.29	24.42	547 922
RG	-0.01	1.60	-0.12	46.15	-35.42	-1.83	-0.46	-0.02	0.42	1.82	28.89	501 752
AS	44.37	5.54	-1.10	3.76	17.43	33.76	40.78	46.23	49.09	49.85	50.24	549 509
AW	38.59	11.97	0.13	3.54	0.00	21.69	30.50	37.70	46.15	59.47	83.27	549 509
ICI	9.64	16.03	2.89	10.99	0.07	0.57	2.08	3.98	7.69	50.98	92.66	549 452
RS	0.31	1.06	2.10	25.26	-8.16	-0.94	-0.08	0.21	0.60	1.80	12.99	419 268
SI	0.33	2.59	-0.34	7.96	-24.80	-3.96	-0.93	0.47	1.72	4.13	32.25	544 920
FD	14.71	8.94	0.95	3.87	0.00	3.00	8.10	13.14	19.54	32.08	50.36	545 094

This table reports summary statistics of nine RBPs and ten HBPs considered in our study. All variables are winsorized at 0.5% on both tails on a period-by-period basis. The average age of a fund is 137 months or 11.4 years; the average size (TNA under management) is 304.9 million US$. The average holding duration (FD) is 14.7 months; the average AS, AW, and ICI of mutual funds are 44%, 39%, and 9.64%, respectively.

We propose an intuitive way of aggregating individual fund predictors to separate outperforming from underperforming mutual funds. Unlike the typical approach in the fund predictors literature, which ranks funds based on a single variable (such as age or past alphas), our approach constructs a composite predictor by averaging rankings across multiple predictors. In essence, the method is similar to the construction of the mispricing score of Stambaugh and Yuan (2017), and the quality score of Asness, Frazzini, and Pedersen (2019). A notable difference in our method is that prior to averaging (i.e., aggregating individual predictors), we identify the expected sign of the relationship between a predictor and the future fund return. Figure 1 shows smoothed slopes computed using Equations (1) and (2), which provides evidence that for many predictors the sign frequently varies over time, and hence, a static sign is not optimal.

Figure 1: Smoothed Slope Estimates of Predictors, 1988-2022

This figure plots loadings from univariate weighted-least-squares (WLS) cross-sectional regressions that are estimated with an expanding window starting from 36 months. First, the dependent variable is R_t^e regressed on a lagged standardized predictor x_{t-1}, as in Equation (1). Second, these beta estimates are then smoothed using an expanding window average starting from 36 months.



Instead of taking a sign motivated by theory or the one reported in the original study, we choose an entirely data-driven approach, which is also implementable as a real-life trading strategy. First, theoretical motivation is not always available, or there exists contradictory empirical evidence. Second, empirical results reported in the original studies are likely to introduce a look-ahead bias since all considered studies share overlapping sample periods with our research. Overall, the empirical simplicity of our composite measure allows us to easily capture time-varying relationships between past individual predictors and future fund returns.

We follow Kelly, Pruitt, and Su (2019) to standardize predictors in a cross-section. Each month, we calculate funds' ranks for each predictor, then divide the ranks by the number of non-missing observations and subtract 0.5. This transforms predictors into a uniform distribution on a $[-0.5, +0.5]$ interval. This standardization prioritizes ordering and ignores the magnitude of observations, which makes it insensitive to outliers. The distribution is centered at zero, which comes in handy when imputing missing values with a cross-sectional mean without any forward-looking bias. The interval length from -0.5 to $+0.5$ also makes it convenient to reverse the distribution by multiplying it by -1 when low (high) values of a predictor are associated with high (low) future returns.

We compute a composite predictor in four steps. First, we run a standard univariate weighted-least-squares (WLS) cross-sectional regression,

$$R^e_{i,t} = \alpha + \beta X_{i,t-1} + \epsilon_{i,t}, \qquad i = 1, 2, \ldots N, \tag{1}$$

where $R^e_{i,t}$ is the return (net of all management expenses and 12b-1 fees) on fund i in month t in excess of the risk-free rate, and $X_{i,t-1}$ is a lagged standardized predictor available in month $t-1$. We use a WLS regression specification to increase the importance of larger funds. For WLS, we use the following square-root form,

$$w_i = \frac{\sqrt{TNA_{i,t-1}}}{\sum_{i=1}^{N} \sqrt{TNA_i}},$$

where $TNA_{i,t-1}$ is the lagged TNA of fund i. Then, we compute smoothed beta estimates using an expanding window average starting from 36 months. We utilize these smoothed estimates ($\bar{\beta}_t$) to determine the expected sign of a given predictor, i.e., implicitly assuming the following relationship between the future fund return and a predictor variable:

$$E[R^e_{i,t+1}] = \bar{\beta}_t X_{i,t}. \tag{2}$$

Second, we multiply the standardized predictor by $+1$ or -1 depending on the sign of the $\bar{\beta}$ from (2). Note that $\bar{\beta}$ is estimated through period $t-1$ to eliminate a potential look-ahead bias. Third, we equal-weight the individual sign-adjusted forecasts to arrive at a composite return prediction for each fund in month $t+1$. This ensures that all signals receive equal weight in the composite score. Finally, we again standardize this aggregate predictor to be uniformly distributed over the $[-0.5, +0.5]$ interval.

Table 3 displays pairwise correlations between the different predictor variables, which provide preliminary indications of their relationships. We find that, on average, most of the variables exhibit correlations that are close to zero or very low in magnitude. This indicates that different predictors are likely to contain alternative information content that one can use to predict future mutual fund performance.

Table 3: Correlations of RBPs and HBPs, 1985-2022, monthly returns

	Age	Size	Flows	Exp	STO	Vola	CA	Tstat	R2	GT	CS	CT	RG	AS	AW	ICI	RS	SI
Age	.	0.41	-0.15	-0.19	0.04	-0.03	-0.00	-0.03	0.06	-0.03	0.00	-0.04	-0.01	-0.14	0.02	-0.03	-0.01	0.00
Size	0.41	.	0.00	-0.39	-0.03	-0.04	0.02	0.15	0.17	-0.02	0.01	-0.04	-0.01	-0.21	-0.07	-0.06	-0.00	0.00
Flows	-0.14	-0.00	.	-0.05	-0.06	-0.04	-0.01	0.25	0.01	-0.01	-0.01	-0.01	-0.00	0.01	-0.01	-0.01	0.03	-0.02
Exp	-0.18	-0.40	0.00	.	-0.01	0.20	-0.02	-0.07	-0.26	0.03	-0.00	0.02	0.01	0.33	0.08	0.21	0.00	0.01
STO	0.04	-0.03	-0.04	0.00	.	0.00	-0.01	-0.10	0.00	0.02	-0.01	0.03	0.00	0.00	0.01	-0.02	0.00	-0.01
Vola	-0.04	-0.05	-0.02	0.20	0.00	.	-0.03	0.02	-0.13	0.02	-0.01	0.01	0.01	0.45	-0.02	0.28	-0.07	0.01
CA	-0.01	0.02	0.01	-0.02	-0.01	-0.03	.	0.17	-0.00	0.32	0.58	0.14	0.07	-0.00	0.01	0.01	0.03	0.53
Tstat	-0.04	0.15	0.16	-0.08	-0.11	0.03	0.16	.	-0.12	0.03	0.10	-0.01	0.02	0.10	0.02	0.13	0.07	0.10
R2	0.06	0.14	-0.01	-0.21	-0.00	-0.12	-0.01	-0.10	.	-0.02	-0.01	-0.02	-0.00	-0.45	-0.16	-0.47	-0.08	-0.04
GT	-0.04	-0.03	0.00	0.03	0.02	0.02	0.34	0.03	-0.01	.	0.45	0.61	-0.21	0.02	0.00	-0.01	0.01	0.58
CS	0.00	0.01	0.01	-0.00	-0.01	-0.01	0.60	0.09	-0.00	0.50	.	0.08	-0.31	-0.00	0.00	0.00	0.03	0.74
CT	-0.05	-0.04	0.00	0.02	0.03	0.00	0.14	-0.01	-0.01	0.61	0.08	.	-0.10	0.01	0.00	-0.01	0.02	0.32
RG	-0.02	-0.02	0.00	0.02	0.00	0.01	0.12	0.02	0.00	-0.29	-0.34	-0.12	.	0.02	0.00	0.01	-0.02	-0.31
AS	-0.12	-0.20	0.02	0.30	-0.00	0.40	-0.00	0.11	-0.36	0.03	0.00	0.02	0.02	.	-0.10	0.43	0.08	0.04
AW	0.02	-0.07	0.00	0.10	0.01	0.01	0.01	0.02	-0.13	0.00	0.00	-0.00	0.00	0.08	.	0.13	0.04	0.00
ICI	-0.02	-0.06	0.01	0.17	-0.01	0.27	0.00	0.09	-0.58	-0.00	-0.00	-0.00	0.01	0.29	-0.01	.	0.10	0.02
RS	-0.01	-0.00	0.02	0.01	0.00	-0.07	0.03	0.07	-0.15	0.01	0.02	0.02	-0.01	0.06	-0.02	0.17	.	0.05
SI	0.00	0.01	0.01	0.01	-0.01	-0.00	0.56	0.09	-0.02	0.62	0.79	0.31	-0.36	0.04	0.00	0.00	0.04	.
FD	0.44	0.24	-0.09	-0.23	-0.05	-0.20	0.01	0.05	0.04	-0.11	-0.00	-0.13	-0.02	-0.22	0.01	0.02	0.02	-0.01

This table reports the time-series averages of the monthly cross-sectional correlations between pairs of variables. Below-diagonal entries present the average Pearson product-moment correlations. Above-diagonal entries present the average Spearman rank correlations. We compute correlations in two steps. First, we calculate the cross-sectional correlation between each pair of variables, for each month t. Second, we take the time-series average of these cross-sectional correlations. All variables are measured contemporaneously and winsorized at 0.5% on both tails. Winsorization is performed on a month-by-month basis.

5 Empirical Results: Combining Predictors

In this section, we present the first set of our main results. To do so, we report the univariate performance of the 19 predictor variables considered in this study. We will later combine these predictors into composite scores.

5.1 Univariate performance

Like Jones and Mo (2021), our objective is not to replicate previous research findings but to reconstruct prominent predictors from the literature within a consistent framework. This approach enables us to combine individual predictors into a composite predictor of fund performance. Given this focus, it is useful to first compare our results for individual predictors with the in-sample alpha spreads reported by Jones and Mo (2021) (pg. 160, Table 1: Alpha spreads in and out of sample). While direct one-to-one comparisons are not possible due to differences in sample construction and the separate reporting of in-sample and out-of-sample spreads by Jones and Mo (2021), a reasonable degree of comparability with our findings can still be expected.

Table 4 shows the performance of individual predictors. For this purpose, we sort the funds in our sample according to a respective predictor variable in month t. We then report the average high-minus-low spread of decile portfolios in month $t+1$. As measures of performance, we utilize excess returns, as well as the intercept from a time-series regression that includes the Fama–French-Carhart four-factor model (α^{FF4}) of Carhart (1997), the Fama–French six-factor model (α^{FF6}) of Fama and French (2018), and the five-factor q^5 model (α^{Q5}) of Hou et al. (2021). Results are reported based on an equal-weighted and value-weighted weighting scheme. For a set of predictors that overlap with Jones and Mo (2021), results for 10 out of 11 predictors agree. Similar to Jones and Mo (2021), we find positive and statistically significant spreads for *CA*, *AW*, and *SI*, significantly negative spreads for *Exp* and *STO*, as well as marginally significant results or no evidence of significance for *AS*, *ICI*, *Size*, *Flows*, and *R2*. *RG* is not significant in our sample, while Jones and Mo (2021) document significantly positive in-sample and out-of-sample alpha spreads.

Table 4: Univariate Performance: Long-Short Decile Portfolios, 1985-2022, monthly returns

	Equal-Weighted				Value-Weighted			
	R^e	α^{FF4}	α^{FF6}	α^{Q5}	R^e	α^{FF4}	α^{FF6}	α^{Q5}
Fund- and Return-Based Predictors (RBPs)								
Age	0.07	0.07*	0.03	-0.03	0.08	0.11*	0.04	-0.07
	[1.52]	[1.67]	[0.87]	[-0.78]	[1.38]	[1.90]	[0.79]	[-1.45]
Size	-0.03	-0.04	0.02	0.02	-0.01	-0.02	0.04	0.02
	[-0.56]	[-1.00]	[0.59]	[0.47]	[-0.20]	[-0.35]	[0.95]	[0.54]
Flows	0.12	0.07	0.12	0.12	0.20*	0.13	0.20*	0.19
	[1.41]	[0.91]	[1.52]	[1.38]	[1.72]	[1.34]	[1.85]	[1.54]
Exp	-0.09	-0.13***	-0.03	0.05	-0.15	-0.21***	-0.04	0.10
	[-1.52]	[-3.57]	[-1.09]	[1.00]	[-1.49]	[-2.75]	[-0.64]	[1.21]
STO	-0.14***	-0.16***	-0.15***	-0.15***	-0.12**	-0.15***	-0.21***	-0.20***
	[-3.03]	[-3.87]	[-3.46]	[-3.26]	[-2.38]	[-2.78]	[-3.57]	[-3.23]
Vola	0.00	-0.21	0.11	0.29*	-0.01	-0.21	0.12	0.27*
	[0.00]	[-1.47]	[0.89]	[1.86]	[-0.04]	[-1.41]	[0.96]	[1.75]
CA	0.21**	0.23**	0.22**	0.20*	0.24**	0.27***	0.27***	0.28***
	[2.18]	[2.44]	[2.14]	[1.95]	[2.33]	[2.78]	[2.62]	[2.62]
Tstat	0.20**	0.24***	0.33***	0.31***	0.23***	0.26***	0.35***	0.34***
	[2.37]	[3.51]	[4.62]	[4.18]	[2.64]	[3.68]	[4.59]	[4.56]
R2	-0.09	-0.08	-0.03	-0.00	-0.07	-0.04	-0.01	0.01
	[-1.04]	[-1.01]	[-0.44]	[-0.03]	[-0.75]	[-0.39]	[-0.06]	[0.09]
Portfolio Holding-Based Predictors (HBPs)								
GT	0.33***	0.37***	0.31***	0.23*	0.25***	0.31***	0.26**	0.21*
	[3.39]	[4.02]	[2.97]	[1.87]	[2.73]	[3.41]	[2.50]	[1.78]
CS	0.26***	0.29***	0.27***	0.18**	0.23**	0.26***	0.24**	0.15
	[2.81]	[3.43]	[3.01]	[2.03]	[2.30]	[2.77]	[2.36]	[1.55]
CT	0.15*	0.17*	0.13	0.09	0.13	0.15	0.12	0.08
	[1.72]	[1.93]	[1.29]	[0.64]	[1.35]	[1.47]	[0.97]	[0.56]
RG	0.05	0.07	0.08*	0.05	-0.02	0.02	0.03	0.06
	[1.14]	[1.48]	[1.75]	[1.03]	[-0.28]	[0.29]	[0.45]	[0.73]
AS	0.05	-0.00	-0.09	-0.02	-0.04	-0.05	-0.17*	-0.09
	[0.36]	[-0.01]	[-1.26]	[-0.19]	[-0.27]	[-0.51]	[-1.85]	[-0.81]
AW	0.11*	0.09*	0.15***	0.15**	0.18**	0.16**	0.23***	0.24***
	[1.74]	[1.79]	[2.74]	[2.32]	[2.12]	[2.32]	[2.80]	[2.72]
ICI	0.06	0.07	0.10	0.16*	-0.07	-0.04	0.10	0.20*
	[0.74]	[0.88]	[1.34]	[1.94]	[-0.75]	[-0.38]	[1.05]	[1.75]
RS	0.04	0.05	0.08	0.09	0.06	0.07	0.10	0.17*
	[0.58]	[0.71]	[1.31]	[1.42]	[0.74]	[0.84]	[1.22]	[1.88]
SI	0.36***	0.45***	0.38***	0.25**	0.30***	0.39***	0.31***	0.26*
	[3.68]	[4.36]	[3.56]	[2.35]	[3.15]	[3.67]	[2.98]	[1.88]
FD	0.01	0.10*	-0.05	-0.14**	0.11	0.18**	0.03	-0.11
	[0.17]	[1.92]	[-1.10]	[-2.12]	[0.94]	[2.12]	[0.34]	[-1.13]

This table reports univariate analysis of RBPs and HBPs. For each predictor, we report a high-minus-low spread: the excess return (R^e) of the top-decile minus the low-decile equal- or value-weighted portfolio, or, accordingly, the intercept from a time-series regression that includes the Fama–French-Carhart four-factor model (α^{FF4}) of Carhart (1997), the Fama–French six-factor model (α^{FF6}) of Fama and French (2018), and the five-factor q^5 model (α^{Q5}) of Hou et al. (2021). t-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

5.2 Combining RBPs and HBPs separately

When considering individual performance, RBPs such as *CA*, *Exp*, *STO*, *Tstat*, and HBPs such as *AW*, *CS*, *GT*, and *SI* are significantly associated with subsequent fund performance. Building on these results, we standardize predictors as described in Section 4 and run a standard univariate WLS cross-sectional regression for each predictor as depicted in Equations (1) and (2). What is important for our methodology is the sign of the relationship between a given predictor and its future return in the subsequent month, as depicted in Figure 1.

5.2.1 RBP-composite score

We first combine nine RBPs into a composite score. Table 5 shows the performance of the RBP-composite score. Our main result is that the RBP-composite score, composed in month t, predicts a value-weighted monthly six-factor alpha spread of 0.23% (2.76% p.a., $t = 2.96$) in month $t+1$. The corresponding monthly q^5 alpha spread amounts to 0.27% (3.24% p.a., $t = 3.65$). Hence, both alpha spreads are statistically significant at the 1% level and economically sizeable. When we consider an alternative factor model and an equal-weighted scheme, the results remain significant. For instance, the equal-weighted monthly q^5 alpha spread is 0.20% (2.40% p.a.) with a t-statistic of 3.21.

We further decompose these alpha spreads into the long- and short-side. Such an analysis is particularly valuable as funds cannot be shorted in reality. We evaluate how low composite score funds and high composite score funds perform against the aggregate portfolio of active mutual funds in our sample, which is denoted as AF in Tables 5, 6, and 7. Starting with low-score funds, we find that the RBP-composite picks funds that underperform the average fund annually by −1.32% ($t = 2.58$; $t = 2.08$). Moving to the high-score funds, the RBP-composite selects funds that outperform the average fund by 1.32% ($t = 2.58$, α^{FF6}) or 2.04% ($t = 2.50$, α^{Q5}) p.a. Notably, our results indicate that the best (worst) funds significantly beat (underperform) the performance of an aggregate portfolio of passive mutual funds in our sample, denoted as PF.

Table 5: RBP-Composite Predictor Decile Portfolios, 1988-2022, monthly returns

	Equal-Weighted				Value-Weighted			
Portfolio	R^e	α^{FF4}	α^{FF6}	α^{Q5}	R^e	α^{FF4}	α^{FF6}	α^{Q5}
Decile Portfolios								
Low (L)	0.57**	-0.13**	-0.14***	-0.14***	0.52**	-0.17***	-0.18***	-0.17***
	[2.57]	[-2.23]	[-3.01]	[-3.90]	[2.32]	[-3.00]	[-3.59]	[-4.08]
2	0.58***	-0.13**	-0.13***	-0.11***	0.54**	-0.14***	-0.13***	-0.10***
	[2.59]	[-2.42]	[-3.05]	[-3.35]	[2.43]	[-2.82]	[-2.82]	[-2.79]
3	0.59***	-0.12***	-0.12***	-0.10***	0.54**	-0.14***	-0.15***	-0.13***
	[2.67]	[-2.99]	[-3.20]	[-3.12]	[2.48]	[-2.85]	[-3.72]	[-3.26]
4	0.63***	-0.10**	-0.10**	-0.08**	0.60***	-0.10**	-0.10**	-0.11**
	[2.81]	[-2.17]	[-2.34]	[-2.47]	[2.72]	[-2.23]	[-2.13]	[-2.46]
5	0.63***	-0.09**	-0.08*	-0.07*	0.55**	-0.14***	-0.12***	-0.12***
	[2.82]	[-1.97]	[-1.91]	[-1.90]	[2.46]	[-3.66]	[-3.30]	[-3.35]
6	0.64***	-0.09**	-0.08**	-0.05	0.59***	-0.10***	-0.09**	-0.08**
	[2.86]	[-2.26]	[-2.38]	[-1.44]	[2.67]	[-2.59]	[-2.43]	[-2.07]
7	0.64***	-0.08*	-0.09**	-0.07**	0.58**	-0.11**	-0.11**	-0.11***
	[2.88]	[-1.95]	[-2.36]	[-2.03]	[2.53]	[-2.35]	[-2.44]	[-2.69]
8	0.67***	-0.07	-0.06	-0.04	0.64***	-0.07*	-0.06	-0.05
	[3.00]	[-1.43]	[-1.48]	[-1.12]	[2.85]	[-1.78]	[-1.45]	[-1.19]
9	0.70***	-0.05	-0.03	-0.00	0.64***	-0.06	-0.05	-0.05
	[3.06]	[-1.15]	[-0.83]	[-0.03]	[2.84]	[-1.29]	[-1.07]	[-1.11]
High (H)	0.75***	-0.02	0.02	0.06	0.71***	-0.02	0.04	0.09
	[3.22]	[-0.40]	[0.38]	[1.19]	[2.87]	[-0.38]	[0.68]	[1.47]
High-minus-low Spread								
H-L	0.18**	0.11*	0.16***	0.20***	0.19**	0.15*	0.23***	0.27***
	[2.51]	[1.65]	[2.59]	[3.21]	[2.17]	[1.85]	[2.96]	[3.65]
Comparison with the Average Active Fund (AF)								
AF	0.64***	-0.09**	-0.10***	-0.10***	0.62***	-0.09***	-0.07**	-0.08***
	[2.87]	[-2.13]	[-4.59]	[-3.55]	[2.75]	[-3.07]	[-2.42]	[-3.21]
H - AF	0.11***	0.07*	0.12***	0.16***	0.09*	0.07	0.11**	0.17**
	[2.61]	[1.81]	[2.79]	[2.57]	[1.81]	[1.44]	[2.42]	[2.50]
L - AF	-0.07**	-0.04	-0.04	-0.04	-0.10**	-0.08*	-0.11**	-0.10**
	[-2.03]	[-1.25]	[-0.98]	[-1.04]	[-1.98]	[-1.66]	[-2.58]	[-2.08]
Comparison with the Average Passive Fund (PF)								
PF	0.65***	-0.07***	-0.08**	-0.06*	0.67***	-0.03	-0.06***	-0.06**
	[2.95]	[-2.60]	[-2.29]	[-1.95]	[3.11]	[-1.47]	[-3.39]	[-2.15]
H - PF	0.10	0.05	0.10***	0.12***	0.04	0.00	0.11*	0.15**
	[1.62]	[1.06]	[2.84]	[3.12]	[0.55]	[0.07]	[1.70]	[3.21]
L - PF	-0.07*	-0.07	-0.06*	-0.08**	-0.14***	-0.15***	-0.12**	-0.11***
	[-1.78]	[-1.48]	[-1.91]	[-2.69]	[-2.83]	[-2.69]	[-2.36]	[-2.89]

This table reports the risk-adjusted performance of decile-portfolios sorted on the RBP-composite predictor. The composite RBP is composed of nine RBPs. The top panel reports the performance of decile-portfolios and a high-minus-low spread. The bottom panel compares the performance of high- and low-decile portfolios to the average fund. The columns of the table represent the excess return (R^e) of an equal- or value-weighted portfolio or, accordingly, the intercept from a time-series regression that includes the Fama–French-Carhart four-factor model (α^{FF4}) of Carhart (1997), the Fama–French six-factor model (α^{FF6}) of Fama and French (2018), and the five-factor q^5 model (α^{Q5}) of Hou et al. (2021). t-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

5.2.2 HBP-composite score

We repeat the analysis for our sample of HBP-predictor variables. Table 6 reports the performance of the HBP-composite score. The HBP-composite score predicts a monthly six-factor alpha spread of 0.29% (3.48% p.a., $t = 2.21$). The q^5 alpha spread amounts to 3.48% p.a. but with slightly lower statistical significance ($t = 1.81$). The alphas are slightly larger and increase in statistical significance for other specifications. Starting with low-score funds, we find that the HBP-composite selects funds that underperform the average AF monthly by −0.19% (−2.28% p.a., $t = -2.70$) or −0.16% (-1.92% p.a., $t = -2.06$) depending on the factor model. When comparing high-score funds against the average AF, the results differ for equal- and value-weighting schemes. Value-weighted results are insignificant, while equal-weighted results show the high-score funds outperform the average fund by 0.16% (1.92% p.a., $t = 2.14$) or 0.19% (2.28% p.a., $t = 1.96$) depending on the factor model. These results remain quantitatively similar when benchmarking against the average PF.

Table 6: HBP-Composite Predictor Decile Portfolios, 1988-2022, monthly returns

	Equal-Weighted				**Value-Weighted**			
Portfolio	R^e	α^{FF4}	α^{FF6}	α^{Q5}	R^e	α^{FF4}	α^{FF6}	α^{Q5}
Decile Portfolios								
Low (L)	0.47**	-0.26***	-0.26***	-0.23***	0.40*	-0.31***	-0.26***	-0.24***
	[2.06]	[-3.93]	[-3.79]	[-2.89]	[1.68]	[-4.24]	[-3.59]	[-2.99]
2	0.55**	-0.17***	-0.17***	-0.15***	0.56**	-0.15***	-0.12**	-0.09
	[2.41]	[-3.49]	[-3.36]	[-2.63]	[2.41]	[-3.13]	[-2.25]	[-1.41]
3	0.58**	-0.15***	-0.14***	-0.11**	0.57**	-0.14***	-0.11***	-0.07
	[2.56]	[-3.52]	[-3.38]	[-2.49]	[2.52]	[-3.50]	[-2.74]	[-1.49]
4	0.59***	-0.14***	-0.14***	-0.11***	0.57**	-0.15***	-0.13***	-0.11***
	[2.59]	[-3.39]	[-3.85]	[-2.92]	[2.47]	[-3.61]	[-3.32]	[-2.72]
5	0.62***	-0.11**	-0.10**	-0.07*	0.61***	-0.09**	-0.07*	-0.04
	[2.72]	[-2.45]	[-2.52]	[-1.91]	[2.72]	[-2.11]	[-1.73]	[-1.28]
6	0.65***	-0.08*	-0.07*	-0.04	0.64***	-0.06	-0.04	-0.01
	[2.92]	[-1.76]	[-1.65]	[-1.02]	[2.87]	[-1.53]	[-0.98]	[-0.30]
7	0.70***	-0.03	-0.03	-0.01	0.68***	-0.04	-0.03	-0.02
	[3.20]	[-0.59]	[-0.59]	[-0.14]	[3.05]	[-0.97]	[-0.61]	[-0.40]
8	0.71***	-0.01	-0.01	-0.00	0.71***	0.02	0.04	0.02
	[3.24]	[-0.20]	[-0.15]	[-0.01]	[3.28]	[0.43]	[0.72]	[0.40]
9	0.75***	0.02	0.02	0.02	0.76***	0.04	0.04	0.05
	[3.31]	[0.37]	[0.26]	[0.28]	[3.23]	[0.47]	[0.59]	[0.55]
High (H)	0.76***	0.05	0.06	0.10	0.70***	0.00	0.03	0.06
	[3.23]	[0.68]	[0.79]	[1.04]	[2.94]	[0.00]	[0.34]	[0.55]
High-minus-low Spread								
H-L	0.29**	0.31***	0.32**	0.32**	0.31**	0.31**	0.29**	0.29*
	[2.35]	[2.64]	[2.56]	[2.08]	[2.33]	[2.53]	[2.21]	[1.81]
Comparison with the Average Active Fund (AF)								
AF	0.64***	-0.09**	-0.10***	-0.10***	0.62***	-0.09***	-0.07**	-0.08***
	[2.87]	[-2.13]	[-4.59]	[-3.55]	[2.75]	[-3.07]	[-2.42]	[-3.21]
H - AF	0.12*	0.14**	0.16**	0.19*	0.09	0.09	0.10	0.13
	[1.89]	[2.33]	[2.14]	[1.96]	[1.15]	[1.46]	[1.32]	[1.21]
L - AF	-0.17***	-0.17***	-0.16**	-0.13*	-0.22***	-0.22***	-0.19***	-0.16**
	[-2.62]	[-2.70]	[-2.39]	[-1.68]	[-3.30]	[-3.06]	[-2.70]	[-2.06]
Comparison with the Average Passive Fund (PF)								
PF	0.65***	-0.07***	-0.08**	-0.06*	0.67***	-0.03	-0.06***	-0.06**
	[2.95]	[-2.60]	[-2.29]	[-1.95]	[3.11]	[-1.47]	[-3.39]	[-2.15]
H - PF	0.11	0.12	0.14**	0.15*	0.04	0.03	0.09	0.12
	[1.36]	[1.64]	[2.30]	[1.89]	[0.39]	[0.37]	[1.07]	[1.24]
L - PF	-0.17***	-0.19***	-0.18***	-0.17**	-0.27***	-0.28***	-0.20***	-0.17**
	[-2.73]	[-3.08]	[-2.60]	[-2.13]	[-3.80]	[-3.86]	[-2.73]	[-2.24]

This table reports the risk-adjusted performance of decile-portfolios sorted on the HBP-composite predictor. The composite HBP is composed of ten HBPs. The top panel reports the performance of decile-portfolios and a high-minus-low spread. The bottom panel compares the performance of high- and low-decile portfolios to the average fund. The columns of the table represent the excess return (R^e) of an equal- or value-weighted portfolio or, accordingly, the intercept from a time-series regression that includes the Fama–French-Carhart four-factor model (α^{FF4}) of Carhart (1997), the Fama–French six-factor model (α^{FF6}) of Fama and French (2018), and the five-factor q^5 model (α^{Q5}) of Hou et al. (2021). t-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

5.3 Combining all predictors

Table 7 displays the performance of the aggregate composite score. It shows that combining RBPs and HBPs into a unified measure of fund performance considerably improves risk-adjusted performance both economically and statistically. When considered separately, the RBP-composite score and the HBP-composite score predict a monthly six-factor alpha spread in future fund performance of 0.23% (2.76% p.a., $t = 2.96$) and 0.29% (3.48% p.a., $t = 2.21$), respectively. In comparison, the aggregate score predicts a six-factor alpha spread in future fund performance by 0.38% (4.56% p.a., $t = 3.23$). This result remains consistent across alternative specifications, such as evaluating different alphas and weighting schemes.

When decomposing these alpha spreads into the long and short portfolios, our results reveal that the improvement in the degree of predictability (in both economic and statistical dimensions) is primarily driven by the long side of the strategy.

For instance, the RBP-composite, HBP-composite, and the *All*-composite select funds that outperform the average active fund monthly by 0.11% (1.32% p.a., $t = 2.58$), 0.10% (1.20% p.a., $t = 1.32$), and 0.16% (1.92% p.a., $t = 2.85$), respectively. For the low-score funds, our results indicate that the RBP-composite, HBP-composite, and the *All*-composite select funds that underperform the average active fund monthly by 0.11% (1.32% p.a., $t = 2.58$), 0.19% (2.28% p.a., $t = 2.70$), and 0.22% (2.64% p.a., $t = 2.91$). Again, this result is practically important as mutual funds cannot be shorted in practice.

To put our findings in perspective with related work using machine learning algorithms for mutual fund selection: Kaniel et al. (2023) report a monthly four-factor alpha spread of 0.48% (5.76% p.a., $t = 5.20$), and DeMiguel et al. (2023) report a statistically significant long-only six-factor alpha of 2.4% p.a. In comparison, we document a six-factor alpha spread of 0.38% (4.56% p.a., $t = 3.23$). In contrast to DeMiguel et al. (2023), we do not find a significant long-only alpha. However, when we benchmark the long-only portfolio against the average *active* and *passive* mutual funds in our sample, our method achieves significant long-only six-factor alphas of 0.16% (1.92% p.a., $t = 2.85$), and 0.15% (1.80% p.a., $t = 2.17$), respectively.

Table 7: *All*-composite Predictor Decile Portfolios, 1988-2022, monthly returns

	Equal-Weighted				Value-Weighted			
Portfolio	R^e	α^{FF4}	α^{FF6}	α^{Q5}	R^e	α^{FF4}	α^{FF6}	α^{Q5}
Decile Portfolios								
Low (L)	0.49**	-0.22***	-0.23***	-0.21***	0.40*	-0.30***	-0.29***	-0.26***
	[2.19]	[-3.42]	[-3.73]	[-3.05]	[1.73]	[-3.82]	[-3.85]	[-3.15]
2	0.55**	-0.18***	-0.18***	-0.16***	0.48**	-0.23***	-0.22***	-0.19***
	[2.42]	[-3.53]	[-3.53]	[-2.93]	[2.04]	[-3.97]	[-3.66]	[-3.01]
3	0.56**	-0.17***	-0.16***	-0.14***	0.54**	-0.15***	-0.13***	-0.13**
	[2.46]	[-3.89]	[-4.09]	[-3.52]	[2.38]	[-3.79]	[-3.16]	[-2.52]
4	0.59***	-0.13***	-0.13***	-0.11***	0.55**	-0.14***	-0.13***	-0.10*
	[2.66]	[-2.91]	[-3.17]	[-2.75]	[2.50]	[-3.33]	[-2.82]	[-1.90]
5	0.60***	-0.12***	-0.12***	-0.10***	0.52**	-0.18***	-0.18***	-0.13***
	[2.68]	[-2.88]	[-3.34]	[-2.91]	[2.31]	[-4.19]	[-4.09]	[-2.91]
6	0.66***	-0.07*	-0.07*	-0.04	0.63***	-0.07*	-0.05	-0.04
	[2.93]	[-1.75]	[-1.69]	[-1.09]	[2.78]	[-1.80]	[-1.35]	[-1.03]
7	0.70***	-0.03	-0.02	-0.01	0.65***	-0.05	-0.03	-0.01
	[3.12]	[-0.62]	[-0.48]	[-0.15]	[2.86]	[-1.15]	[-0.63]	[-0.15]
8	0.73***	-0.01	-0.00	0.02	0.67***	-0.04	-0.01	0.01
	[3.25]	[-0.27]	[-0.03]	[0.34]	[2.93]	[-0.90]	[-0.27]	[0.24]
9	0.73***	-0.00	0.00	0.02	0.69***	-0.02	0.02	0.05
	[3.20]	[-0.07]	[0.06]	[0.40]	[2.99]	[-0.24]	[0.31]	[0.65]
High (H)	0.79***	0.06	0.08	0.12	0.76***	0.05	0.09	0.12
	[3.35]	[0.82]	[1.03]	[1.36]	[3.22]	[0.68]	[1.31]	[1.47]
High-minus-low Spread								
H-L	0.30***	0.28**	0.30***	0.33**	0.36***	0.34***	0.38***	0.38***
	[2.62]	[2.52]	[2.68]	[2.34]	[3.01]	[2.85]	[3.23]	[2.63]
Comparison with the Average Active Fund (AF)								
AF	0.64***	-0.09**	-0.10***	-0.10***	0.62***	-0.09***	-0.07**	-0.08***
	[2.87]	[-2.13]	[-4.59]	[-3.55]	[2.75]	[-3.07]	[-2.42]	[-3.21]
H - AF	0.15**	0.15**	0.18**	0.21**	0.15**	0.14**	0.16***	0.19**
	[2.39]	[2.57]	[2.50]	[2.32]	[2.45]	[2.47]	[2.85]	[2.21]
L - AF	-0.15***	-0.13**	-0.13**	-0.11**	-0.22***	-0.20***	-0.22***	-0.19**
	[-2.60]	[-2.18]	[-2.15]	[-1.68]	[-2.96]	[-2.65]	[-2.91]	[-2.31]
Comparison with the Average Passive Fund (PF)								
PF	0.65***	-0.07***	-0.08**	-0.06*	0.67***	-0.03	-0.06***	-0.06**
	[2.95]	[-2.60]	[-2.29]	[-1.95]	[3.11]	[-1.47]	[-3.39]	[-2.15]
H - PF	0.14*	0.12*	0.16***	0.18**	0.10	0.08	0.15**	0.18**
	[1.75]	[1.87]	[2.65]	[2.33]	[1.15]	[1.05]	[2.17]	[2.54]
L - PF	-0.16***	-0.15**	-0.15**	-0.15**	-0.27***	-0.27***	-0.23***	-0.20**
	[-2.80]	[-2.58]	[-2.42]	[-2.14]	[-3.76]	[-3.62]	[-3.03]	[-2.34]

This table reports the risk-adjusted performance of decile-portfolios sorted on the *All*-composite predictor. The composite predictor is composed of all 19 predictors considered in this paper. The top panel reports the performance of decile-portfolios and a high-minus-low spread. The bottom panel compares the performance of high- and low-decile portfolios to the average fund. The columns of the table represent the excess return (R^e) of an equal- or value-weighted portfolio or, accordingly, the intercept from a time-series regression that includes the Fama–French-Carhart four-factor model (α^{FF4}) of Carhart (1997), the Fama–French six-factor model (α^{FF6}) of Fama and French (2018), and the five-factor q^5 model (α^{Q5}) of Hou et al. (2021). t-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

5.4 Explaining the performance of the composite predictors

In this section, we aim to explain the performance of the RBP-composite, HBP-composite, and *All*-composite predictors with additional risk factors on top of the six-factor model (α^{FF6}) of Fama and French (2018), and the five-factor q^5 model (α^{Q5}) of Hou et al. (2021). We apply the following additional risk factors: market-wide liquidity factor (LIQUI) introduced by Pástor and Stambaugh (2003), long-run reversal effect (LTREV) found by De Bondt and Thaler (1985), a factor that exploits the return reversal at a horizon of one month (STREV) by Jegadeesh (1990), betting-against-beta (BAB) found by Frazzini and Pedersen (2014), and the quality-minus-junk (long high-quality stocks and short low-quality stocks) factor by Asness, Frazzini, and Pedersen (2019).[15]

Table 8 reports the results of contemporaneous time-series regressions that include additional risk factors. Starting with the HBP-composite, we observe that the monthly spreads vary from 0.27% (3.24% p.a., $t = 1.95$) to 0.38% (4.56% p.a., $t = 3.19$). If the *BAB* and *QMJ* factors are added to the FF6-model, the spread of the HBP-composite becomes insignificant. Moving to the RBP-composite, the alphas vary less and are economically smaller; however, they are statistically more significant: 0.21% (2.52% p.a., $t = 2.58$) to 0.30% (3.60% p.a., $t = 3.18$). The RBP-composite remains significant across all specifications. Finally, the *All*-composite outperforms other composite scores in terms of both economic and statistical significance. The *All*-composite monthly spreads range from 0.35% (4.20% p.a., $t = 2.89$) to 0.44% (5.28% p.a., $t = 4.13$), with the highest spread being achieved if we add the short-term reversal factor to the asset pricing model (see column 8). The *All*-composite remains significant at 1% across all specifications and outperforms the RBP- and HBP-composites. Hence, both RBPs and HBPs matter for the prediction exercise of future mutual fund performance. These findings remain stable when we use different model specifications to measure risk-adjusted performance. We obtain qualitatively similar

[15] We obtain: LIQUI factor from https://finance.wharton.upenn.edu/~stambaug/; STREV and LTREV from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/; BAB and QMJ from https://www.aqr.com/Insights/Datasets/.

results when we utilize the Hou et al. (2021) five-factor q^5 asset pricing model as our benchmark. Results are shown in Appendix C, Table C1.

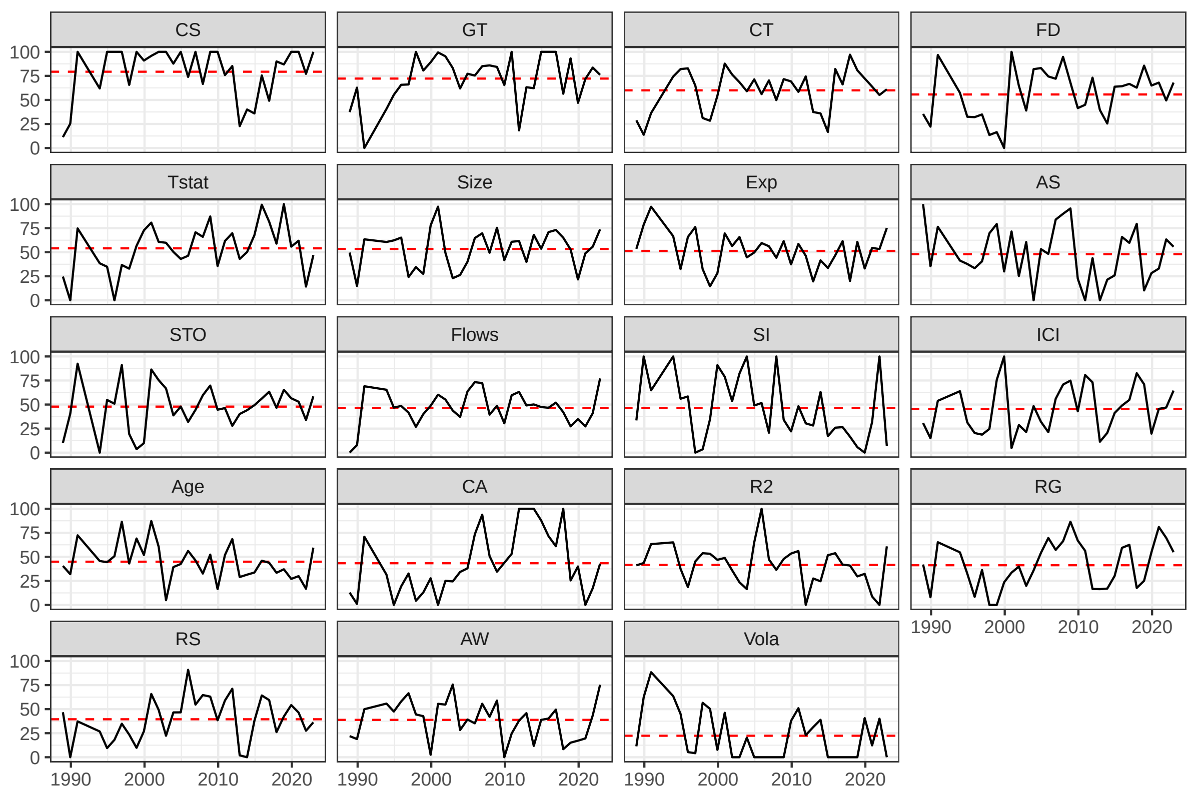
Table 8: Time-Series Regressions - Additional Risk Factors, 1988-2022, monthly returns

	HBP			RBP			All		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
FF6 factors									
Mkt	-0.06	0.06**	-0.04	0.02	0.06***	-0.01	-0.02	0.09***	-0.02
	[-1.44]	[2.01]	[-0.91]	[0.87]	[2.72]	[-0.47]	[-0.50]	[2.85]	[-0.46]
SMB	0.12*	0.12**	0.13**	-0.00	0.04	-0.01	0.11	0.13**	0.12*
	[1.73]	[2.25]	[1.98]	[-0.04]	[0.80]	[-0.17]	[1.49]	[2.06]	[1.71]
HML	-0.04	-0.01	-0.01	-0.15***	-0.11***	-0.17***	-0.11	-0.07	-0.09
	[-0.47]	[-0.09]	[-0.08]	[-3.22]	[-2.77]	[-3.38]	[-1.45]	[-1.33]	[-1.08]
RMW	-0.01	0.05	-0.03	-0.13***	-0.14***	-0.01	-0.10	-0.05	-0.04
	[-0.10]	[0.74]	[-0.20]	[-3.29]	[-4.03]	[-0.18]	[-0.91]	[-1.15]	[-0.31]
CMA	0.10	-0.13	0.10	-0.03	-0.07	-0.05	0.05	-0.14**	0.05
	[0.88]	[-1.64]	[1.04]	[-0.68]	[-1.50]	[-1.06]	[0.47]	[-2.14]	[0.55]
UMD	0.04	-0.04	0.05	0.07*	0.06**	0.10***	0.07	-0.00	0.09
	[0.51]	[-0.82]	[0.52]	[1.95]	[2.09]	[2.81]	[1.00]	[-0.10]	[1.24]
Additional Risk Factors									
LIQUI	0.01			0.08**			0.03		
	[0.23]			[2.11]			[0.71]		
STREV		-0.61***			-0.17***			-0.53***	
		[-8.96]			[-6.45]			[-10.91]	
LTREV		0.15*			-0.06			0.09	
		[1.82]			[-1.04]			[1.27]	
BAB			-0.08			-0.01			-0.09*
			[-0.98]			[-0.26]			[-1.71]
QMJ			0.09			-0.18**			-0.00
			[0.57]			[-2.16]			[-0.02]
Intercept									
α	0.27*	0.38***	0.28	0.21**	0.26***	0.30***	0.35***	0.44***	0.39***
	[1.95]	[3.19]	[1.59]	[2.58]	[3.04]	[3.18]	[2.89]	[4.13]	[2.62]
Details									
$\bar{R}^2$	0.02	0.51	0.02	0.30	0.38	0.29	0.07	0.51	0.07
T	384	384	384	384	384	384	384	384	384

This table reports the time-series analysis of the RBP-, HBP-, and *All*-composite predictors. We regress the high-minus-low spread of the three composite predictors on various factor models. Our baseline model is the Fama–French six-factor model (α^{FF6}) of Fama and French (2018); then we extend the model with the liquidity factor (LIQUI) from Pástor and Stambaugh (2003), long-run reversal effect (LTREV) from De Bondt and Thaler (1985), stock returns reversal at short horizons (STREV) from Jegadeesh (1990), betting-against-beta (BAB) from Frazzini and Pedersen (2014), and quality-minus-junk (QMJ) from Asness, Frazzini, and Pedersen (2019). *t*-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Next, we investigate how individual predictors contribute to the joint forecasting power of the *All*-composite predictor. Figure 2 depicts how the importance of each predictor changes over time. The figure shows that the importance of predictors varies substantially each year, which is consistent with DeMiguel et al. (2023). Considering the average importance over time, as shown by the red dashed line, most of the predictors contribute equally to the *All*-composite, with *CS* and *GT* being the most important predictors and *Vola* being the least important.

Figure 2: Time Evolution of Predictors Importance for *All*-composite Score, 1988-2022
This figure plots the time evolution of the importance of each individual predictor for the *All*-composite score. We measure the importance of each predictor as the average cross-sectional (FMB) beta over the last 12 months. First, we run contemporaneous multivariate weighted-least-squares (WLS) cross-sectional regressions for each period, where the dependent variable All_t is regressed on the 19 standardized predictors X_t. Note that all variables are on the same scale ranging from -0.5 to +0.5, i.e., standardized as described in Section 4. Second, the corresponding beta estimates are then smoothed using a rolling window average over the previous 12 months. We scale these smoothed betas so that they range between zero for the least important predictor and 100 for the most important predictor. We report relative importance for each year from 1988 to 2022. The red dashed line is the average scaled importance over the entire period.



5.5 Quantifying noise diversification benefits

We further explore the noise diversification explanation of our main findings. For this purpose, we compute realized alphas of individual predictors and RBP-, HBP-, and *All*-composite scores for the decile portfolios and corresponding long-short spreads on a monthly basis. We estimate $\alpha_{j,t}$ as the one-month risk-adjusted return from the Fama–French six-factor model of Fama and French (2018) or, as an alternative, the five-factor q^5 model of Hou et al. (2021). The estimate we obtain is the monthly realized alpha computed as

$$\begin{aligned} \alpha_{j,t} = R^e_{i,t} &- \hat{\beta}^{MKT}_{j,t-1} MKT_t - \hat{\beta}^{SMB}_{j,t-1} SMB_t - \hat{\beta}^{HML}_{j,t-1} HML_t \\ &- \hat{\beta}^{RMW}_{j,t-1} RMW_t - \hat{\beta}^{CMA}_{j,t-1} CMA_t - \hat{\beta}^{UMD}_{j,t-1} UMD_t, \end{aligned} \tag{3}$$

where MKT_t, SMB_t, HML_t, RMW_t, CMA_t, and UMD_t are the returns of the five Fama-French and momentum factors in month t, and $\hat{\beta}^{MKT}_{j,t-1}$, $\hat{\beta}^{SMB}_{j,t-1}$, $\hat{\beta}^{HML}_{j,t-1}$, $\hat{\beta}^{RMW}_{j,t-1}$, $\hat{\beta}^{CMA}_{j,t-1}$, and $\hat{\beta}^{UMD}_{j,t-1}$ are the factor loadings of the high-minus-low spread of the j^{th} predictor with respect to the FF6-factors estimated using the 36-month estimation window (with a minimum requirement of 30 observations) ending in month $t-1$.

Once we have a time series of monthly realized alphas for each individual predictor and the composite scores, spanning from 1991 to 2022, we compute the cumulative abnormal returns of a one-dollar portfolio invested at the beginning of 1991. Figure 3 contrasts the average cumulative abnormal performance of all individual predictors against the performance of the *All*-composite predictor, which aggregates the information from 19 predictors using our method. Our results reveal that the spread between the bottom and top decile portfolios is more pronounced when we construct the aggregate predictor that contains information from 19 predictors. Precisely, the spread is more than three times wider for the plot on the right-hand side. Moreover, the top-decile portfolio, although volatile, is consistently above the one-dollar line. We provide additional results in Appendix C. Figures C3 and C4 show similar results when the RBPs and HBPs are considered separately. Overall, the *All*-composite score results in the most sizeable gains.

Figure 3: Cumulative Abnormal Returns of Prediction Deciles for *All*, 1991-2022
The figure shows the cumulative risk-adjusted performance of the average individual predictor (on the left-hand side) against the performance of the *All*-composite (on the right-hand side). To obtain the left side of the plot, we simply compute, for each decile, the average alpha across 19 predictors. The right side of the plot shows the performance of deciles based on a single sorting variable, which is the *All-Composite* score. The risk-adjusted performance is calculated using the Fama and French (2018) six-factor model.

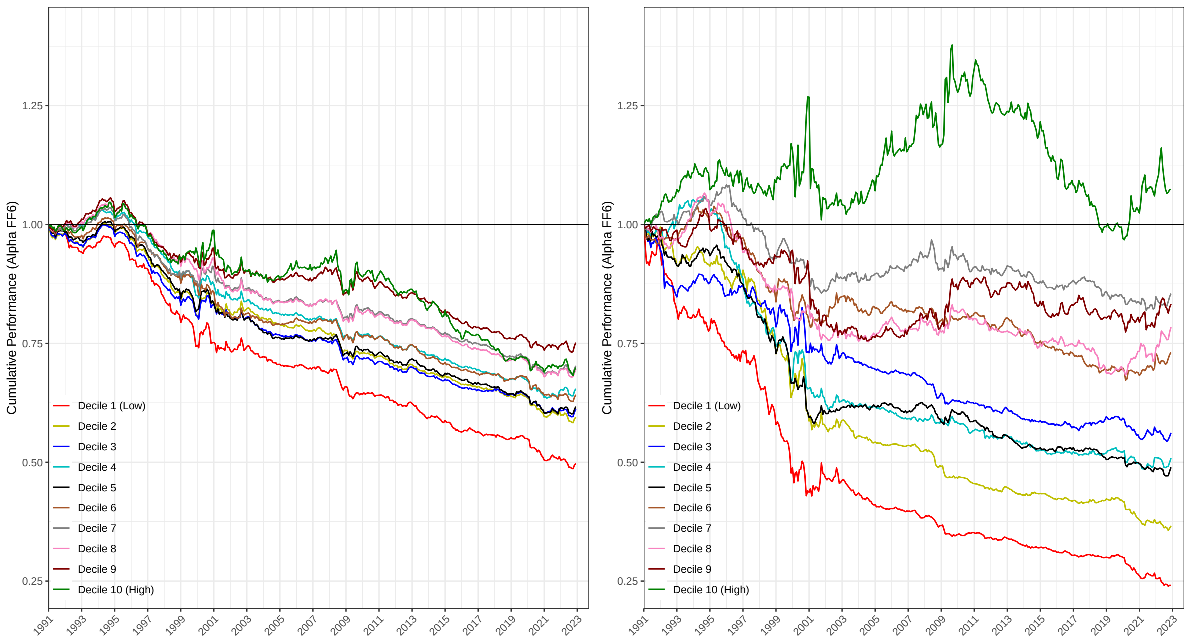


Table 9 reports the difference between average individual predictor returns and the corresponding composite predictors. The average monthly long-short spread for 19 individual predictors is 0.10% (1.20% p.a., $t = 2.71$) compared to 0.35% (4.2% p.a., $t = 2.96$) for the *All*-composite. The difference amounts to 0.25% (3% p.a., $t = 2.91$) and is statistically significant at the 1% level. Hence, our simple, yet effective method of aggregating fund predictors results in noise diversification benefits that are both economically and statistically significant. We show that individual fund predictors are not uniquely valuable in capturing managerial skill or future fund performance. Thus, averaging across predictor rankings performs better since some noise is diversified away. As a result, the composite predictor more precisely indicates which funds to avoid and, more importantly, which funds to buy.

Table 9: Quantifying Noise Reduction Benefits, 1991-2022

	RBP			HBP			All		
Portfolio	**Comp.**	**Avg.**	**Diff.**	**Comp.**	**Avg.**	**Diff.**	**Comp.**	**Avg.**	**Diff.**
Decile Portfolios									
Low (L)	-0.19***	-0.13***	-0.06*	-0.26***	-0.16***	-0.09**	-0.27***	-0.15***	-0.12**
	[-3.48]	[-3.41]	[-1.77]	[-3.37]	[-3.61]	[-1.97]	[-3.56]	[-4.01]	[-2.09]
2	-0.14***	-0.11***	-0.03	-0.12**	-0.10***	-0.02	-0.24***	-0.10***	-0.13***
	[-2.73]	[-3.19]	[-0.97]	[-2.22]	[-2.73]	[-0.62]	[-3.70]	[-3.15]	[-2.70]
3	-0.16***	-0.10***	-0.06*	-0.12***	-0.09***	-0.02	-0.13***	-0.10***	-0.04
	[-3.72]	[-2.74]	[-1.85]	[-2.96]	[-2.95]	[-0.72]	[-3.00]	[-2.97]	[-1.03]
4	-0.12**	-0.07**	-0.04	-0.13***	-0.09***	-0.04	-0.17***	-0.08**	-0.08**
	[-2.22]	[-2.04]	[-1.10]	[-3.15]	[-2.82]	[-1.27]	[-3.49]	[-2.54]	[-2.03]
5	-0.12***	-0.09**	-0.03	-0.07*	-0.09***	0.02	-0.18***	-0.09***	-0.09**
	[-3.05]	[-2.43]	[-1.06]	[-1.73]	[-2.76]	[0.75]	[-3.96]	[-2.75]	[-2.31]
6	-0.10**	-0.08**	-0.03	-0.05	-0.08**	0.03	-0.05	-0.08**	0.03
	[-2.52]	[-2.23]	[-0.77]	[-1.29]	[-2.40]	[0.99]	[-1.17]	[-2.46]	[0.98]
7	-0.12**	-0.06**	-0.05*	-0.04	-0.07*	0.03	-0.03	-0.07**	0.03
	[-2.58]	[-2.12]	[-1.67]	[-0.77]	[-1.95]	[0.98]	[-0.79]	[-2.07]	[1.10]
8	-0.07	-0.09***	0.03	0.04	-0.04	0.08***	-0.02	-0.06*	0.04
	[-1.63]	[-2.82]	[0.98]	[0.68]	[-1.09]	[2.63]	[-0.44]	[-1.85]	[0.98]
9	-0.05	-0.05	0.00	0.04	-0.04	0.08	0.02	-0.05	0.06
	[-1.12]	[-1.63]	[0.04]	[0.52]	[-1.01]	[1.62]	[0.25]	[-1.32]	[1.32]
High (H)	0.04	-0.05	0.09**	0.02	-0.03	0.05	0.09	-0.04	0.13***
	[0.60]	[-1.39]	[2.13]	[0.19]	[-0.59]	[1.02]	[1.20]	[-0.93]	[2.92]
High-minus-low Spread									
H-L	0.23***	0.07***	0.16***	0.28**	0.13**	0.15*	0.35***	0.10***	0.25***
	[2.83]	[2.63]	[2.72]	[1.97]	[2.04]	[1.76]	[2.96]	[2.71]	[2.91]

This figure reports the difference between the average risk-adjusted performance of individual predictors and the corresponding composite scores that aggregate the individual predictors into a single predictor of future fund performance. Comp.: is the RBP-, HBP-, or *All*-composite predictor. Avg.: is the average of the comparable set of predictors. For instance, for RBPs, the average would be based on the nine individual return-based predictors. Diff.: is the spread between the composite return and the average return of the comparable set of individual predictors. The risk-adjusted performance is calculated using the Fama and French (2018) six-factor model. t-statistics are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

6 Empirical Results: More Predictors, Better Performance

Expected returns are noisy, and forecasting errors have a significant impact on ex-post performance; hence, $1/N$ portfolios perform well and are difficult to beat (see, e.g., DeMiguel, Garlappi, and Uppal (2009)). The same logic applies to the context of our paper. Since we use cross-sectional regressions solely to forecast the sign, and predictors in the composite score are equally weighted (following the $1/N$ analogy), all signals receive equal weight in the composite. This is also why we do not use multivariate FMB regressions (as in Lewellen (2015)), because FMB betas would serve as weights for the predictors.

6.1 Chasing best predictors

In this section, we perform a horse race for different practically relevant investment strategies. It is assumed that an investor is chasing fund predictors based on their past observed performance. We compare the future alphas of such an investment strategy by altering the number of predictor variables from one (i.e., the variable that performed best in selecting funds in the past) to two (i.e., the predictor based on the two best-performing variables), up to 19 (i.e., our composite predictor based on all variables). We show that increasing the number of variables in such an exercise leads to improved future performance of the investment strategy.

We estimate $\alpha_{j,t}$ as a one-month risk-adjusted return from the Fama–French six-factor model of Fama and French (2018). The estimate we obtain is the monthly realized alpha that we use to rank each predictor each month. Since such ranking can be sensitive to the choice of the factor model, we use both FF6 and q^5 factor models for robustness. The monthly realized alpha is computed as in Equation (3). For our hypothetical strategy, we use $\alpha_{j,t-1}$ to rank predictors to avoid look-ahead bias and to mimic a real-time investment strategy.

Our hypothetical performance chaser ranks individual predictors each month based on the most recently realized alpha that he or she estimates using the last 36 months of data. Based on the rankings, the investor decides to construct a composite predictor and invest accordingly. For instance, when $N = 1$, the composite predictor is based on the single best predictor; when $N = 3$, the composite predictor is constructed using the top three best-ranked predictors, and so forth. By construction, using all available predictors is not impacted by ranking.

Table 10 reports the results of a hypothetical investment strategy where an investor forms composite scores based on the past performance of individual predictors. Clearly, in all cases (the RBP-, HBP-, and *All*-composites), chasing the best predictors is sub-optimal because the past performance of a predictor is not highly indicative of future performance, and a predictor-performance relationship is most likely time-varying. Furthermore, the importance of predictors varies substantially over time, as shown in Figure 2, and individual fund predictors are not uniquely valuable in capturing managerial skill or future fund performance, as shown in Figure 3. We find that the optimal strategy is to combine all available predictors. Such an approach should also lead to the greatest diversification of noise when ranking funds.

6.2 Examining all possible combinations of predictors

Finally, we examine all possible ways to compute the composite predictor for a given fund. For example, a set of nine predictors gives rise to $2^9 - 1 = 511$ possible ways to build an RBP-composite predictor. Similarly, 19 predictors give rise to 524,287 possible ways to construct a composite predictor by combining 9 RBPs and 10 HBPs indiscriminately.

Table 11 reveals the performance of all possible combinations of predictors that can be achieved with a set of 19 individual predictors. We observe how the number of predictors used to compute a composite predictor influences its performance. Clearly, as we increase the number of predictors used to compute the composite predictor, the performance improves on average. Moreover, the dispersion in performance decreases, and the t-statistic distribution shifts to the right. The long-short spread becomes larger both economically and statistically. This pattern is best illustrated by Figure 4, which again documents the benefits

Table 10: Performance of a Hypothetical Performance Chaser, 1991-2022, monthly returns

	Equal-Weighted						**Value-Weighted**					
	RBP		**HBP**		**All**		**RBP**		**HBP**		**All**	
	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}
N = 1	0.12	0.20	0.08	0.16	0.12	0.28	0.07	0.22	0.16	0.21	0.20	0.30
	[1.12]	[1.39]	[0.42]	[0.82]	[0.72]	[1.36]	[0.49]	[1.46]	[0.86]	[1.04]	[1.04]	[1.42]
N = 2	0.07	0.16	0.21	0.12	0.18	0.20	0.09	0.25*	0.26	0.13	0.20	0.19
	[0.72]	[1.33]	[1.26]	[0.64]	[1.21]	[1.08]	[0.82]	[1.83]	[1.47]	[0.67]	[1.29]	[0.95]
N = 3	0.07	0.14	0.21	0.17	0.20	0.24	0.05	0.20	0.22	0.22	0.23	0.26
	[0.69]	[1.18]	[1.30]	[0.92]	[1.36]	[1.26]	[0.44]	[1.50]	[1.27]	[1.11]	[1.42]	[1.37]
N = 4	0.08	0.12	0.27*	0.21	0.18	0.17	0.11	0.06	0.24	0.22	0.20	0.22
	[0.90]	[1.14]	[1.73]	[1.12]	[1.25]	[0.94]	[1.13]	[0.54]	[1.51]	[1.18]	[1.39]	[1.18]
N = 5	0.06	0.13	0.20	0.13	0.16	0.23	0.10	0.12	0.16	0.09	0.15	0.20
	[0.71]	[1.37]	[1.33]	[0.75]	[1.23]	[1.29]	[0.96]	[1.13]	[0.95]	[0.51]	[1.08]	[1.05]
N = 6	0.08	0.14	0.22	0.18	0.15	0.22	0.14	0.18*	0.20	0.18	0.18	0.21
	[0.89]	[1.54]	[1.51]	[1.05]	[1.13]	[1.26]	[1.35]	[1.73]	[1.36]	[0.96]	[1.26]	[1.16]
N = 7	0.05	0.13	0.29**	0.26	0.23*	0.19	0.06	0.19*	0.27*	0.26	0.26*	0.18
	[0.57]	[1.43]	[2.03]	[1.54]	[1.76]	[1.10]	[0.61]	[1.81]	[1.74]	[1.49]	[1.87]	[1.02]
N = 8	0.06	0.15*	0.31**	0.28*	0.21*	0.20	0.13	0.20**	0.30**	0.25	0.23*	0.22
	[0.83]	[1.74]	[2.32]	[1.66]	[1.67]	[1.23]	[1.52]	[2.02]	[1.99]	[1.37]	[1.73]	[1.23]
N = 9	0.17**	0.20***	0.32**	0.31*	0.22*	0.23	0.23***	0.27***	0.33**	0.31*	0.30**	0.24
	[2.57]	[3.06]	[2.46]	[1.88]	[1.67]	[1.46]	[2.83]	[3.36]	[2.26]	[1.81]	[2.12]	[1.42]
N = 10			0.32**	0.32**	0.23*	0.21			0.28**	0.28*	0.24*	0.24
			[2.43]	[2.05]	[1.75]	[1.30]			[1.97]	[1.67]	[1.74]	[1.39]
N = 11					0.22*	0.24					0.28**	0.25
					[1.75]	[1.45]					[2.07]	[1.42]
N = 12					0.23*	0.24					0.25*	0.27
					[1.86]	[1.50]					[1.86]	[1.61]
N = 13					0.22*	0.25					0.31**	0.24
					[1.86]	[1.61]					[2.37]	[1.43]
N = 14					0.22*	0.25					0.29**	0.25
					[1.85]	[1.57]					[2.25]	[1.43]
N = 15					0.21*	0.27*					0.27**	0.26
					[1.70]	[1.70]					[1.99]	[1.47]
N = 16					0.23*	0.28*					0.26**	0.28*
					[1.91]	[1.78]					[2.03]	[1.67]
N = 17					0.26**	0.27*					0.30**	0.28*
					[2.15]	[1.76]					[2.43]	[1.72]
N = 18					0.28**	0.30**					0.32***	0.33**
					[2.39]	[2.04]					[2.63]	[2.04]
N = 19					0.29**	0.32**					0.35***	0.36**
					[2.58]	[2.28]					[2.96]	[2.40]

The column labels of the table α^{FF6} and α^{Q5} refer to both at the same time: first, the model that was used to compute realized alpha and subsequently rank predictors; second, the model that is used to measure the risk-adjusted performance, i.e., the intercept from a time-series regression that includes the Fama–French six-factor model of Fama and French (2018), and the five-factor q^5 model of Hou et al. (2021), respectively. The row labels of the table represent the number of predictors used to construct the composite score. For instance, when $N = 1$, the composite predictor is based on a single best predictor; when $N = 3$, the composite predictor is constructed using the top three best-ranked predictors, and so forth. t-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

of diversifying mutual fund selection across various predictor variables. As we increase the number of predictors that we average across, the degree of mutual fund performance predictability, on average, improves. The distribution of the long-short alpha spread shifts to the right, and the information ratios of both the short and long legs improve consistently. We compute the information ratio of the top (bottom) decile of funds against the average active fund in our sample.

Figure 5 illustrates how each individual predictor variable, on average, contributes to the long-short alpha spread of a composite predictor relative to other predictor variables. We compare the average risk-adjusted long-short spread of a composite score when a given predictor is included in the computation of the composite score. For better interpretability and presentation, we rescale long-short alpha spreads conditional on the number of predictors in a composite score (within columns). Thus, the comparison *between* predictors can only be made within each column. Additionally, we can evaluate the evolution of the marginal contribution of a given predictor by moving along the x-axis (within rows).

For interpretation within columns, consider the example of 18 on the x-axis and *CS* (*AS*) on the y-axis. In this case, the square for *CS* is red, while the square for *AS* is yellow, indicating that when a composite score is constructed using 18 predictors, the mean alpha is highest with the inclusion of *CS* and lowest with the inclusion of *AS*. In other words, *CS* contributes the most to performance, while *AS* contributes the least when all other predictors are considered. For interpretation within rows, take expenses (*Exp*) on the y-axis and follow its progression along the x-axis. As the number of predictors used to construct a composite score increases, the relative importance of *Exp* improves, likely due to the unique information it provides compared to the other predictors. When return-based predictors (RBPs) and holdings-based predictors (HBPs) are combined, predictors such as *CS*, *GT*, *CA*, *Tstat*, *Exp*, *AW*, and *SI* generally contribute the most to performance.

We emphasize once again the importance of combining HBPs with RBPs and increasing the number of predictors. Additionally, our results underscore the significance and potential benefits of considering interaction effects across predictors in future research.

Table 11: Summary of All Possible Combinations of the Composite Predictor, 1991-2022

${}_nC_k$	α^{FF6} Mean	Median	SD	IQR	*t-stat* Mean	Median	SD	IQR
$\binom{19}{1} = 19$	0.10	0.05	0.20	0.29	0.96	0.77	1.66	2.51
$\binom{19}{2} = 171$	0.13	0.11	0.17	0.26	1.28	1.28	1.58	2.58
$\binom{19}{3} = 969$	0.16	0.15	0.16	0.25	1.53	1.66	1.50	2.33
$\binom{19}{4} = 3,876$	0.18	0.19	0.16	0.23	1.73	1.88	1.42	2.15
$\binom{19}{5} = 11,628$	0.20	0.21	0.15	0.22	1.90	2.02	1.34	1.93
$\binom{19}{6} = 27,132$	0.22	0.23	0.14	0.20	2.03	2.13	1.25	1.75
$\binom{19}{7} = 50,388$	0.23	0.24	0.13	0.19	2.14	2.23	1.16	1.61
$\binom{19}{8} = 75,582$	0.25	0.26	0.12	0.17	2.23	2.32	1.08	1.48
$\binom{19}{9} = 92,378$	0.26	0.27	0.11	0.16	2.32	2.39	1.00	1.35
$\binom{19}{10} = 92,378$	0.27	0.28	0.11	0.15	2.39	2.45	0.92	1.23
$\binom{19}{11} = 75,582$	0.28	0.29	0.10	0.13	2.45	2.51	0.84	1.12
$\binom{19}{12} = 50,388$	0.29	0.30	0.09	0.12	2.52	2.56	0.77	1.02
$\binom{19}{13} = 27,132$	0.30	0.31	0.08	0.11	2.58	2.61	0.69	0.92
$\binom{19}{14} = 11,628$	0.31	0.32	0.07	0.10	2.63	2.66	0.62	0.83
$\binom{19}{15} = 3,876$	0.32	0.32	0.06	0.09	2.68	2.70	0.53	0.72
$\binom{19}{16} = 969$	0.33	0.33	0.05	0.07	2.73	2.74	0.45	0.61
$\binom{19}{17} = 171$	0.33	0.34	0.05	0.06	2.75	2.78	0.38	0.53
$\binom{19}{18} = 19$	0.34	0.35	0.04	0.04	2.79	2.84	0.30	0.31
$\binom{19}{19} = 1$	0.35	0.35	0.00	0.00	2.96	2.96	0.00	0.00

This table reports summary statistics of the high-minus-low FF6-spreads across all possible combinations used to construct a composite predictor. For example, the first row reports the mean, median, standard deviation (SD), and interquartile range (IQR) of the H-L spread of 19 composite predictors constructed using only one constituent. Similarly, the second row refers to 171 composite predictors constructed with only two constituents at a time. ${}_nC_k$ denotes the number of combinations where n is the total number of predictors considered and k is the number of predictors used to construct the composite predictor. t-statistics are computed using Newey and West (1987) standard errors with six lags.

Figure 4: Relative Contribution of Predictors to the Performance, 1991-2022
The figure plots the distributions of FF6-alphas of the long-short spread and the corresponding information ratios of both the short and long legs, as functions of the number of predictors in a composite predictor. We compute the information ratio for the top (bottom) decile of funds against the average active fund in our sample.

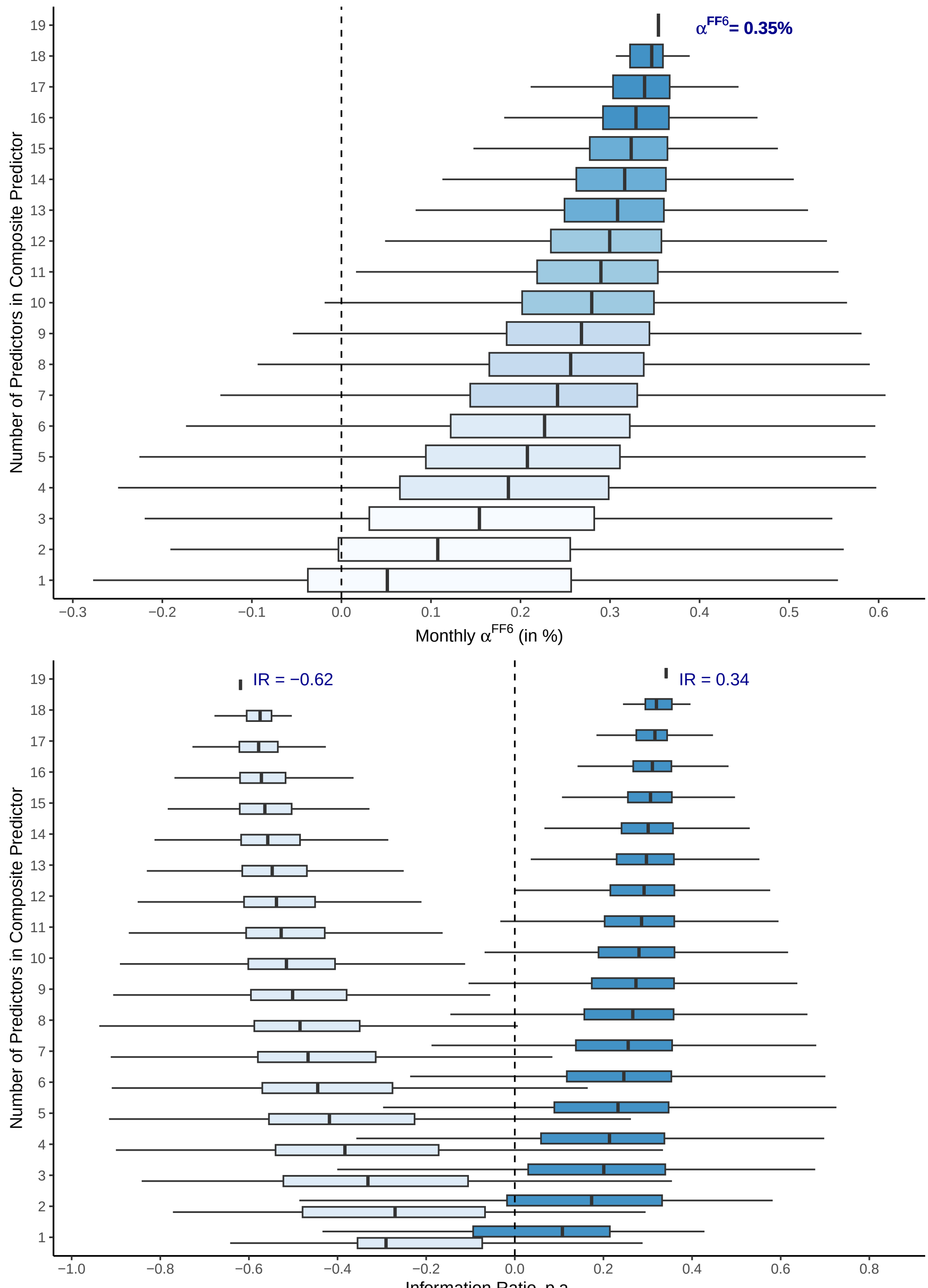
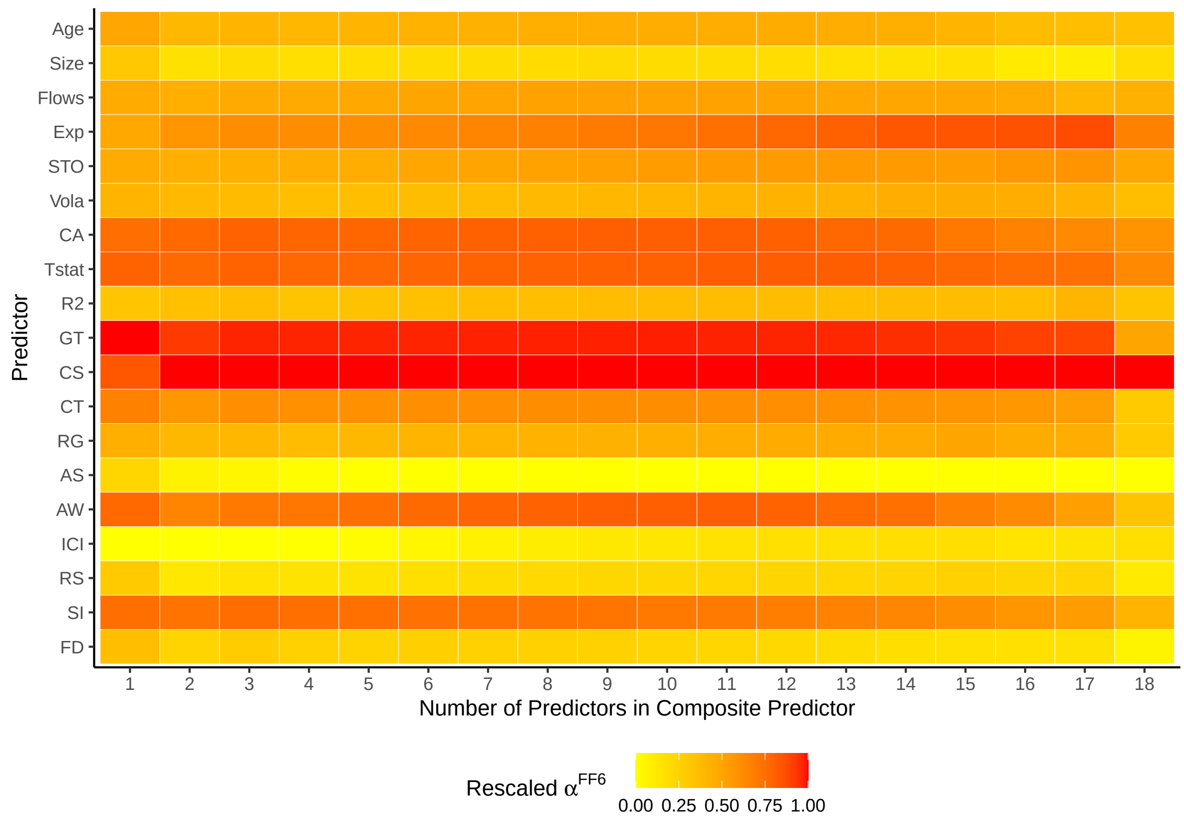


Figure 5: Relative Contribution of Predictors to the Performance, 1991-2022
The figure shows the relative contribution of a given predictor to the performance of the composite predictor conditional on the number of predictors used to construct the composite. We compare the average risk-adjusted long-short spread of a composite score when a given predictor is included in the computation of the composite score. For instance, consider the intersection of 18 (x-axis) and *CS* (y-axis). The square is red, indicating that when a composite score is constructed using 18 predictors, the mean alpha is the highest when *CS* is included. We rescale mean alpha spreads to enhance the interpretability and visual appeal of the heatmap. Note that comparisons between predictors can only be made within each column. Lastly, we omit the column with the maximum number of predictors, i.e., 19. When all predictors are used, there is only one possible combination. Consequently, the average alpha conditioned on a given predictor being included in the composite score becomes uninformative.



7 Robustness Checks

In this section, we perform robustness checks of our main findings. Table 12 displays how the high-minus-low spreads of RBP-, HBP-, and *All*-composite predictors vary across sub-sample splits by the fund size, fund style, and time sub-periods. Starting with the fund size split, we find that the *All*-composite predictor consistently delivers the highest performance except for the small fund sample where HBPs perform the best (while RBPs do not yield a significant performance). With regard to medium-size funds, the spread significantly improves when we combine both sets of predictors. For instance, composite scores solely based on RBPs and HBPs predictors yield monthly spreads of 0.19% (2.28% p.a., $t = 2.91$) and 0.16% (1.92% p.a., $t = 1.87$), respectively, while *All*-composite earns 0.32% (3.84% p.a., $t = 2.79$). Therefore, our approach performs relatively well across size bins, especially for large mutual funds. This contrasts with stock market return prediction where predictability is usually concentrated in small stocks and deteriorates when stock size increases. This underscores the practical importance of our findings as the performance of larger funds is arguably more important to predict.

Moving to style splits, the *All*-composite predictor is consistently greater than composite scores solely based on RBPs and HBPs. Our predictor appears to work better for Growth funds 0.33% (3.96% p.a., $t = 3.01$) than for Growth funds, with a spread of 0.33% (3.96% p.a., $t = 3.01$), compared to Growth & Income funds, which have a spread of 0.20% (2.40% p.a., $t = 2.93$).

Lastly, we split the sample into two sub-periods, each spanning 17 years. Interestingly, the HBPs outperform the RBPs in the earlier period with a 5.64% ($t = 2.34$) against insignificant 2% p.a. However, this changes in the later part of the sample where the RBPs outperform the HBPs: 3.36% ($t = 3.05$) compared to an insignificant 2% p.a., respectively. Overall, the *All*-composite delivers the highest performance in the earlier period. To test whether the performance decline in the later part of the sample is statistically significant, we estimate a regression with a time dummy variable. We set the time dummy to equal 1 in the later part of the sample (years 2006 to 2022). Neither the increase in the predictability of RBPs nor the decrease of HBPs and *All* is statistically different from zero. Hence, we cannot conclude that our method significantly deteriorates in the post-2006 period. Our results here differ from studies analyzing composite predictors for stock selection, which often find a significant decline in performance for more recent periods (see e.g., Green, Hand, and Zhang, 2017).

Table 12: Robustness Tests of the Composite Predictors, 1988-2022

Panel A: Size Split

	Small			**Medium**			**Large**		
	RBP	**HBP**	**All**	**RBP**	**HBP**	**All**	**RBP**	**HBP**	**All**
α^{FF6}	0.08	0.36***	0.31***	0.19***	0.16*	0.32***	0.28***	0.28**	0.33***
	[1.34]	[2.81]	[2.65]	[2.91]	[1.87]	[2.79]	[3.59]	[1.97]	[2.76]
α^{Q5}	0.12*	0.34**	0.33**	0.20***	0.17*	0.33**	0.34***	0.27	0.35**
	[1.93]	[2.24]	[2.30]	[2.71]	[1.89]	[2.28]	[4.09]	[1.62]	[2.31]

Panel B: Style Split

	Growth			**Growth & Income**			**Other**		
	RBP	**HBP**	**All**	**RBP**	**HBP**	**All**	**RBP**	**HBP**	**All**
α^{FF6}	0.26***	0.21*	0.33***	0.09**	0.16*	0.20***	0.22***	0.32**	0.31**
	[3.78]	[1.73]	[3.01]	[2.06]	[1.87]	[2.93]	[2.67]	[2.11]	[2.42]
α^{Q5}	0.29***	0.21	0.33**	0.11**	0.17*	0.24***	0.25***	0.30	0.29*
	[4.07]	[1.50]	[2.49]	[2.26]	[1.89]	[3.17]	[2.87]	[1.64]	[1.75]

Panel C: Time Split

	1989 - 2005			**2006 - 2022**			**Difference**		
	RBP	**HBP**	**All**	**RBP**	**HBP**	**All**	**RBP**	**HBP**	**All**
α^{FF6}	0.17	0.47**	0.56***	0.28***	0.16	0.24*	0.11	-0.24	-0.23
	[1.33]	[2.34]	[2.93]	[3.05]	[1.15]	[1.88]	[0.77]	[-0.96]	[-1.04]
α^{Q5}	0.19	0.32	0.43	0.33***	0.24*	0.33**	0.07	-0.24	-0.24
	[1.41]	[1.00]	[1.38]	[3.64]	[1.84]	[2.59]	[0.44]	[-0.81]	[-0.91]

This table reports the high-minus-low spreads of RBP-, HBP-, and *All*-composite predictors across various sub-samples and time periods. Size splits: assign funds to terciles based on the last month's *TNA*. Style splits: are based on the CRSP Objective Code (*crsp_obj_cd*), which is CRSP mapping of Strategic Insights, Wiesenberger, and Lipper objective codes. To locate growth funds, we use *EDYG* code, for growth and income *EDYB* and *EDYI*, respectively (where $E = Equity$, $D = Domestic$, $Y = Style$, $G = Growth$, $B = Growth\ \&\ Income$, and $I = Income$). The other category is composed of all other funds. Time splits: the difference column reports the loading on a time dummy variable added to the risk factors on the right-hand side. The dummy equals 1 when we are in the later part of the sample (years 2006 to 2022). The risk-adjusted performance is calculated using the Fama and French (2018) six-factor model and the Hou et al. (2021) five-factor q^5 model. t-statistics are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Next, we test alternative ways to standardize the individual predictors before aggregating them. Table 13 reports results if we were to use the z-score or the rank z-score standardization instead of normalizing individual predictors to range from −0.5 to 0.5. To compute the z-score, each predictor is demeaned in a cross-section and divided by the corresponding standard deviation, resulting in a variable with a mean of zero and a standard deviation of one. To limit the effect of outliers, the z-score is trimmed at 3 standard deviations at each tail. To compute the rank z-score: for each predictor, first, all observations are ranked in ascending order (in a cross-section). Second, the rank values are standardized using the z-score method. The rank z-score method is similar to the quality score of Asness, Frazzini, and Pedersen (2019). Regardless of the approach, our main result—that combining all predictors yields the highest performance—is once again supported. If we consider the value-weighted results, the long-short spread systematically improves both economically and statistically when we aggregate RBPs and HBPs. The rank z-score results in extremely similar predictor rankings as our preferred method because both approaches transform individual predictors into a uniform distribution. Hence, portfolio sorts yield close results.

Additionally, we test whether changing the estimation approach and the window size used to smooth betas would impact the prediction accuracy of our method. Our method uses loadings from univariate cross-sectional regressions (that are estimated using WLS with an expanding window starting from 36 months) to estimate the sign of the relationship between a predictor and next month's fund return. Table 14 reports the intersection of the weighting schemes (WLS and OLS) and the estimation window length. First, both WLS and OLS estimation approaches yield extremely similar results across different specifications. In fact, when using OLS, the performance of *All*-composite score even slightly improves. Moving to the estimation window length, we see that the performance improves as we increase the window length. The betas are extremely noisy, hence, our method benefits from the expanding window smoothing that stabilizes the estimates. Similar to Lewellen (2015), we document that predictions based on longer histories of cross-sectional regression slopes work best.

Table 13: Alternative Composite Predictors, 1988-2022

	Equal-Weighted				Value-Weighted			
	R^e	α^{FF4}	α^{FF6}	α^{Q5}	R^e	α^{FF4}	α^{FF6}	α^{Q5}
Z-score (trimmed at -3 and +3)								
RBP	0.13*	0.07	0.14**	0.21***	0.12	0.10	0.17*	0.25**
	[1.68]	[1.06]	[2.09]	[2.82]	[1.21]	[1.06]	[1.89]	[2.53]
HBP	0.33**	0.34***	0.33**	0.28*	0.29**	0.28**	0.25*	0.26
	[2.50]	[2.88]	[2.55]	[1.75]	[2.09]	[2.19]	[1.81]	[1.48]
All	0.33***	0.30***	0.31**	0.31**	0.33***	0.30**	0.30**	0.31**
	[2.60]	[2.78]	[2.58]	[2.12]	[2.79]	[2.55]	[2.44]	[2.03]
Rank Z-score								
RBP	0.18**	0.12*	0.16**	0.20***	0.19**	0.16*	0.23***	0.27***
	[2.46]	[1.67]	[2.55]	[3.05]	[2.10]	[1.85]	[2.83]	[3.34]
HBP	0.29**	0.31**	0.32**	0.32**	0.29**	0.29**	0.28**	0.28*
	[2.25]	[2.53]	[2.43]	[2.04]	[2.05]	[2.27]	[1.97]	[1.66]
All	0.30**	0.27**	0.29**	0.32**	0.34***	0.32***	0.35***	0.35**
	[2.51]	[2.47]	[2.58]	[2.28]	[2.76]	[2.63]	[2.95]	[2.38]

This table reports the high-minus-low spreads of RBP-, HBP-, and *All*-composite predictors constructed using different predictor standardization approaches. The Z-score: each predictor is demeaned in a cross-section and divided by the corresponding standard deviation, resulting in a variable with a mean of zero and a standard deviation of one. To limit the effect of outliers, the z-score is trimmed at 3 standard deviations at each tail. The Rank Z-score: for each predictor, first, all observations are ranked in ascending order (in a cross-section). Second, the rank values are standardized using the z-score method. The rank z-score method is similar to the quality score of Asness, Frazzini, and Pedersen (2019). The risk-adjusted performance is calculated using the Carhart (1997) four-factor model, the Fama and French (2018) six-factor model, and the Hou et al. (2021) five-factor q^5 model. t-statistics are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Lastly, we repeat the analysis from Subsection 6.1 where we introduce a hypothetical performance chaser who selects individual predictors based on their historical risk-adjusted performance before aggregating them into a composite score. Figure 6 demonstrates that previously discussed findings from Table 10 are robust to the estimation window length and risk-adjustment model. The best-performing strategy, in both economic and statistical terms, is to construct the composite predictor using the entire set of fund predictors. This method systematically results in the highest noise-diversification benefits, which improve risk-adjusted performance.

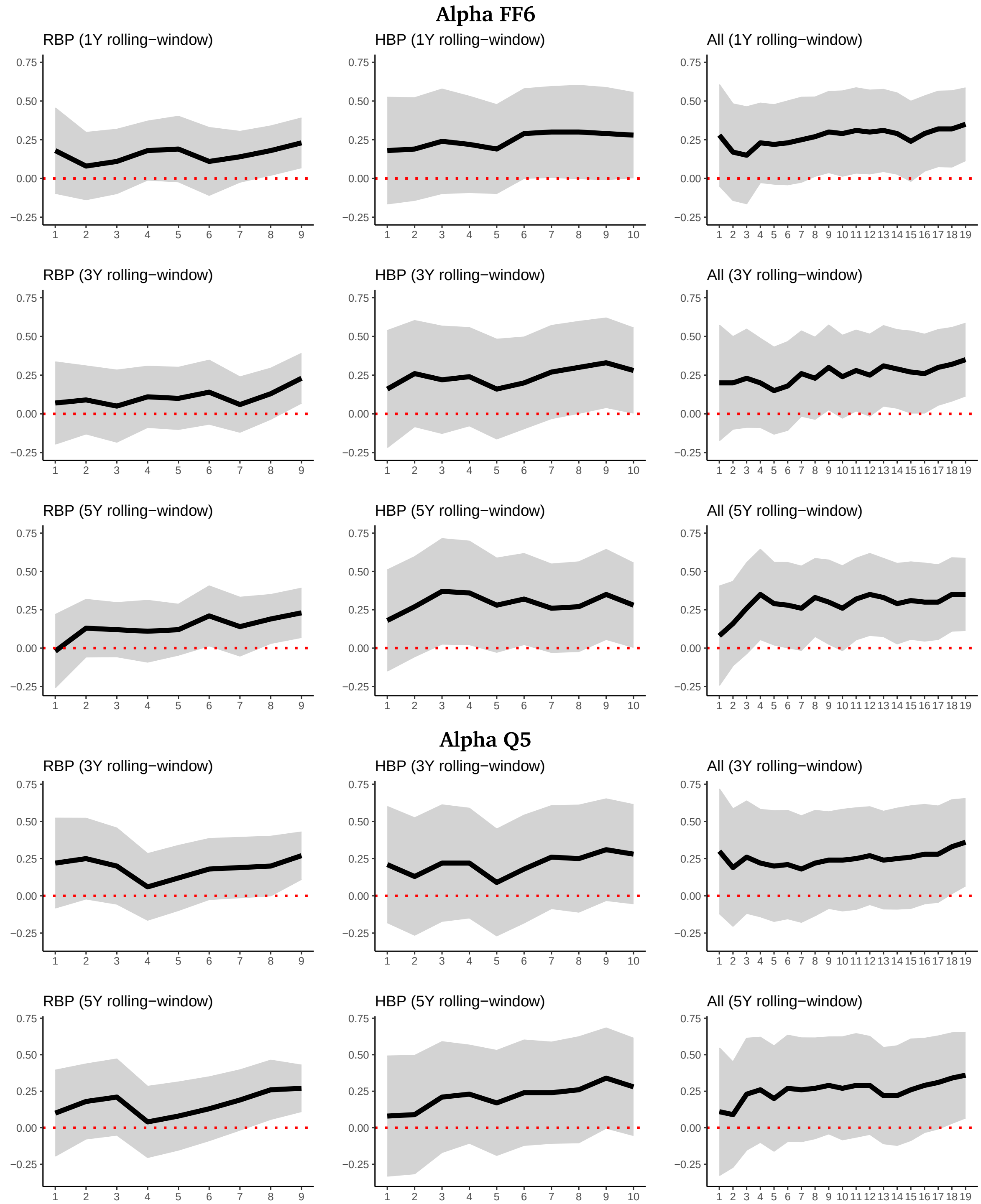
Table 14: Slope Estimates of Predictors: Altering Weighting Scheme and Estimation Window Length, 1988-2022

	WLS						OLS					
	RBP		HBP		All		RBP		HBP		All	
	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}	α^{FF6}	α^{Q5}
12-months	0.08	0.15	0.14	0.16	0.17	0.20	0.06	0.14	0.24	0.26	0.25*	0.30*
	[0.68]	[1.05]	[1.01]	[0.97]	[1.10]	[1.08]	[0.52]	[1.02]	[1.64]	[1.60]	[1.74]	[1.67]
36-months	0.14	0.20*	0.29**	0.26*	0.28***	0.29**	0.23**	0.28**	0.24*	0.20	0.28***	0.28**
	[1.30]	[1.91]	[2.36]	[1.85]	[2.73]	[2.42]	[2.27]	[2.45]	[1.85]	[1.44]	[2.63]	[2.22]
60-months	0.16	0.19*	0.26*	0.22	0.28**	0.27**	0.18	0.22*	0.26*	0.20	0.35***	0.31**
	[1.44]	[1.88]	[1.95]	[1.43]	[2.46]	[2.11]	[1.61]	[1.84]	[1.90]	[1.26]	[2.98]	[2.20]
120-months	0.20	0.26*	0.13	0.13	0.30**	0.31**	0.23*	0.28*	0.17	0.17	0.34**	0.35**
	[1.60]	[1.71]	[0.84]	[0.77]	[2.50]	[2.11]	[1.85]	[1.80]	[0.99]	[0.91]	[2.35]	[2.06]
Expanding	0.23***	0.27***	0.29**	0.29*	0.38***	0.38***	0.20**	0.25***	0.31**	0.29*	0.39***	0.37**
	[2.96]	[3.65]	[2.21]	[1.81]	[3.23]	[2.63]	[2.53]	[3.34]	[2.24]	[1.67]	[3.31]	[2.50]

This table reports the high-minus-low spreads of RBP-, HBP-, and *All*-composite predictors across two estimation frameworks (weighted least squares (WLS) vs. ordinary least squares (OLS)) and different estimation window lengths. First, the dependent variable R_t^e is regressed on a lagged standardized predictor x_{t-1}, as in Equation (1). We run a standard univariate WLS or OLS cross-sectional regression. Second, the beta estimates are then smoothed using either a rolling window or an expanding window average starting from 36 months. The risk-adjusted performance is calculated using the Fama and French (2018) six-factor model and the Hou et al. (2021) five-factor q^5 model. t-statistics are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Figure 6: Chasing Performance: Alternative Estimation Windows, 1991-2022

The figure plots FF6- and Q5-alphas (and corresponding 99% confidence intervals) that represent the performance of a hypothetical performance chaser that invests in the exact same manner as described in Table 10 from Subsection 6.1. The grey shaded area is the 99% confidence interval computed using Newey and West (1987) standard errors with six lags. The y-axis is alpha per month, and the x-axis is the number of predictors. For presentation, we omit Q5-alphas at the 1Y rolling-window.



8 Conclusion

In this paper, we show that there is a noise diversification effect from averaging predictor rankings for future mutual fund performance. Individual fund predictors are most likely not uniquely valuable in capturing managerial skill or future fund performance. Therefore, when we combine individual predictors into a composite score, the aggregate predictor performs better as noise is diversified away. Our empirical results support this rationale since the composite predictor more precisely indicates which funds to avoid and, more importantly, which funds to buy for investors.

Our research is closely related to the work of DeMiguel et al. (2023) as well as Kaniel et al. (2023), which utilize state-of-the-art ML algorithms for fund performance prediction. Both papers show that ML is able to consistently differentiate high from low-performing mutual funds, before and after fees. We achieve the same goal using a much simpler method. First, we incorporate HBPs and demonstrate that they add sizeable value when predicting future fund returns and alphas. Second, we show that ML alone is not necessary to achieve these sizeable gains through combining individual predictors. The more predictors we include in a composite score, the better, on average, the performance becomes, hence, it is especially important to include as many predictors as possible. Lastly, both RBPs and HBPs matter for the prediction exercise of future mutual fund performance.

Our method of predicting fund performance also holds for large mutual funds, and it also successfully selects funds that significantly outperform (or underperform relative to) the average *passive* fund on the long side and the short side. This underscores the practical importance of our findings, as large funds are essential in fund selection strategies, and investors are primarily concerned with the long side of investing. Notably, results remain significant in different parts of the sample period. Altogether, our study suggests a new and distinct approach to predict mutual fund performance using a simple averaging technique of predictor variables.

References

Agarwal, V., S. Ruenzi, and F. Weigert. 2023. Unobserved performance of hedge funds. Journal of Finance, Forthcoming .

Amihud, Y., and R. Goyenko. 2013. Mutual fund's r^2 as predictor of performance. Review of Financial Studies 26:667–94.

Asness, C. S., A. Frazzini, and L. H. Pedersen. 2019. Quality minus junk. Review of Accounting Studies 24:34–112.

Barber, B. M., T. Odean, and L. Zheng. 2005. Out of sight, out of mind: The effects of expenses on mutual fund flows. Journal of Business 78:2095–120.

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

Bergstresser, D., J. M. Chalmers, and P. Tufano. 2008. Assessing the costs and benefits of brokers in the mutual fund industry. Review of Financial Studies 22:4129–56.

Berk, J. B., and R. C. Green. 2004. Mutual fund flows and performance in rational markets. Journal of Political Economy 112:1269–95.

Brown, S. J., and W. N. Goetzmann. 1995. Performance persistence. Journal of Finance 50:679–98.

Busse, J. A., A. Goyal, and S. Wahal. 2010. Performance and persistence in institutional investment management. Journal of Finance 65:765–90.

Busse, J. A., and P. J. Irvine. 2006. Bayesian alphas and mutual fund persistence. Journal of Finance 61:2251–88.

Carhart, M. M. 1997. On persistence in mutual fund performance. Journal of Finance 52:57–82.

Chan, L. K., H.-L. Chen, and J. Lakonishok. 2002. On mutual fund investment styles. Review of Financial Studies 15:1407–37.

Chauvet, M., and J. Piger. 2008. A comparison of the real-time performance of business cycle dating methods. Journal of Business & Economic Statistics 26:42–9.

Chen, A. Y., and T. Zimmermann. 2021. Open source cross-sectional asset pricing. Critical Finance Review .

Chen, H.-L., N. Jegadeesh, and R. Wermers. 2000. The value of active mutual fund management: An examination of the stockholdings and trades of fund managers. Journal of Financial and Quantitative Analysis 35:343–68.

Chen, J., H. Hong, M. Huang, and J. D. Kubik. 2004. Does fund size erode mutual fund performance? the role of liquidity and organization. American Economic Review 94:1276–302.

Chevalier, J., and G. Ellison. 1997. Risk taking by mutual funds as a response to incentives. Journal of Political Economy 105:1167–200.

Cochrane, J. H. 2011. Presidential address: Discount rates. Journal of Finance 66:1047–108.

Cohen, R. B., J. D. Coval, and L. Pástor. 2005. Judging fund managers by the company they keep. Journal of Finance 60:1057–96.

Cremers, K. M., and A. Petajisto. 2009. How active is your fund manager? a new measure that predicts performance. Review of Financial Studies 22:3329–65.

Cremers, M., and A. Pareek. 2015. Short-term trading and stock return anomalies: Momentum, reversal, and share issuance. Review of Finance 19:1649–701.

———. 2016. Patient capital outperformance: The investment skill of high active share managers who trade infrequently. Journal of Financial Economics 122:288–306.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance 52:1035–58.

De Bondt, W. F., and R. Thaler. 1985. Does the stock market overreact? Journal of Finance 40:793–805.

DeMiguel, V., L. Garlappi, and R. Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Review of Financial Studies 22:1915–53.

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

Doshi, H., R. Elkamhi, and M. Simutin. 2015. Managerial activeness and mutual fund performance. Review of Asset Pricing Studies 5:156–84.

Edelen, R. M. 1999. Investor flows and the assessed performance of open-end mutual funds. Journal of Financial Economics 53:439–66.

Elton, E. J., M. J. Gruber, and C. R. Blake. 1996. The persistence of risk-adjusted mutual fund performance. Journal of Business 133–57.

———. 2001. A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases. Journal of Finance 56:2415–30.

———. 2011. Holdings data, security returns, and the selection of superior mutual funds. Journal of Financial and Quantitative Analysis 46:341–67.

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

Evans, R. B. 2010. Mutual fund incubation. Journal of Finance 65:1581–611.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. Journal of Finance 25:383–417.

Fama, E. F., and K. R. French. 2010. Luck versus skill in the cross-section of mutual fund returns. Journal of Finance 65:1915–47.

———. 2018. Choosing factors. Journal of Financial Economics 128:234–52.

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

Fischer, B. R., and R. Wermers. 2012. Performance evaluation and attribution of security portfolios. Academic Press.

Frazzini, A., and L. H. Pedersen. 2014. Betting against beta. Journal of Financial Economics 111:1–25.

Gil-Bazo, J., and P. Ruiz-Verdú. 2009. The relation between price and performance in the mutual fund industry. Journal of Finance 64:2153–83.

Green, J., J. R. Hand, and X. F. Zhang. 2017. The characteristics that provide independent information about average us monthly stock returns. Review of Financial Studies 30:4389–436.

Grinblatt, M., and S. Titman. 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. Journal of Business 393–416.

———. 1992. The persistence of mutual fund performance. Journal of Finance 47:1977–84.

———. 1993. Performance measurement without benchmarks: An examination of mutual fund returns. Journal of Business 47–68.

———. 1994. A study of monthly mutual fund returns and performance evaluation techniques. Journal of Financial and Quantitative Analysis 29:419–44.

Grinblatt, M., S. Titman, and R. Wermers. 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. American Economic Review 85:1088–105.

Gruber, M. J. 1996. Another puzzle: The growth in actively managed mutual funds. Journal of Finance 51:783–810.

Gupta-Mukherjee, S. 2014. Investing in the "new economy": Mutual fund performance and the nature of the firm. Journal of Financial and Quantitative Analysis 49:165–91.

Harvey, C. R. 2017. Presidential address: The scientific outlook in financial economics. Journal of Finance 72:1399–440.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. Review of Financial Studies 29:5–68.

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

Hoberg, G., N. Kumar, and N. Prabhala. 2018. Mutual fund competition, managerial skill, and alpha persistence. Review of Financial Studies 31:1896–929.

Hou, K., H. Mo, C. Xue, and L. Zhang. 2021. An Augmented q-Factor Model with Expected Growth*. Review of Finance 25:1–41.

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

———. 2020. Replicating Anomalies. Review of Financial Studies 33:2019–133.

Huang, J., C. Sialm, and H. Zhang. 2011. Risk shifting and mutual fund performance. Review of Financial Studies 24:2575–616.

Hunter, D., E. Kandel, S. Kandel, and R. Wermers. 2014. Mutual fund performance evaluation with active peer benchmarks. Journal of Financial Economics 112:1–29.

Ippolito, R. A. 1989. Efficiency with costly information: A study of mutual fund performance, 1965–1984. Quarterly Journal of Economics 104:1–23.

Jacobs, H., and S. Müller. 2020. Anomalies across the globe: Once public, no longer existent? Journal of Financial Economics 135:213–30.

Jegadeesh, N. 1990. Evidence of predictable behavior of security returns. Journal of Finance 45:881–98.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945–1964. Journal of Finance 23:389–416.

Jensen, T. I., B. Kelly, and L. H. Pedersen. 2023. Is there a replication crisis in finance? Journal of Finance 78:2465–518.

Jones, C. S., and H. Mo. 2021. Out-of-sample performance of mutual fund predictors. Review of Financial Studies 34:149–93.

Jordan, B. D., and T. B. Riley. 2015. Volatility and mutual fund manager skill. Journal of Financial Economics 118:289–98.

Kacperczyk, M., S. V. Nieuwerburgh, and L. Veldkamp. 2014. Time-varying fund manager skill. Journal of Finance 69:1455–84.

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

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

———. 2008. Unobserved actions of mutual funds. Review of Financial Studies 21:2379–416.

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

Kelly, B. T., S. Pruitt, and Y. Su. 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134:501–24.

Kozak, S., S. Nagel, and S. Santosh. 2020. Shrinking the cross-section. Journal of Financial Economics 135:271–92.

Lewellen, J. 2015. The Cross-section of Expected Stock Returns. Critical Finance Review 4:1–44.

Lou, D. 2012. A Flow-Based Explanation for Return Predictability. Review of Financial Studies 25:3457–89.

Malkiel, B. G. 1995. Returns from investing in equity mutual funds 1971 to 1991. Journal of Finance 50:549–72.

Mamaysky, H., M. Spiegel, and H. Zhang. 2007. Improved forecasting of mutual fund alphas and betas. Review of Finance 11:359–400.

Morey, M. R. 2015. Predicting Mutual Fund Performance. In Mutual Funds and Exchange-Traded Funds: Building Blocks to Wealth. Oxford University Press.

Newey, W. K., and K. D. West. 1987. Hypothesis testing with efficient method of moments estimation. International Economic Review 777–87.

Pástor, L., and R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111:642–85.

Pástor, L., R. F. Stambaugh, and L. A. Taylor. 2015. Scale and skill in active management. Journal of Financial Economics 116:23–45.

———. 2017. Do funds make more when they trade more? Journal of Finance 72:1483–528.

Pedersen, L. H. 2018. Sharpening the arithmetic of active management. Financial Analysts Journal 74:21–36.

Pollet, J. M., and M. Wilson. 2008. How Does Size Affect Mutual Fund Behavior? Journal of Finance 63:2941–69.

Roll, R. 1978. Ambiguity when performance is measured by the securities market line. Journal of finance 33:1051–69.

Sharpe, W. F. 1966. Mutual fund performance. Journal of Business 39:119–38.

———. 1991. The arithmetic of active management. Financial Analysts Journal 47:7–9.

Simutin, M. 2014. Cash Holdings and Mutual Fund Performance*. Review of Finance 18:1425–64.

Sirri, E. R., and P. Tufano. 1998. Costly search and mutual fund flows. Journal of Finance 53:1589–622.

Stambaugh, R. F., and Y. Yuan. 2017. Mispricing factors. Review of Financial Studies 30:1270–315.

Wermers, R. 1999. Mutual fund herding and the impact on stock prices. Journal of Finance 54:581–622.

———. 2000. Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses. Journal of Finance 55:1655–95.

Yan, X. S. 2008. Liquidity, investment style, and the relation between fund size and fund performance. Journal of Financial and Quantitative Analysis 43:741–67.

Zheng, L. 1999. Is money smart? a study of mutual fund investors' fund selection ability. Journal of Finance 54:901–33.

Appendices

A Appendix Predictors Construction

This appendix describes in detail how we construct each mutual fund performance predictor considered in our study. We consider nine RBPs and ten HBPs documented in the mutual fund literature. Note that for HBPs, for a more convenient notation throughout this subsection, we omit the i subscript that is supposed to represent fund i. Each holdings-based measure covered below is computed on a fund (fund-portfolio) level for a given month t and aggregated across all stocks that are held by a fund (in a portfolio).

A.1 Fund age

We calculate the *Age* of a fund as the difference in months between the current month t and the month when the fund was first offered or started its first share class (denoted as *first_offer_dt* in CRSP MF). Subsequently, we take the natural logarithm of this difference.

A.2 Fund size

We calculate the *Size* of a fund as the natural logarithm of TNA under management, which are in millions of US dollars.

A.3 Fund flows

We calculate *Flows* as a monthly percentage growth in TNA, net of the internal growth assuming reinvestment of dividends and distributions:

$$Flows_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}(1+R_{i,t})}{TNA_{i,t-1}} = \frac{TNA_{i,t}}{TNA_{i,t-1}} - (1+R_{i,t}), \tag{4}$$

where $TNA_{i,t}$ is the total net assets of fund i at the end of month t and $R_{i,t}$ is the CRSP net return of fund i (i.e., return net of expenses).

A.4 Fund expenses

We calculate the *Exp* of a fund as the expense ratio divided by 12. The expense ratio is from CRSP MF; it is reported as an annual number that represents total shareholder charges.

A.5 Fund excess turnover

In their model, Pástor, Stambaugh, and Taylor (2017) use fund fixed effects to capture within-fund time variation in turnover. Thus, we consider standardized excess turnover (henceforth *STO*) to capture the abnormal trading of a given fund. To limit the effect of outliers, we trim monthly turnover at 1% before computing the *STO*. We compute *STO* of a fund as a monthly abnormal turnover in excess of the mean monthly turnover standardized by its standard deviation:

$$STO_{i,t} = \frac{TO_{i,t} - \mu^{TO}_{i,t:t-59}}{\sigma^{TO}_{i,t:t-59}}, \tag{5}$$

where $TO_{i,t}$ is fund i's monthly turnover (i.e., annual turnover ratio from CRSP MF divided by 12) trimmed at 1%, $\mu^{TO}_{i,t:t-59}$ is the mean turnover over the most recent 60 months (i.e., t:t-59), and $\sigma^{TO}_{i,t:t-59}$ is the sample standard deviation of turnover over the same period.

A.6 Volatility of fund returns

We calculate *Vola* of a fund as the sample standard deviation of monthly net returns over the most recent 12 months. Jordan and Riley (2015) use daily returns in their main result and monthly returns as a robustness check. To be consistent with other predictors in our paper, we use the monthly frequency of fund returns to compute *Vola*.

A.7 Realized Carhart's alpha

We estimate *CA* of a fund as the one-month realized Carhart's alpha:

$$\alpha_{i,t} = R^e_{i,t} - \hat{\beta}^{MKT}_{i,t-1} MKT_t - \hat{\beta}^{SMB}_{i,t-1} SMB_t - \hat{\beta}^{HML}_{i,t-1} HML_t - \hat{\beta}^{UMD}_{i,t-1} UMD_t, \tag{6}$$

where MKT_t, SMB_t, HML_t, and UMD_t are the returns of the three Fama-French and momentum factors in month t, and $\hat{\beta}^{MKT}_{i,t-1}$, $\hat{\beta}^{SMB}_{i,t-1}$, $\hat{\beta}^{HML}_{i,t-1}$, and $\hat{\beta}^{UMD}_{i,t-1}$ are the factor loadings of the net excess returns of a fund with respect to the FF4-factors estimated using the 36-month estimation window (with a minimum requirement of 30 observations) ending in month $t-1$.

A.8 t-statistic of Carhart's alpha

We estimate *Tstat* of a fund by obtaining the *t*-statistic of the intercept (alpha) from a regression where we regress fund net returns in excess of risk-free on the Fama–French-Carhart four-factors from Carhart (1997). The regression is estimated on a rolling basis with a window width set to 36 months (i.e., 3 years as in Elton, Gruber, and Blake (1996)). We require a fund to have at least 30 months of observations. *Tstat* is computed using Newey and West (1987) standard errors with six lags.

A.9 R-squared

We estimate *R2* of a fund by obtaining the R^2 from a regression where we regress fund net returns in excess of the risk-free rate on the Fama–French-Carhart four-factors from Carhart (1997). The regression is estimated on a rolling basis with a window width set to 24 months (i.e., 2 years as in Amihud and Goyenko (2013)). We require a fund to have at least 20 months of observations to compute *R2*.

A.10 Grinblatt and Titman performance measure

We measure *GT* of a fund as the difference between the return of the current portfolio and a historical portfolio:

$$GT_t = \sum_{j=1}^{N} (\tilde{w}_{j,t-1} - \tilde{w}_{j,t-k-1})\tilde{R}_{j,t}, \tag{7}$$

where $\tilde{w}_{j,t-1}$ is the portfolio weight of stock j held by the fund at the end of month $t-1$ (beginning of month t), $\tilde{w}_{j,t-k-1}$ is the weight of the same stock j lagged k months, and $\tilde{R}_{j,t}$ is return of stock j in month t. We follow Grinblatt and Titman (1993) to set $k = 12$, which corresponds to the one-year lagged portfolio weights.

A.11 Characteristic selectivity

We measure *CS* of a fund as the difference between the return of the current portfolio adjusted for the current return of corresponding characteristic-based benchmarks:

$$CS_t = \sum_{j=1}^{N} \tilde{w}_{j,t-1}(\tilde{R}_{j,t} - \tilde{R}_t^{b_{j,t-1}}), \tag{8}$$

where $\tilde{w}_{j,t-1}$ is the portfolio weight of stock j held by the fund at the end of month $t-1$ (beginning of month t), $\tilde{R}_{j,t}$ is buy-and-hold return of stock j in month t, and $\tilde{R}_t^{b_{j,t-1}}$ is the month t buy-and-hold value-weighted return of a characteristic-matched benchmark portfolio for stock j based on its characteristics at the beginning of the quarter (i.e., the most recently available quarter).

A.12 Characteristic timing

We measure *CT* of a fund as the difference between the current return of the most recent corresponding characteristic-based benchmarks and the current return of the historical corresponding characteristic-based benchmarks:

$$CT_t = \sum_{j=1}^{N} (\tilde{w}_{j,t-1} \tilde{R}_t^{b_{j,t-1}} - \tilde{w}_{j,t-k-1} \tilde{R}_t^{b_{j,t-k-1}}), \tag{9}$$

where $\tilde{w}_{j,t-1}$ is the portfolio weight of stock j held by the fund at the end of month $t-1$ (beginning of month t), $\tilde{R}_t^{b_{j,t-1}}$ is the month t buy-and-hold value-weighted return of a characteristic-matched benchmark portfolio for stock j based on its characteristics at the beginning of the quarter (i.e., most recently available quarter), $\tilde{w}_{j,t-k-1}$ is the weight of the same stock j lagged k months, and $\tilde{R}_t^{b_{j,t-k-1}}$ is the month t buy-and-hold value-weighted return of a characteristic-matched benchmark portfolio to stock j based on its characteristics at $t-k-1$ month.

A.13 Return gap

We measure *RG* of a fund as the difference between the reported net returns and the net return on a hypothetical buy-and-hold portfolio of the most recently disclosed portfolio holdings. The latter is computed as:

$$RH_t = \sum_{j=1}^{N} (\tilde{w}_{j,t-1} \tilde{R}_{j,t}), \tag{10}$$

where $\tilde{w}_{j,t-1}$ is the portfolio weight of stock j held by the fund at the end of month $t-1$ (beginning of month t), and $\tilde{R}_{j,t}$ is the month t return of stock j.

The final measure is:

$$RG_t = RF_t - (RH_t - Exp_t), \tag{11}$$

where RF_t is the month t net return of the fund, and Exp_t is the fund's monthly expenses, which are simply the annual expense ratio divided by 12.

A.14 Active share

We measure AS differently than Cremers and Petajisto (2009). We set our data requirements such that each predictor must be constructible using standard mutual funds databases mentioned in Sections 2 and 3, which do not include data on specific benchmarks. Hence, we use the market portfolio that includes all U.S.-based common stocks in the CRSP database as the ultimate benchmark for all funds in our sample. We compute AS as the share of a fund's portfolio holdings that differ from the market portfolio holdings:

$$AS_t = \frac{1}{2}\sum_{j=1}^{N}|\tilde{w}_{j,t} - \tilde{w}_{j,t}^{MKT}|, \tag{12}$$

where $\tilde{w}_{j,t}$ and $\tilde{w}_{j,t}^{MKT}$ are the month t portfolio weights of stock j in the fund and in the market index, respectively.

A.15 Active weight

We measure AW of a fund as the sum of absolute differences between a fund's actual portfolio weights and the weights of a hypothetical cap-weighted portfolio, i.e., the value weights of the most recent disclosed portfolio holdings:

$$AW_t = \frac{1}{2}\sum_{j=1}^{N}|\tilde{w}_{j,t} - \tilde{w}_{j,t}^{VW}|, \tag{13}$$

where $\tilde{w}_{j,t}$ and $\tilde{w}_{j,t}^{VW}$ are the month t portfolio weight of stock j in the fund and the weight had the manager market cap-weighted fund's equity portfolio, respectively.

A.16 Industry concentration index

We measure *ICI* of a fund as the sum of the squared differences between a fund's industry weights and the industry weights of the market portfolio:

$$ICI_t = \sum_{i=1}^{10} (\tilde{w}_{i,t} - \tilde{w}_{i,t}^{MKT})^2, \quad (14)$$

where $\tilde{w}_{i,t}$ and $\tilde{w}_{i,t}^{MKT}$ are the month t portfolio weights of industry i in the fund and in the market portfolio, respectively.

A.17 Risk shifting

We measure *RS* of a fund as the difference between the volatility inferred from a fund's most recent holdings and the past realized volatility of the fund's returns:

$$RS_t = \sigma_{t:t-35}^{H} - \sigma_{t:t-35}^{R}, \quad (15)$$

where $\sigma_{t:t-35}^{H}$ is the current holdings volatility computed as a square root of VAR(RH_t) as in Equation (10), and $\sigma_{t:t-35}^{R}$ is the realized volatility of the fund's returns. Both volatilities are computed as the sample standard deviation of the corresponding monthly returns series over the most recent 36 months (i.e., t:t-35).

A.18 Skill index

SI is computed in multiple steps. First, we compute the beta coefficient with respect to the market:

$$\beta_{j,t} = \frac{COV(\tilde{R}_{j,t:t-11}^{e}, \tilde{R}_{t:t-11}^{MKT})}{VAR(\tilde{R}_{t:t-11}^{MKT})}, \quad (16)$$

where $\tilde{R}^{e}_{j,t:t-11}$ is the return of stock j in excess of the risk-free rate and $\tilde{R}^{MKT}_{t:t-11}$ is the market excess return. $\beta_{j,t}$ is estimated with a rolling-window regression model, using monthly returns series over the most recent 12 months (i.e., t:t-11).

Timing measures how well a fund captures the systematic component of the stock return relative to the market:

$$Timing_t = \sum_{j=1}^{N} (\tilde{w}_{j,t} - \tilde{w}^{MKT}_{j,t})(\beta_{j,t}\tilde{R}^{MKT}_{t+1}), \tag{17}$$

where the first term is the deviation of a fund's weight in stock j from stock j's weight in the market portfolio, and the second term is the systematic component of stock j's return. Since the term $\tilde{R}^{MKT}_{t+1}$ is unknown at time t, *Timing* will be positive for a fund that overweights (underweights) stock j anticipating an increase (decrease) in the market return.

Picking measures how well a fund captures the idiosyncratic component of the stock return relative to the market:

$$Picking_t = \sum_{j=1}^{N} (\tilde{w}_{j,t} - \tilde{w}^{MKT}_{j,t})(\tilde{R}^{e}_{j,t+1} - \beta_{j,t}\tilde{R}^{MKT}_{t+1}). \tag{18}$$

Similarly, the terms $\tilde{R}^{e}_{j,t+1}$ and $\tilde{R}^{MKT}_{t+1}$ are unknown at time t. Hence, *Picking* will be positive for a fund that overweights (underweights) stocks with subsequently high (low) idiosyncratic returns.

Finally, we measure *SI* of a fund as the equally weighted combination of the fund's $Timing$ and $Picking$ abilities:

$$SI_t = 0.5 \cdot Timing_{t-1} + 0.5 \cdot Picking_{t-1}. \tag{19}$$

Note that $Timing$ and $Picking$ are lagged, so the predictor is based on information available at time t. In Kacperczyk, Nieuwerburgh, and Veldkamp (2014) the weight on $Timing$ (wt) is equal to the real-time recession probability of Chauvet and Piger (2008). The weight on $Picking$ is $1 - wt$, where wt can take values between 0 and 1. We deviate from the

Kacperczyk, Nieuwerburgh, and Veldkamp (2014) in terms of the weighting scheme. The reason is additional data requirements outside of the scope of our study mentioned in Sections 2 and 3.

A.19 Fund duration

We measure *FD* of a fund as the average number of months that a fund holds its current portfolio, weighted across the percentages of stock owned by the fund over the last five years:

$$FD_t = \sum_{j=1}^{N} \tilde{w}_{j,t} \left[\sum_{t=T-W}^{T-1} \left(\frac{(T-t-1)\Delta_{j,t}}{H_j + B_j} \right) + \frac{(W-1)H_j}{H_j + B_j} \right], \tag{20}$$

where $\tilde{w}_{j,t}$ is the month t portfolio weights of stock j in the fund, $\Delta_{j,t}$ is the percentage of total shares outstanding of stock j bought or sold by a fund between month $t-1$ and t (for buys $\Delta_{j,t} > 0$ and for sells $\Delta_{j,t} < 0$), B_j is the total percentage of shares of stock j bought by a fund between $t = T-W$ and $t = T-1$, and H_j is the percentage of total shares outstanding of stock j held by a fund at time $t = T-W$. Cremers and Pareek (2015, 2016) choose $W = 20$ quarters, i.e., five years. Hence, we set $W = 60$ months accordingly as we perform the analysis on monthly frequency (Section 3 explains our set of assumptions required to reconcile the frequency of holdings from quarterly to monthly).

B Appendix Data Sample Summary Statistics

Table B1: Sample Summary Statistics, 1985-2022

Year	Number of funds	Fund-month observations	Mean **R** (%)	SD **R** (%)	Mean TNA	SD TNA	Mean Exp (%)	Mean TO (%)
1985	208	2 056	2.17	4.14	324	513	0.93	72.83
1986	252	2 512	1.05	5.06	392	726	0.98	73.65
1987	297	2 993	0.28	9.53	400	867	1.03	76.53
1988	308	3 230	1.14	3.52	342	763	1.11	86.97
1989	342	3 536	2.01	3.57	380	922	1.23	75.76
1990	353	3 747	-0.30	5.91	394	1 007	1.27	72.84
1991	438	4 430	2.72	5.15	428	1 134	1.28	82.44
1992	510	5 040	0.79	3.35	466	1 295	1.27	74.39
1993	644	6 041	1.12	3.12	496	1 544	1.24	72.43
1994	721	7 347	-0.11	3.38	520	1 738	1.23	75.87
1995	865	8 664	2.33	2.82	620	2 205	1.24	80.57
1996	1 021	10 126	1.53	4.11	767	2 608	1.28	88.42
1997	1 196	12 530	1.91	5.03	928	3 127	1.30	92.52
1998	1 316	13 755	1.39	7.47	1 146	3 930	1.28	92.69
1999	1 498	15 151	2.14	5.95	1 439	5 100	1.29	95.15
2000	1 757	17 575	0.16	8.45	1 530	5 252	1.30	113.16
2001	1 837	19 235	-0.75	7.71	1 211	4 173	1.35	111.86
2002	1 861	19 994	-1.99	6.32	998	3 505	1.40	108.60
2003	1 899	20 439	2.55	3.97	913	3 086	1.41	98.17
2004	1 916	21 163	1.08	3.50	1 051	3 535	1.36	88.90
2005	1 933	21 422	0.66	3.38	1 102	3 688	1.33	86.14
2006	1 884	21 372	1.05	3.11	1 210	4 030	1.30	88.07
2007	1 996	21 817	0.57	3.42	1 329	4 428	1.24	87.83
2008	2 049	21 761	-3.72	7.50	1 070	3 608	1.22	99.59
2009	1 941	21 338	2.65	6.82	824	2 669	1.22	98.13
2010	1 824	20 337	1.65	5.86	1 027	3 172	1.19	83.16
2011	1 794	19 721	-0.07	5.61	1 205	3 586	1.16	76.69
2012	1 669	19 000	1.22	3.62	1 202	3 186	1.14	68.91
2013	1 624	17 994	2.44	2.98	1 434	3 794	1.11	67.37
2014	1 591	17 895	0.71	3.41	1 638	4 341	1.10	65.51

2015	1 611	17 821	-0.12	4.07	1 661	4 461	1.08	65.40
2016	1 558	17 307	1.04	4.28	1 598	4 456	1.07	66.63
2017	1 579	17 553	1.43	2.05	1 851	5 812	1.05	63.87
2018	1 633	18 191	-0.63	4.88	2 138	7 397	1.02	62.35
2019	1 631	18 168	2.16	4.49	2 426	9 292	1.01	60.18
2020	2 111	21 336	2.05	8.40	2 308	9 061	0.98	63.47
2021	2 051	22 545	1.78	3.91	2 818	11 188	0.95	56.50
2022	1 803	20 869	-1.52	6.96	2 492	9 817	0.94	52.49
Mean	1 218	13 118	0.94	4.96	1 053	3 411	1.15	78.56
Total	4 416	564 082	.	.	.	.	.	.

R is the monthly return, TNA is the total net assets in millions of US dollars, Exp is the annual expense ratio, and TO is the annual turnover ratio. Mean values are the arithmetic average of all observations in a given year, and SD is the standard deviation, respectively. The sample period is 1985-2022. The number of funds is the number of distinct funds in a given year.

C Appendix Additional Results

Table C1: Time-Series Regressions - Additional Risk Factors, 1988-2022, monthly returns

	HBP			RBP			All		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
Q5 factors									
Mkt	-0.07*	0.05**	-0.05	-0.01	0.04	-0.04	-0.04	0.07**	-0.04
	[-1.86]	[2.08]	[-1.30]	[-0.45]	[1.53]	[-1.38]	[-1.09]	[2.54]	[-1.05]
Me	0.10	0.06	0.12	0.03	0.05	0.01	0.12	0.09	0.13
	[1.34]	[1.01]	[1.54]	[0.40]	[0.89]	[0.22]	[1.25]	[1.61]	[1.41]
I/A	0.05	-0.07	0.07	-0.23***	-0.22***	-0.26***	-0.11	-0.19***	-0.09
	[0.40]	[-0.75]	[0.67]	[-4.39]	[-5.62]	[-4.35]	[-0.98]	[-2.82]	[-0.85]
Roe	-0.05	-0.02	-0.08	-0.11*	-0.12***	-0.02	-0.13	-0.10**	-0.09
	[-0.58]	[-0.35]	[-0.75]	[-1.79]	[-2.92]	[-0.38]	[-1.58]	[-2.19]	[-0.98]
Eg	0.07	-0.09	0.05	0.07	-0.00	0.11*	0.13	-0.03	0.13
	[0.58]	[-1.30]	[0.43]	[1.07]	[-0.03]	[1.65]	[1.19]	[-0.37]	[1.28]
Additional Risk Factors									
LIQUI	0.02			0.08*			0.04		
	[0.35]			[1.79]			[0.71]		
STREV		-0.60***			-0.20***			-0.54***	
		[-10.02]			[-7.35]			[-11.38]	
LTREV		0.12			-0.07			0.07	
		[1.53]			[-1.18]			[1.00]	
BAB			-0.06			-0.01			-0.08*
			[-1.02]			[-0.24]			[-1.71]
QMJ			0.11			-0.18***			-0.00
			[0.93]			[-2.59]			[-0.00]
Intercept									
α	0.27*	0.44***	0.28	0.24***	0.33***	0.32***	0.34**	0.50***	0.39**
	[1.66]	[3.48]	[1.57]	[3.12]	[3.96]	[3.74]	[2.39]	[4.27]	[2.44]
Details									
$\bar{R}^2$	0.01	0.50	0.01	0.17	0.30	0.16	0.03	0.51	0.04
T	384	384	384	384	384	384	384	384	384

This table reports the time-series analysis of the RBP-, HBP-, and *All*-composite predictors. We regress the high-minus-low spread of the three composite predictors on various factor models. Our baseline model is the Fama–French six-factor model (α^{FF6}) of Fama and French (2018); then we extend the model by adding the liquidity factor (LIQUI) from Pástor and Stambaugh (2003), the long-run reversal effect (LTREV) from De Bondt and Thaler (1985), the stock returns reversal at short horizons (STREV) from Jegadeesh (1990), the betting-against-beta (BAB) from Frazzini and Pedersen (2014), and the quality-minus-junk (QMJ) from Asness, Frazzini, and Pedersen (2019). *t*-statistics, in square brackets, are computed using Newey and West (1987) standard errors with six lags. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Figure C1: Performance of All Possible Combinations of RBPs and HBPs, 1991-2022

The figure plots the distributions of FF6-alphas and corresponding t-stats of high-minus-low spreads, which are a function of the number of predictors in a composite predictor.

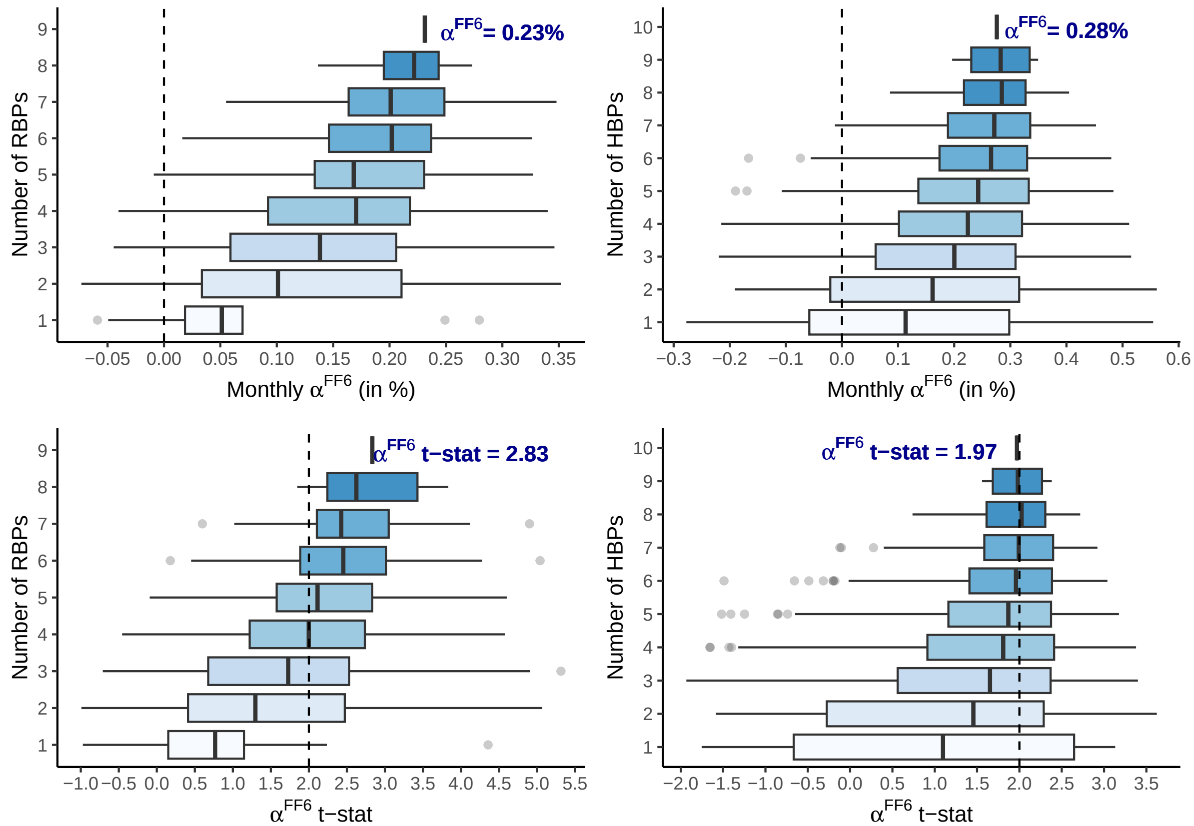
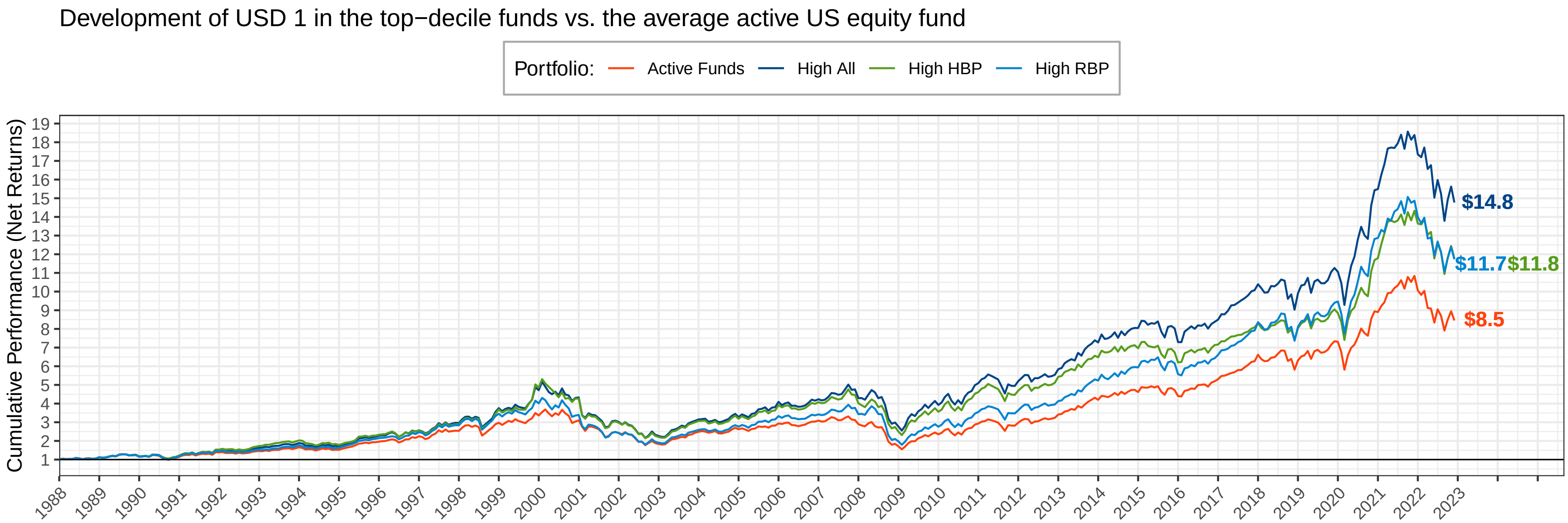


Table C2: Summary of All Possible Combinations of RBPs and HBPs, 1991-2022

	α^{FF6}				*t-stat*			
${}_nC_k$	Mean	Median	SD	IQR	Mean	Median	SD	IQR
Possible Combinations of RBPs								
$\binom{9}{1} = 9$	0.07	0.05	0.12	0.05	0.97	0.77	1.57	1.00
$\binom{9}{2} = 36$	0.11	0.10	0.11	0.18	1.44	1.30	1.49	2.06
$\binom{9}{3} = 84$	0.13	0.14	0.10	0.15	1.66	1.73	1.34	1.85
$\binom{9}{4} = 126$	0.16	0.17	0.09	0.13	1.98	2.00	1.14	1.52
$\binom{9}{5} = 126$	0.18	0.17	0.07	0.10	2.17	2.11	1.00	1.26
$\binom{9}{6} = 84$	0.19	0.20	0.07	0.09	2.42	2.45	0.94	1.13
$\binom{9}{7} = 36$	0.20	0.20	0.06	0.08	2.51	2.43	0.85	0.94
$\binom{9}{8} = 9$	0.22	0.22	0.04	0.05	2.79	2.63	0.71	1.19
$\binom{9}{9} = 1$	0.23	0.23	0.00	0.00	2.83	2.83	0.00	0.00
Possible Combinations of HBPs								
$\binom{10}{1} = 10$	0.13	0.11	0.26	0.36	0.95	1.10	1.82	3.31
$\binom{10}{2} = 45$	0.15	0.16	0.20	0.34	1.12	1.46	1.50	2.56
$\binom{10}{3} = 120$	0.18	0.20	0.17	0.25	1.37	1.65	1.32	1.81
$\binom{10}{4} = 210$	0.21	0.22	0.16	0.22	1.56	1.81	1.14	1.50
$\binom{10}{5} = 252$	0.22	0.24	0.14	0.20	1.67	1.87	0.96	1.21
$\binom{10}{6} = 210$	0.25	0.27	0.12	0.16	1.80	1.96	0.79	0.98
$\binom{10}{7} = 120$	0.26	0.27	0.10	0.15	1.91	1.99	0.62	0.81
$\binom{10}{8} = 45$	0.27	0.28	0.08	0.11	1.94	2.03	0.49	0.69
$\binom{10}{9} = 10$	0.28	0.28	0.06	0.10	1.98	1.98	0.33	0.58
$\binom{10}{10} = 1$	0.28	0.28	0.00	0.00	1.97	1.97	0.00	0.00

This table reports summary statistics of the high-minus-low FF6-spreads across all possible combinations used to construct a composite predictor using only RBPs or HBPs. For example, the first row reports the mean, median, standard deviation (SD), and interquartile range (IQR) of the H-L spread of nine composite predictors constructed using only one RBP. Similarly, the second row refers to thirty-six composite predictors constructed with only two RBPs at a time. ${}_nC_k$, where n is the number of predictors considered and k is the number of predictors used to construct the joint predictor. t-statistics are computed using Newey and West (1987) standard errors with six lags.

Figure C2: Cumulative Performance (Net Returns), 1988-2022



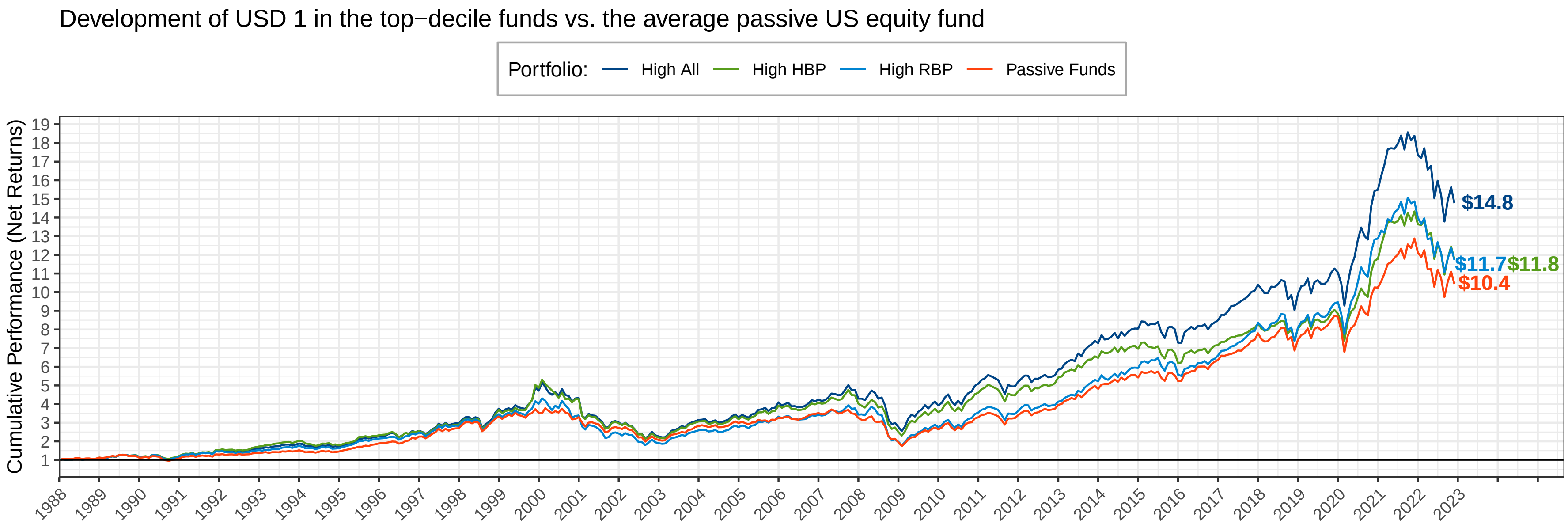


Figure C3: Cumulative Abnormal Returns of Prediction Deciles for RBPs, 1991-2022
The figure shows the cumulative risk-adjusted performance of the average individual return-based predictor (on the left-hand side) against the performance of the RBP-composite (on the right-hand side). To obtain the left side of the plot, we simply compute, for each decile, the average alpha across nine RBPs. The right side of the plot shows the performance of deciles based on a single sorting variable that is the RBP-composite score. The risk-adjusted performance is calculated using the Fama and French (2018) six-factor model.

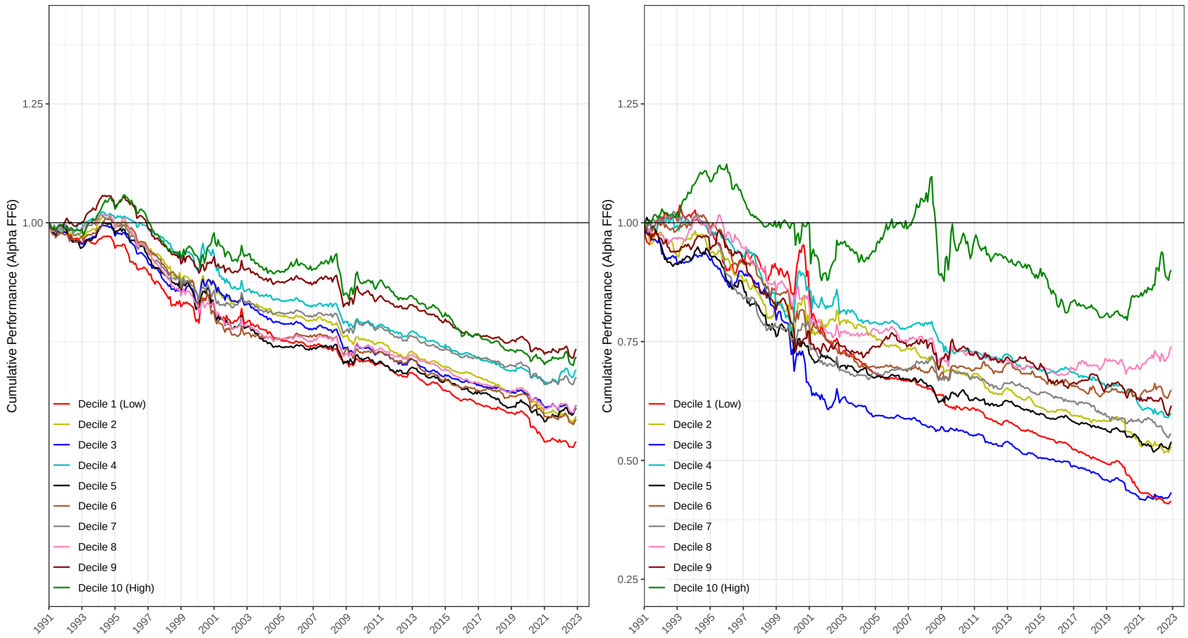
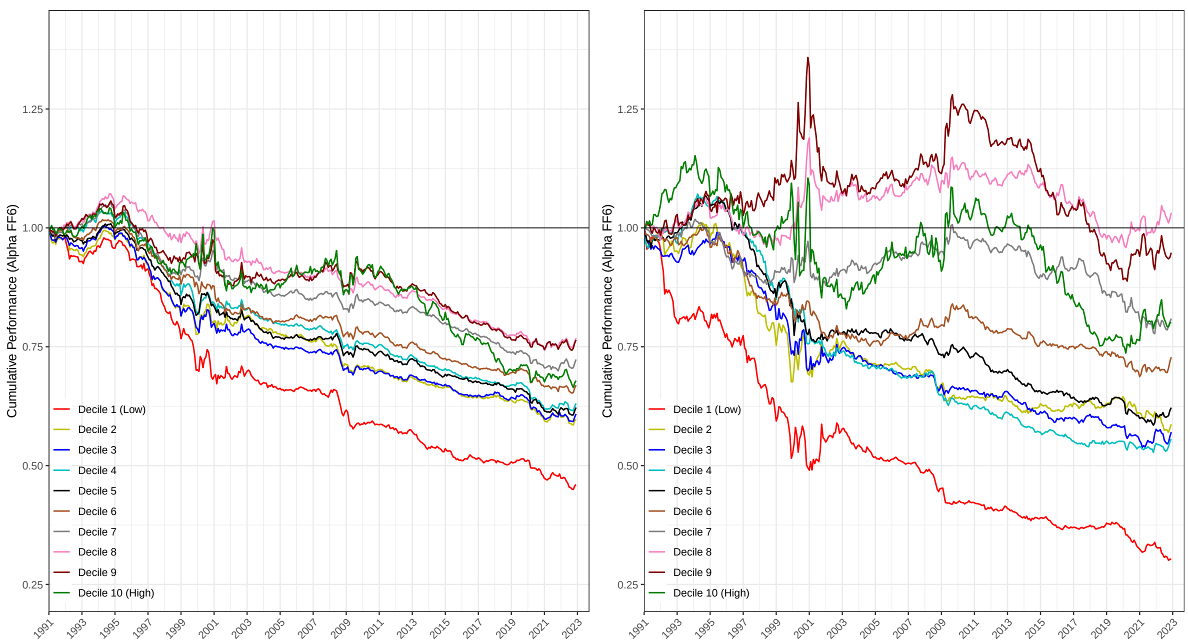


Figure C3 plots the cumulative risk-adjusted performance of the average individual return-based predictor (on the LHS) against the performance of the composite RBP. To obtain the left side of the plot, we simply compute, for each decile, the average alpha across nine RBPs. The right side of the plot shows the performance of deciles based on a single sorting variable that is the RBP-composite score. Clearly, the spread between the bottom and top decile portfolios is more pronounced when we construct the aggregate predictor that contains information across nine RBPs.

Moving to the set of HBPs (Figure C4), we observe a similar pattern that appears to be more pronounced. In fact, the underperformance of the bottom decile is more pronounced for both the average HBP (left-hand-side chart) and the HBP-composite score (right-hand-side chart). However, the top decile is not the best performing one and is frequently below the 8th and the 9th deciles.

Figure C4: Cumulative Abnormal Returns of Prediction Deciles for HBPs, 1991-2022
The figure shows the cumulative risk-adjusted performance of the average individual holdings-based predictor (on the left-hand side) against the performance of the HBP-composite (on the right-hand side). To obtain the left side of the plot, we simply compute, for each decile, the average alpha across ten HBPs. The right side of the plot shows performance of deciles based on a single sorting variable that is the HBP-composite score. The risk-adjusted performance is calculated using the Fama and French (2018) six-factor model.



cfr working paper series

centre for financial research
cologne

cfr working papers are available for download from www.cfr-cologne.de.

2026

No.	Author(s)	Title
26-01	S. Müller, N. Pugachyov, F. Weigert	Forecasting Mutual Fund Performance – Combining Return-Based with Portfolio Holdings-Based Predictors

2025

No.	Author(s)	Title
25-11	D. Aobdia, G. Köchling, P. Limbach, A. Yoon	Emissions Restatements after the SEC's Request for Public Input on Climate-Related Disclosures: Evidence from Carbon Disclosure Project Filings
25-10	V. Agarwal, J.-P. Gomez, K. Hosseini, M. Jha	ESG Metrics in Executive Compensation: A Multitasking Approach
25-09	V. Agarwal, P. Ghosh, N. Prabhala, H. Zhao	Animal Spirits on Steroids: Evidence from Retail Options Trading in India
25-08	V. Agarwal, S. Cao, S. Huang, M. Kim	Incentive Realignment: Mutual Funds' Influence on Executive Compensation Contracts
25-07	V. Agarwal, Y. E. Arisoy, T. Trinh	Eponymous Hedge Funds
25-06	D. Hess, F. Simon, S. Weibels	Interpretable Machine Learning for Earnings Forecasts: Leveraging High-Dimensional Financial Statement Data
25-05	B. F. Ballensiefen	Collateral Choice
25-04	C. Andres, F. Brochet, P. Limbach, N. Schumacher	Sell-Side Analysts with Accounting Experience
25-03	W. Bazley, G. Cici, J. Liao	Conflicts of Interest among Affiliated Financial Advisors in 401(k) Plans: Implications for Plan Participants
25-02	A. T. Maître, N. Pugachyov, F. Weigert	Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns
25-01	N. Käfer, M. Mörke, F. Weigert, T. Wiest	A Bayesian Stochastic Discount Factor for the Cross-Section of Individual Equity Options

2024

No.	Author(s)	Title
24-06	V. Beyer, T. Bauckloh	Non-Standard Errors in Carbon Premia
24-05	C. Achilles, P. Limbach, M. Wolff, A. Yoon	Inside the Blackbox of Firm Environmental Efforts: Evidence from Emissions Reduction Initiatives
24-04	Ivan T. Ivanov, T. Zimmermann	The “Privatization” of Municipal Debt
24-03	T. Dyer, G. Köchling, P. Limbach	Traditional Investment Research and Social Networks: Evidence from Facebook Connections
24-02	A. Y. Chen, A. Lopez-Lira, T. Zimmermann	Does Peer-Reviewed Research Help Predict Stock Returns?
24-01	G. Cici, P. Schuster, F. Weishaupt	Once a Trader, Always a Trader: The Role of Traders in Fund Management

2023

No.	Author(s)	Title
23-08	A. Braun, J. Braun, F. Weigert	Extreme Weather Risk and the Cost of Equity
23-07	A. G. Huang, R. Wermers, J. Xue	“Buy the Rumor, Sell the News”: Liquidity Provision by Bond Funds Following Corporate News Events
23-06	J. Dörries, O. Korn, G. J. Power	How Should the Long-term Investor Harvest Variance Risk Premiums?
23-05	V. Agarwal, W. Jiang, Y. Luo, H. Zou	The Real Effect of Sociopolitical Racial Animus: Mutual Fund Manager Performance During the AAPI Hate
23-04	V. Agarwal, B. Barber, S. Cheng, A. Hameed, H. Shanker, A. Yasuda	Do Investors Overvalue Startups? Evidence from the Junior Stakes of Mutual Funds
23-03	A. Höck, T. Bauckloh, M. Dumrose, C. Klein	ESG Criteria and the Credit Risk of Corporate Bond Portfolios
23-02	T. Bauckloh, J. Dobrick, A. Höck, S. Utz, M. Wagner	In partnership for the goals? The level of agreement between SDG ratings
23-01	F. Simon, S. Weibels, T. Zimmermann	Deep Parametric Portfolio Policies

This document only covers the most recent CFR Working Papers. A full list can be found at www.cfr-cologne.de.

centre for financial research
cfr/university of cologne
Albertus-Magnus-Platz
D-50923 cologne
fon +49(0)221-470-6995
fax +49(0)221-470-3992
kempf@cfr-cologne.de
www.cfr-cologne.de