Université de Neuchâtel
Faculté des sciences économiques

THREE ESSAYS ON MUTUAL FUND PERFORMANCE PREDICTORS AND CRYPTOCURRENCIES

PhD thesis submitted to the Faculty of Economics and Business

Institute of Financial Analysis

University of Neuchâtel

For the PhD degree in Finance

by

Nikolay PUGACHYOV

Approved by the dissertation committee:

Prof. Florian WEIGERT, University of Neuchâtel, thesis director

Prof. Michael HASLER, University of Neuchâtel, jury president

Prof. Sebastian MÜLLER, Technical University of Munich, Germany

Prof. Tim A. KRÖNCKE, Fachhochschule Nordwestschweiz

Defended on 22 January 2025

Faculté des sciences économiques
Avenue du 1er-Mars 26
CH-2000 Neuchâtel

www.unine.ch/seco

Université de Neuchâtel
Faculté des sciences économiques

IMPRIMATUR POUR LA THÈSE

Three Essays on Mutual Fund Performance Predictors and Cryptocurrencies

Nikolay PUGACHYOV

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

La Faculté des sciences économiques,

sur le rapport des membres du jury

Prof. Florian WEIGERT, Université de Neuchâtel, directeur de thèse

Prof. Michael HASLER, Université de Neuchâtel

Prof. Tim A. KRÖNCKE, Fachhochschule Nordwestschweiz (FHNW) Basel

Prof. Sebastian MÜLLER, Technical University of Münich, Germany

autorise l'impression de la présente thèse.

Neuchâtel, le 11 février 2025

Le doyen
Peter Fiechter

Faculté des sciences économiques
Avenue du 1er-Mars 26
CH-2000 Neuchâtel
Email: secretariat.seco@unine.ch
Tel: +41 32 718 15 00
www.unine.ch/seco

Acknowledgments

No journey is ever undertaken alone and this PhD is no exception. Along this path, I have been fortunate to be guided and supported by people who believed in me and inspired me.

I am deeply grateful to my PhD supervisor, Florian Weigert, for the guidance, encouragement and unwavering support throughout this journey. Thank you very much for greatly contributing to my growth as a researcher and believing in my potential even during the most challenging moments!

I owe a deep gratitude to Tim A. Kroencke, who first sparked my curiosity for research during his Research in Financial Analysis course in my master's studies. His passion for empirical asset pricing was truly inspiring and influential in shaping my academic interests. To David Ardia, to whom I am equally grateful for nurturing my interest in research, being a great master thesis supervisor and encouraging me to deepen my understanding. I am profoundly thankful to both for their mentorship that inspired me to pursue a PhD.

I am grateful to my dissertation committee, Michael Hasler, Tim A. Kroencke, Sebastian Mueller, and Florian Weigert, for dedicating their time and effort to reviewing this manuscript and providing valuable contributions to this thesis. To Sebastian Mueller, for collaboration, guidance and great research ideas.

To the University of Neuchatel, for providing the infrastructure, support and resources necessary for my research, including funding for courses and conferences worldwide. I was lucky to pursue my PhD at the University of Neuchatel.

I would like to especially thank Kira Facchinetti for her outstanding administrative help and warm support. To the professors Michel Dubois, Peter Fiechter, Michael Hasler, Tim A. Kroencke, Carolina Salva, and Florian Weigert for maintaining a stimulating research environment within the Institute of Financial Analysis.

A special thanks to Peter Fiechter for sharing insights into academia and for the engaging research discussions during lunch breaks and beers.

To my PhD colleagues — Arnaud (especially for being a great co-author!), Albian, Axel, Christophe, Constantin, Daniele, Emanuele, Fanol, Loïc, Louis, Manuel, Moreno (for discussions on mutual funds research), Nadine, Polina, Sidita, and Sofiya (for being a supportive sister and colleague) — thank you for fostering a friendly and collaborative atmosphere that made this journey more enjoyable.

I am beyond grateful to my family for being my greatest support! They have instilled in me the values of perseverance and the importance of lifelong learning. I dedicate this thesis to the memory of my grandparents, whose resilience, wisdom, and unwavering belief in me continue to inspire my journey.

Abstract

The first chapter introduces 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.

The second chapter utilizes the non-standard errors approach to identify the set of robust mutual fund performance predictors. Menkveld et al. (2024) formulate the concept of non-standard errors, which is a measure of the uncertainty that stems from the evidence-generating process, i.e., the variation generated by the methodological choices made by researchers. I show that the differences in research design lead to a substantial influence on the statistical inference when forecasting future mutual fund performance. Moreover, identifying critical design choices and measuring their importance facilitates the prediction exercise of future mutual fund performance and sheds light on the numerous mixed results in the literature.

The third chapter investigates how investors' abnormal attention affects the cross-section of cryptocurrency returns in the period from 2018 to 2022. We capture abnormal attention using the (log) number of Twitter posts on individual cryptocurrencies on the current day minus a 30-day average. Our results reveal that abnormal attention is positively associated with contemporaneous and one-day ahead crypto performance. Among the different Twitter tweets, return predictability arises due to *Ticker*-tweets from investors, but not due to tweets from the cryptocurrency channel. These *Official*-tweets, however, are able to forecast technological innovations on the blockchain.

Keywords: Mutual funds, US fund performance, performance prediction, performance predictors, composite predictor, non-standard errors, cryptocurrencies, Bitcoin, Twitter attention, textual sentiment.

Résumé

Le premier chapitre présente une méthode simple mais puissante pour améliorer la prédiction de la performance des fonds communs de placement en combinant des prédicteurs individuels en un prédicteur composite. Cette approche composite intègre des informations provenant de 19 prédicteurs bien établis de la littérature, basés sur le rendement et sur les avoirs en portefeuille. Elle identifie efficacement les fonds du décile supérieur qui surperforment les fonds du décile inférieur de 4.56% par an, corrigés du risque. En outre, il obtient une surperformance statistiquement significative pour les investissements en fonds long-only par rapport à la moyenne des fonds actifs et passifs. Les prédicteurs basés sur le rendement (par exemple, l'alpha du fonds et la statistique t de l'alpha) et les prédicteurs basés sur les avoirs (par exemple, l'indice de compétence et la pondération active) contribuent de manière égale au succès du prédicteur composite.

Le deuxième chapitre utilise l'approche des erreurs non standards pour identifier l'ensemble des prédicteurs robustes de la performance des fonds communs de placement. Menkveld et al. (2024) formule le concept d'erreurs non standards, qui est une mesure de l'incertitude qui découle du processus de production de preuves, c'est-à-dire la variation générée par les choix méthodologiques faits par les chercheurs. Je montre que les différences dans la conception de la recherche ont une influence substantielle sur l'inférence statistique lors de la prévision de la performance future des fonds communs de placement. En outre, l'identification des choix critiques en matière de conception et la mesure de leur importance facilitent l'exercice de prévision de la performance future des fonds communs de placement et éclairent les nombreux résultats mitigés de la littérature.

Le troisième chapitre étudie comment l'attention anormale des investisseurs affecte la section transversale des rendements des crypto-monnaies entre 2018 et 2022. Nous capturons l'attention anormale en utilisant le nombre (log) de messages Twitter sur les crypto-monnaies individuelles le jour en cours moins une moyenne de 30 jours. Nos résultats révèlent que l'attention anormale est positivement associée à la performance contemporaine et à celle à un jour des crypto-monnaies. Parmi les différents tweets, la prévisibilité des rendements est due aux tweets *Ticker* des investisseurs, mais pas aux tweets du canal des crypto-monnaies. Ces tweets *Officiel* sont toutefois capables de prévoir les innovations technologiques sur la blockchain.

Mots-clés: Fonds communs de placement, performance des fonds américains, prédiction de per-

formance, prédicteurs de performance, prédicteur composite, erreurs non standard, cryptocurrencies, Bitcoin, attention de Twitter, sentiment textuel.

Contents

List of Tables

List of Figures

Chapter I

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

(with Sebastian Müller and Florian Weigert)

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 per-

[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.

formance (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.

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,

[3]As RBPs, we select a fund's age, size, flows, expense ratio, turnover, volatility, alpha, t-statistic of alpha and R^2. 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.

our aim is to obtain a diversified and less noisy proxy of (hidden) manager skill, which 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 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.

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

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-

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

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 I.1 lists alpha predictors covered in our study.

[Insert Table I.1 here]

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

[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.

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 (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 alphas improve

[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).

[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.

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.

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

[10] Similar evidence is documented by Agarwal, Ruenzi, and Weigert (2024) in the hedge fund literature.

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*, and *EDY* (where *E* = *Equity*, *D* = *Domestic*, *S* = *Sector*, *C* = *Cap-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

[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.

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%.

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 I.A.

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

[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.

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 I.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 I.B for details on the predictors' construction.

We provide summary statistics of the RBPs and HBPs in this study in Table I.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.

[Insert Table I.2 here]

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 I.1 shows smoothed slopes computed using Equations (I.1) and (I.2), which provides evidence that for many predictors the sign frequently varies over time, and hence, a static sign is not optimal.

[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).

[Insert Figure I.1 here]

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,

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

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:

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

Second, we multiply the standardized predictor by $+1$ or -1 depending on the sign of the $\bar{\beta}$ from (I.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 I.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.

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 I.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.

[Insert Table I.4 here]

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 (I.1) and (I.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 I.1.

5.2.1. RBP-composite score

We first combine nine RBPs into a composite score. Table I.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 I.5, I.6, and I.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.

[Insert Table I.5 here]

5.2.2. HBP-composite score

We repeat the analysis for our sample of HBP-predictor variables. Table I.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.

[Insert Table I.6 here]

5.3. Combining all predictors

Table I.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 long-short 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.

[Insert Table I.7 here]

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]

[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/.

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

[Insert Table I.8 here]

Next, we investigate how individual predictors contribute to the joint forecasting power of the *All*-composite predictor. Figure I.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.

[Insert Figure I.2 here]

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

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

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 I.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 I.C. Figures I.C1 and I.C2 show similar results when the RBPs and HBPs are considered separately. Overall, the *All*-composite score results in the most sizeable gains.

[Insert Figure I.3 here]

Table I.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.

[Insert Table I.9 here]

6. Empirical Results: More Predictors, Better Performance

Expected returns are noisy, and forecasting errors have a significant impact on ex-post performance; hence, $\frac{1}{N}$ 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 $\frac{1}{N}$ 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 (I.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 I.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 I.2, and individual fund predictors are not uniquely valuable in capturing managerial skill or future fund performance, as shown in Figure I.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.

[Insert Table I.10 here]

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 I.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 I.4, which again documents the benefits 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.

[Insert Table I.11 here]

[Insert Figure I.4 here]

Figure I.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).

[Insert Figure I.5 here]

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.

7. Robustness Checks

In this section, we perform robustness checks of our main findings. Table I.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).

[Insert Table I.12 here]

Next, we test alternative ways to standardize the individual predictors before aggregating them. Table I.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.

[Insert Table I.13 here]

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 I.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.

[Insert Table I.14 here]

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 I.6 demonstrates that previously discussed findings from Table I.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.

[Insert Figure I.6 here]

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.

Table I.1: **Mutual Fund Performance Predictors Considered in the Study**

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.

	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*)	(+)

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

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.

	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

Figure I.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 (I.1). Second, these beta estimates are then smoothed using an expanding window average starting from 36 months.

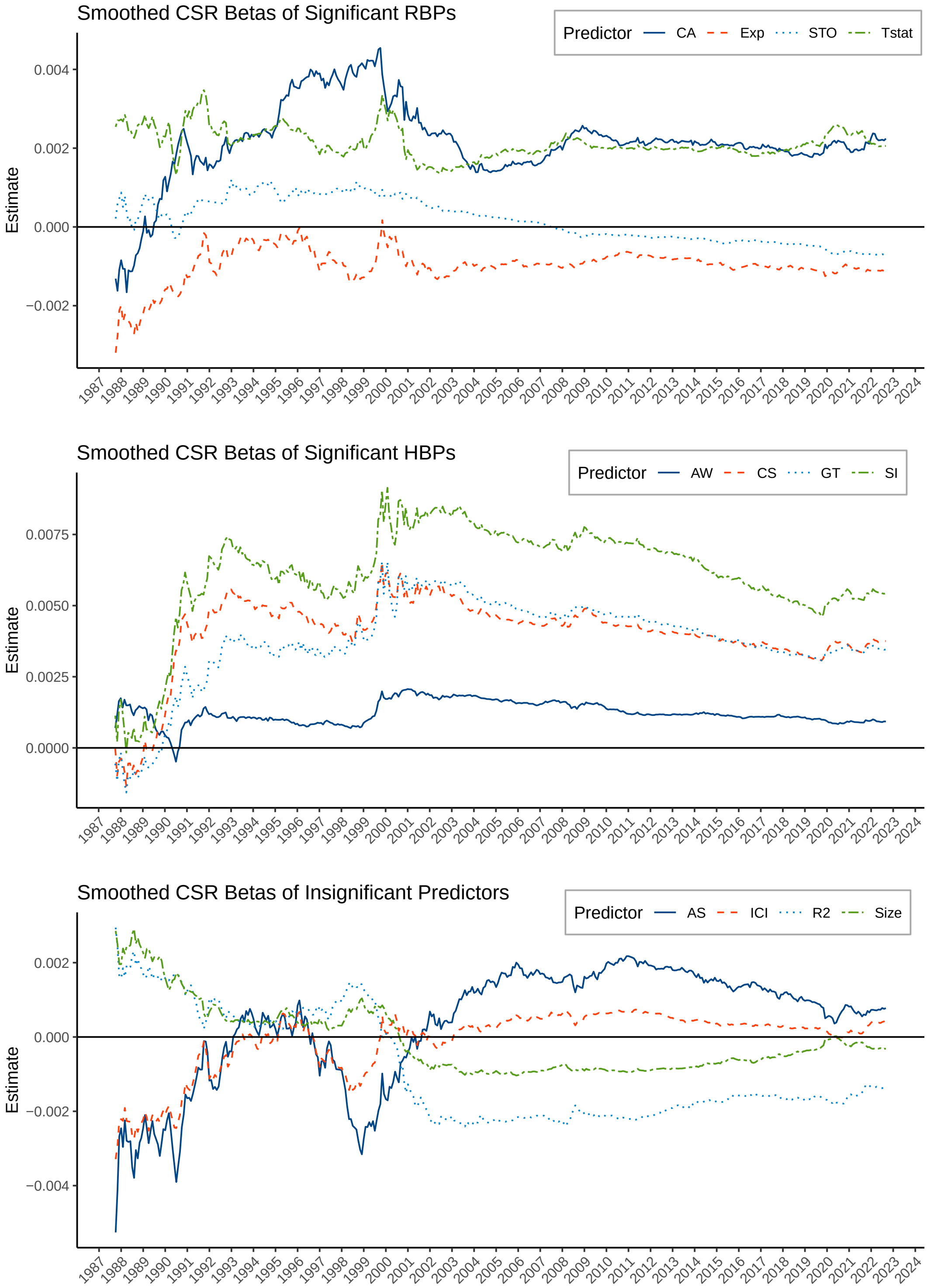


Table I.3: **Correlations of RBPs and HBPs, 1985-2022, monthly returns**

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.

	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

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

This table reports a 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.

	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]

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

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.

	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]

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

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.

	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]

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

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.

	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]

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

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.

	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

Figure I.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.

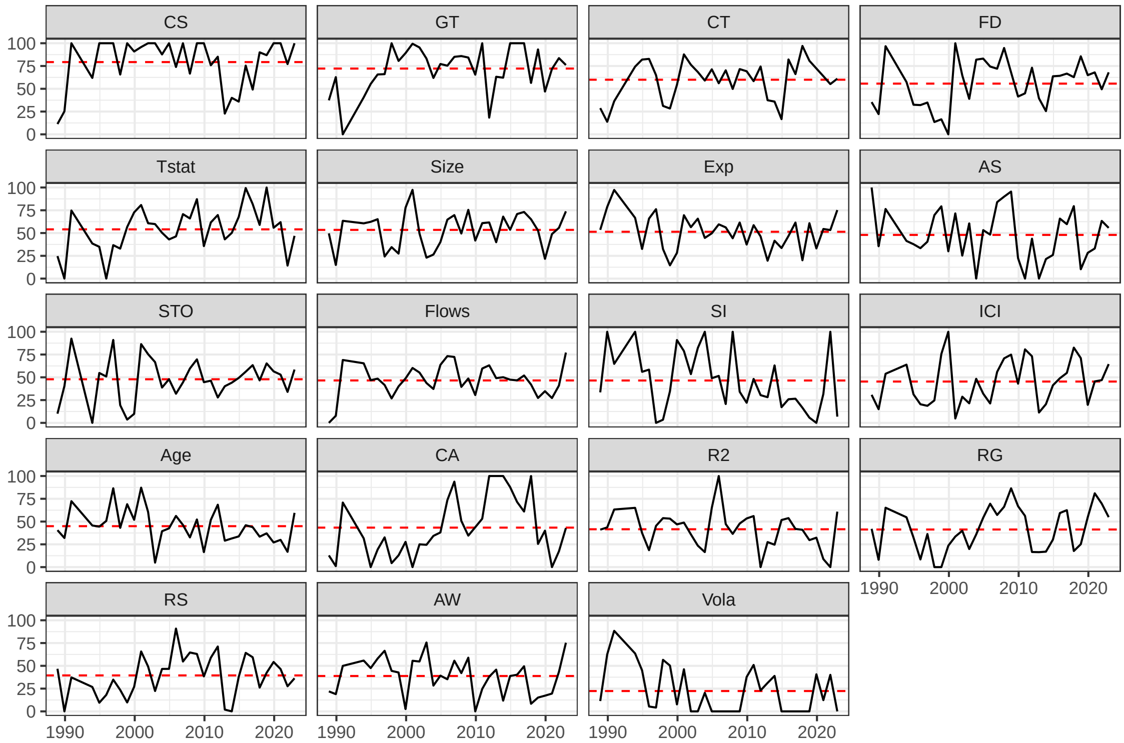


Figure I.3: **Cumulative Abnormal Returns of Prediction Deciles for *All*, 1991-2022**

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.

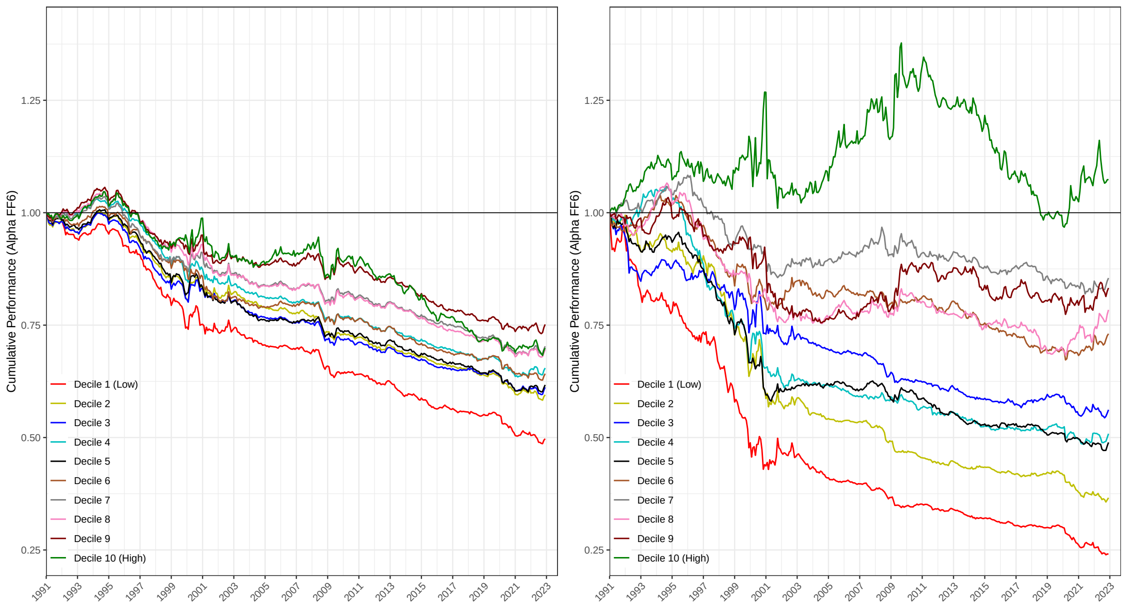


Table I.9: **Quantifying Noise Reduction Benefits, 1991-2022**

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.

	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]

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

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.

	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]

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

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.

	α^{FF6}				*t-stat*			
${}_nC_k$	**Mean**	**Median**	**SD**	**IQR**	**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

Figure I.4: **Performance of All Possible Combinations of Predictors, 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.

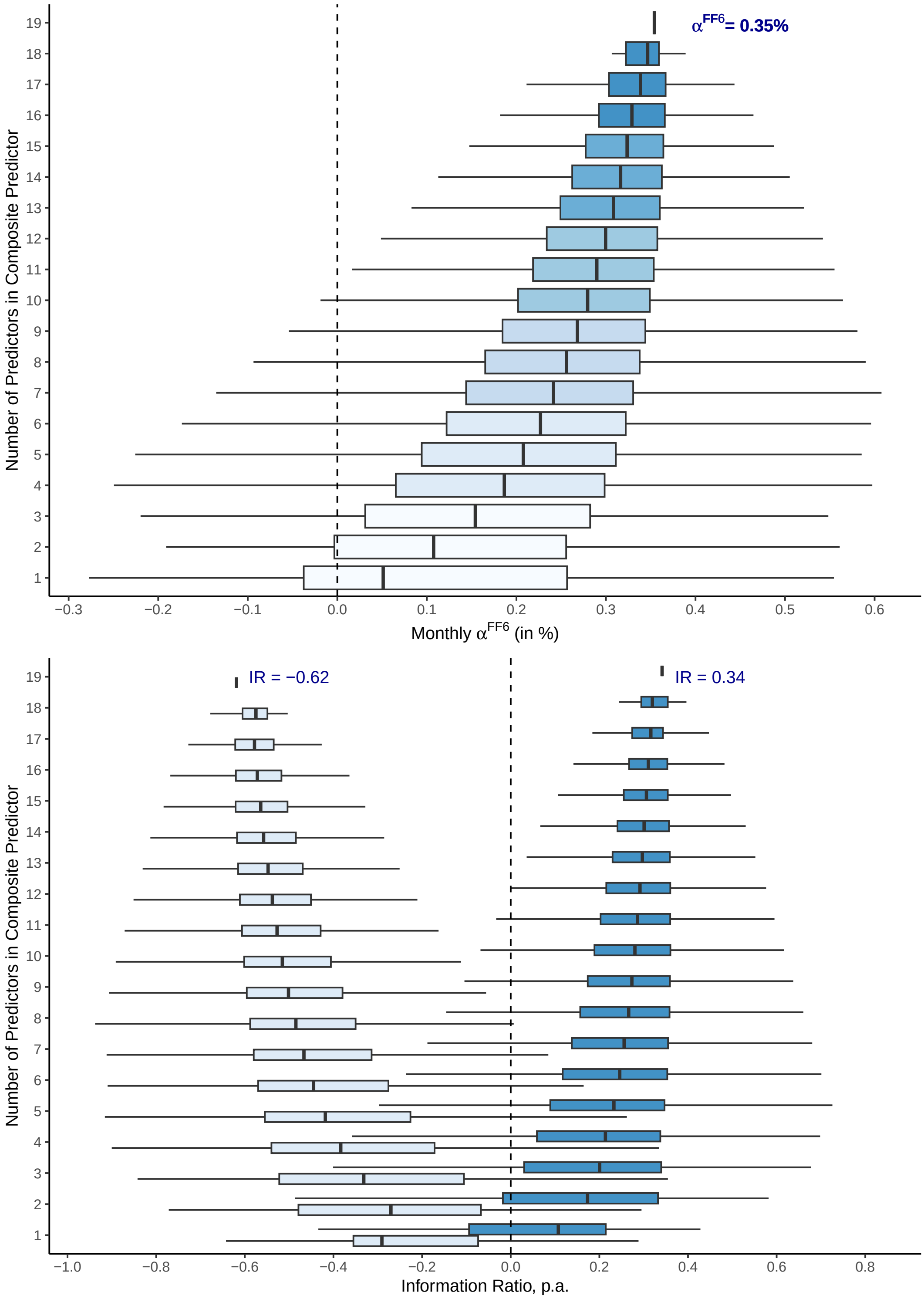


Figure I.5: **Relative Contribution of Predictors to the Performance, 1991-2022**

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.

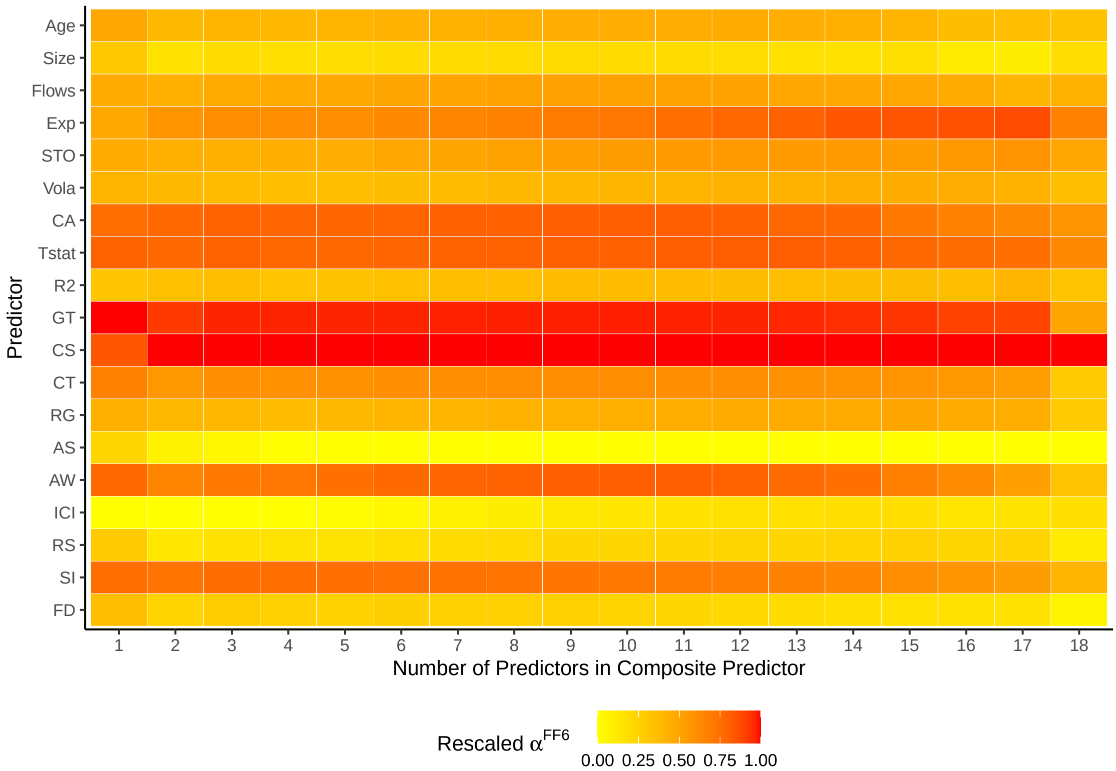


Table I.12: **Robustness Tests of the Composite Predictors, 1988-2022**

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.

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]

Table I.13: **Alternative Composite Predictors, 1988-2022**

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.

	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]

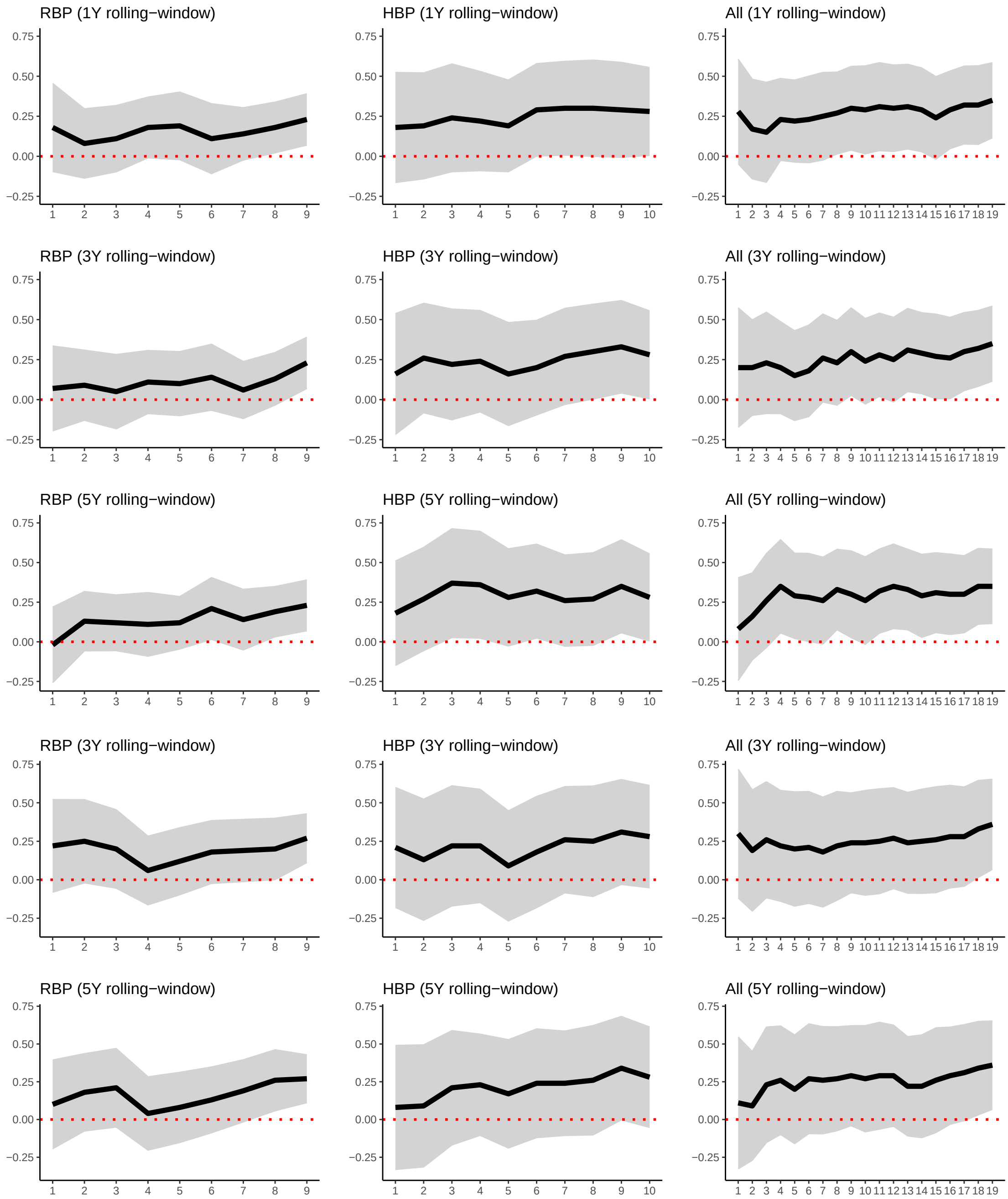
Table I.14: **Slope Estimates of Predictors: Altering Weighting Scheme and Estimation Window Length, 1988-2022**

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 (I.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.

	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]

Figure I.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 I.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.



Appendix

I.A. Data Sample Summary Statistics

Table I.A1: **Sample Summary Statistics, 1985-2022**

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.

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	.	.	.	.	.	.

I.B. 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).

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.

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

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

$$\begin{aligned} Flows_{i,t} &= \frac{TNA_{i,t} - TNA_{i,t-1}(1 + R_{i,t})}{TNA_{i,t-1}} \\ &= \frac{TNA_{i,t}}{TNA_{i,t-1}} - (1 + R_{i,t}), \end{aligned} \tag{I.4}$$

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).

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.

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:

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

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.

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*.

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

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

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 $t-1$.

***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.

***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*.

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:

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

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.

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:

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

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).

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:

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

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.

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:

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

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:

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

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.

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:

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

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.

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:

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

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.

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:

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

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.

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:

$$RS_t = \sigma^H_{t:t-35} - \sigma^R_{t:t-35}, \tag{I.15}$$

where $\sigma^H_{t:t-35}$ is the current holdings volatility computed as a square root of $\text{VAR}(RH_t)$ as in Equation (I.10), and $\sigma^R_{t:t-35}$ 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).

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

$$\beta_{j,t} = \frac{COV(\tilde{R}^e_{j,t:t-11}, \tilde{R}^{MKT}_{t:t-11})}{VAR(\tilde{R}^{MKT}_{t:t-11})}, \tag{I.16}$$

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:

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

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:

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

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:

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

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.

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:

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

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).

I.C. Additional Results

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

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 Hou et al. (2021) five-factor q^5 model; 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.

	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

Figure I.C1: **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.

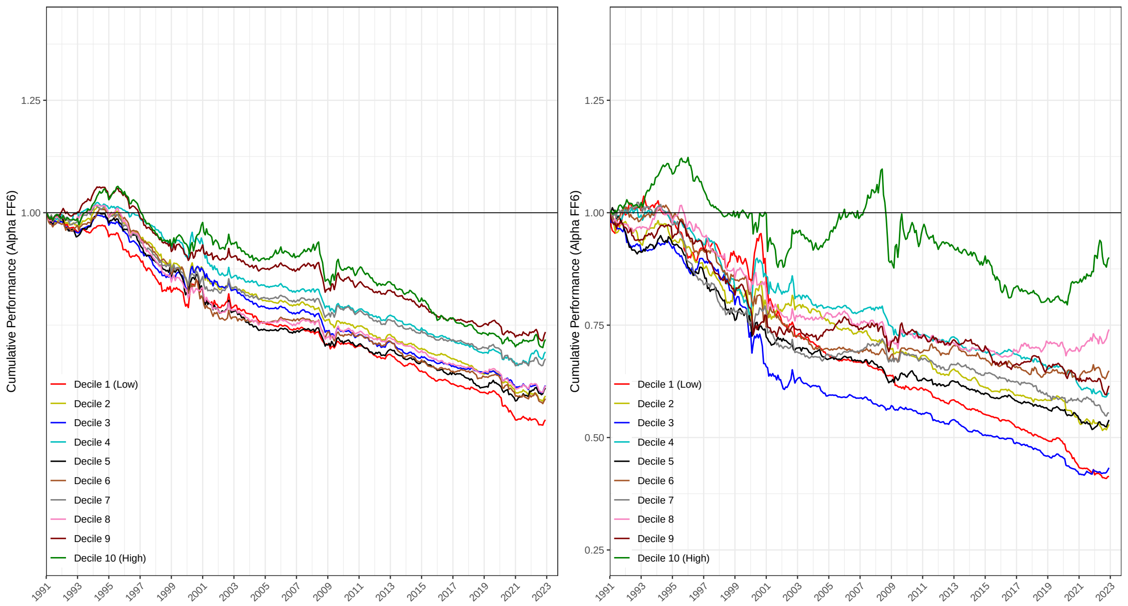
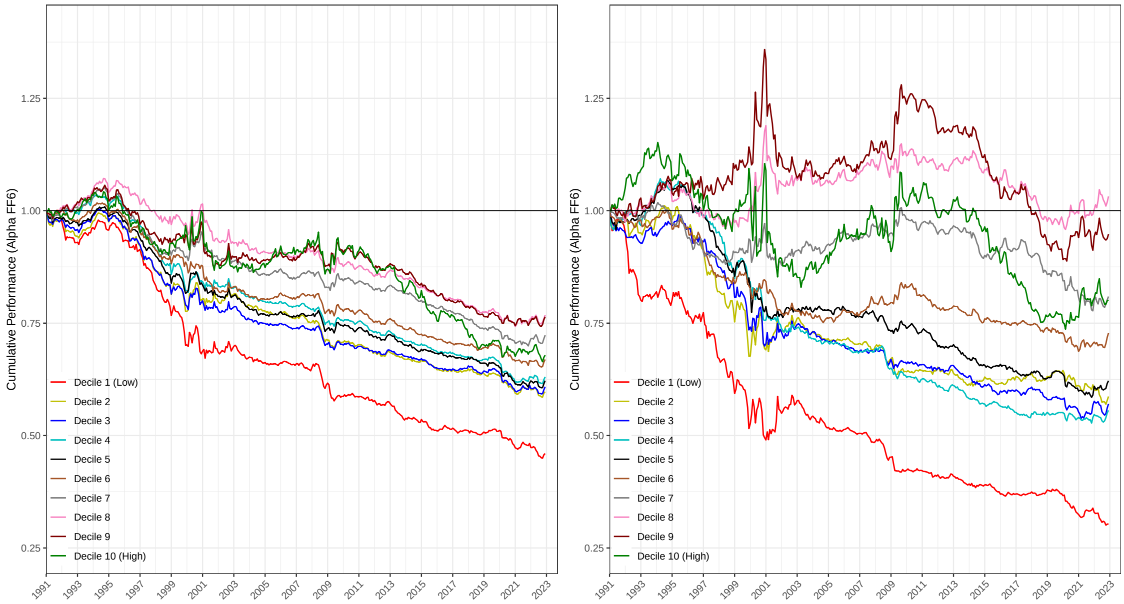


Figure I.C1 plots the cumulative risk-adjusted performance of the average individual return-based predictor (on the LHS) against the performance of the composite RBP. To obtain the left side of the plot, we simply compute, for each decile, the average alpha across nine RBPs. The right side of the plot 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.

Figure I.C2: **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.



Moving to the set of HBPs (Figure I.C2), 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.

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

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.

	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

Figure I.C3: **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.

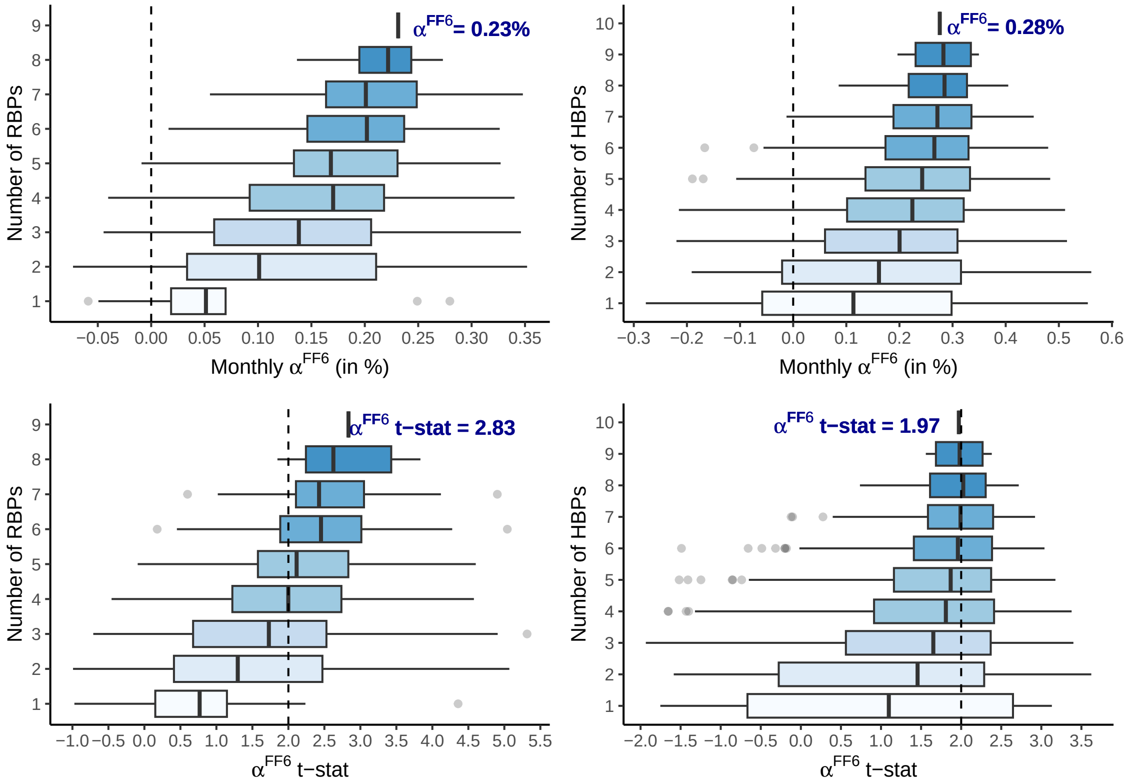
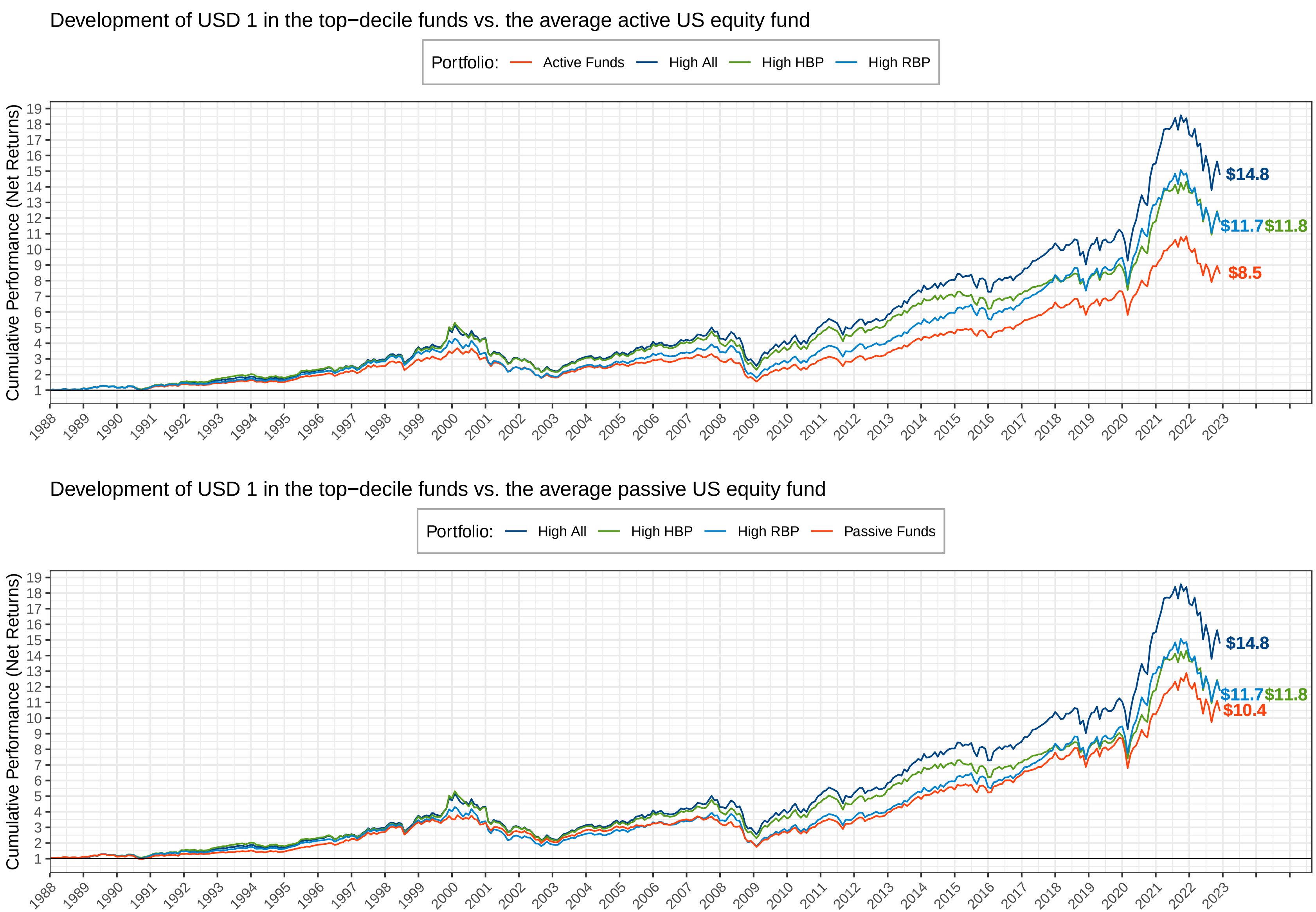


Figure I.C4: **Cumulative Performance (Net Returns), 1988-2022**



Chapter II

Non-Standard Errors in Mutual Fund Performance Predictors

1. Introduction

Mutual funds play a significantly important role in financial markets. The U.S. market for actively managed funds has exceeded 13.7 trillion US$ AUM, as of the end of 2023. Given the industry's importance, it is not surprising that the financial literature has proposed numerous measures to differentiate between well- and poor-performing mutual funds ex-ante. Moreover, academic research on mutual funds exerts an influence on the industry, and mutual fund performance predictors proposed by academics are being used to select fund managers and allocate investors' capital. However, not all researchers agree on the usefulness of certain measures. Consequently, numerous mixed results are open for interpretation.

In this study, I attempt to identify a set of robust predictors of mutual fund performance, which is of high practical value for both asset managers and investors. I intend to achieve this by answering the following questions:

1. How large are NSEs in mutual fund performance predictors?
2. To what extent are predictors of mutual fund performance sensitive to methodological choices in research design, specifically predictor construction choices?
3. Which methodological choices induce the largest variation in estimated premia?
4. Which predictors are most robust to methodological choices?

My analysis of methodological uncertainty in mutual fund performance is inspired by the studies of Walter, Weber, and Weiss (2024) and Jones and Mo (2021). The former work provides a comprehensive analysis of methodological uncertainty in portfolio sorts for stock return anomalies. The latter work studies the out-of-sample predictability of mutual fund alpha predictors. I follow a procedure similar to Jones and Mo (2021) for selecting predictors to be included in my study and base my analysis on the open-source tools provided by Walter, Weber, and Weiss (2024).[1]

While Walter, Weber, and Weiss (2024) investigate portfolio sorts, my focus is on the effect of methodological uncertainty induced by predictor construction. Unlike the majority of sorting variables in stock return anomalies, the construction of mutual fund performance predictors introduces an additional level of uncertainty. Frequently, stocks are sorted on some firm characteristic that is observable (such as the value of market equity, past stock returns, or book-to-market, to name a few). In contrast, mutual fund performance predictors have to be estimated (such as a fund's alpha, t-statistic of alpha, fund's R^2, etc.).

As a motivating example, consider the interplay between the benchmark-choice problem of Roll (1978) and the non-standard errors in factor models identified by Soebhag, Van Vliet, and Verwijmeren (2024). The former work demonstrates that regression-based performance measurement is sensitive to benchmark choice, while the latter demonstrates that the performance of factor models is sensitive to research design.

Now let us consider a fund's alpha, t-statistic of alpha, and R^2, which are typically estimated from the four-factor Carhart (1997) model. As proposed in the original papers, a fund's alpha and corresponding t-statistic are estimated using the prior 36 months of returns (Carhart, 1997), while the R^2 is based on the prior 24 months of returns (Amihud and Goyenko, 2013). Arguably, the difference between estimates from a regression based on the 24 and 36 months of returns should not be distorting. However, there are many alternative factor models that can be used to estimate these predictors. For instance, a combination of just four prominent factor models and three estimation windows leads to 12 possible ways to construct each predictor.[2]

Ultimately, the intersection of sample, predictor, and portfolio construction decisions may create substantial variation in estimates, creating room for conflicting results and significant variation in estimated premia.[3] In particular, I comprehensively investigate 4288 premia return series arising from the combination of predictor and portfolio construction choices that span over the period 1985 to 2022.

[1] I thank Patrick Weiss, Dominik Walter, and Rüdiger Weber for providing an open-source tool and a detailed R code for the replication of Walter, Weber, and Weiss (2024) on GitHub (https://github.com/patrick-weiss/PortfolioSorts_NSE/). Additionally, I followed some of the routines from Scheuch, Voigt, and Weiss (2023) in their code when constructing my analysis.

[2] The Carhart (1997) four-factor model, the Hou et al. (2021) five-factor q^5 model, the Fama and French (2018) six-factor model, and the Barillas and Shanken (2017) six-factor model. Three estimation windows for estimating factor-loadings: 24-, 36-, 60-month rolling windows.

[3] I follow Walter, Weber, and Weiss (2024) in using the term premium somewhat loosely to denote the expected long-short return spread generated by a portfolio sort.

My first main result is that the average monthly non-standard error is 0.12% and is statistically significant at the 5% level, which is comparable to 0.19% per month for stock return anomalies in Walter, Weber, and Weiss (2024). Furthermore, the non-standard errors are the same size as the standard errors, i.e., a ratio of 1.04.[4] Although the non-standard errors are sizable and statistically significant, most premia estimates (66%) are positive, alleviating concerns about the replicability of mutual fund predictors. Nonetheless, I show that methodology-induced variation matters for the prediction exercise of future mutual fund performance.

Furthermore, I demonstrate that predictor construction is an important source of methodologically induced variation and that predictor sensitivity to this source of variation varies across predictors. Moreover, I show that predictor construction can have a substantial impact on performance predictability even in cases where non-standard errors are insignificant and smaller in magnitude than standard errors. For instance, although non-standard errors are insignificant for *active weight* (Doshi, Elkamhi, and Simutin, 2015) and *active share* (Cremers and Petajisto, 2009), both measures are sensitive to predictor construction choice. When altering between the equal- and value-weighted benchmarks, results change substantially.[5]

On the other hand, there are predictors that are robust to predictor construction choices but for which methodologically induced variation stems from the portfolio construction choices. Consequently, it is important to consider both sources of methodological uncertainty.

Finally, I (i) identify a smaller subset of extremely robust mutual fund performance predictors, and (ii) analyze which practical considerations expand the set of well-performing predictors.[6] Firstly, *active weight* and *expense ratio* appear as safe bets. Their return spreads are stable and extremely persistent, thus allowing for infrequent rebalancing, which makes these measures appropriate for retail investors.[7] Secondly, given that an investor has access to the latest information and can rebalance at least semiannually, *skill index*, the Grinblatt and Titman (1993) performance measure, *characteristic selectivity*, and *past alpha* are suitable for a mu-

[4]To put these results into perspective with related studies on non-standard errors, the magnitude is generally lower than ratios of 1.60 in Menkveld et al. (2024), 1.06 in Soebhag, Van Vliet, and Verwijmeren (2024), 1.10 in Walter, Weber, and Weiss (2024), 1.55 in Fieberg et al. (2024), and 1.59 in Chen, Hanauer, and Kalsbach (2024).

[5]This finding supports Frazzini, Friedman, and Pomorski (2016), who show that the ability of *active share* to predict mutual fund performance is sensitive to the benchmark type. Overall, *active weight* could be a more robust alternative to *active share* since Doshi, Elkamhi, and Simutin (2015) show that *active weight* is not subsumed or explained by the *active share*.

[6]It is worth mentioning that conclusions regarding well-performing predictors are broadly consistent with Müller, Pugachyov, and Weigert (2024).

[7]The spreads remain positive even for a 12-month sorting variable lag with annual rebalancing, which signals that *expense ratio* and *active weight* persist for up to two years.

tual funds selection strategy, as they consistently generate positive return spreads. Overall, the final set of results allows us to evaluate the practical aspects of a predictor-based trading strategy and highlights important practical considerations for investors when they intend to actively invest in mutual funds.

This paper is closely related to research of Menkveld et al. (2024) that formulate the concept of non-standard errors. The non-standard errors measure the uncertainty that stems from the evidence-generating process, i.e., variation in methodological choices made by researchers. Although non-standard errors is a novel concept in finance literature, it is gaining popularity and has already been applied in the context of factor models Soebhag, Van Vliet, and Verwijmeren (2024), stock return predictors Walter, Weber, and Weiss (2024), cryptocurrencies Fieberg et al. (2024), and machine learning in finance Chen, Hanauer, and Kalsbach (2024). My study is the first systematic attempt to examine the robustness of mutual fund performance predictors through the use of the non-standard errors approach.

The remainder of the paper is as follows: Section 2 describes the empirical methodology used to estimate methodological uncertainty in mutual fund performance predictors. Section 3 provides details on the sample construction. Section 4 answers the four key questions of this paper by presenting and discussing the main results. Finally, the last section concludes.

2. Empirical Methodology

The empirical methodology adopted from Walter, Weber, and Weiss (2024). I use the term "premium" to stay consistent with Walter, Weber, and Weiss (2024). For each predictor and each path, I compute a premium, which is effectively a long-short spread of a hypothetical trading strategy that invests in mutual funds based on a given predictor, as follows:

$$r_{v,p} = \frac{1}{T} \sum_{t=1}^{T} \left(r_{v,p,t}^{\text{Long}} - r_{v,p,t}^{\text{Short}} \right), \tag{II.1}$$

where $r_{v,p,t}^{\text{Long}}$ and $r_{v,p,t}^{\text{Short}}$ are the returns of the long and the short legs of a hypothetical trading strategy. I normalize the sign such that it yields a positive premium in line with the direction documented in the original paper.

In accordance with Menkveld et al. (2024), to measure methodological uncertainty, I compute non-standard errors as the interquartile range of estimates across paths. For each predictor v, the non-standard error is:

$$\text{NSE}_v = Q_{0.75}(r_v) - Q_{0.25}(r_v), \tag{II.2}$$

where $Q_{0.75}(r_v)$ is the α-quantile of the distribution of r_v, which corresponds to the aggregated premia $r_{v,p}$ across all paths p.

As in Menkveld et al. (2024), to test whether methodological uncertainty is statistically significant for a given predictor v, I test the following hypothesis:

$$H_0 : r_{v,p} = Median(r_v) \qquad \forall p \in \{1, \ldots, P\}. \tag{II.3}$$

In accordance with Walter, Weber, and Weiss (2024) and Menkveld et al. (2024), non-standard errors for a given predictor v are considered significant if at least one of the tests in Equation II.3 rejects the null hypothesis. Note that throughout this paper, I do not adjust results for multiple hypothesis testing.

Lastly, to compare the size of non-standard errors to the size of standard errors, I compute the ratio of the standard deviation of the distribution of premia (across all paths) to the average standard error:

$$\text{Ratio}_v = \frac{\sqrt{\frac{1}{P-1}\sum_{p=1}^{P}\left(r_{v,p} - \overline{r_v}\right)^2}}{\frac{1}{P}\sum_{p=1}^{P}\sigma_{v,p}}, \tag{II.4}$$

where $\overline{r_v}$ is the average premium across all paths and $\sigma_{v,p}$ is the estimated time-series standard error of $r_{v,p}$.

2.1. *Performance predictors and predictor construction nodes*

In this section, I split predictors into four categories based on their construction methodology. For a detailed review of the predictors covered in this study, please see Müller, Pugachyov, and Weigert (2024) or Jones and Mo (2021).

2.1.1. *Fund characteristics*

Fund characteristics such as age (henceforth *Age*), size (henceforth *Size*), flows (henceforth *Flows*), and expenses (henceforth *Exp*) are observable and straightforward to compute, hence, they do not have predictor construction uncertainty. Thus, the number of predictor construction choices equals one for these predictors. For

Size (Chen et al., 2004) and *Flows* (Zheng, 1999), *high* values predict outperformance. For *Age* (Pástor, Stambaugh, and Taylor, 2015) and *Exp* (Gil-Bazo and Ruiz-Verdú, 2009), *low* values predict outperformance.

2.1.2. Regression-based predictors

The regression-based predictors are measures of future fund performance that are derived from factor models: fund's past alpha (henceforth *PA*), t-statistic of alpha (henceforth *Tstat*), and R^2 (henceforth *R2*). *High* values predict outperformance (Carhart, 1997; Elton, Gruber, and Blake, 1996), except for *R2* (Amihud and Goyenko, 2013).

These predictors are typically estimated from the four-factor Carhart (1997) model, which remains a workhorse model in mutual funds research (Kaniel et al., 2023). However, there are many other competing models. I consider four factor models and three estimation windows, which lead to 12 possible ways to construct each predictor. The factor models are: the Carhart (1997) four-factor model, the Hou et al. (2021) five-factor q^5 model, the Fama and French (2018) six-factor model, and the Barillas and Shanken (2017) six-factor model. The estimation windows for estimating factor-loadings are: 24-, 36-, and 60-month** rolling windows.

2.1.3. Portfolio holdings-based predictors

The portfolio holdings-based predictors utilize information on mutual fund portfolio holdings to serve as a better proxy for managerial skill. The estimation approaches vary, hence, the number of predictor construction choices varies as well.

Predictors with a single predictor construction choice are: characteristic selectivity (henceforth *CS*), fund duration (henceforth *FD*), return gap (henceforth *RG*), and risk shifting (henceforth *RS*). *High* values predict outperformance (Daniel et al., 1997; Cremers and Pareek, 2016; Kacperczyk, Sialm, and Zheng, 2008), except for *RS* (Kacperczyk, Sialm, and Zheng, 2005).

Both active share (henceforth *AS*) and active weight (henceforth *AW*) have two alternative versions for construction (against an equal- or value-weighted benchmark). *High* values predict outperformance (Cremers and Petajisto, 2009; Doshi, Elkamhi, and Simutin, 2015).

Grinblatt and Titman (1993) performance measure (henceforth *GT*) and characteristic timing (henceforth *CT*) can be estimated over different horizons to compute returns. The original research uses a 12-month horizon, while I additionally include 3-month and 6-month horizons, which are plausible alternatives. For both measures, *high* values predict outperformance (Grinblatt and Titman, 1993; Daniel et al., 1997).

The industry concentration index (henceforth *ICI*) can be computed using different industry definitions. I consider the three most common approaches to define industries: two-SIC code, Fama-French ten, or 48 industries. *High* values predict outperformance (Kacperczyk, Sialm, and Zheng, 2005).

Finally, the skill index (henceforth *SI*) relies on a market beta estimation, which is estimated using 1, 2, 3, or 5 years of monthly data. *High* values predict outperformance (Kacperczyk, Nieuwerburgh, and Veldkamp, 2014).

2.1.4. Other predictors

The last two predictors do not fit into any of the categories above. In their model, Pástor, Stambaugh, and Taylor (2017) use fund fixed effects to capture within-fund time variation in turnover. Thus, I consider standardized excess turnover (henceforth *STO*) to capture the abnormal trading of a given fund. I compute *STO* of a fund as the monthly abnormal turnover in excess of the mean monthly turnover standardized by its standard deviation, where both statistics can be computed over the last 24-, 36-, or 60-month period. For *STO*, *high* values predict outperformance.

The last predictor is the volatility of monthly net returns (henceforth *Vola*). Jordan and Riley (2015) use daily returns in their main result and monthly returns as a robustness check. I calculate *Vola* using monthly data for the last 12-, 24-, and 36-month periods. For *Vola*, *low* values predict outperformance.

2.2. Sample construction nodes

I isolate the impact of the sample construction choices. First, sample construction decisions are well motivated in the literature and their effect is documented. For instance, possible decision nodes could include controlling for the first offer date, which is known as the incubation bias documented by Evans (2010). Further, one could consider a fund size exclusion filter, i.e., whether to keep or exclude small size funds (TNA below 15 million US$). Once again, Elton, Gruber, and Blake (2001) document the effect of keeping small funds, whose returns tend to be biased upward in the CRSP MF database. Ultimately, I argue for a unified procedure by following Jones and Mo (2021) to construct a mutual funds sample. Therefore, I fix sample construction nodes to focus on the effect of the predictor and portfolio construction choices.

2.3. *Portfolio construction nodes*

Walter, Weber, and Weiss (2024) show empirically that the number of portfolios, the portfolio weighting scheme, the time lag between information arrival, i.e., sorting variable lag, and portfolio rebalancing frequency are the most impactful drivers of non-standard errors in portfolio sorts. Considering the importance of these choices, I account for the following portfolio construction decisions: Sorting variable lags of 1-, 3-, 6-, and 12-months; monthly, quarterly, semiannual, and annual rebalancing frequencies; quintile and decile portfolios; and lastly, equal- or value-weighting schemes. This implies 64 possible paths at the portfolio construction level (4x4x2x2). Therefore, for a predictor that has a single predictor construction choice, I have 64 premia return series and up to 768 return series for the regression-based predictors, for instance. All in all, I comprehensively investigate 4288 premia return series for 19 predictors arising from the combination of predictor and portfolio construction choices that span over the period 1985 to 2022.

3. Data and Sample Construction

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 are 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 Compustat North America database to obtain additional information on stock returns and accounting variables.

I follow Jones and Mo (2021) and apply data filters that are common among prominent studies that use holdings data (Daniel et al., 1997; Wermers, 2000; Kacperczyk, Sialm, and Zheng, 2005, 2008; Doshi, Elkamhi, and Simutin, 2015).

Most funds in the sample have multiple share classes. Once the filters are applied, share-class level data are merged to arrive at fund-level variables. Qualitative variables (such as the fund name or fund style) are aggregated by utilizing information from the largest share class in terms of TNA. Quantitative variables such as the net return, expense, and turnover ratios, are summed and weighted by the lagged TNA of the corresponding share classes. Other quantitative variables, such as TNA and NAV, are simply summed up. Once at the fund level, funds that are below US$15 million of TNA are dropped since returns on small funds are biased upwards in the CRSP MF database (Elton, Gruber, and Blake, 2001)). The final filter 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. The 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.

4. Results: Identifying Robust Mutual Fund Performance Predictors

In this section, I present my main results that closely address the four main research questions this study aims to answer. Each subsection answers one of the main questions that I investigate.

4.1. Non-standard errors across mutual fund performance predictors

I start by studying the magnitude of the non-standard errors across mutual fund performance predictors. For each of the 19 predictors, the premia are generated by varying methodological choices in predictor construction and portfolio construction. As a result, for each predictor, methodological variation results in 64 to 768 long-short spreads for predictors with 1 to 12 predictor construction choices, respectively. Figure II.1 presents the resulting distributions in the form of box plots, where boxes cover the interquartile range and thus correspond to the definition of non-standard errors from Equation II.2.

[Insert Figure II.1 here]

Clearly, Figure II.1 reveals that methodologically induced variation results in sizeable non-standard errors, which substantially differ across predictors. To quantify the magnitude of non-standard errors for different groups of predictors and provide statistical evidence, I conduct a set of tests in accordance with Walter, Weber, and Weiss (2024). Table II.1 provides key statistics for each predictor.

[Insert Table II.1 here]

Starting with the summary statistics, we observe that the four groups of predictors differ substantially in terms of the generated distributions. Fund characteristics can be considered the simplest set of predictors as they have a single construction choice, which likely explains the lowest kurtosis (2.44) among the four groups. In contrast, holdings-based predictors have the most extreme outliers, which are primarily on the right side of the premia distribution, i.e., positively skewed with a fat right tail ($Skew. = 0.86$ and $Kurt. = 5.46$). Surprisingly, the number of predictor construction choices does not necessarily lead to larger non-standard errors.

The magnitude of the non-standard errors is measured by the ratio of the size of non-standard errors to the size of standard errors. Among the predictor groups, fund characteristics and holdings-based predictors have the largest ratio of 1.13 for both, followed by regression-based predictors (*Ratio* = 1.00) and other predictors (*Ratio* = 0.42). Arguably, the regression-based predictors face the greatest methodological uncertainty as they have considerably more predictor construction choices compared to other predictor groups. Nonetheless, the relative number of significant deviations to the left and right of the median (*Left-right*) is among the lowest for the regression-based predictors. Additionally, 75% of their premia estimates are positive compared to 66%, 62%, and 15% for holdings-based predictors, fund characteristics, and other predictors, respectively.[8]

Regarding the size of non-standard errors within groups, there is considerable heterogeneity. For instance, *PA* has a ratio of 1.18, while two other regression-based predictors (*R2* and *Tstat*) have considerably lower ratios of 0.72 and 0.89, respectively. It is worth mentioning that the non-standard error ratio for *PA* is among the highest ratios across 19 predictors. This might be related to the fact that out-of-sample forecasts of mutual fund performance that are based on past alphas are likely driven primarily by estimation error (Mamaysky, Spiegel, and Zhang, 2007).

Speaking of which predictors are the most stable, both the visual (Figure II.1) and quantitative evidence (Table II.1) indicate *Exp*, *Size*, and *AW* as more robust predictors than others. All observations for these measures are entirely positive; *AW* and *Exp* have the highest monthly mean returns of 0.10% and 0.09%, respectively (0.05% for *Size*). Moreover, *AW* has by far the largest fraction of statistically significant estimates: 38%. Interestingly, these results may point towards the outperformance of high *AW*, low *Exp*, and smaller-size funds (for both *Exp* and *Size*, low values are associated with high performance).

The non-standard error ratios for *Exp* and *Size* are among the lowest: 0.55 and 0.32, respectively (1.01 for *AW*). Furthermore, *Exp* and *Size* face the lowest

[8]Here, I do not focus on the risk adjustment of the premia. Furthermore, one could expect the long-short spread to increase with risk adjustment. It is common for the risk-adjusted performance of mutual fund predictors to be largely driven by the short leg of a hypothetical strategy, i.e., it is far easier to identify funds that would underperform in the future rather than those that would outperform (Carhart, 1997).

methodological uncertainty in terms of predictor construction, i.e., zero degrees of freedom. Similarly, AW has a single degree of freedom. However, in the section to follow, we will see how even a single choice can induce substantial variation in the estimated premia.

Overall, the average monthly non-standard error is 0.12% and is statistically significant at the 5% level, which is comparable to 0.19% per month for stock return anomalies in Walter, Weber, and Weiss (2024). Furthermore, the non-standard errors are the same size as the standard errors, i.e., a ratio of 1.04. To put my results into perspective with related studies on non-standard errors, the magnitude is generally lower than ratios of 1.60 in Menkveld et al. (2024), 1.06 in Soebhag, Van Vliet, and Verwijmeren (2024), 1.10 in Walter, Weber, and Weiss (2024), 1.55 in Fieberg et al. (2024), and 1.59 in Chen, Hanauer, and Kalsbach (2024). Although the non-standard errors are sizable and statistically significant, most premia estimates (66%) are positive, alleviating concerns about the replicability of mutual fund predictors. Nonetheless, it is important to consider methodology-induced variation.

4.2. Detailed view on predictor construction

In this section, I examine to what extent predictors of mutual fund performance are sensitive to predictor construction choices. Figure II.2 illustrates how results might change when we move from a predictor-level analysis to the predictor-version level, where each version of a predictor arises from fixing a certain predictor construction choice.

[Insert Figure II.2 here]

One can see that predictor construction is an important source of methodologically induced variation. However, predictor sensitivity to this source of variation varies across predictors. For instance, *AS* has two alternative versions for construction (against an EW or VW benchmark), while *SI* has four (1, 2, 3, or 5 years to estimate β). Nonetheless, the distribution of *AS* varies substantially between the two alternatives, while the distribution of estimated premia for *SI* is stable across the four alternatives. Likewise, regression-based predictors such as *PA* and *R2* seem to be more stable in response to the interaction of a factor model and estimation window length choices; meanwhile, *Tstat* flips the sign when estimated using the five-factor q^5 model of Hou et al. (2021). Table II.2 provides key statistics for each version of a given predictor.

[Insert Table II.2 here]

Although non-standard errors are insignificant for *AS*, the measure is sensitive to predictor construction choice. When altering between the equal- and value-weighted benchmarks, results change drastically. *AS* computed against the equal-weighted market portfolio has a mean return of -0.08% per month, a non-standard error ratio of 0.45, and zero positive premia estimates. In contrast, *AS* computed against the value-weighted market portfolio has a mean return of 0.05% per month, a non-standard error ratio of 0.27, and 94% of positive premia estimates. This finding supports Frazzini, Friedman, and Pomorski (2016), who show that the ability of *AS* to predict mutual fund performance is sensitive to the benchmark type.

AW is somewhat similar to *AS*. However, unlike *AS*, which is computed against the market portfolio as a benchmark, *AW* benchmarks fund portfolio holdings against either a hypothetical cap-weighted portfolio (VW) or a hypothetical naive portfolio (EW).[9] Figure II.2 reveals that the choice of a benchmark impacts *AW*. Although both VW and EW versions are positive, the results for the VW (EW) version are substantially stronger (weaker): a monthly mean return of 0.13% (0.07%) with 73% (2%) of statistically significant premia estimates. Overall, *AW* could be a more robust alternative to *AS* considering that Doshi, Elkamhi, and Simutin (2015) show that *AW* is not subsumed or explained by the *AS*.

Returning to the regression-based predictors, the *PA* measure stands out as one of the most variable predictors. The non-standard error ratio spans from 0.94 to 1.35 and the average monthly non-standard error ranges from 0.12% to 0.20%. The majority of these errors are statistically significant as indicated by the relative number of significant deviations to the left and right of the median (*Left-right*). This finding echoes well with the study of Soebhag, Van Vliet, and Verwijmeren (2024) that stresses the importance of the non-standard errors in factor models, as well as the study of (Mamaysky, Spiegel, and Zhang, 2007) that shows how estimation error in the past alphas impacts out-of-sample predictability of mutual fund performance.

Continuing with the rest of the regression-based predictors (*R2* and *Tstat*), the non-standard error ratio spans from 0.52 to 0.93 and from 0.40 to 0.87, respectively. Moreover, the vast majority of the non-standard errors are statistically insignificant, with a few exceptions arising from the five-factor q^5 model of Hou et al. (2021). For instance, for *Tstat*, the only specification that has significant deviations to the left of the median is (Q5, 2Y), which is 3%. Similarly, for *R2*, only two specifications—(Q5, 3Y) and (Q5, 5Y)—show significant deviations to the left of the median, which are

[9] To avoid confusion between *AW* and *AS*, note that for *AW*, the portfolio of a given fund and a hypothetical portfolio are identical in terms of their constituents.

2% and 8%, respectively. In both cases, these deviations could arise by chance. The results are not adjusted for multiple hypothesis testing, i.e., some true null hypotheses could be rejected purely by chance. Overall, *R2* is more stable than *PA* and *Tstat*, which flips the sign when estimated using the q^5 model.

Lastly, it is interesting to investigate *SI* as a predictor that has sizable non-standard errors ranging from 0.17% to 0.20% per month. The non-standard error ratios span from 1.10 to 1.20, and most importantly, *SI* has one of the highest numbers of significant deviations from the median. This implies that the non-standard errors are most likely statistically significant. However, Figure II.2 and Table II.2 show that the distributions of *SI* are very stable and have similar distributional statistics, such as the mean, skewness, and kurtosis, across predictor construction choices. Thus, it is likely that *SI* is robust to predictor construction choices and that methodologically induced variation is likely to stem from the portfolio construction choices. Consequently, it is important to consider what research design choices in portfolio construction are behind this variation.

4.3. Impact of fixing specific portfolio construction choices

At this point, I have identified a smaller subset of predictors such as *Exp*, *Size*, and *AW* that are robust to methodological choices. Moreover, there is a subset of predictors for which the distribution of the estimated premia exhibits substantial variation, yet is relatively stable across predictor construction choices. Such predictors include *SI*, *GT*, *CT*, and *ICI*. Therefore, in this section, I focus on the predictors that have a median greater than zero and investigate how they perform under certain conditions.

4.3.1. Impact of fixing sorting variable lag

I first investigate the impact of keeping sorting variable lags fixed to assess the longevity of a signal. Fixing this choice is informative in determining whether certain predictors are persistent or short-lived. Figure II.3 visualizes the impact of the arrival of information for the sorting variable for 13 selected predictors, while Table II.A3 in Appendix II.A provides related quantitative evidence for all predictors.

[Insert Figure II.3 here]

First, surprisingly, the performance of *Exp* improves with shorter information lags, doubling from a monthly 0.06% (12 months) to 0.12% (1 month). In contrast, the performance of *Size* drops from a monthly 0.06% (12 months) to 0.03% (1 month). It is surprising to observe patterns related to the signal recency for slow-moving variables such as *Exp* and *Size*. Moving to *AW*, the predictor is persistent, i.e., *AW* remains largely stable across information lags, although it slightly improves with the recency of the signal.

The regression-based predictors respond differently to the recency of a signal. Starting with *R2*, the predictor appears to be the most persistent within its predictor group. Overall, *R2* behaves similarly to *AW*. However, an important difference is the pronounced left tail, which probably stems from the predictor's sensitivity to the q^5 factor model discussed earlier. Moving to *PA* and *Tstat*, both predictors are clearly short-lived with the best (worst) performance being achieved at the shortest (longest) lag. The fact that *R2* is a persistent measure, while *PA* and *Tstat* are short-lived, is unsurprising. Amihud and Goyenko (2013) argue that *R2* is a measure of selectivity, while *PA* and *Tstat* are measures of past performance, which is rarely persistent for mutual funds (Carhart, 1997; Fama and French, 2010; Barras, Scaillet, and Wermers, 2010; Busse, Goyal, and Wahal, 2010; Ferreira et al., 2013).

Regarding holdings-based measures, such predictors as *CS*, *CT*, *GT*, *RG*, and *SI* perform the best (worst) at the shortest (longest) lag. The vast majority of premia estimates are positive: 75% for *CS*, 81% for *CT*, 90% for *GT*, 94% for *RG*, and 78% for *SI*. Arguably, *CS*, *GT*, and *SI* are related to past performance; hence, they exhibit a similar pattern to *PA*. However, on average, the median of premia estimates is higher for these holdings-based measures, and extreme positive outcomes are more likely given the long right tail. Lastly, *ICI* is mostly stable across information lags. However, its monthly mean is low.

4.3.2. Impact of fixing rebalancing frequency

Rebalancing frequency refers to the holding period, which is an important practical consideration. Walter, Weber, and Weiss (2024) argue that allowing for rebalancing choices makes the most sense for sorting variables that are updated yearly, since annual rebalancing for monthly or quarterly updated predictors would imply the use of stale information. I argue that, unlike stock anomalies, where monthly rebalancing could be feasible depending on a strategy, a mutual funds selection strategy with a holding period horizon shorter than six months (semiannual rebalancing) is practically not feasible or is probably reserved for a limited set of investors. Thus, keeping rebalancing frequency fixed allows us to evaluate the practical aspects of a predictor-based trading strategy.

According to Figure II.4, *Exp*, *Size*, *AW*, *R2*, *ICI*, and *RS* are largely stable to portfolio formation frequency. Meanwhile, short-lived predictors such as *PA*, *Tstat*, *CS*, *CT*, *GT*, and *SI* perform the best (worst) when portfolios are formed monthly (annually). Table II.A4 in Appendix II.A further supports these findings. Moreover, for the second set of predictors, the variation within the monthly rebalancing bin is largest. Therefore, I further investigate the impact of the intersection between rebalancing frequency and the most recent predictor information, i.e., sorting variable lags of 1- and 3-months.

[Insert Figure II.4 here]

4.3.3. Impact of fixing rebalancing frequency and sorting variable lag

Lastly, I limit the predictor set to short-lived predictors to assess whether trading strategies based on these predictors are implementable. Figure II.5 considers information lags up to three months (top pane) and an information lag of precisely one month (bottom pane).

[Insert Figure II.5 here]

In the first case (1- and 3-month lags), only *CT*, *GT*, and *SI* remain positive up to quarterly rebalancing. Implementing strategies based on these predictors could be suitable for funds of funds as they could potentially rebalance that frequently, while for retail investors quarterly rebalancing appears to be unrealistic. In the second case (1-month lag), *CS*, *GT*, and *SI* remain positive up to semiannual rebalancing. Moreover, the monthly mean returns are economically significant for these predictors. For instance, the median monthly returns for *GT* and *SI* are around 0.23% (2.75% p.a.) and 0.27% (3.24% p.a.), respectively. The same mostly applies to *PA*; however, it has a longer left tail. *Tstat* would probably perform similarly if we were to drop the q^5 based versions of the predictor.

5. Conclusion

Overall, given that an investor has access to the latest information on the following predictors and can rebalance at least semiannually, *SI*, *GT*, *CS*, and *PA* are suitable for a mutual funds selection strategy as they consistently generate positive return spreads. Alternatively, *AW* and *Exp* appear as safe bets. Their return spreads are stable, and extremely persistent, thus allowing for infrequent rebalancing, which makes these measures appropriate for retail investors. The spreads remain positive even for a 12-month sorting variable lag with annual rebalancing, which signals that *Exp* and *AW* persist for up to two years.

Finally, it is worth mentioning that conclusions regarding well-performing predictors are broadly consistent with Müller, Pugachyov, and Weigert (2024). These results are timely given the rise of machine-learning research in mutual funds domain (Li and Rossi, 2020; Kaniel et al., 2023; DeMiguel et al., 2023; Fausch et al., 2024).

Figure II.1: **Non-standard Errors Across Mutual Fund Performance Predictors, 1985-2022**

This figure shows the estimated premia (in % per month) in box plots for all mutual fund performance predictors across all paths. The boxes cover the interquartile range and thus correspond to the definition of non-standard errors from Equation II.2. The lines at the ends of each box indicate the maximum and minimum premium of each distribution. The vertical axis shows the associated predictor. Note that the color scheme is consistent with Figure II.2.

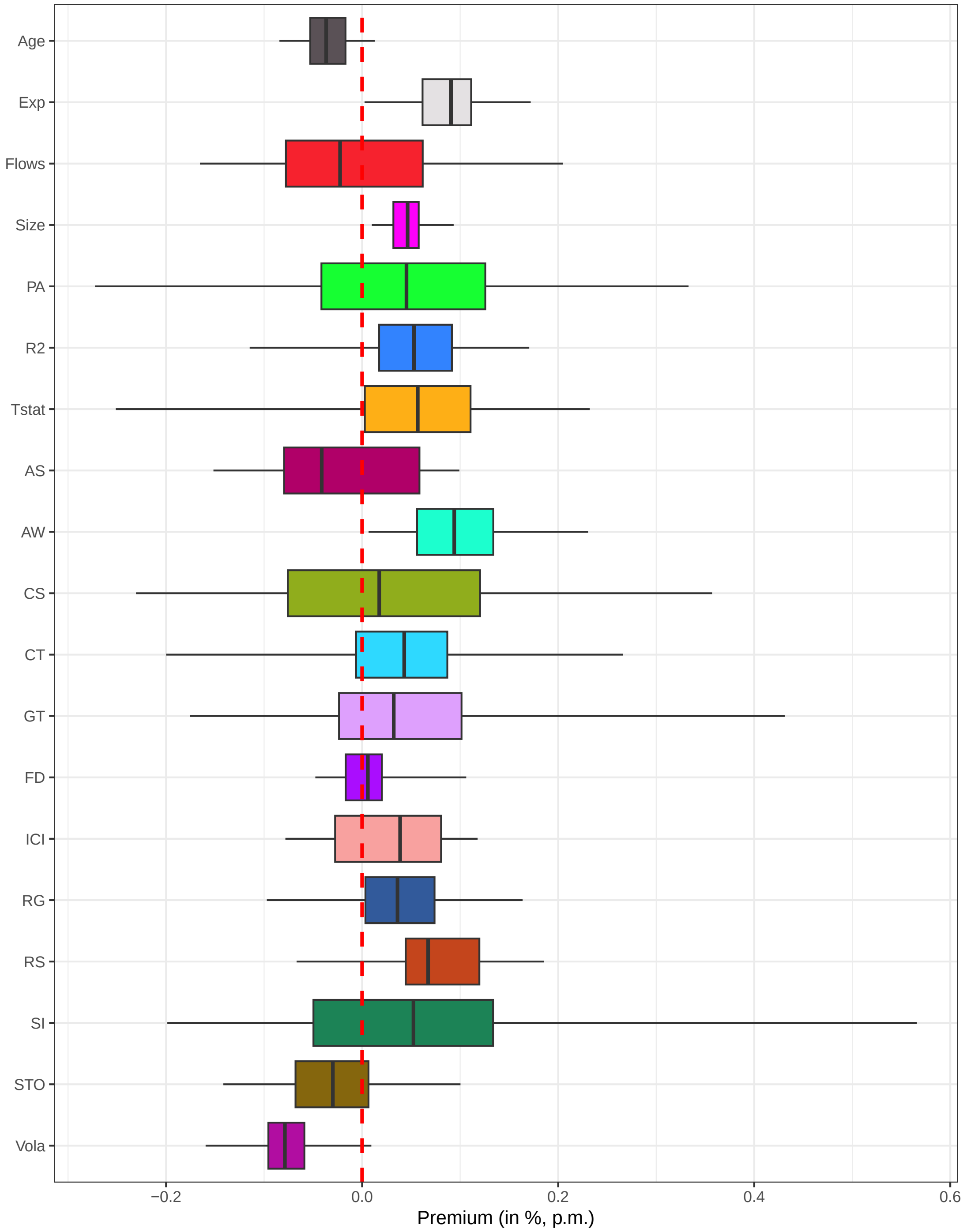


Table II.1: **Methodology-induced Variation in Mutual Fund Performance Predictors, 1985-2022**

This table shows key statistics for each predictor across all paths. Predictors are split into four groups: Fund characteristics, regression-based, holdings-based, and other predictors. For each group, the last row (**Total**) represents the statistics computed at a group level, i.e., aggregates predictors within a group. These summary statistics include the mean (**Mean**, in % per month), skewness (**Skew.**), and kurtosis (**Kurt.**) of premia across all paths for each predictor or a group of predictors. Column **NSE** contains the non-standard error (in % per month; see II.2). Column **Left-right** is the relative number of significant deviations to the left and right of the median using a 5% significance level (see II.3). Column **Ratio** is the ratio of the dispersion of premia relative to the average time-series standard error (**Ratio**$_{iid}$ is computed using *iid* standard errors; see II.4). Columns **Pos.** and **Sig.** show the relative number of positive premia and t-statistics larger than 1.96. Standard errors used in the computation of **Ratio**, **Left-right**, and **Sig.** columns are computed using Newey and West (1987) standard errors with six lags.

	Mean	**NSE**	**Ratio**	**Ratio**$_{iid}$	**Left-right**	**Skew.**	**Kurt.**	**Pos.**	**Sig.**
Fund characteristics									
Age	-0.04	0.04	0.50	0.54	(0.00, 0.00)	0.10	2.31	0.06	0.00
Exp	0.09	0.05	0.55	0.57	(0.00, 0.00)	-0.03	2.53	1.00	0.03
Flows	-0.01	0.14	1.21	1.27	(0.00, 0.08)	0.46	2.34	0.44	0.02
Size	0.05	0.03	0.32	0.42	(0.00, 0.00)	0.21	3.01	1.00	0.00
Total	0.02	0.10	1.13	1.24	(0.05, 0.00)	-0.07	2.44	0.62	0.01
Regression-based predictors									
PA	0.05	0.17	1.18	1.28	(0.03, 0.06)	0.07	2.62	0.65	0.12
R2	0.05	0.07	0.72	0.79	(0.01, 0.00)	-0.32	2.81	0.83	0.09
Tstat	0.05	0.11	0.89	1.12	(0.04, 0.01)	-0.74	3.47	0.76	0.06
Total	0.05	0.11	1.00	1.14	(0.03, 0.02)	-0.25	3.74	0.75	0.09
Holdings-based predictors									
AS	-0.02	0.14	0.84	0.91	(0.00, 0.00)	0.08	1.50	0.47	0.00
AW	0.10	0.08	1.01	1.11	(0.01, 0.00)	0.56	2.42	1.00	0.38
CS	0.03	0.20	1.14	1.26	(0.00, 0.06)	0.40	2.64	0.53	0.06
CT	0.04	0.09	0.97	1.02	(0.04, 0.00)	0.10	3.44	0.72	0.04
GT	0.05	0.12	1.16	1.23	(0.01, 0.08)	0.84	3.67	0.66	0.12
FD	0.00	0.04	0.31	0.33	(0.00, 0.00)	0.72	4.89	0.58	0.00
ICI	0.03	0.11	0.74	0.76	(0.00, 0.00)	-0.19	1.46	0.55	0.00
RG	0.04	0.07	0.98	1.04	(0.00, 0.02)	-0.01	2.90	0.75	0.11
RS	0.08	0.08	0.93	1.01	(0.00, 0.00)	-0.11	2.42	0.89	0.23
SI	0.06	0.18	1.16	1.25	(0.07, 0.06)	0.78	3.53	0.61	0.09
Total	0.04	0.11	1.13	1.21	(0.06, 0.04)	0.86	5.46	0.66	0.10
Other predictors									
STO	-0.03	0.07	1.20	1.29	(0.07, 0.09)	0.09	2.57	0.29	0.02
Vola	-0.08	0.04	0.16	0.17	(0.00, 0.00)	0.14	3.36	0.02	0.00
Total	-0.05	0.07	0.42	0.44	(0.01, 0.11)	0.61	3.10	0.15	0.01
All predictors									
Total	**0.04**	**0.12**	**1.04**	**1.15**	**(0.05, 0.03)**	**0.34**	**4.27**	**0.66**	**0.08**

Figure II.2: **Non-standard Errors Across Mutual Fund Performance Predictors: Detailed View on Predictor Construction, 1985-2022**

This figure shows the estimated premia (in % per month) in box plots for all mutual fund performance predictors across all paths with details on predictor construction. Predictors are split into four groups: Fund characteristics (Chars.), regression-based, holdings-based, and other predictors (Other). For each predictor (where applicable), the results are shown at a predictor-version level, i.e., an alternative predictor construction as described in Section 2.1. The boxes cover the interquartile range and thus correspond to the definition of non-standard errors from Equation II.2. The lines at the ends of each box indicate the maximum and minimum premium of each distribution. The vertical axis shows the associated predictor construction version, while the color scheme connects each version to the respective predictor. Note that this color scheme is consistent with Figure II.1.

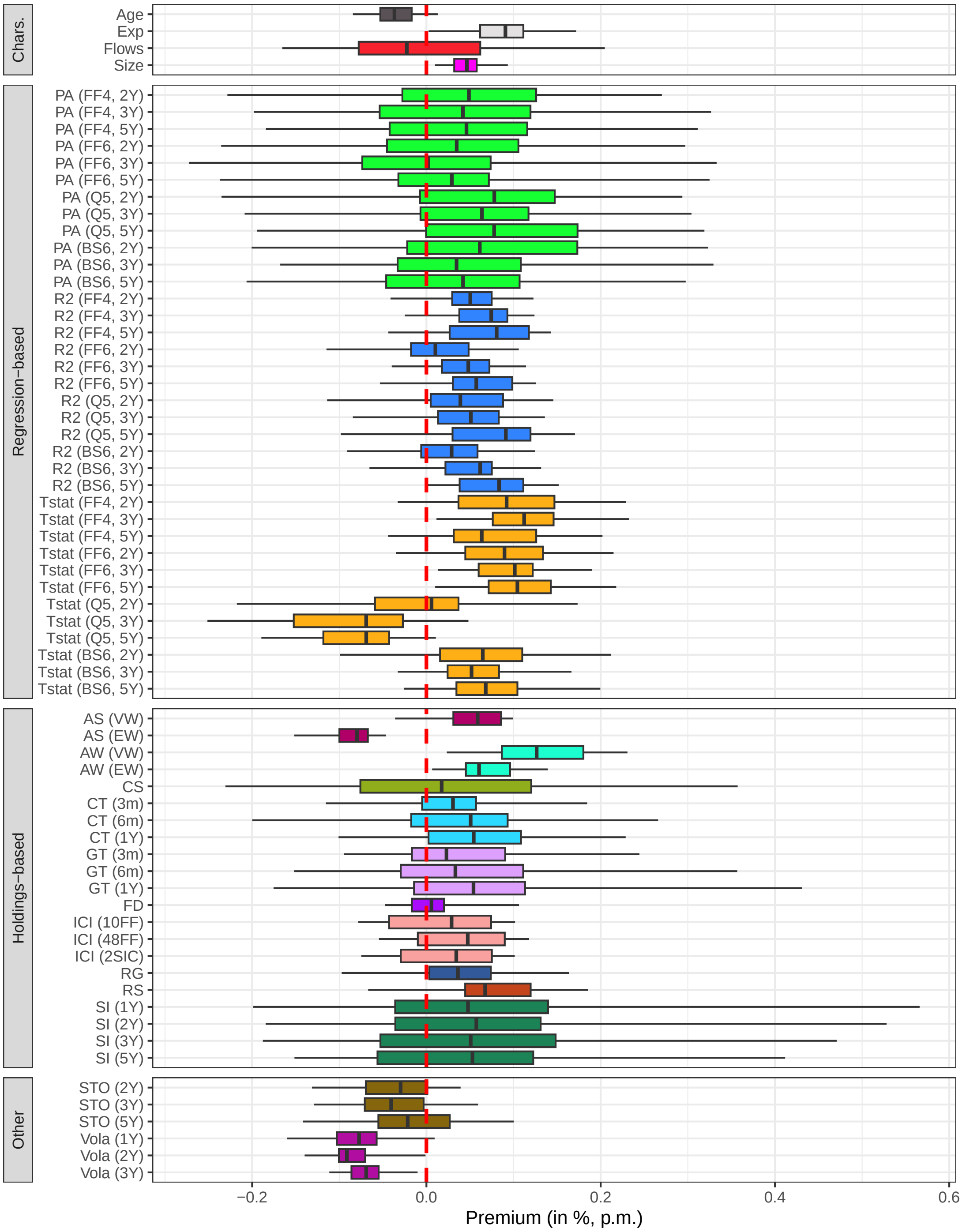


Table II.2: **Methodology-induced Variation in Mutual Fund Performance Predictors: Detailed View on Predictor Construction, 1985-2022**

This table shows key statistics for each predictor across all paths with details on predictor construction. For each predictor, the first row (in bold) represents the statistics computed at a predictor level, followed by (where applicable) statistics computed at a predictor-version level, i.e., an alternative predictor construction as described in Section 2.1. These summary statistics include the mean (**Mean**, in % per month), skewness (**Skew.**), and kurtosis (**Kurt.**) of premia across all paths for each predictor or a group of predictors. Column **NSE** contains the non-standard error (in % per month; see II.2). Column **Left-right** is the relative number of significant deviations to the left and right of the median using a 5% significance level (see II.3). Column **Ratio** is the ratio of the dispersion of premia relative to the average time-series standard error (see II.4). Columns **Pos.** and **Sig.** show the relative number of positive premia and t-statistics larger than 1.96. Standard errors used in the computation of **Ratio**, **Left-right**, and **Sig.** columns are computed using Newey and West (1987) standard errors with six lags.

	Mean	**NSE**	**Ratio**	**Left-right**	**Skew.**	**Kurt.**	**Pos.**	**Sig.**
Fund characteristics								
Age	**-0.04**	**0.04**	**0.50**	**(0.00, 0.00)**	**0.10**	**2.31**	**0.06**	**0.00**
Exp	**0.09**	**0.05**	**0.55**	**(0.00, 0.00)**	**-0.03**	**2.53**	**1.00**	**0.03**
Flows	**-0.01**	**0.14**	**1.21**	**(0.00, 0.08)**	**0.46**	**2.34**	**0.44**	**0.02**
Size	**0.05**	**0.03**	**0.32**	**(0.00, 0.00)**	**0.21**	**3.01**	**1.00**	**0.00**
Regression-based predictors								
PA	**0.05**	**0.17**	**1.18**	**(0.03, 0.06)**	**0.07**	**2.62**	**0.65**	**0.12**
PA (FF4, 2Y)	0.05	0.15	1.21	(0.05, 0.08)	-0.11	2.72	0.66	0.16
PA (FF4, 3Y)	0.04	0.17	1.30	(0.02, 0.08)	0.32	2.40	0.59	0.16
PA (FF4, 5Y)	0.05	0.16	1.23	(0.00, 0.09)	0.28	2.49	0.61	0.16
PA (FF6, 2Y)	0.03	0.15	1.34	(0.06, 0.09)	0.00	3.09	0.61	0.17
PA (FF6, 3Y)	0.01	0.15	1.35	(0.06, 0.12)	0.44	3.28	0.50	0.12
PA (FF6, 5Y)	0.03	0.10	1.20	(0.03, 0.06)	0.34	3.69	0.64	0.12
PA (Q5, 2Y)	0.06	0.15	1.00	(0.03, 0.00)	-0.46	2.56	0.73	0.02
PA (Q5, 3Y)	0.05	0.12	0.94	(0.02, 0.00)	-0.25	2.49	0.73	0.03
PA (Q5, 5Y)	0.07	0.17	1.02	(0.00, 0.00)	-0.25	2.19	0.75	0.08
PA (BS6, 2Y)	0.08	0.20	1.32	(0.05, 0.12)	0.14	2.19	0.67	0.25
PA (BS6, 3Y)	0.04	0.14	1.21	(0.03, 0.09)	0.41	2.84	0.64	0.11
PA (BS6, 5Y)	0.03	0.15	1.17	(0.05, 0.05)	0.02	2.69	0.61	0.09
R2	**0.05**	**0.07**	**0.72**	**(0.01, 0.00)**	**-0.32**	**2.81**	**0.83**	**0.09**
R2 (FF4, 2Y)	0.05	0.05	0.55	(0.00, 0.00)	-0.21	2.59	0.89	0.00
R2 (FF4, 3Y)	0.07	0.06	0.52	(0.00, 0.00)	-0.75	2.96	0.92	0.00
R2 (FF4, 5Y)	0.07	0.09	0.80	(0.00, 0.00)	-0.37	1.88	0.86	0.34
R2 (FF6, 2Y)	0.01	0.07	0.63	(0.00, 0.00)	-0.20	2.91	0.62	0.00
R2 (FF6, 3Y)	0.04	0.05	0.58	(0.00, 0.00)	-0.25	2.23	0.81	0.00
R2 (FF6, 5Y)	0.06	0.07	0.68	(0.00, 0.00)	-0.22	2.23	0.97	0.23
R2 (Q5, 2Y)	0.04	0.08	0.73	(0.00, 0.00)	-0.37	2.86	0.78	0.00
R2 (Q5, 3Y)	0.05	0.07	0.65	(0.02, 0.00)	-0.46	2.67	0.81	0.00
R2 (Q5, 5Y)	0.07	0.09	0.93	(0.08, 0.00)	-0.68	2.45	0.75	0.25
R2 (BS6, 2Y)	0.03	0.06	0.65	(0.00, 0.00)	-0.20	2.58	0.72	0.00
R2 (BS6, 3Y)	0.05	0.05	0.62	(0.00, 0.00)	-0.50	2.99	0.88	0.02
R2 (BS6, 5Y)	0.08	0.07	0.64	(0.00, 0.00)	0.08	1.73	0.98	0.25

Table continued on the next page

Table continued

	Mean	NSE	Ratio	Left-right	Skew.	Kurt.	Pos.	Sig.
Tstat	**0.05**	**0.11**	**0.89**	**(0.04, 0.01)**	**-0.74**	**3.47**	**0.76**	**0.06**
Tstat (FF4, 2Y)	0.09	0.11	0.87	(0.00, 0.00)	0.06	2.03	0.95	0.27
Tstat (FF4, 3Y)	0.11	0.07	0.65	(0.00, 0.00)	0.10	2.33	1.00	0.22
Tstat (FF4, 5Y)	0.07	0.09	0.79	(0.00, 0.00)	0.13	2.18	0.88	0.11
Tstat (FF6, 2Y)	0.09	0.09	0.66	(0.00, 0.00)	0.08	2.34	0.95	0.08
Tstat (FF6, 3Y)	0.10	0.06	0.40	(0.00, 0.00)	0.07	2.46	1.00	0.00
Tstat (FF6, 5Y)	0.11	0.07	0.46	(0.00, 0.00)	0.14	2.29	1.00	0.00
Tstat (Q5, 2Y)	-0.01	0.10	0.68	(0.03, 0.00)	-0.50	2.78	0.58	0.00
Tstat (Q5, 3Y)	-0.09	0.13	0.63	(0.00, 0.00)	-0.49	2.29	0.11	0.00
Tstat (Q5, 5Y)	-0.08	0.08	0.47	(0.00, 0.00)	-0.44	2.27	0.03	0.00
Tstat (BS6, 2Y)	0.06	0.09	0.65	(0.00, 0.00)	0.01	2.71	0.84	0.00
Tstat (BS6, 3Y)	0.05	0.06	0.40	(0.00, 0.00)	0.07	2.65	0.86	0.00
Tstat (BS6, 5Y)	0.07	0.07	0.50	(0.00, 0.00)	0.32	2.59	0.91	0.00
Holdings-based predictors								
AS	**-0.02**	**0.14**	**0.84**	**(0.00, 0.00)**	**0.08**	**1.50**	**0.47**	**0.00**
AS (VW)	0.05	0.05	0.27	(0.00, 0.00)	-0.60	2.62	0.94	0.00
AS (EW)	-0.08	0.03	0.45	(0.00, 0.00)	-0.60	2.98	0.00	0.00
AW	**0.10**	**0.08**	**1.01**	**(0.01, 0.00)**	**0.56**	**2.42**	**1.00**	**0.38**
AW (VW)	0.13	0.09	0.99	(0.08, 0.00)	0.06	1.89	1.00	0.73
AW (EW)	0.07	0.05	0.63	(0.00, 0.00)	0.38	2.17	1.00	0.02
CS	**0.03**	**0.20**	**1.14**	**(0.00, 0.06)**	**0.40**	**2.64**	**0.53**	**0.06**
CT	**0.04**	**0.09**	**0.97**	**(0.04, 0.00)**	**0.10**	**3.44**	**0.72**	**0.04**
CT (3m)	0.03	0.06	0.87	(0.02, 0.00)	0.33	3.85	0.72	0.03
CT (6m)	0.04	0.11	1.13	(0.08, 0.00)	-0.10	3.39	0.67	0.06
CT (1Y)	0.05	0.11	0.85	(0.00, 0.00)	0.02	2.58	0.77	0.03
GT	**0.05**	**0.12**	**1.16**	**(0.01, 0.08)**	**0.84**	**3.67**	**0.66**	**0.12**
GT (3m)	0.04	0.11	1.11	(0.00, 0.08)	0.70	2.91	0.67	0.12
GT (6m)	0.05	0.14	1.20	(0.03, 0.06)	0.69	3.06	0.62	0.12
GT (1Y)	0.06	0.13	1.11	(0.03, 0.05)	0.79	3.36	0.69	0.12
FD	**0.00**	**0.04**	**0.31**	**(0.00, 0.00)**	**0.72**	**4.89**	**0.58**	**0.00**
ICI	**0.03**	**0.11**	**0.74**	**(0.00, 0.00)**	**-0.19**	**1.46**	**0.55**	**0.00**
ICI (10FF)	0.02	0.12	0.77	(0.00, 0.00)	-0.12	1.32	0.52	0.00
ICI (48FF)	0.04	0.10	0.71	(0.00, 0.00)	-0.18	1.39	0.61	0.00
ICI (2SIC)	0.02	0.10	0.72	(0.00, 0.00)	-0.21	1.42	0.53	0.00
RG	**0.04**	**0.07**	**0.98**	**(0.00, 0.02)**	**-0.01**	**2.90**	**0.75**	**0.11**
RS	**0.08**	**0.08**	**0.93**	**(0.00, 0.00)**	**-0.11**	**2.42**	**0.89**	**0.23**
SI	**0.06**	**0.18**	**1.16**	**(0.07, 0.06)**	**0.78**	**3.53**	**0.61**	**0.09**
SI (1Y)	0.07	0.18	1.20	(0.09, 0.06)	0.85	3.71	0.62	0.08
SI (2Y)	0.06	0.17	1.19	(0.09, 0.06)	0.80	3.59	0.59	0.08
SI (3Y)	0.06	0.20	1.16	(0.05, 0.06)	0.66	3.05	0.59	0.09
SI (5Y)	0.05	0.18	1.10	(0.05, 0.06)	0.66	3.03	0.61	0.09
Other predictors								
STO	**-0.03**	**0.07**	**1.20**	**(0.07, 0.09)**	**0.09**	**2.57**	**0.29**	**0.02**
STO (2Y)	-0.04	0.07	1.06	(0.09, 0.06)	-0.40	2.12	0.30	0.00
STO (3Y)	-0.04	0.07	1.12	(0.03, 0.11)	0.12	2.23	0.23	0.00
STO (5Y)	-0.02	0.08	1.36	(0.05, 0.16)	0.04	2.43	0.34	0.05
Vola	**-0.08**	**0.04**	**0.16**	**(0.00, 0.00)**	**0.14**	**3.36**	**0.02**	**0.00**
Vola (1Y)	-0.08	0.05	0.20	(0.00, 0.00)	0.07	2.87	0.05	0.00
Vola (2Y)	-0.09	0.03	0.15	(0.00, 0.00)	0.67	3.87	0.00	0.00
Vola (3Y)	-0.07	0.03	0.12	(0.00, 0.00)	0.33	2.85	0.00	0.00

Figure II.3: **Impact of Fixing Sorting Variable Lag, 1985-2022**

This figure shows the estimated premia (in % per month) in box plots for a subset of predictors across all paths, holding the individual choices of the portfolio construction nodes (here, *sorting variable lag*) constant. The boxes cover the interquartile range and thus correspond to the definition of non-standard errors from Equation II.2. The lines at the ends of each box indicate the maximum and minimum premium of each distribution. The vertical axis shows the associated predictor, while the color scheme connects to the respective sorting variable lag.

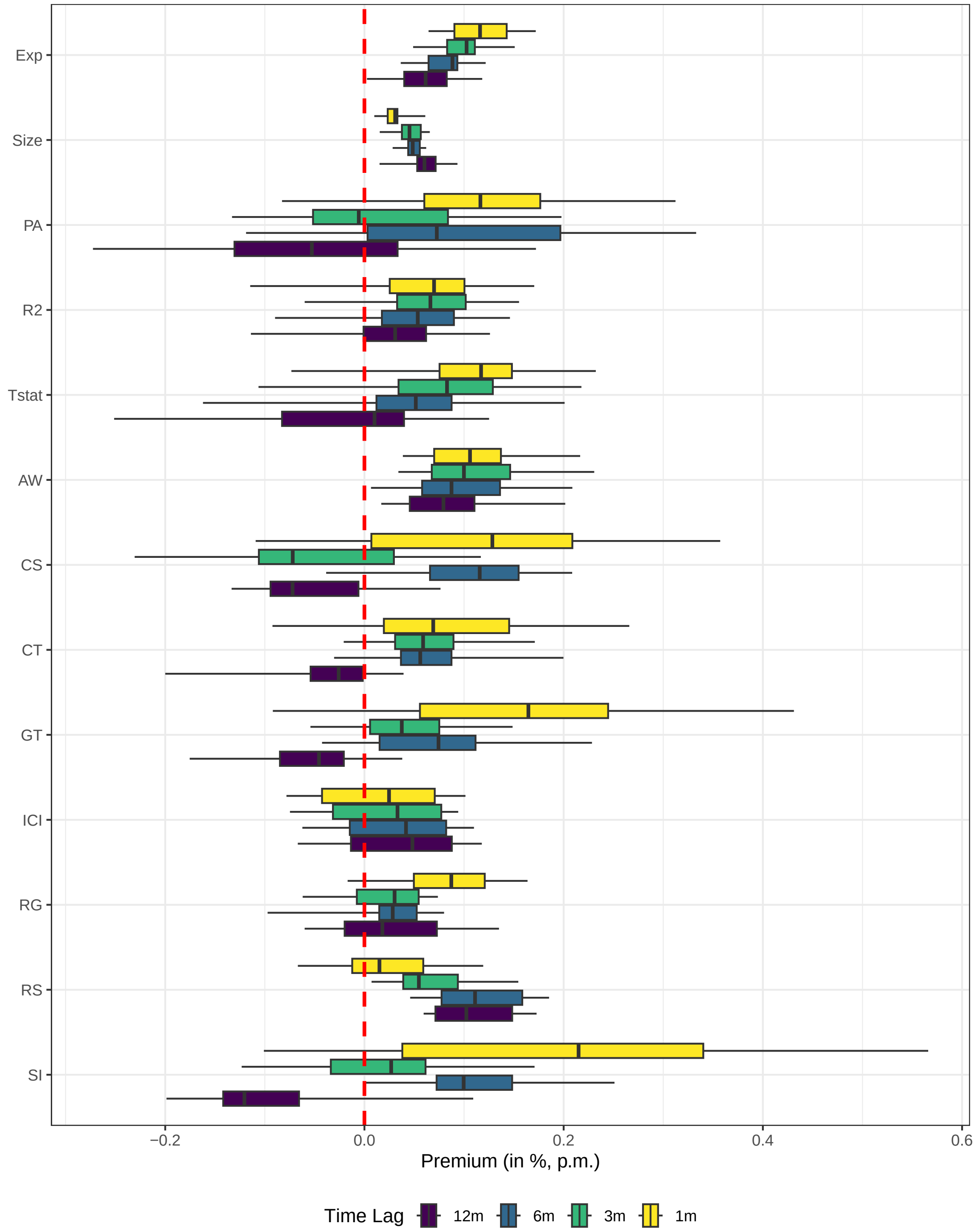


Figure II.4: **Impact of Fixing Rebalancing Frequency, 1985-2022**

This figure shows the estimated premia (in % per month) in box plots for a subset of predictors across all paths, holding the individual choices of the portfolio construction nodes (here, *rebalancing frequency*) constant. The boxes cover the interquartile range and thus correspond to the definition of non-standard errors from Equation II.2. The lines at the ends of each box indicate the maximum and minimum premium of each distribution. The vertical axis shows the associated predictor, while the color scheme connects to the respective rebalancing frequency.

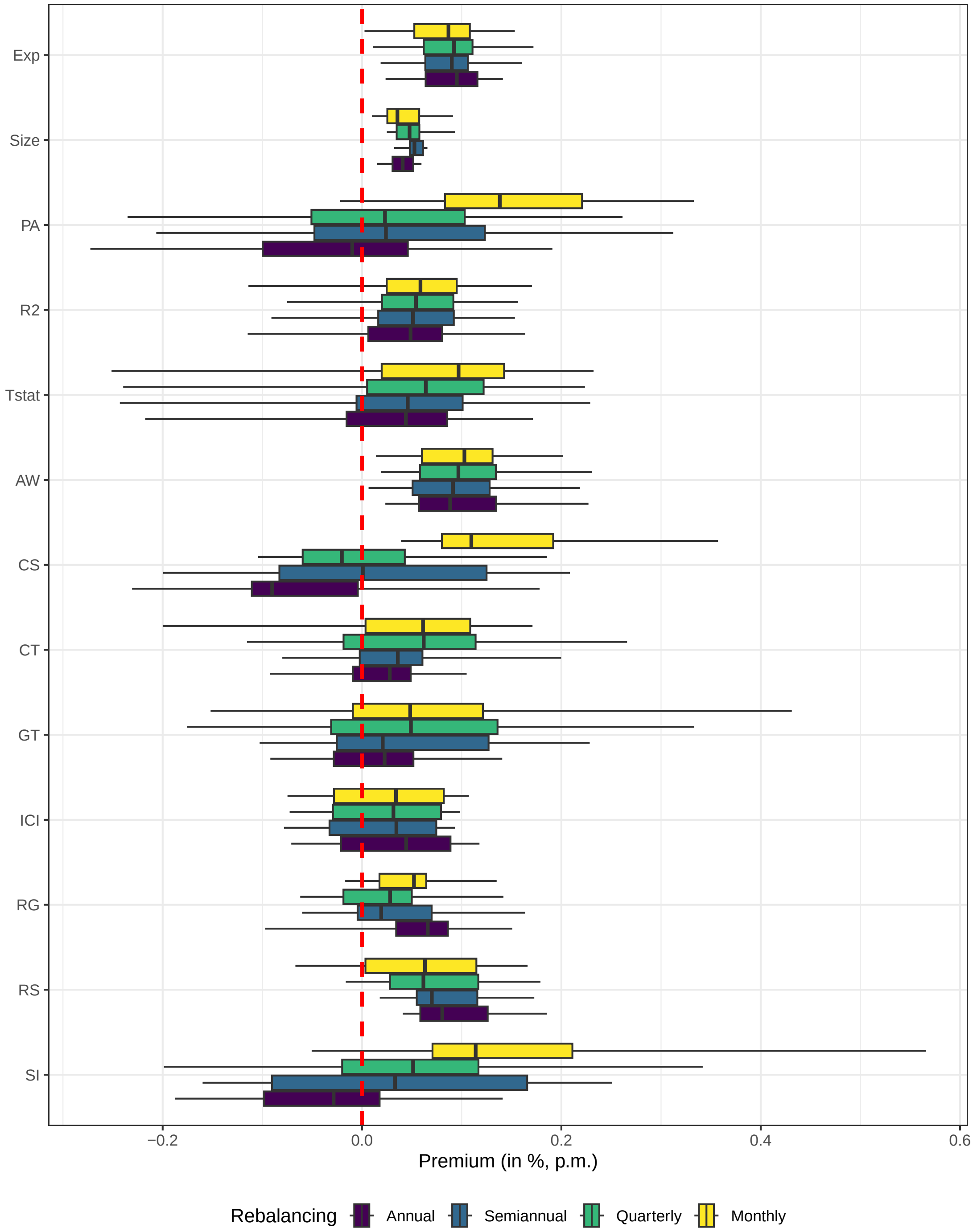
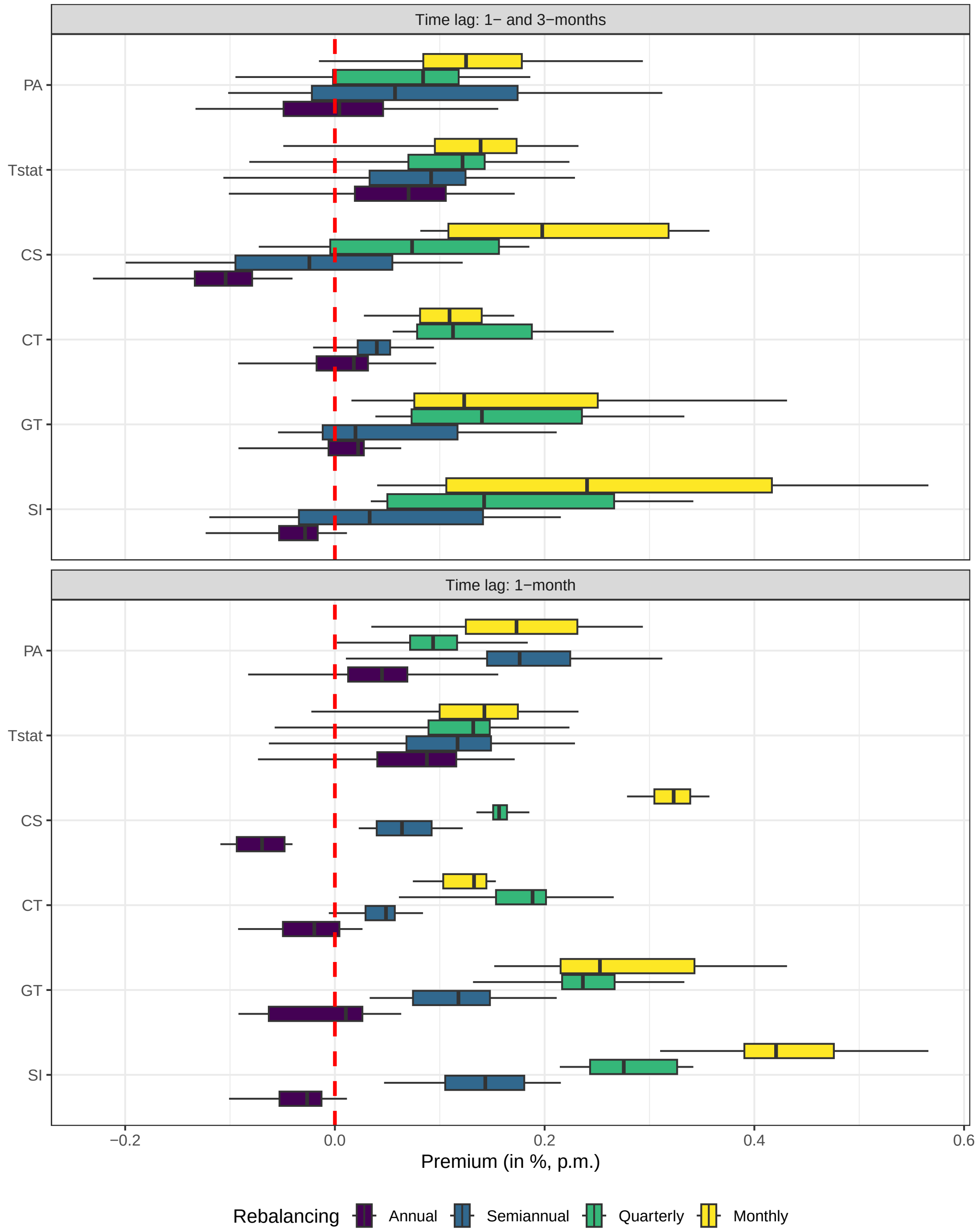


Figure II.5: **Impact of Fixing Rebalancing Frequency and Sorting Variable Lag, 1985-2022**

This figure shows the estimated premia (in % per month) in box plots for a subset of predictors across all paths, holding the individual choices of the portfolio construction nodes (here, *rebalancing frequency* and *sorting variable lag*) constant. The boxes cover the interquartile range and thus correspond to the definition of non-standard errors from Equation II.2. The lines at the ends of each box indicate the maximum and minimum premium of each distribution. The vertical axis shows the associated predictor, while the color scheme connects to the respective rebalancing frequency.



Appendix

II.A. Additional Results

Table II.A1: **Impact of Fixing Specific Portfolio Construction Choices: Number of Portfolios, 1985-2022**

This table shows key statistics for each predictor across all paths holding the individual choices of the portfolio construction nodes constant. For each predictor, the first row (in bold) represents the statistics computed at a predictor level, followed by the statistics computed for a given *number of portfolios*: Quintile-portfolios (5) and decile-portfolios (10) held constant. These summary statistics include the mean (**Mean**, in % per month) of premia across all paths for each predictor or a group of predictors. Column **NSE** contains the non-standard error (in % per month; see II.2). Column **Left-right** is the relative number of significant deviations to the left and right of the median using a 5% significance level (see II.3). Columns **Pos.** and **Sig.** show the relative number of positive premia and t-statistics larger than 1.96. Standard errors used in the computation of **Left-right** and **Sig.** columns are computed using Newey and West (1987) standard errors with six lags.

	Mean	NSE	Left-right	Pos.	Sig.		Mean	NSE	Left-right	Pos.	Sig.
Age	**-0.04**	**0.04**	**(0.00,0.00)**	**0.06**	**0.00**	**Exp**	**0.09**	**0.05**	**(0.00,0.00)**	**1.00**	**0.03**
5	-0.03	0.02	(0.00,0.00)	0.06	0.00	5	0.07	0.05	(0.00,0.00)	1.00	0.06
10	-0.05	0.02	(0.00,0.00)	0.06	0.00	10	0.11	0.03	(0.00,0.00)	1.00	0.00
Flows	**-0.01**	**0.14**	**(0.00,0.08)**	**0.44**	**0.02**	**Size**	**0.05**	**0.03**	**(0.00,0.00)**	**1.00**	**0.00**
5	-0.01	0.12	(0.00,0.03)	0.44	0.00	5	0.04	0.02	(0.00,0.00)	1.00	0.00
10	-0.00	0.17	(0.00,0.12)	0.44	0.03	10	0.05	0.02	(0.00,0.00)	1.00	0.00
PA	**0.05**	**0.17**	**(0.03,0.06)**	**0.65**	**0.12**	**R2**	**0.05**	**0.07**	**(0.01,0.00)**	**0.83**	**0.09**
5	0.04	0.15	(0.02,0.06)	0.64	0.11	5	0.08	0.06	(0.00,0.00)	0.97	0.17
10	0.06	0.19	(0.05,0.07)	0.65	0.13	10	0.03	0.07	(0.00,0.00)	0.70	0.02
Tstat	**0.05**	**0.11**	**(0.04,0.01)**	**0.76**	**0.06**	**AS**	**-0.02**	**0.14**	**(0.00,0.00)**	**0.47**	**0.00**
5	0.04	0.10	(0.05,0.00)	0.74	0.05	5	-0.01	0.14	(0.03,0.00)	0.50	0.00
10	0.06	0.12	(0.04,0.00)	0.78	0.06	10	-0.02	0.14	(0.00,0.00)	0.44	0.00
AW	**0.10**	**0.08**	**(0.01,0.00)**	**1.00**	**0.38**	**CS**	**0.03**	**0.20**	**(0.00,0.06)**	**0.53**	**0.06**
5	0.08	0.05	(0.00,0.08)	1.00	0.30	5	0.03	0.19	(0.00,0.06)	0.56	0.06
10	0.13	0.05	(0.00,0.00)	1.00	0.45	10	0.03	0.22	(0.00,0.06)	0.50	0.06
CT	**0.04**	**0.09**	**(0.04,0.00)**	**0.72**	**0.04**	**GT**	**0.05**	**0.12**	**(0.01,0.08)**	**0.66**	**0.12**
5	0.04	0.08	(0.02,0.00)	0.72	0.04	5	0.04	0.11	(0.01,0.04)	0.65	0.11
10	0.05	0.10	(0.02,0.00)	0.72	0.04	10	0.07	0.14	(0.02,0.11)	0.68	0.14
FD	**0.00**	**0.04**	**(0.00,0.00)**	**0.58**	**0.00**	**ICI**	**0.03**	**0.11**	**(0.00,0.00)**	**0.55**	**0.00**
5	-0.00	0.03	(0.00,0.00)	0.41	0.00	5	0.03	0.09	(0.00,0.00)	0.58	0.00
10	0.01	0.02	(0.00,0.00)	0.75	0.00	10	0.02	0.13	(0.00,0.00)	0.52	0.00
RG	**0.04**	**0.07**	**(0.00,0.02)**	**0.75**	**0.11**	**RS**	**0.08**	**0.08**	**(0.00,0.00)**	**0.89**	**0.23**
5	0.04	0.06	(0.00,0.00)	0.81	0.12	5	0.08	0.10	(0.00,0.00)	0.91	0.34
10	0.04	0.08	(0.03,0.00)	0.69	0.09	10	0.07	0.07	(0.00,0.00)	0.88	0.12
SI	**0.06**	**0.18**	**(0.07,0.06)**	**0.61**	**0.09**	**STO**	**-0.03**	**0.07**	**(0.07,0.09)**	**0.29**	**0.02**
5	0.05	0.15	(0.10,0.06)	0.60	0.06	5	-0.02	0.07	(0.07,0.10)	0.34	0.00
10	0.07	0.23	(0.05,0.06)	0.61	0.11	10	-0.04	0.07	(0.05,0.09)	0.24	0.03
Vola	**-0.08**	**0.04**	**(0.00,0.00)**	**0.02**	**0.00**						
5	-0.08	0.03	(0.00,0.00)	0.01	0.00						
10	-0.08	0.04	(0.00,0.00)	0.02	0.00						

Table II.A2: **Impact of Fixing Specific Portfolio Construction Choices: Weighting Scheme, 1985-2022**

This table shows key statistics for each predictor across all paths holding the individual choices of the portfolio construction nodes constant. For each predictor, the first row (in bold) represents the statistics computed at a predictor level, followed by the statistics computed for a given portfolio *weighting scheme*: equal-weighted (EW) and value-weighted (VW) held constant. These summary statistics include the mean (**Mean**, in % per month) of premia across all paths for each predictor or a group of predictors. Column **NSE** contains the non-standard error (in % per month; see II.2). Column **Left-right** is the relative number of significant deviations to the left and right of the median using a 5% significance level (see II.3). Columns **Pos.** and **Sig.** show the relative number of positive premia and t-statistics larger than 1.96. Standard errors used in the computation of **Left-right** and **Sig.** columns are computed using Newey and West (1987) standard errors with six lags.

	Mean	NSE	Left-right	Pos.	Sig.		Mean	NSE	Left-right	Pos.	Sig.
Age	**-0.04**	**0.04**	**(0.00,0.00)**	**0.06**	**0.00**	**Exp**	**0.09**	**0.05**	**(0.00,0.00)**	**1.00**	**0.03**
EW	-0.02	0.03	(0.00, 0.00)	0.12	0.00	EW	0.07	0.05	(0.00, 0.00)	1.00	0.00
VW	-0.05	0.02	(0.00, 0.00)	0.00	0.00	VW	0.11	0.05	(0.00, 0.00)	1.00	0.06
Flows	**-0.01**	**0.14**	**(0.00,0.08)**	**0.44**	**0.02**	**Size**	**0.05**	**0.03**	**(0.00,0.00)**	**1.00**	**0.00**
EW	-0.01	0.13	(0.00, 0.00)	0.44	0.00	EW	0.05	0.02	(0.00, 0.00)	1.00	0.00
VW	-0.00	0.14	(0.00, 0.16)	0.44	0.03	VW	0.04	0.03	(0.00, 0.00)	1.00	0.00
PA	**0.05**	**0.17**	**(0.03,0.06)**	**0.65**	**0.12**	**R2**	**0.05**	**0.07**	**(0.01,0.00)**	**0.83**	**0.09**
EW	0.06	0.17	(0.05, 0.07)	0.66	0.16	EW	0.08	0.05	(0.00, 0.00)	0.98	0.18
VW	0.03	0.16	(0.03, 0.05)	0.63	0.09	VW	0.02	0.06	(0.00, 0.00)	0.69	0.00
Tstat	**0.05**	**0.11**	**(0.04,0.01)**	**0.76**	**0.06**	**AS**	**-0.02**	**0.14**	**(0.00,0.00)**	**0.47**	**0.00**
EW	0.04	0.12	(0.05, 0.01)	0.72	0.07	EW	0.01	0.15	(0.00, 0.00)	0.50	0.00
VW	0.06	0.10	(0.04, 0.00)	0.80	0.04	VW	-0.04	0.13	(0.00, 0.00)	0.44	0.00
AW	**0.10**	**0.08**	**(0.01,0.00)**	**1.00**	**0.38**	**CS**	**0.03**	**0.20**	**(0.00,0.06)**	**0.53**	**0.06**
EW	0.07	0.05	(0.00, 0.00)	1.00	0.23	EW	0.04	0.19	(0.00, 0.06)	0.53	0.06
VW	0.13	0.09	(0.02, 0.00)	1.00	0.52	VW	0.02	0.21	(0.00, 0.06)	0.53	0.06
CT	**0.04**	**0.09**	**(0.04,0.00)**	**0.72**	**0.04**	**GT**	**0.05**	**0.12**	**(0.01,0.08)**	**0.66**	**0.12**
EW	0.05	0.08	(0.01, 0.00)	0.79	0.02	EW	0.07	0.12	(0.01, 0.09)	0.74	0.14
VW	0.03	0.11	(0.04, 0.01)	0.65	0.06	VW	0.04	0.12	(0.00, 0.09)	0.58	0.11
FD	**0.00**	**0.04**	**(0.00,0.00)**	**0.58**	**0.00**	**ICI**	**0.03**	**0.11**	**(0.00,0.00)**	**0.55**	**0.00**
EW	-0.01	0.03	(0.00, 0.00)	0.38	0.00	EW	0.08	0.02	(0.00, 0.00)	1.00	0.00
VW	0.02	0.03	(0.00, 0.00)	0.78	0.00	VW	-0.03	0.04	(0.00, 0.00)	0.10	0.00
RG	**0.04**	**0.07**	**(0.00,0.02)**	**0.75**	**0.11**	**RS**	**0.08**	**0.08**	**(0.00,0.00)**	**0.89**	**0.23**
EW	0.05	0.04	(0.00, 0.00)	0.88	0.19	EW	0.05	0.04	(0.00, 0.00)	0.88	0.06
VW	0.03	0.09	(0.00, 0.03)	0.62	0.03	VW	0.10	0.09	(0.03, 0.00)	0.91	0.41
SI	**0.06**	**0.18**	**(0.07,0.06)**	**0.61**	**0.09**	**STO**	**-0.03**	**0.07**	**(0.07,0.09)**	**0.29**	**0.02**
EW	0.07	0.18	(0.05, 0.06)	0.65	0.09	EW	-0.02	0.07	(0.16, 0.08)	0.40	0.03
VW	0.05	0.18	(0.06, 0.06)	0.56	0.08	VW	-0.04	0.07	(0.03, 0.04)	0.19	0.00
Vola	**-0.08**	**0.04**	**(0.00,0.00)**	**0.02**	**0.00**						
EW	-0.08	0.03	(0.00, 0.00)	0.01	0.00						
VW	-0.08	0.04	(0.00, 0.00)	0.02	0.00						

Table II.A3: **Impact of Fixing Specific Portfolio Construction Choices: Sorting Variable Lag, 1985-2022**

This table shows key statistics for each predictor across all paths holding the individual choices of the portfolio construction nodes constant. For each predictor, the first row (in bold) represents the statistics computed at a predictor level, followed by the statistics computed for a given *sorting variable lag*: 1-month (1m), 3-months (3m), 6-months (6m), and 12-months (12m) held constant. These summary statistics include the mean (**Mean**, in % per month) of premia across all paths for each predictor or a group of predictors. Column **NSE** contains the non-standard error (in % per month; see II.2). Column **Left-right** is the relative number of significant deviations to the left and right of the median using a 5% significance level (see II.3). Columns **Pos.** and **Sig.** show the relative number of positive premia and t-statistics larger than 1.96. Standard errors used in the computation of **Left-right** and **Sig.** columns are computed using Newey and West (1987) standard errors with six lags.

	Mean	NSE	Left-right	Pos.	Sig.		Mean	NSE	Left-right	Pos.	Sig.
Age	**-0.04**	**0.04**	**(0.00,0.00)**	**0.06**	**0.00**	**Exp**	**0.09**	**0.05**	**(0.00,0.00)**	**1.00**	**0.03**
1m	-0.06	0.02	(0.00,0.00)	0.00	0.00	1m	0.12	0.05	(0.00,0.00)	1.00	0.12
3m	-0.04	0.02	(0.00,0.00)	0.00	0.00	3m	0.10	0.03	(0.00,0.00)	1.00	0.00
6m	-0.03	0.03	(0.00,0.00)	0.00	0.00	6m	0.08	0.03	(0.00,0.00)	1.00	0.00
12m	-0.02	0.04	(0.00,0.00)	0.25	0.00	12m	0.06	0.04	(0.00,0.00)	1.00	0.00
Flows	**-0.01**	**0.14**	**(0.00,0.08)**	**0.44**	**0.02**	**Size**	**0.05**	**0.03**	**(0.00,0.00)**	**1.00**	**0.00**
1m	0.11	0.06	(0.00,0.00)	1.00	0.06	1m	0.03	0.01	(0.00,0.00)	1.00	0.00
3m	-0.01	0.11	(0.00,0.00)	0.44	0.00	3m	0.05	0.02	(0.00,0.00)	1.00	0.00
6m	-0.03	0.06	(0.00,0.00)	0.31	0.00	6m	0.05	0.01	(0.00,0.00)	1.00	0.00
12m	-0.09	0.03	(0.00,0.00)	0.00	0.00	12m	0.06	0.02	(0.00,0.00)	1.00	0.00
PA	**0.05**	**0.17**	**(0.03,0.06)**	**0.65**	**0.12**	**R2**	**0.05**	**0.07**	**(0.01,0.00)**	**0.83**	**0.09**
1m	0.12	0.12	(0.01,0.00)	0.97	0.21	1m	0.06	0.08	(0.01,0.00)	0.88	0.09
3m	0.02	0.14	(0.00,0.05)	0.48	0.04	3m	0.06	0.07	(0.00,0.00)	0.86	0.09
6m	0.10	0.19	(0.00,0.10)	0.76	0.22	6m	0.05	0.07	(0.00,0.00)	0.86	0.09
12m	-0.05	0.16	(0.00,0.09)	0.36	0.02	12m	0.03	0.06	(0.00,0.00)	0.73	0.10
Tstat	**0.05**	**0.11**	**(0.04,0.01)**	**0.76**	**0.06**	**AS**	**-0.02**	**0.14**	**(0.00,0.00)**	**0.47**	**0.00**
1m	0.10	0.07	(0.00,0.00)	0.86	0.13	1m	-0.02	0.12	(0.00,0.00)	0.41	0.00
3m	0.07	0.09	(0.00,0.00)	0.83	0.08	3m	-0.02	0.12	(0.00,0.00)	0.47	0.00
6m	0.04	0.08	(0.00,0.00)	0.79	0.01	6m	-0.01	0.13	(0.00,0.00)	0.50	0.00
12m	-0.02	0.12	(0.08,0.00)	0.56	0.00	12m	-0.01	0.17	(0.03,0.00)	0.50	0.00
AW	**0.10**	**0.08**	**(0.01,0.00)**	**1.00**	**0.38**	**CS**	**0.03**	**0.20**	**(0.00,0.06)**	**0.53**	**0.06**
1m	0.11	0.07	(0.00,0.00)	1.00	0.31	1m	0.12	0.20	(0.00,0.00)	0.75	0.25
3m	0.11	0.08	(0.00,0.00)	1.00	0.44	3m	-0.05	0.14	(0.00,0.00)	0.38	0.00
6m	0.10	0.08	(0.03,0.00)	1.00	0.41	6m	0.10	0.09	(0.00,0.00)	0.75	0.00
12m	0.09	0.06	(0.00,0.03)	1.00	0.34	12m	-0.05	0.09	(0.00,0.00)	0.25	0.00
CT	**0.04**	**0.09**	**(0.04,0.00)**	**0.72**	**0.04**	**GT**	**0.05**	**0.12**	**(0.01,0.08)**	**0.66**	**0.12**
1m	0.08	0.13	(0.00,0.00)	0.81	0.12	1m	0.16	0.19	(0.06,0.02)	0.90	0.48
3m	0.06	0.06	(0.00,0.00)	0.92	0.00	3m	0.04	0.07	(0.00,0.00)	0.81	0.00
6m	0.06	0.05	(0.00,0.00)	0.92	0.04	6m	0.07	0.10	(0.00,0.00)	0.83	0.02
12m	-0.04	0.05	(0.00,0.00)	0.23	0.00	12m	-0.05	0.06	(0.00,0.00)	0.10	0.00
FD	**0.00**	**0.04**	**(0.00,0.00)**	**0.58**	**0.00**	**ICI**	**0.03**	**0.11**	**(0.00,0.00)**	**0.55**	**0.00**
1m	-0.01	0.03	(0.00,0.00)	0.44	0.00	1m	0.01	0.11	(0.00,0.00)	0.50	0.00
3m	0.01	0.02	(0.00,0.00)	0.56	0.00	3m	0.02	0.11	(0.00,0.00)	0.52	0.00
6m	0.01	0.04	(0.00,0.00)	0.62	0.00	6m	0.03	0.10	(0.00,0.00)	0.56	0.00
12m	0.01	0.03	(0.00,0.00)	0.69	0.00	12m	0.04	0.10	(0.00,0.00)	0.62	0.00
RG	**0.04**	**0.07**	**(0.00,0.02)**	**0.75**	**0.11**	**RS**	**0.08**	**0.08**	**(0.00,0.00)**	**0.89**	**0.23**
1m	0.08	0.07	(0.00,0.00)	0.94	0.38	1m	0.02	0.07	(0.00,0.00)	0.56	0.00
3m	0.02	0.06	(0.00,0.00)	0.69	0.00	3m	0.07	0.05	(0.00,0.00)	1.00	0.12
6m	0.02	0.04	(0.00,0.00)	0.75	0.06	6m	0.12	0.08	(0.00,0.00)	1.00	0.44
12m	0.02	0.09	(0.00,0.00)	0.62	0.00	12m	0.11	0.08	(0.00,0.00)	1.00	0.38
SI	**0.06**	**0.18**	**(0.07,0.06)**	**0.61**	**0.09**	**STO**	**-0.03**	**0.07**	**(0.07,0.09)**	**0.29**	**0.02**
1m	0.20	0.30	(0.03,0.03)	0.78	0.34	1m	-0.08	0.06	(0.00,0.06)	0.04	0.00
3m	0.02	0.10	(0.00,0.00)	0.52	0.00	3m	-0.05	0.05	(0.02,0.04)	0.10	0.00
6m	0.11	0.08	(0.00,0.00)	1.00	0.00	6m	-0.02	0.04	(0.00,0.00)	0.25	0.00
12m	-0.09	0.08	(0.00,0.03)	0.12	0.00	12m	0.03	0.05	(0.00,0.00)	0.77	0.06
Vola	**-0.08**	**0.04**	**(0.00,0.00)**	**0.02**	**0.00**						
1m	-0.08	0.03	(0.00,0.00)	0.06	0.00						
3m	-0.07	0.04	(0.00,0.00)	0.00	0.00						
6m	-0.08	0.02	(0.00,0.00)	0.00	0.00						
12m	-0.09	0.05	(0.00,0.00)	0.00	0.00						

Table II.A4: **Impact of Fixing Specific Portfolio Construction Choices: Rebalancing Frequency, 1985-2022**

This table shows key statistics for each predictor across all paths holding the individual choices of the portfolio construction nodes constant. For each predictor, the first row (in bold) represents the statistics computed at a predictor level, followed by the statistics computed for a given *rebalancing frequency*: Monthly (Mon.), quarterly (Qrt.), semiannually (Semi.), and annually (Ann.) held constant. These summary statistics include the mean (**Mean**, in % per month) of premia across all paths for each predictor or a group of predictors. Column **NSE** contains the non-standard error (in % per month; see II.2). Column **Left-right** is the relative number of significant deviations to the left and right of the median using a 5% significance level (see II.3). Columns **Pos.** and **Sig.** show the relative number of positive premia and t-statistics larger than 1.96. Standard errors used in the computation of **Left-right** and **Sig.** columns are computed using Newey and West (1987) standard errors with six lags.

	Mean	NSE	Left-right	Pos.	Sig.		Mean	NSE	Left-right	Pos.	Sig.
Age	**-0.04**	**0.04**	**(0.00,0.00)**	**0.06**	**0.00**	**Exp**	**0.09**	**0.05**	**(0.00,0.00)**	**1.00**	**0.03**
Mon.	-0.04	0.03	(0.00,0.00)	0.06	0.00	Mon.	0.09	0.06	(0.00,0.00)	1.00	0.06
Qrt.	-0.04	0.03	(0.00,0.00)	0.12	0.00	Qrt.	0.09	0.05	(0.00,0.00)	1.00	0.06
Semi.	-0.04	0.03	(0.00,0.00)	0.00	0.00	Semi.	0.09	0.04	(0.00,0.00)	1.00	0.00
Ann.	-0.03	0.03	(0.00,0.00)	0.06	0.00	Ann.	0.09	0.05	(0.00,0.00)	1.00	0.00
Flows	**-0.01**	**0.14**	**(0.00,0.08)**	**0.44**	**0.02**	**Size**	**0.05**	**0.03**	**(0.00,0.00)**	**1.00**	**0.00**
Mon.	0.03	0.10	(0.00,0.00)	0.75	0.00	Mon.	0.04	0.03	(0.00,0.00)	1.00	0.00
Qrt.	-0.00	0.10	(0.00,0.12)	0.44	0.00	Qrt.	0.05	0.02	(0.00,0.00)	1.00	0.00
Semi.	-0.02	0.09	(0.00,0.06)	0.31	0.00	Semi.	0.05	0.01	(0.00,0.00)	1.00	0.00
Ann.	-0.04	0.07	(0.00,0.25)	0.25	0.06	Ann.	0.04	0.02	(0.00,0.00)	1.00	0.00
PA	**0.05**	**0.17**	**(0.03,0.06)**	**0.65**	**0.12**	**R2**	**0.05**	**0.07**	**(0.01,0.00)**	**0.83**	**0.09**
Mon.	0.15	0.14	(0.01,0.01)	0.98	0.34	Mon.	0.06	0.07	(0.01,0.00)	0.85	0.09
Qrt.	0.02	0.15	(0.00,0.00)	0.57	0.01	Qrt.	0.05	0.07	(0.00,0.00)	0.85	0.09
Semi.	0.04	0.17	(0.00,0.11)	0.57	0.14	Semi.	0.05	0.08	(0.00,0.00)	0.83	0.09
Ann.	-0.03	0.15	(0.03,0.00)	0.46	0.00	Ann.	0.05	0.07	(0.01,0.00)	0.80	0.10
Tstat	**0.05**	**0.11**	**(0.04,0.01)**	**0.76**	**0.06**	**AS**	**-0.02**	**0.14**	**(0.00,0.00)**	**0.47**	**0.00**
Mon.	0.08	0.12	(0.05,0.00)	0.81	0.12	Mon.	-0.02	0.14	(0.00,0.00)	0.47	0.00
Qrt.	0.05	0.12	(0.04,0.00)	0.77	0.07	Qrt.	-0.02	0.14	(0.00,0.00)	0.44	0.00
Semi.	0.04	0.11	(0.04,0.01)	0.74	0.03	Semi.	-0.01	0.14	(0.00,0.00)	0.47	0.00
Ann.	0.03	0.10	(0.04,0.00)	0.71	0.00	Ann.	-0.01	0.13	(0.00,0.00)	0.50	0.00
AW	**0.10**	**0.08**	**(0.01,0.00)**	**1.00**	**0.38**	**CS**	**0.03**	**0.20**	**(0.00,0.06)**	**0.53**	**0.06**
Mon.	0.10	0.07	(0.03,0.00)	1.00	0.31	Mon.	0.15	0.11	(0.00,0.00)	1.00	0.25
Qrt.	0.10	0.08	(0.00,0.00)	1.00	0.41	Qrt.	0.01	0.10	(0.00,0.00)	0.38	0.00
Semi.	0.10	0.08	(0.03,0.00)	1.00	0.41	Semi.	0.01	0.21	(0.00,0.00)	0.50	0.00
Ann.	0.10	0.08	(0.00,0.03)	1.00	0.38	Ann.	-0.05	0.11	(0.00,0.19)	0.25	0.00
CT	**0.04**	**0.09**	**(0.04,0.00)**	**0.72**	**0.04**	**GT**	**0.05**	**0.12**	**(0.01,0.08)**	**0.66**	**0.12**
Mon.	0.05	0.10	(0.06,0.00)	0.75	0.04	Mon.	0.08	0.13	(0.04,0.15)	0.71	0.25
Qrt.	0.06	0.13	(0.10,0.00)	0.71	0.08	Qrt.	0.06	0.17	(0.12,0.00)	0.67	0.23
Semi.	0.04	0.06	(0.00,0.00)	0.71	0.04	Semi.	0.05	0.15	(0.00,0.00)	0.60	0.02
Ann.	0.02	0.06	(0.00,0.00)	0.71	0.00	Ann.	0.02	0.08	(0.00,0.00)	0.67	0.00
FD	**0.00**	**0.04**	**(0.00,0.00)**	**0.58**	**0.00**	**ICI**	**0.03**	**0.11**	**(0.00,0.00)**	**0.55**	**0.00**
Mon.	0.01	0.03	(0.00,0.00)	0.56	0.00	Mon.	0.03	0.11	(0.00,0.00)	0.56	0.00
Qrt.	0.01	0.03	(0.00,0.00)	0.69	0.00	Qrt.	0.02	0.11	(0.00,0.00)	0.58	0.00
Semi.	-0.00	0.04	(0.00,0.00)	0.44	0.00	Semi.	0.02	0.11	(0.00,0.00)	0.52	0.00
Ann.	0.00	0.03	(0.00,0.00)	0.62	0.00	Ann.	0.03	0.11	(0.00,0.00)	0.54	0.00
RG	**0.04**	**0.07**	**(0.00,0.02)**	**0.75**	**0.11**	**RS**	**0.08**	**0.08**	**(0.00,0.00)**	**0.89**	**0.23**
Mon.	0.04	0.05	(0.00,0.00)	0.81	0.06	Mon.	0.06	0.11	(0.00,0.00)	0.75	0.19
Qrt.	0.02	0.07	(0.00,0.00)	0.62	0.06	Qrt.	0.07	0.09	(0.00,0.06)	0.81	0.25
Semi.	0.03	0.07	(0.00,0.06)	0.69	0.12	Semi.	0.08	0.06	(0.00,0.00)	1.00	0.19
Ann.	0.06	0.05	(0.12,0.00)	0.88	0.19	Ann.	0.09	0.07	(0.00,0.00)	1.00	0.31
SI	**0.06**	**0.18**	**(0.07,0.06)**	**0.61**	**0.09**	**STO**	**-0.03**	**0.07**	**(0.07,0.09)**	**0.29**	**0.02**
Mon.	0.17	0.14	(0.00,0.20)	0.88	0.25	Mon.	-0.05	0.08	(0.10,0.10)	0.21	0.00
Qrt.	0.06	0.14	(0.05,0.00)	0.75	0.09	Qrt.	-0.04	0.07	(0.06,0.15)	0.21	0.02
Semi.	0.04	0.26	(0.11,0.00)	0.52	0.00	Semi.	-0.03	0.08	(0.02,0.12)	0.29	0.02
Ann.	-0.03	0.12	(0.00,0.00)	0.28	0.00	Ann.	-0.01	0.05	(0.00,0.02)	0.46	0.02
Vola	**-0.08**	**0.04**	**(0.00,0.00)**	**0.02**	**0.00**						
Mon.	-0.07	0.04	(0.00,0.00)	0.06	0.00						
Qrt.	-0.08	0.03	(0.00,0.00)	0.00	0.00						
Semi.	-0.09	0.03	(0.00,0.00)	0.00	0.00						
Ann.	-0.07	0.04	(0.00,0.00)	0.00	0.00						

Chapter III

Twitter-Based Attention and the Cross-Section of Cryptocurrency Returns

(with Arnaud T. Maître and Florian Weigert)

1. Introduction

Social media has significant effects on financial markets, as illustrated by the involvement of the WallStreetBets subreddit in the Gamestop short squeeze and the infamous "to the moon" tweets, leading to subsequent Dogecoin price hikes.[1] Internet platforms are particularly important for cryptocurrencies, as they constitute the preferred medium of information exchange between cryptocurrency market participants, with Twitter arguably being the most important platform.[2] To put the importance of social media, and especially Twitter, for cryptocurrencies in context, about 90% of cryptocurrencies have Twitter accounts, whereas this proportion is only 50% for US public firms (Hosseini et al., 2020).

Given the size of the cryptocurrency market and the significance of Twitter as an information source, studying their relationship is essential for understanding the cross-section of expected cryptocurrency returns.[3] Recent literature provides mixed evidence on the link between Twitter activity and cryptocurrency prices. In the cross-section, Benedetti and Kostovetsky (2021) provide evidence supportive of an overreaction channel consistent with Barber and Odean (2008) and Da, Engelberg, and Gao (2011), while Borri et al. (2022) identify a negative risk premium for investor attention. In the time-series, Liu and Tsyvinski (2021) and Borri et al. (2022) report a positive relationship between the number of tweets and future cumulative cryptocurrency returns. Altogether, the link between Twitter-based investor

[1]The topic has received wide attention in the empirical finance literature. Previous research shows that social media content has predictive power over expected stock returns and expected earnings (Chen et al., 2014; Bartov, Faurel, and Mohanram, 2018; Broadstock and Zhang, 2019; Gu and Kurov, 2020).

[2]The first Bitcoin transaction was arranged on a forum (source) followed by the first-ever Bitcoin-related tweet as early as January 11$^{\text{th}}$, 2009 (source).

[3]According to Forbes, August 2024, the global cryptocurrency market cap is US$ 2.25 trillion.

attention and cryptocurrency expected returns remains unclear.

Inspired by Da, Engelberg, and Gao (2011), we inspect the relation between investor attention and cryptocurrency returns and capture abnormal investor attention computed as the (log) number of tweets during the current day minus the (log) mean number of tweets during the previous 30 days. The novelty of our paper is that we collect several different samples of tweets for each cryptocurrency. We consider both tweets written by the organization developing the cryptocurrency (*Official*-tweets) and tweets written by all other users, i.e., user-generated tweets. We further separate user-generated tweets into tweets on the cryptocurrency's ticker (*Ticker*-tweets) and tweets sent to the cryptocurrency's official account (*Mention*-tweets).

Our motivation to categorize tweets is that different categories have unique features, characteristics, and usage motivations. For instance, Barber et al. (2022) find that unique features of the Robinhood app partly drive attention-induced trading among the platform's investors. Likewise, *Ticker*-tweets possess the unique feature of being clickable and are commonly employed for discussing trading strategies. Once a user clicks on these tweets, they are presented with the latest tweets related to the associated financial security, making *Ticker*-tweets easily seen by users who are not following the tweet's author. Therefore, *Ticker*-tweets reach a wider audience than the other types of tweets. *Mention*-tweets can be viewed as public messages, as the account tagged by this type of tweet receives a notification. *Official*-tweets, on the other hand, represent corporate announcements. It is particularly interesting to investigate the possible impact of these features on crypto investors' trading behavior through the lens of attention-induced trading.

In our empirical analysis, we first consider the aggregated set of tweets, i.e., *All*-tweets. For each cryptocurrency, an abnormal attention measure similar to Da, Engelberg, and Gao (2011) is created. We show that abnormal attention is positively related to contemporaneous returns and next-day returns. On average, a one cross-sectional standard deviation rise in Twitter-based abnormal attention is associated with an increase in cryptocurrency excess returns by 0.70% contemporaneously and 0.11% on the following day. By utilizing our Twitter samples separately, we show that this effect is driven entirely by the user-generated tweets, with *Ticker*-tweets showing a stronger impact than *Mention*-tweets. Thus, unlike Benedetti and Kostovetsky (2021), we do not find a link between *Official*-tweets and excess returns in the cross-section, suggesting that the predictability of Twitter attention arises mainly from user-generated content.

Although the existing literature offers several possible explanations, we show that tweets written by Twitter users predict returns consistently in line with the continued overreaction channel. Boosts in tweet posting activity grab the attention of retail investors who, due to limited cognitive ability, become more likely to purchase these cryptocurrencies (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). This mechanism results in a temporary positive price pressure. Usually, price rises due to overreaction tend to correct over the following days after the surge in attention, which is not the case in our empirical data. In this regard, Schmeling, Schrimpf, and Todorov (2023) point out that limits-to-arbitrage are more important in the crypto market than for other asset classes. Hence, we hypothesize that, e.g., the difficulty of short-selling cryptocurrencies is likely to explain the absence of reversal in returns.

To better understand this difference in predictability power, we create a new lexicon to capture the textual characteristics of user-generated content following the methodology of Renault (2017). We apply this new lexicon to compute a score on the cryptocurrency-day level and show that this score predicts the cross-section of expected returns in the same fashion as abnormal attention. We interpret this finding as evidence that the predictability of user-generated content arises in part from its unique textual content. Furthermore, we show that our results are not driven by the GameStop short-squeeze in 2021, which was surrounded by high social media activity and positive performance for cryptocurrencies.

Given that social media is frequently utilized to report bugs, hacks, or technical problems with blockchain technology, a concurrent explanation of our results is that Twitter abnormal attention predicts returns through its link with future development activities. This hypothesis is motivated by the results of Cong, Li, and Wang (2021); Liu, Sheng, and Wang (2022) who reveal that cryptocurrency valuations and ICO success are linked to the quality of their underlying technology. By using the daily number of commits on GitHub as a proxy for technological innovation, we show that *Official*-tweets is the only Twitter sample predicting future technological improvements in the cross-section. As *Official*-tweets do not have predictive power for future returns, the technological innovation channel is not supported by our findings. This difference in predictability between the different Twitter samples supports the notion that social media content is indeed heterogeneous. It also highlights the need to carefully select the appropriate social media data that is best suited for the desired application.

Our paper contributes to two strands of literature. First and foremost, we add to quantify the impact of influential social media users on their followers (Benetton et al., 2024; Pedersen, 2022). We document that the posting activity of influential users increases the magnitude of the continued overreaction effect. Our results reveal that even small influencers can have a significant impact on cryptocurrency prices (Benetton et al., 2024). We argue that, similar to media coverage (Hillert, Jacobs, and Müller, 2014), influential tweets exacerbate behavioral biases. Therefore, our paper contributes to the nascent literature that explores behavioral biases induced by social media activity.[4]

Second, we add to the emerging literature that studies the impact of information salience on investor behavior (Bose et al., 2022; Kumar, Ruenzi, and Ungeheuer, 2021; Frydman and Wang, 2020). For instance, Barber et al. (2022) find that Robinhood users are more likely to purchase stocks displayed in the Top-Movers list than stocks with similar returns absent from the list. In our context, we document that *Ticker*-tweets are seen on average by a wider audience than the other types of tweets after controlling for their number of likes and retweets. Despite this increased visibility of *Ticker*-tweets over *Mention*-tweets, both Twitter samples predict returns similarly in the cross-section. This suggests that our results are not driven by information salience alone. Instead, we provide evidence that the difference in predictability between user-generated content and tweets written by cryptocurrencies is also driven by content differences.

Sections 2 and 3 describe the data and methods used in the paper. We investigate the drivers of Twitter-based attention in Section 4. Section 5 discusses the interplay between aggregate attention and cryptocurrency returns by breaking down attention into several components using the specificity of Twitter. The robustness of our main findings is addressed in Section 6. Finally, Section 7 concludes the paper.

2. Data and Sample Construction

We apply several datasets to construct our sample. Our primary datasets are data on cryptocurrency returns from CoinMarketCap and data on cryptocurrency-related posts on Twitter. We explain the construction of our Twitter sample in detail in Section 2.2. Additionally, we gather data on each cryptocurrency from GitHub, the leading platform for software development and project collaboration. We use data sampled at a daily frequency in our empirical tests, as Twitter effects documented by the literature are generally short-lived (Benedetti and Kostovetsky, 2021; Gu and Kurov, 2020).

[4]For example, social media users tend to self-expose to information in line with their beliefs (Cookson, Engelberg, and Mullins, 2023) and are influenced by investment returns experienced by users they are following (Bailey et al., 2018; Pedersen, 2022).

2.1. *Cryptocurrency data*

CoinMarketCap (henceforth CMC) is a widely accepted data source for cryptocurrency market data. However, the data from the CMC website is subject to survivorship bias because it only provides information on currently listed cryptocurrencies. To address this, we use an API to download survivorship bias-free data from CMC, following the methodology proposed by Ammann et al. (2022) that corrects the bias.[5]

Our dataset is at a daily frequency for the period from 2018 to 2022. We drop assets with missing volume data and exclude cryptocurrencies with erroneously reported data. Additionally, we drop stablecoins, which are cryptocurrencies whose value is pegged to other assets such as USD or gold. Our final dataset includes both cryptocoins and tokens.[6] Figure III.1 displays the average monthly number of cryptocurrencies meeting our criteria.

[Insert Figure III.1 here]

2.2. *Twitter data*

Given the download restrictions imposed by Twitter, we limit our sample to the 165 largest cryptocurrencies as of the end of 2017.[7] We choose this year because it includes a large number of cryptocurrencies meeting our criteria. Our sample excludes cryptocurrencies that have changed names during our sample period or those that do not have a Twitter account.[8] For each selected cryptocurrency, we separately collect tweets written by the organization developing the cryptocurrency (henceforth *Official*-tweets) and tweets written by all other users, i.e., user-generated tweets. While collecting *Official*-tweets is straightforward (simply by requesting tweets posted by the developers' verified accounts), collecting user-generated tweets is more complex. The complexity arises from the fact that users have several ways to signal that their

[5]We access the API through the R package named ***crypto2***, which can be found at https://CRAN.R-project.org/package=crypto2. The authors of this package are Sebastian Stöckl and Jesse Vent. More information about the package can be found on the author's personal website: https://www.sebastianstoeckl.com/

[6]Crypto coins primarily act as a medium of exchange and a store of value. Some well-known examples include Bitcoin (BTC), Ethereum (ETH), and Dogecoin (DOGE). Crypto tokens, on the other hand, are created to serve various purposes, such as utility, ownership, and governance rights. For instance, the Edgeless token can be used to play in an online casino.

[7]Due to the abrupt stop of the Twitter academic API in 2023, we restrict the Twitter samples to the 165 cryptocurrencies for which data could be fully collected.

[8]While it is true that excluding cryptocurrencies that changed their names during the sample period introduces a look-ahead bias, we believe that this bias is unlikely to have a substantial impact on our study's findings. The difference in mean returns between the two samples is not statistically significant from zero.

tweet is about a specific cryptocurrency. For instance, users can utilize the name (e.g., #*bitcoin* or #*BTC*), the ticker (e.g., \$*BTC*), or tag (mention) the official account of the cryptocurrency (e.g., @*Bitcoin*). We choose to collect tweets on the cryptocurrency's ticker (henceforth *Ticker*-tweets) and tweets sent to the cryptocurrency's official account (henceforth *Mention*-tweets) to exclude tweets with ambiguous hashtags.[9] Ultimately, we categorize tweets into three categories in our Twitter sample: *Official*-tweets, *Mention*-tweets, and *Ticker*-tweets.

Our motivation to categorize tweets is that different categories contain unique features, characteristics, and usage motivations. For instance, Barber et al. (2022) find that unique features of the Robinhood app partly drive attention-induced trading among the platform's investors. *Ticker*-tweets, commonly employed for discussing trading strategies, possess the unique feature of being clickable. When a user clicks on these tweets, they can view the latest tweets related to the associated financial security. Therefore, they have a potentially higher reach than other types of tweets, as they can be easily seen by users who are not following the tweet's author, thanks to the clickable feature. It is particularly interesting to investigate the possible impact of this feature on crypto investors' trading behavior through the lens of attention-induced trading. *Mention*-tweets can be viewed as public messages, as the account tagged by this type of tweet receives a notification. *Official*-tweets, on the other hand, represent corporate announcements.

Table III.1 summarizes the information about the sample size and the characteristics of each sample. The number of cryptocurrencies in the *Official* sample is lower than in the other samples because we could not retrieve any *Official*-tweets for eleven cryptocurrencies. We chose to leave the history of *Official*-tweets as missing for those eleven cryptocurrencies because it is hard to tell whether those eleven cryptocurrencies had never posted any tweets during the sample period or had retroactively deleted their tweets.

[Insert Table III.1 here]

Figure III.2 reports the number of tweets posted per sample per day. In all our samples, the number of Twitter posts published per day and per cryptocurrency corresponds to the actual number of tweets published on that date. Twitter activity has increased over our sample period, with the only exception being the number of *Official*-tweets, which has remained stable over time. For all samples except the *Official* sample, we only download the 100 most relevant tweets per day over the

[9] For instance, consider two cryptocurrencies named ICON (ICX) and TRON (TRX). Both #*icon* and #*tron* can be associated with different meanings, not just cryptocurrencies.

period from 2018 to 2022.[10] The *Official* sample includes every tweet posted by the cryptocurrency's official account during the sample period.

[Insert Figure III.2 here]

Unfortunately, as Twitter provides us with tweets based on textual matching of keywords, *Ticker*-tweets may contain measurement error, as tickers are generally not unique. We do not expect this bias to be substantial for two reasons. First, we only consider tweets written in English, which limits the number of ticker homonyms in our sample. Second, as cryptocurrencies are a popular topic on social media, we expect the *Ticker* sample to be primarily composed of tweets about cryptocurrencies.[11] Our Twitter samples are also subject to survivorship bias, as tweets and Twitter accounts can be deleted by their creators. This problem is more severe for the *Official* sample, as it is impossible to retrieve tweets from deleted accounts. Consequently, we miss *Official* tweets from cryptocurrencies with deleted Twitter accounts, which are likely to be defunct. However, also in this case, the bias is assumed to be small, as we could retrieve *Official* tweets from 35 out of the 44 defunct cryptocurrencies included in our Twitter samples.

2.3. GitHub data

To better identify the channels through which Twitter attention predicts the cross-section of cryptocurrency returns, we also collect data on each cryptocurrency from GitHub, the leading platform for software development collaboration. Specifically, we collect the list of all historical contributions (commits) made by developers on all repositories owned by the organization developing the cryptocurrency. Similar to *Official*-tweets, GitHub data is also subject to survivorship bias, as cryptocurrencies with missing data are likely defunct. Our GitHub data covers 143 cryptocurrencies and includes data for 32 out of the 44 defunct cryptocurrencies in our sample.[12]

[10] As determined by Twitter, the exact methodology is not disclosed but considers, among other factors, the degree of keyword matching, tweet engagement, and the author's popularity.

[11] We also report the results if ambiguous tickers are removed from our sample as a robustness check. We judge a ticker as ambiguous if a stock present in the CRSP database has a similar ticker during our sample period. For example, consider a cryptocurrency named Neo (ticker: $*NEO*; https://x.com/neo_blockchain) and a publicly traded company on NASDAQ, NeoGenomics, Inc. (ticker: $*NEO*; https://x.com/NeoGenomics).

[12] We miss data on 22 cryptocurrencies. Three cryptocurrencies use other platforms than GitHub to share their code. We could not find information about 18 cryptocurrencies, and one cryptocurrency's repository contains no commits.

3. Empirical Methodology

3.1. Abnormal Twitter-based attention

Inspired by Da, Engelberg, and Gao (2011), we define our main measure of investor attention as abnormal attention, which we compute as the (log) number of tweets during the current day minus the (log) mean number of tweets during the previous 30 days.[13] Specifically,

$$Abn\,Attention_{i,t} = Log(NT_t) - Log(E[NT_{t-30}, NT_{t-29}, ..., NT_{t-1}]), \qquad \text{(III.1)}$$

where NT_t is the number of tweets on day t. Abnormal attention has the advantage of being less sensitive to large spikes in the number of tweets than the (log) number of tweets because it takes into account the average level of attention through its rolling mean component.

3.2. Twitter lexicon

The literature has not reached a consensus on which type of social media content is more relevant to empirical studies. Some authors focus on *Ticker*-tweets or other social media platforms oriented towards investing to capture the attention and sentiment of financially savvy users (Chen et al., 2014; Renault, 2017; Ardia and Bluteau, 2024; Gu and Kurov, 2020), whereas others look at content published by companies or aggregated content (Benedetti and Kostovetsky, 2021; Da, Engelberg, and Gao, 2011; Borri et al., 2022). Despite most studies focusing on one type of social media data, there is limited evidence documenting the effects of this decision. To fill this gap, we construct a new lexicon aimed at capturing the differences in textual content between *Ticker*-tweets and other tweets.

We construct our lexicon using a methodology similar to Renault (2017). First, we clean and process the text following common practices. We remove stop words from the tweets and replace words with bigrams when possible, using our own list of bigrams supplemented by the list of Renault (2017).[14] We remove all punctuation. Email addresses, web links, and emojis are replaced by keywords (mailtag, linktag, emojipos, or emojineg).[15] Following Renault (2017), we also add a prefix (negtag) to

[13]Da, Engelberg, and Gao (2011) rely on Google's Search Volume Index (SVI) to measure investor attention. Their main variable is abnormal SVI, which is defined as the (log) SVI during the current week minus the (log) median SVI during the previous eight weeks.

[14]For forming the bigrams, we use 300,000 randomly selected tweets over the complete sample.

[15]We manually classify the most frequent emojis as either being positive or negative.

tokens directly following a negation. Finally, we create a training sample composed of 180,000 tweets equally distributed across *Ticker*-tweets, *Mention*-tweets, and *Official*-tweets.

Importantly, the training sample does not contain any retweets. To avoid any look-ahead bias, the training sample is only composed of tweets published in 2017. We also filter out irrelevant unigrams and bigrams by restricting our sample to n-grams appearing in at least 0.01% of tweets. To prevent our final lexicon from being too influenced by prolific authors, cryptocurrencies, or seasonal topics, we require n-grams to be used by at least 10 different authors, to be used in discussions about at least 10 different cryptocurrencies, and to be used in at least three distinct months. Our last filtering step is to remove tickers and user mentions from the selected tokens to prevent our lexicon from loading on terms that are specific to one sample of tweets.

To spot unigrams and bigrams that are specific to *Ticker*-tweets, we compute a coefficient for each token i which is defined as:

$$Coefficient_i = \frac{Occurence_i^T - \max(Occurence_i^M ; Occurence_i^O)}{Occurence_i^T + \max(Occurence_i^M ; Occurence_i^O)}, \tag{III.2}$$

where T, M, and O stand for Ticker, Mention, and Official, respectively.

Like Renault (2017), we only keep in the lexicon n-grams with a coefficient in the first or the last quintile. The resulting lexicon is composed of 2,833 n-grams. Table III.2 presents a selection of the lexicon's vocabulary. We observe that the lexicon loads positively on terms related to trading such as 'bull–flag' or 'sold–all' regardless of their sentiment and negatively on terms related to technology or news such as 'support–team' or 'fully-decentralized'.

The lexicon confirms the intuition that different types of users use different types of tweets. Users tend to use mainly *Ticker*-tweets to discuss cryptocurrency investing. In contrast, *Official*-tweets and *Mention*-tweets are used for corporate announcements and to signal and solve issues related to the cryptocurrency. In the rest of the paper, we compute the lexicon score per tweet as the equal-weighted average coefficient of all n-grams present in the tweet that belong to the lexicon. We then compute the cryptocurrency-day lexicon score as the average score of all tweets published about that cryptocurrency on a given date. To illustrate our methodology, the hypothetical tweet '*@username please contact support team*' would get a score of $-0.8189 = \frac{-0.7353 + -0.9024}{2}$ where -0.7353 and -0.9024 are respectively the score assigned by the lexicon for the n-grams 'contact' and 'support-team'.

[Insert Table III.2 here]

3.3. Variables description and summary statistics

As in our Twitter data, we observe the number of likes, text, and author identity for a subset of the tweets posted per cryptocurrency-date; however, we are not able to identify the complete activity of tweet authors. Consequently, we choose to measure influential user activity in an indirect manner by checking if at least one influential user has tweeted on this cryptocurrency-date. We construct a dummy variable ($Popular\,Tweet^{All}_{i,t}$) equal to one when at least one tweet published per cryptocurrency-date has an aggregated number of likes and replies equal to or above 100 across all of our Twitter samples.[16] The number 100 corresponds to the 93$^{\text{th}}$ percentile of the distribution of the sum of the number of comments and likes. Classifying tweets with as few as 100 likes and/or comments as influential might seem optimistic. However, the actual number of users reached by a tweet is often much larger than its number of likes and replies. Using the number of views of each tweet, we find that the median number of views for tweets with at least 100 likes is 17,523 views and this number does not include the views obtained by the activity induced by popular tweets.[17] We choose to compute this variable across all Twitter samples simultaneously because we are interested in how influential users' activity, regardless of their tweeting preferences, impacts other Twitter users.

Sentiment- and lexicon-based variables are constructed at the tweet level and then averaged at the cryptocurrency-date level. To compute sentiment, we use a bag-of-words approach and consider the dictionary of Renault (2017), which is the main lexicon used for quantifying the sentiment of texts from social media in the financial academic literature.[18]

We follow Liu, Tsyvinski, and Wu (2022) to construct the cryptocurrency market factor. Throughout the paper, we control for a wide range of variables impacting asset prices. We control for illiquidity and liquidity fluctuations using the measure of Amihud (2002); Chordia, Subrahmanyam, and Anshuman (2001) and measures of codependence with the cryptocurrency market using co-skewness (Kraus and Litzenberger, 1976; Harvey and Siddique, 2000) and co-kurtosis (Fang and Lai, 1997; Jondeau and Rockinger, 2006). In addition, we also include controls related to volatility, higher-order moments of returns, and tail risk using variables such as idiosyncratic volatility (Ang et al., 2006), skewness, kurtosis, and value-at-risk.

[16] The number of retweets is excluded from this total. We do so because the number of retweets is shared across the retweets and the original tweet. Therefore, using the number of retweets to spot popular tweets can be misleading.

[17] We do not directly use the number of views to spot popular tweets in our analysis because this variable is only available for a limited number of tweets.

[18] Our results do not change if we use the lexicon of Loughran and McDonald (2011) instead.

Momentum is included in our set of controls, as momentum strategies are profitable for a wide range of asset classes (Asness, Moskowitz, and Pedersen, 2013) and are related to investor attention (Daniel, Hirshleifer, and Subrahmanyam, 1998; Hillert, Jacobs, and Müller, 2014). Hillert, Jacobs, and Müller (2014) find that media coverage of larger firms attracts investor attention (Solomon, Soltes, and Sosyura, 2014) and intensifies investors' overconfidence and self-attribution biases (Daniel, Hirshleifer, and Subrahmanyam, 1998), which result in temporary improvements in momentum returns. However, evidence from Liu and Tsyvinski (2021) suggests that the relationship between momentum returns and investor attention is different between cryptocurrencies and stocks. The authors find that the return predictability of attention and momentum does not encompass each other.

In addition, we add a control variable capturing the 'MAX effect' (Bali, Cakici, and Whitelaw, 2011; Fong and Toh, 2014) which is related to investor sentiment and gambling preferences (Fong and Toh, 2014). The GameStop short-squeeze attracted a significant number of retail speculators and triggered a surge in investor attention (Lyócsa, Baumöhl, and Vỳrost, 2022). This event suggests that the 'Max effect' and investor attention could be related. Table III.3 provides a description of all variables used in our paper.

[Insert Table III.3 here]

Given the suspicious returns documented by Ammann et al. (2022), we follow their approach and trim the daily cryptocurrency returns at the 99% level. The remaining variables are winsorized at the 1% level. Table III.4 displays descriptive statistics for the variables used in our paper. In our data, abnormal attention is negative on average and is negatively skewed across all samples. Furthermore, we can see that tweets tend to have positive sentiment and to focus on terms related to trading as indicated by the positive mean of sentiment and lexicon scores. In terms of popularity on Twitter, cryptocurrencies differ significantly. For instance, Bitcoin gets a median number of tweets of 11,236 per day, while the median across all cryptocurrencies is 60 tweets. The low number of tweets for some assets is not purely driven by size; the smallest cryptocurrency we consider as a market cap of $47 million as of the start of our sample. Even if the median number of tweets seems low, one has to remember that the actual number of people reached by tweets is larger. The median number of views per tweet is about 118 and it does not include the number of times that tweets were collected by web-scrapping algorithms. Anecdotal evidence suggests that automatized data collection efforts can be sizeable. For instance, 27% of hedge funds surveyed by Ernst and Young in their 'EY Global Hedge Fund and Investor Survey 2017' reports that they are using or planning to use social media data as part of their investment process.[19] Furthermore, the low popularity of some cryptocurrencies goes against us finding any return predictability for Twitter attention.

[Insert Table III.4 here]

4. Characterizing Twitter-Based Investor Attention

4.1. Drivers of Twitter-based investor attention

We start our empirical analysis by investigating the drivers of Twitter-based investor attention. For this purpose, Table III.5 reports the results on contemporaneous relationships between the different attention measures. We use changes in attention rather than abnormal attention because we are interested in what causes variation in attention more generally. The panel regression models control for the variables defined in Panel A of Table III.3.

Table III.5 reveals that attention variables are positively related to each other at the 1% or 5% significance level. Daily excess returns are also significantly linked to contemporaneous changes in attention for *Mention*-tweets and *Ticker*-tweets at the 1% level, but do not appear to be linked to the change in the number of *Official*-tweets. The presence of $Popular\,Tweet_{i,t}^{All}$ is strongly positively related to changes in the number of tweets across all three samples. This is in line with the results of Benetton et al. (2024) who document that celebrity tweeting activity influences their followers' behavior. Variables derived from volume, $\Delta\,Volume_{i,t}$ and $Volume\,Vola_{i,t}$, are significantly linked with the dependent variables. Changes in volume have a similar regression coefficient to excess returns, and volatility of volume is negatively linked with changes in the number of tweets from each sample. Interestingly, momentum is negatively related to attention changes, which suggests that the association between attention and momentum is less strong for cryptocurrencies compared to equities (Liu and Tsyvinski, 2021). We do not find strong links between change in tweeting activity and the other control variables described in Panel A of Table III.3.

[Insert Table III.5 here]

[19]https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2017/12/ey-how-will-you-embrace-innovation.pdf

We run similar tests using future change in tweeting activity in Table III.6. Future changes in the different changes in attention measures are inversely related to their own one-day lags at the 1% significance level. Interestingly, we observe differences in cross-one-day auto-correlations between attention variables. For instance, *Official*-tweets and *Ticker*-tweets positively predict future changes in the number of tweets for the other sample at the 1% significance level. On the other hand, *Mention*-tweets are negatively related to future changes in *Official*-tweets. As in Table III.5, daily cryptocurrency returns and volume changes are strong predictors of the number of tweets, even for *Official*-tweets. Furthermore, our results reveal that coefficient estimates of $Popular\,Tweet_{i,t}^{All}$ become negative. We interpret these results as indicative that attention tends to revert to a normal level following large changes; Holding other factors equal, a *Popular*-tweet posted at time t stimulates the number of tweets posted at time t and reinforces the attention reversal at time $t+1$. Similarly to Table III.5, momentum is negatively related to future changes in tweeting activity in all models.

Taken together, Table III.5 and Table III.6 both suggest that tweeting activity tends to spike and revert partially over the following day. As expected, the link between past returns and Twitter attention is sizeable (Da, Engelberg, and Gao, 2011; Liu and Tsyvinski, 2021) and is the strongest for *Ticker*-tweets, both in terms of statistical and economic significance. Intuitively, this observation makes sense given that *Ticker*-tweets are mainly used for discussing trading and investments as shown in Table III.2. Despite Twitter users reacting to past and contemporaneous returns, change in attention is inversely related to momentum as documented by Liu and Tsyvinski (2021). The authors argue that if momentum arises from underreaction to news (Hong and Stein, 1999), then momentum and change in attention should be negatively correlated as observed in our data.

While the different attention variables are positively related contemporaneously, we observe different patterns at $t+1$, especially for *Mention*-tweets, which become negatively related to the other samples. One could wonder why *Official*-tweets and *Ticker*-tweets are positively associated with future tweeting activity in other samples. We posit that those two samples of tweets are on average more visible and therefore induce more people to tweet. *Ticker*-tweets have a unique clickable feature which should increase their visibility, hence their impact. Alternatively, *Official*-tweets are posted by a recognized Twitter user and constitute an important source of information for cryptocurrency investors. We investigate our hypothesis in the following section.

[Insert Table III.6 here]

4.2. Differences between the Twitter samples

The three categories of tweets used in this study have different characteristics. Their most important difference is how they are disseminated on Twitter. By default, a tweet is displayed on the profile page of the author and in the timeline of users following the author.[20] However, both *Ticker*-tweets and *Mention*-tweets differ from this default behavior. *Ticker*-tweets also appear in the respective financial security timeline. *Mention*-tweets can either follow the default behavior or be customized such that these tweets only appear in the timeline of users following both the author and the user being tagged. Therefore, *Ticker*-tweets should be more visible than classic tweets, whereas *Mention*-tweets should be less visible.

Several studies show that information salience impacts investor behavior (Barber and Odean, 2008; Barber et al., 2022). For instance, Tan, Wang, and Zhou (2015) find that better readability helps investors to better incorporate new information. Therefore, it is reasonable to expect that the Twitter samples predict returns differently in the cross-section based on their visibility. As a first step, we test in Table III.7 whether the category of the tweet indeed impacts its visibility and other tweet metrics. The different models use date and cryptocurrency fixed effects to control for unobserved invariant characteristics and cluster standard errors by authors. In addition, we exclude retweets from each estimated model. As original tweets and their retweets all share the same retweet count, keeping retweets in the sample may lead to spurious relationships.

Our results reveal that being a *Ticker*-tweet or an *Official*-tweet positively relates to the number of views obtained by the tweet. This effect is statistically significant at the 1% and 5% levels for *Official*-tweets and *Ticker*-tweets, respectively. In addition, *Ticker*-tweets are also significantly associated with a larger number of retweets. In contrast, we find a negative relationship between *Mention*-tweets and the number of views obtained by the tweet at the 1% level. Textual sentiment is positively related to the number of likes, replies, and retweets at the 1% significance level, but negatively correlated with the number of views at the 1% level. The lexicon score, which captures trading jargon, is positively linked with the number of views at the 1% significance level, but negatively associated with the other dependent variables at the 1% significance level.

The results support our intuition that *Ticker*-tweets are more visible than classic tweets thanks to their clickable feature. In addition, we find similar results for *Official*-tweets, which is in line with our hypothesis that investors value the news announced by *Official*-tweets. In contrast, the regression coefficients of *Mention*-

[20] https://help.twitter.com/en/using-x/types-of-posts

tweets in the last model are negative. As expected, *Mention*-tweets are generally written with the intent to reach a specific set of users. Interestingly, $Lexicon_{i,n,t}$ is positively associated with the number of views and negatively associated with the number of retweets and likes. Social media users do not seem to actively share tweets with high lexicon scores, but they still search for them as indicated by the positive correlation with the number of views. We observe similar patterns for the effects of tweets with negative sentiment on tweet visibility. One potential interpretation could be that Twitter users actively search for bad news and other users' opinions as part of their investment process. Such behavior would be in line with the 'Do your own research' (DYOR) advice which is frequently addressed to new cryptocurrency investors on social media.[21] Overall, the evidence contained in Table III.7 confirms that the different characteristics of tweets matter and impact their salience on Twitter, thus providing evidence that the choice of social media is not innocuous.

[Insert Table III.7 here]

5. Twitter-Based Investor Attention and the Cross-Section of Cryptocurrency Returns

5.1. Twitter-based abnormal attention

This section studies the link between abnormal attention and the cross-section of cryptocurrency returns. In Table III.8, we present the results of panel regressions of excess returns on abnormal attention and various controls as defined in Panel A of Table III.3. Model (1) investigates the association between Twitter attention and returns contemporaneously. The other models replace contemporaneous returns with future returns using different time horizons. One potential issue with *Ticker*-tweets is that their trading symbol is not unique, as trading symbols are attributed by the exchange where assets are traded. Given that cryptocurrencies are traded on their own exchanges, they can possibly share their ticker with other financial securities traded elsewhere such as US stocks. Therefore, for robustness, we replicate the same regressions with a different sample in Panel B; when doing so, we drop cryptocurrencies that share their ticker with a firm covered by the CRSP database during our sample period.[22]

[21] The unigram 'dyor' has a score of 0.5673 in the lexicon.

[22] The cryptocurrencies that are dropped in Panel B include both large and small cryptocurrencies. For instance, both Bitcoin and Ethereum are dropped in Panel B.

Our results reveal that the regression coefficient of $Abn\,Attention_t^{All}$ is statistically significant and positive in models (1) and (2) at the 1% significance level. Evaluating its economic significance, the effect is large; a one standard-deviation increase in $Abn\,Attention_t^{All}$ is related to an increase of 0.72% in contemporaneous daily excess returns and an increase in future daily excess returns of 0.11%. The association between abnormal attention at day t and excess returns at $t+1$ hence amounts to an annual 39%. The predictability of Twitter activity is robust to a wide range of popular predictors used in the literature. Among those predictors, only past returns, size, and change in volume are all significantly related to contemporaneous and future cryptocurrency returns. Volume change predicts returns in a similar way as abnormal attention. Past returns predict future returns up to two periods ahead, and size is negatively related to returns in all model specifications, consistent with previous literature. Furthermore, we document similar results in Panel B where cryptocurrencies with ambiguous trading symbols have been removed, suggesting that our results are not driven by tweets that should not be included in our sample.

The positive connection between $Abn\,Attention_t^{All}$ and excess returns is consistent with the results of Liu and Tsyvinski (2021), who also document a positive relationship between investor attention and cryptocurrency returns. However, we need to be cautious when interpreting the results of Table III.8, as several theories could explain why attention and (expected) returns are positively correlated. For instance, as Twitter is used to discuss the technology underlying cryptocurrencies, Twitter attention could predict technological improvements and therefore returns (Cong, Li, and Wang, 2021; Lyandres, Palazzo, and Rabetti, 2022; Liu, Sheng, and Wang, 2022). Another possibility is that attention predicts increases in the user base, which would have positive network externalities for existing cryptocurrency users (Cong, Li, and Wang, 2021; Sockin and Xiong, 2023).

Alternatively, the results could also be explained by an overreaction channel (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). Increases in tweet posting activity grab the attention of retail investors, who, due to limited cognitive ability, become more likely to purchase those attention-grabbing cryptocurrencies. This mechanism results in temporary positive price pressure. Usually, price rises caused by overreaction tend to correct over the following days after the surge in attention. However, we do not observe such price correction (as revealed in models (3) and (4)). In this regard, Schmeling, Schrimpf, and Todorov (2023) point out that limits-to-arbitrage are important in the cryptocurrency market and prevent some arbitrage strategies, which are commonly used for other asset classes, from being profitable for cryptocurrencies. The difficulty of short-selling cryptocurrencies could therefore explain the absence of reversal.[23]

As documented in Tables III.2 and III.7, each Twitter sample has some unique properties; Tweets are not equally salient between our samples, nor do they share similar textual content. Moreover, the samples also exhibit qualitative differences. For instance, *Official*-tweets are mainly constituted of announcements about their associated cryptocurrency and thus differ fundamentally from tweets written by the crowd. As a result, *Official*-tweets represent a more credible source of information than the other types of tweets for cryptocurrency investors. Therefore, in the following section, we analyze whether the three Twitter samples similarly predict returns.

[Insert Table III.8 here]

5.2. Refinements of Twitter-based abnormal attention

As each category of tweets differs in terms of types of authors, functionalities, or reach, we expect that the relationship between expected returns and Twitter-based attention may change depending on the Twitter sample being used. Compared to general tweets, *Mention*-tweets have the particularity of triggering a notification for the recipient of the mention. *Ticker*-tweets, upon being clicked, display the most recent *Ticker*-tweets about the corresponding financial asset.

We now study if the qualitative and quantitative differences between our Twitter samples translate into different relationships in the cross-section of cryptocurrency expected returns. In Table III.9, we estimate panel regressions of excess returns on the Twitter samples while controlling for our set of control variables. *Mention*-tweets are positively linked with excess returns in t and $t+1$ at the 1% and 10% significance levels, respectively. In contrast, we do not find any link between *Official*-tweets and returns. The regression coefficients of *Ticker*-tweets are positive and statistically significant at the 1% significance level in both models (1) and (2). The results of both panels are similar.

The results of Table III.9 illustrate that the return predictability of Twitter activity is mainly derived from user-generated content and not from the announcements made by cryptocurrencies. This evidence provides additional support for

[23] A risk-based explanation of our results is also possible. Andrei and Hasler (2015) find that variation in investor attention affects the volatility through its impact on the speed of information incorporation into asset prices. Therefore, variation in attention is compensated with similar risk premia variation. We provide additional tests in the appendix showing that our results are best explained by behavioral reasons rather than risk-based mechanisms.

a behavioral-based explanation of our results, as we would expect *Official*-tweets to be associated with returns if our results were driven by a technological innovation channel (Cong, Li, and Wang, 2021). Lastly, we note that *Ticker*-tweets are more strongly related to excess returns than *Mention*-tweets. As the reason behind this difference in predictability power is not clear, we identify two non-exclusive candidate explanations. *Ticker*-tweets could predict returns better because they are more visible and therefore reach more users than *Mention*-tweets. A relatively larger number of users reached would correspond to larger (future) returns in both an overreaction channel (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011) and a network growth channel (Sockin and Xiong, 2023; Cong, Li, and Wang, 2021). Alternatively, investors who use Twitter to inform their investment decisions could choose to primarily consume *Ticker*-tweets as their content is geared toward cryptocurrencies' financial characteristics. The idea that *Ticker*-tweets predict returns because they are more representative of investor attention would be more consistent with an overreaction channel. Our rationale is that the attention of all types of users should matter in a network growth channel and not just investors' attention.

[Insert Table III.9 here]

5.3. Twitter textual content and investor attention

In this section, we examine why user-generated content-based attention is able to forecast returns in the cross-section of expected returns, whereas cryptocurrency-generated content does not show predictability. As both *Mention*-tweets and *Ticker*-tweets are able to predict returns despite their different visibility, it seems that investors are concerned about the tweets' content. To test this hypothesis, we now investigate the relationship between the lexicon score computed over the aggregated sample of tweets and daily excess cryptocurrency returns. We also control for textual sentiment to verify that abnormal attention doesn't predict returns because it originates from positive fundamental news. In addition, we also include an interaction term to test whether the return predictability is stronger when an attention-grabbing tweet is posted. We report the regression results in Table III.10.

As displayed in Table III.8, the regression coefficients of $Abn\ Attention_{i,t}^{All}$ are statistically significant in models (1) and (2) at the 1% significance level. $Sentiment_{i,t}^{All}$, $Lexicon_{i,t}^{All}$, and $Popular\ Tweet_{i,t}^{All}$ are also positively linked with excess returns at time t and $t+1$ at the 1% significance level. The interaction term between Twitter activity and popular tweets is positively connected with returns in (1) and (2) at the 1% and 5% significance levels, respectively.

We find $Sentiment_{i,t}^{All}$ to be positively associated with contemporaneous and next-day returns, which is in line with previous literature on Twitter sentiment (Gu and Kurov, 2020; Jiang et al., 2023). The effect of popular users posting on (future) returns is consistent with Benetton et al. (2024); Pedersen (2022), who document that influential users have a strong impact on the investment behavior of their followers and on asset returns. Compared to Benetton et al. (2024), which focus on a set of 75 celebrities, our results show that even small influencers can have a significant impact on asset prices. When we interact Twitter abnormal attention with $Popular\,Tweet_{i,t}^{All}$, the effect of Twitter activity on returns is even stronger. This finding also echoes with the results of Hillert, Jacobs, and Müller (2014), who find that media coverage can amplify behavioral biases. In the context of our study, any users with sufficient popularity on social media seem to be able to exacerbate behavioral biases through their influence on their followers. This latter observation provides more support for the overreaction interpretation of our results.

The regression coefficients of $Abn\,Attention_{i,t}^{All}$ are robust to the inclusion of the new independent variables. We interpret this result as indicative that the number of tweets posted is not subsumed by the textual content nor the sentiment of tweets. This provides support to our explanation that *Ticker*-tweets predict better returns than the other Twitter samples due to their better visibility. Given that lexicon score also forecasts returns, we conclude that both tweet volume and textual content matter in grabbing users' attention. In addition, the positive predictability of the lexicon on (future) returns casts doubts on the interpretation that abnormal attention predicts returns because of its relationship with future user base's growth (Cong, Li, and Wang, 2021; Sockin and Xiong, 2023). In fact, the lexicon loads negatively on tokens which we expect to be frequently used by newcomers, such as terms related to assistance or troubleshooting.

[Insert Table III.10 here]

5.4. Technological innovations and investor attention

In this section, we test whether Twitter-based attention predicts expected returns through a technological innovation channel. Twitter content could potentially forecast future returns through its predictability of technological improvement in the blockchain. This channel is plausible as social media is frequently used by developers to exchange ideas about potential improvements or to signal cybersecurity breaches. In the literature, Cong, Li, and Wang (2021) theoretically show that cryptocurrencies valuations are influenced by technological improvements. Empirically, technological innovation is linked negatively to delisting probability (Liu, Sheng, and Wang, 2022) and positively to ICO success (Lyandres, Palazzo, and Rabetti, 2022).

To proxy for technological improvements, we utilize the log-difference of the number of commits published for each date and cryptocurrency on GitHub + 1. We believe that this proxy for technological improvements is appropriate because commits capture all code revisions made by developers. Therefore, any new feature or improvement made on the underlying technology used by a specific cryptocurrency will be reflected in its commit history.

Regression results are reported in Table III.11. We find that contemporaneous and future cryptocurrency's technological development is strongly associated with *Official*-tweets. The contemporaneous relationship is positive and statistically significant at the 1% level. The regression coefficients of *Official*-tweets in the other models are all negative and statistically significant at the 1% or 10% level. We do not observe any links between logarithmic change in daily commit and the other Twitter samples. We interpret the results as indicative that *Ticker*-tweets and *Mention*-tweets do not predict returns through a technology innovation channel. In contrast, the significant association between the number of *Official*-tweets and technological innovation further confirms that tweets posted by cryptocurrencies can be interpreted as news.

[Insert Table III.11 here]

6. Additional Empirical Tests

6.1. Robustness

To provide robustness of our results between abnormal attention and future cryptocurrency returns, we estimate panel regression models with slight modifications compared to Table III.8. The robustness checks are reported in Table III.12. In column one, we consider raw attention in the regression setup which is defined as:

$$Raw\,Attention_{i,t} = Log(1 + Number\,Of\,Tweets_{i,t}) \tag{III.3}$$

In the other columns, we use abnormal attention as our main independent variable as in Table III.8. In models (2) and (3), we filter the sample to keep only cryptocurrencies classified as coins or as tokens, respectively, to verify that our results are not driven by characteristic differences between coins and tokens. We also check if our results are influenced by the rally of meme stocks led by R/WallStreetBets in

2021 which also affected some cryptocurrencies. For this purpose, we restrict our sample in models (4) and (5) to observations occurring before and after the first three months of 2021. In models (6) and (7), we explore if our results hold for both large and small assets by restricting our sample to assets with a market capitalization, respectively, above or below the median market capitalization. Finally, model (8) excludes pump and dump events to verify that our results are not driven by deliberate price manipulation schemes. We use data from Ardia and Bluteau (2024) to identify and filter out pump and dump events from our sample.[24] The pump and dump data lists both successful and unsuccessful events. By conservatism, we choose to remove the 3237 cryptocurrency-weeks in our sample concerned by such an event.

In all model specifications of Panel A, the regression coefficients of Twitter attention are significantly positive which is consistent with our main model specification presented in Table III.8. When regressing expected returns at $t+1$ in Panel B, we see that the regression coefficients are statistically significant for all specifications, except when small coins are removed from the sample. This observation is consistent with a limits-to-arbitrage explanation, as larger cryptocurrencies are easier to short-sell than smaller assets.

[Insert Table III.12 here]

6.2. Additional tests and alternative explanation of the results

In this section, we conduct complementary tests to better characterize the empirical links between Twitter abnormal attention and cryptocurrency market variables.

To provide additional evidence that our results are indeed driven by an overreaction narrative, we investigate how Twitter-based attention predicts contemporaneous and future change in trading volume in a panel regression setting. Results are reported in Table III.13. *Mention*-tweets and *Ticker*-tweets are both strongly linked with contemporaneous and future change in volume up to $t+3$. Both types of tweets are positively linked with contemporaneous volume at the 1% significance level and negatively linked with future trading volume. The signs of the regression coefficients of *Ticker*-tweets and *Mention*-tweets make intuitive sense, as an overreaction channel is characterized by an increased buying pressure that decreases over the subsequent days.

[Insert Table III.13 here]

[24]The data has been made publicly available by the authors at https://doi.org/10.5281/zenodo.12019080

To verify that the absence of reversal is plausibly due to limits-to-arbitrage (Schmeling, Schrimpf, and Todorov, 2023), we investigate the relationship between abnormal attention and cryptocurrency squared returns. Schmeling, Schrimpf, and Todorov (2023) argue that due to cryptocurrency futures exchanges rules about maximum losses on futures positions, even small price fluctuations can easily trigger the liquidation of an entire future position. This special set of rules impedes the ability of sophisticated investors to implement common arbitrage strategies, as highlighted by Schmeling, Schrimpf, and Todorov (2023). To align with our results, rises in abnormal attention need to predict increases in volatility to discourage sophisticated investors from arbitraging away the short term positive price pressure. This is indeed what we find in Table III.14. Abnormal attention based on *Ticker*-tweets strongly predicts future squared returns up to $t + 4$ making it risky for arbitrageurs to bet on price decreases. Furthermore, price increases seem to be driven by behavioral factors.

[Insert Table III.14 here]

Our results could also be explained by the model Andrei and Hasler (2015), who document similar relationships between attention, expected returns, and volatility as in this paper. The authors argue and show that increases in investor attention accelerate the incorporation of new information into asset prices, which leads to stronger price variations and risk premia. To shed light on whether our results indeed relate to Andrei and Hasler (2015), we investigate how abnormal attention relates to idiosyncratic squared returns. Idiosyncratic squared returns have the advantage of capturing large changes in returns that are not driven by variation in risk factors. The results are reported in Table III.15.

We document a strong relationship between daily idiosyncratic squared returns and abnormal attention measured on *Mention*-tweets and *Ticker*-tweets. The regression coefficients are statistically significant at the 5% or 1% in all model specifications. As the dependent variable accounts for risk factor variations, Twitter attention's relationship with expected returns is not purely driven by variation in priced risk. This finding provides support for the overreaction channel to be the most likely mechanism behind our results.

[Insert Table III.15 here]

7. Conclusion

Given the size of the cryptocurrency market and the importance of Twitter as a source of information, studying their interplay is essential for understanding the cross-section of cryptocurrency returns. In the literature, the evidence on the link between Twitter activity and cryptocurrency prices is mixed. Consistent with Benedetti and Kostovetsky (2021), we show that Twitter impacts the cross-section of cryptocurrency expected returns through an overreaction channel (Barber and Odean, 2008; Da, Engelberg, and Gao, 2011). However, we do not document any price reversals, which is plausibly due to limits-to-arbitrage, as already noted by (Schmeling, Schrimpf, and Todorov, 2023). We further find that the tweeting activity of popular users exacerbates behavioral biases and therefore the overreaction effect (Hillert, Jacobs, and Müller, 2014). We interpret this evidence as a warning sign about the ability of influential users to manipulate asset prices.

To better understand the relationship of Twitter attention with expected returns, we refine our abnormal attention by utilizing different samples of tweets. We show that the return predictability power of Twitter activity mainly arises from *Ticker*-tweets. In contrast, we find no association between tweets posted by official cryptocurrency channels and future returns, despite *Official*-tweets being able to forecast future innovations in the implementation code of each cryptocurrency (unlike the other Twitter samples). This empirical finding does not align with the predictions of the theoretical model of Cong, Li, and Wang (2021) that cryptocurrency valuations should be linked with technological improvements. Lastly, we document that the return predictability of *Ticker*-tweets is partly due to their salience and unique textual content, which caters to the preferences of retail investors. Overall, our results emphasize the heterogeneity of social media content, highlighting the need for researchers and practitioners to carefully consider which types of social media content best suit their needs.

Figure III.1: **Number of Cryptocurrencies, 2018-2022**

This figure shows the evolution of the number of cryptocurrencies, coins, and tokens over time at a monthly frequency that meet our criteria. The increase in the number of cryptocurrencies is mainly driven by an increase in tokens. The number of coins has even slightly decreased over the sample period.

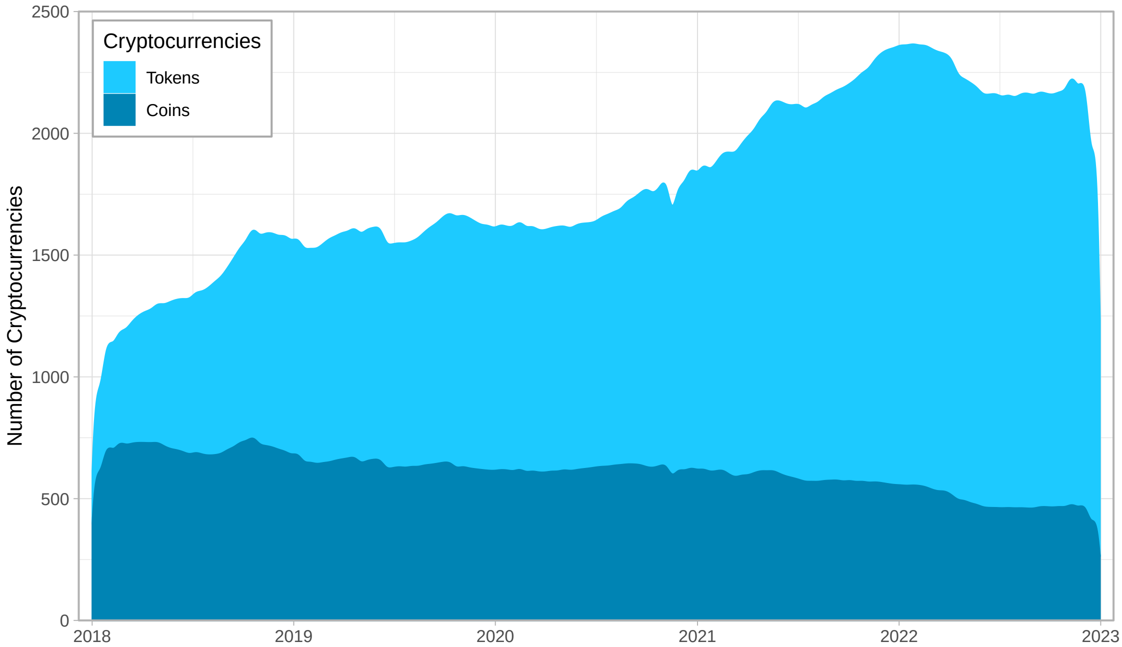


Table III.1: **Description of Twitter Samples**

This table presents the different samples of tweets used in the paper along with their number of constituents. The column Description describes which types of tweets are contained in each sample. Note that the Twitter samples have some overlap and are not mutually exclusive, since users can tweet simultaneously about several cryptocurrencies and can mix the types of tweets.

Sample	**Description**	**Number of Cryptos**	**Defunct Cryptos**
Ticker	Any tweet that contains the ticker of the cryptocurrency.	165	44
Mention	Any tweet that tags (mentions) the official account of the crypto.	165	44
Official	Tweets posted by the official account.	154	35
All	The three samples above aggregated.	165	44

Figure III.2: **Average Number of Tweets, 2018-2022**

This figure displays the evolution of the average number of tweets for cryptocurrencies over time. The average number of tweets is scaled by the number of cryptocurrencies in our sample at each point in time. For scaling purposes, the series is shown on a logarithmic scale. Panel A displays the evolution of the total number of tweets. Panel B shows the number of tweets separated across the different subcategories.

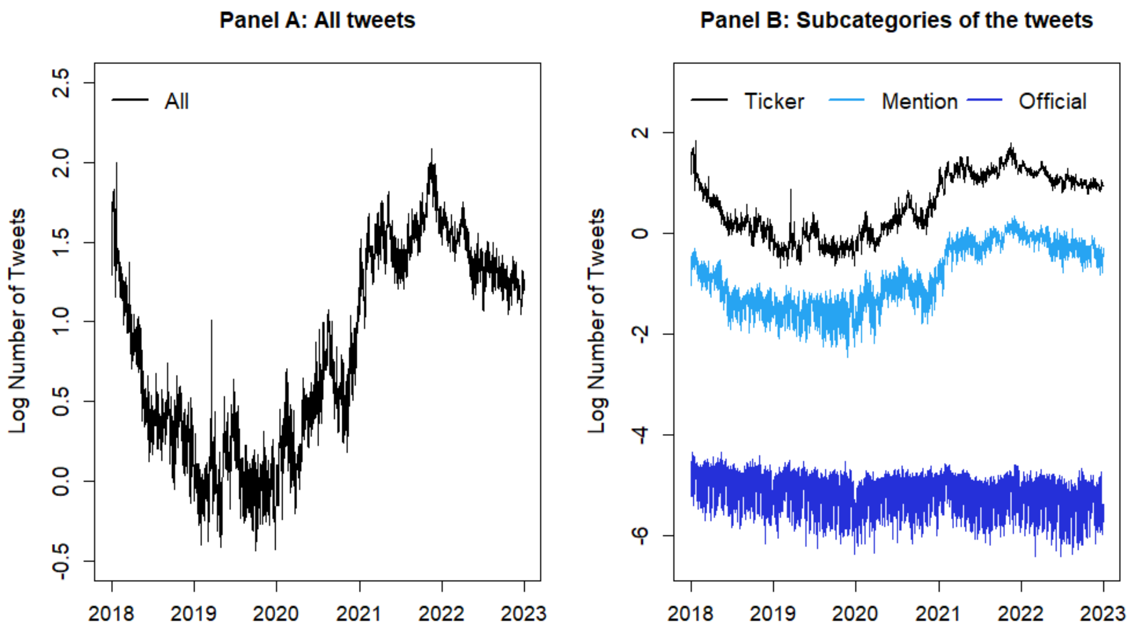


Table III.2: **Selected N-grams from Lexicon**

This table contains selected n-grams from the lexicon along with their number of occurrences across the different Twitter samples.

N-gram count:	***Mention***	***Official***	***Ticker***	**Coefficient**
answering–questions	2	16	0	-1.00
full–stack	6	12	0	-1.00
sorry–inconvenience	5	35	1	-0.94
patch	9	32	1	-0.94
thank–patience	4	30	1	-0.93
assistance	16	99	4	-0.92
support–team	7	39	2	-0.90
pull–request	8	33	2	-0.89
token–sale	53	371	29	-0.86
opensource	49	122	13	-0.81
pleased–announce	22	42	5	-0.79
source–code	13	24	6	-0.60
vulnerabilities	6	11	3	-0.57
fully–decentralized	7	11	3	-0.57
withdrawing	4	11	3	-0.57
trading	522	669	1626	0.42
sold–all	11	0	28	0.44
take–profits	7	0	26	0.58
sell–orders	4	6	24	0.60
still–cheap	14	0	59	0.62
buy–hold	15	2	66	0.63
resistance	46	29	394	0.79
daily–chart	2	0	21	0.83
overbought	2	0	24	0.85
pattern	11	6	166	0.88
buys–numbertag	2	2	36	0.89
stop–loss	4	0	83	0.91
bull–flag	1	0	46	0.96
made–returntag	10	3	692	0.97
break–above	0	0	25	1.00

Table III.3: **Variables Definition**

This table contains a description of the variables used in the paper. CMC stands for CoinMarketCap's website. "Author Homepage" indicates that the data described in the variable definition can be found on the website of the respective authors. KF is Kenneth French's website.

Variable	Definition	Source
Panel A: Cryptocurrency returns and characteristics		
$Excess\,Return_{i,t}$	Excess return on day t for cryptocurrency i.	CMC, KF
$Size_{i,t}$	Logarithmic market capitalization on day t for cryptocurrency i.	CMC
$\Delta Volume_{i,t}$	Logarithmic daily change in trading volume on day t for cryptocurrency i.	CMC
	The variables listed below until the start of Panel B are all computed over a rolling window of 60 days with a minimum of 30 days of non-missing observations.	
$Beta_{i,t}$	Regression coefficient of daily excess return of cryptocurrency i on the daily cryptocurrency market excess return.	CMC
$Momentum_{i,t}$	Compounded return of cryptocurrency i.	CMC
$Volatility_{i,t}$	Standard deviation of returns of cryptocurrency i.	CMC
$Idio\,Vola_{i,t}$	Standard deviation of the residuals when daily excess returns of cryptocurrency i are regressed on daily cryptocurrency market excess returns.	CMC
$Max\,Ret_{i,t}$	Average of the five highest daily excess return of cryptocurrency i.	CMC
$Volume\,Vola_{i,t}$	Standard deviation of Log transformed trading volume of cryptocurrency i.	CMC
$Illiquidity_{i,t}$	Ratio of illiquidity of cryptocurrency i, see Amihud (2002). $Amihud = \frac{1}{T}\Sigma_t^T \frac{\lvert r_{i,t}\rvert}{Volume_{i,t}}$	CMC
$Skewness_{i,t}$	Skewness of cryptocurrency i daily excess returns.	CMC
$Kurtosis_{i,t}$	Kurtosis of cryptocurrency i daily excess returns.	CMC
$Co-skewness_{i,t}$	The Co-Skewness of cryptocurrency i daily excess returns with daily cryptocurrency market excess return. $Coskew = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^2]}{\sigma_{r_i}\sigma^2_{r_m}}$	CMC
$Co-kurtosis_{i,t}$	The Co-Kurtosis of cryptocurrency i daily excess returns with daily cryptocurrency market excess return. $Cokurt = \frac{E[(R_i-\mu_i)(R_m-\mu_m)^3]}{\sigma_{r_i}\sigma^3_{r_m}}$	CMC
$VaR_{i,t}$	The 5% percentile of daily cryptocurrency i excess returns.	CMC
Panel B: Attention and sentiment measures		
$Abn\,Attention^{All}_{i,t}$	Difference between the (log) number of *All*-tweets of cryptocurrency i at time t and the (log) mean number of *All*-tweets of cryptocurrency i during the previous 30 days. See equation III.1. Abnormal attention for other Twitter samples is defined analogously.	Twitter
$\Delta Attention^{All}_{i,t}$	Difference between the (log) total number of *All*-tweets of cryptocurrency i at time t and the (log) number of *All*-tweets of cryptocurrency i at time $t-1$.	Twitter
$Sentiment^{All}_{i,t}$	Sentiment of the tweets published at date t on cryptocurrency i. We use the Renault (2017) lexicon to compute the sentiment of tweets. Sentiments of individual tweets are then averaged to get a sentiment score at a daily frequency. Sentiment is computed using the sample *All*-tweets.	Twitter, Author Homepage
$Popular\,Tweet^{All}_{i,t}$	Dummy variable equals one if at least one tweet from any sample published about cryptocurrency i at time t has an aggregated number of likes and replies equal to or higher than 100 across all Twitter samples. $Popular\,Tweet^{All}_{i,t}$ is computed using the sample *All*-tweets.	Twitter
$Lexicon^{All}_{i,t}$	Lexicon score of the tweets published at date t on cryptocurrency i. The lexicon is made to capture terms that are specific to *Ticker*-tweets. Lexicon score is computed using the sample *All*-tweets.	Twitter

Table continued on the next page

Table continued

Panel C: Tweet characteristics		
$RT_{i,n,t}$	(Log) sum of the number of retweets and the number of quotes + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$PM_{i,n,t}$	(Log) sum of the number of likes and the number of bookmarks + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$Views_{i,n,t}$	(Log) number of times the tweet has been seen on Twitter + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
$LengthTweet_{i,n,t}$	(Log) length of the tweet + 1 that a tweet n posted on cryptocurrency i at time t get.	Twitter
Panel D: Technology improvement measures		
$\Delta Commit_{i,t}$	(Log) daily change of the number of commits + 1 published on GitHub at time t for each repository of the organization developing the respective cryptocurrency i.	GitHub

Table III.4: **Summary Statistics**

This table contains the summary statistics of the variables defined in Table III.3. All variables are winsorized at the 1% level, except for $Excess\,Returns_t$ which is trimmed at the 99% level. The variables are expressed in decimal points. All variables are at a daily frequency.

	Mean	25%	**Median**	75%	**StdDev**
Panel A: Cryptocurrency returns and characteristics					
$Excess\,Return_{i,t}$	−0.0006	−0.038	−0.002	0.032	0.092
$Beta_{i,t}$	0.93	0.76	0.96	1.14	0.35
$Size_{i,t}$	17.43	15.75	17.12	18.84	2.53
$Momentum_{i,t}$	0.008	−0.422	−0.161	0.175	0.780
$Volatility_{i,t}$	0.080	0.052	0.069	0.096	0.041
$Idio\,Vola_{i,t}$	0.064	0.035	0.051	0.079	0.044
$Max\,Ret_{i,t}$	0.161	0.096	0.134	0.199	0.094
$\Delta Volume_{i,t}$	−0.005	−0.308	−0.023	0.255	0.731
$Volume\,Vola_{i,t}$	0.783	0.462	0.689	0.968	0.481
$Illiquidity_{i,t}$	0.002	0	0	0	0.021
$Skewness_{i,t}$	0.375	−0.216	0.280	0.872	1.025
$Kurtosis_{i,t}$	6.173	3.612	4.723	7.042	4.073
$Co\text{-}Skewness_{i,t}$	−0.345	−0.575	−0.279	−0.034	0.515
$Co\text{-}Kurtosis_{i,t}$	3.108	1.701	2.569	3.667	2.509
$VaR_{i,t}$	−0.113	−0.131	−0.099	−0.076	0.062
Panel B: Attention and sentiment measures					
$Abn\,Attention_{i,t}^{All}$	−0.210	−0.571	−0.179	0.176	0.675
$Abn\,Attention_{i,t}^{Mention}$	−0.315	−0.790	−0.210	0.111	0.834
$Abn\,Attention_{i,t}^{Official}$	−0.119	−0.383	−0.065	0.034	0.487
$Abn\,Attention_{i,t}^{Ticker}$	−0.201	−0.546	−0.162	0.166	0.665
$Sentiment_{i,t}^{All}$	0.041	0	0.044	0.084	0.062
$Popular\,Tweet_{i,t}^{All}$	0.222	0	0	0	0.416
$Lexicon_{i,t}^{All}$	0.097	−0.014	0.099	0.236	0.184
$\#\,Tweets_{i,t}^{All}$	812	16	60	233	5127
Panel C: Tweet characteristics					
$RT_{i,n,t}$	1.594	0	1.099	2.565	1.657
$PM_{i,n,t}$	1.004	0	0	1.609	1.380
$Views_{i,n,t}$	4.547	2.773	4.771	6.528	2.934
$Length\,Tweet_{i,n,t}$	2.613	2.197	2.639	3.044	0.553
Panel D: Technology improvement measures					
$\Delta Commit_{i,t}$	−0.0008	−0.29	0	0.24	0.90

Table III.5: **Contemporaneous Determinants of Refined Twitter-based Attention**

The dependent variable is the change in attention using several Twitter samples. The regression spans the period from 2018 to 2022 for a sample of 154 cryptocurrencies. Control variables are defined in Panel A of Table III.3. The regression coefficients are expressed in percentage points. Standard errors are clustered along time and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1) $\Delta Attention_t^{Mention}$	**(2)** $\Delta Attention_t^{Official}$	**(3)** $\Delta Attention_t^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$		22.95^{***}	10.34^{***}
		(26.33)	(16.75)
$\Delta Attention_{i,t}^{Official}$	46.46^{***}		1.3^{**}
	(21.75)		(2.54)
$\Delta Attention_{i,t}^{Ticker}$	20.84^{***}	1.3^{**}	
	(15.33)	(2.53)	
$Excess\,Return_{i,t}$	24.39^{***}	-2.23	53.75^{***}
	(8.44)	(-1.55)	(10.93)
$Popular\,Tweet_{i,t}^{All}$	24.77^{***}	12.05^{***}	14.38^{***}
	(12.01)	(9.23)	(13.38)
$Beta_{i,t}$	-0.46	0.22	-0.44
	(-1.35)	(0.82)	(-1.07)
$Size_{i,t}$	-1.55^{***}	-0.72^{***}	-0.96^{***}
	(-5.5)	(-4.13)	(-4.87)
$Momentum_{i,t}$	-0.32^{**}	-0.34^{***}	-0.69^{***}
	(-2.12)	(-3.54)	(-3.43)
$Volatility_{i,t}$	-18.81	-24.29	14.55
	(-0.85)	(-1.55)	(0.6)
$Idio\,Vola_{i,t}$	10.07	25.32^{*}	3.93
	(0.59)	(1.88)	(0.2)
$Max\,Ret_{i,t}$	-2.81	1.01	-9.37^{**}
	(-0.77)	(0.51)	(-2.11)
$\Delta Volume_{i,t}$	4^{***}	-0.39^{*}	10.36^{***}
	(9.9)	(-1.95)	(12.03)
$Volume\,Vola_{i,t}$	-0.9^{***}	-0.37^{***}	-0.91^{***}
	(-3.33)	(-2.71)	(-3.93)
$Illiquidity_t$	-7.15	-5.65	-1.02
	(-0.8)	(-1.52)	(-0.14)
$Skewness_{i,t}$	0.21	0	0.24^{**}
	(1.26)	(0.04)	(2.12)
$Kurtosis_{i,t}$	0	-0.01	-0.06^{**}
	(0.15)	(-0.45)	(-2.16)
$Co-Skewness_{i,t}$	-0.79^{**}	-0.21	0.14
	(-2.52)	(-0.79)	(0.45)
$Co-Kurtosis_{i,t}$	0.2^{*}	0.05	0.25^{**}
	(1.84)	(0.6)	(2.36)
$VaR_{i,t}$	-7.16^{*}	2.34	2.56
	(-1.68)	(1.04)	(0.68)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	246957	246957	246957

Table III.6: **Predictive Determinants of Refined Twitter-based Attention**

The dependent variables are changes in attention using several Twitter samples. The regression spans the period from 2018 to 2022 for a sample of 154 cryptocurrencies. Control variables are defined in Panel A of Table III.3. The regression coefficients are expressed in percentage points. Standard errors are clustered along time and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	**(2)**	**(3)**
	$\Delta Attention_{t+1}^{Mention}$	$\Delta Attention_{t+1}^{Official}$	$\Delta Attention_{t+1}^{Ticker}$
$\Delta Attention_{i,t}^{Mention}$	-36.6^{***}	-1.58^{***}	-0.08
	(-82.21)	(-9.04)	(-0.47)
$\Delta Attention_{i,t}^{Official}$	10.33^{***}	-41.78^{***}	1.31^{***}
	(15.02)	(-78.76)	(5.4)
$\Delta Attention_{i,t}^{Ticker}$	2.44^{***}	0.66^{***}	-32.52^{***}
	(6.22)	(2.83)	(-66.33)
$Excess\,Return_{i,t}$	19.56^{***}	2.81^{**}	25.93^{***}
	(6.91)	(2.19)	(8.06)
$Popular\,Tweet_{i,t}^{All}$	-7.38^{***}	-9.97^{***}	-2.99^{***}
	(-7.52)	(-11.26)	(-6.58)
$Beta_{i,t}$	-0.35	0.08	-0.66^{*}
	(-0.97)	(0.24)	(-1.76)
$Size_{i,t}$	-0.05	0.51^{***}	-0.36^{**}
	(-0.31)	(2.93)	(-2.27)
$Momentum_{i,t}$	-0.73^{***}	-0.2^{*}	-1.31^{***}
	(-4.13)	(-1.85)	(-5.95)
$Volatility_{i,t}$	44.97^{**}	18.2	33.2
	(2.06)	(0.99)	(1.6)
$Idio\,Vola_{i,t}$	-50.16^{***}	-15.43	-35.68^{**}
	(-2.99)	(-0.93)	(-2.09)
$Max\,Ret_{i,t}$	-2.96	0.73	-2.68
	(-0.64)	(0.34)	(-1.03)
$\Delta Volume_{i,t}$	1.9^{***}	0.48^{***}	1.57^{***}
	(6.95)	(2.7)	(7.24)
$Volume\,Vola_{i,t}$	0.09	0.18	-0.22
	(0.58)	(1.24)	(-0.96)
$Illiquidity_{t}$	7.73	7.37^{**}	1.84
	(1.51)	(2.01)	(0.32)
$Skewness_{i,t}$	0.11	-0.1	0.08
	(0.63)	(-1.09)	(0.7)
$Kurtosis_{i,t}$	-0.07^{***}	-0.02	-0.09^{***}
	(-2.62)	(-0.99)	(-4.12)
$Co-Skewness_{i,t}$	0.08	0.19	0.61^{*}
	(0.27)	(0.74)	(1.82)
$Co-Kurtosis_{i,t}$	0.1	-0.06	0.22^{**}
	(0.98)	(-0.68)	(2.01)
$VaR_{i,t}$	-1.06	3.07	-0.36
	(-0.24)	(1.36)	(-0.1)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	246844	246844	246844

Table III.7: **Visibility of Tweets**

The dependent variables are the number of retweets, the number of likes and replies, and the number of views for each tweet. The dependent variables are log-transformed. The sample covers the period from December 15, 2022, to December 31, 2022, for a sample of 151 cryptocurrencies. The regression coefficients are expressed in percentage points. Standard errors are clustered by authors. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	**(2)**	**(3)**
	$RT_{i,n,t}$	$PM_{i,n,t}$	$Views_{i,n,t}$
$Mention\,Tweet_{i,n,t}$	0.32	0.04	-34.12^{***}
	(0.08)	(0.01)	(-6.09)
$Official\,Tweet_{i,n,t}$	105.59^{***}	7.45	50.21^{***}
	(10.84)	(1.19)	(6.64)
$Ticker\,Tweet_{i,n,t}$	31.75^{***}	3.5	9.89^{**}
	(8.62)	(1.05)	(2.05)
$RT_{i,n,t}$		41.24^{***}	-65.53^{***}
		(63.18)	(-64.82)
$PM_{i,n,t}$	77.03^{***}		158.85^{***}
	(59.06)		(104.52)
$Views_{i,n,t}$	-29.47^{***}	38.25^{***}	
	(-50.86)	(118.99)	
$Length\,Tweet_{i,n,t}$	21.07^{***}	-4.64^{***}	25.87^{***}
	(15.64)	(-2.7)	(7.66)
$Sentiment_{i,n,t}$	20.41^{***}	16.04^{***}	-79.69^{***}
	(5.5)	(4.71)	(-11.84)
$Lexicon_{i,n,t}$	-13.72^{***}	-10.62^{***}	28.64^{***}
	(-8.08)	(-5.11)	(7.95)
Time FE	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes
Number of observations	63221	63221	63221

Table III.8: **Investor Attention and Cryptocurrency Returns, 2018-2022**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 165 (103) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	**(2)**	**(3)**	**(4)**
	Daily Excess Return$_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention^{All}_{i,t}$	1.06***	0.16***	0.01	0.02
	(18.57)	(4.29)	(0.37)	(0.89)
$Excess\ Return_{i,t}$		−18.94***	−2.65***	−0.19
		(−16.85)	(−5.26)	(−0.45)
$Excess\ Return_{i,t-1}$	−18.55***	−6.11***	−0.62	−0.46
	(−16.46)	(−9.14)	(−1.5)	(−0.89)
$Excess\ Return_{i,t-2}$	−5.28***	−1.93***	−0.79	−0.25
	(−7.66)	(−4.04)	(−1.39)	(−0.57)
$Excess\ Return_{i,t-3}$	−1.65***	−1.04**	−0.18	0.71
	(−3.77)	(−2.06)	(−0.42)	(1.45)
$Beta_{i,t}$	0.04	0.04	0.15	0.08
	(0.34)	(0.34)	(1.18)	(0.63)
$Size_{i,t}$	−0.34***	−0.3***	−0.23***	−0.21***
	(−10.51)	(−10.22)	(−8.99)	(−8.29)
$Momentum_{i,t}$	0	−0.06	−0.08	−0.13**
	(0.01)	(−0.84)	(−1.4)	(−2.11)
$Volatility_{i,t}$	0.42	0.14	−3.21	−0.87
	(0.04)	(0.01)	(−0.55)	(−0.13)
$Idio\,Vola_{i,t}$	−7.73	−7.35	−2.96	−4.65
	(−1.6)	(−1.58)	(−0.64)	(−1.03)
$Max\ Ret_{i,t}$	2.92	2.63	2.1	2.41
	(0.63)	(0.71)	(1.36)	(1.16)
$\Delta Volume_{i,t}$	1.81***	0.26***	−0.04	−0.01
	(12.43)	(6.7)	(−1.18)	(−0.32)
$Volume\,Vola_{i,t}$	−0.06	−0.08	−0.08	−0.05
	(−0.91)	(−1.13)	(−1.4)	(−0.78)
$Illiquidity_{i,t}$	−0.45	1.37	1.89	1.3
	(−0.31)	(0.9)	(1.53)	(1.15)
$Skewness_{i,t}$	0.02	0.01	0.02	0.02
	(0.22)	(0.09)	(0.42)	(0.47)
$Kurtosis_{i,t}$	−0.01	−0.01	−0.01	−0.01
	(−0.48)	(−0.63)	(−1.19)	(−1.36)
$Co\text{-}Skewness_{i,t}$	−0.03	−0.01	0.01	−0.04
	(−0.35)	(−0.14)	(0.14)	(−0.58)
$Co\text{-}Kurtosis_{i,t}$	0.02	0.02	0.01	0.02
	(0.63)	(0.63)	(0.54)	(0.74)
$VaR_{i,t}$	−2.24	−1.49	0.67	1.32
	(−1.25)	(−1.02)	(0.45)	(0.78)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	255684	254850	254431	254206
Panel B: Filtered ticker				
$Abn\ Attention^{All}_{i,t}$	1.08***	0.14***	0.03	0.04
	(15.58)	(2.96)	(1.13)	(1.21)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	160712	160246	159991	159846

Table III.9: **Refined Investor Attention and Cryptocurrency Returns, 2018-2022**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10% levels, respectively.

	(1)	**(2)**	**(3)**	**(4)**
	Daily Excess Return$_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\,Attention^{Mention}_{i,t}$	0.31***	0.04*	0.01	−0.02
	(10.85)	(1.92)	(0.32)	(−1.3)
$Abn\,Attention^{Official}_{i,t}$	−0.02	0.02	−0.02	0.04
	(−0.42)	(0.67)	(−0.61)	(1.22)
$Abn\,Attention^{Ticker}_{i,t}$	1.11***	0.17***	0	0.03
	(16.9)	(4.35)	(0.16)	(0.92)
$Excess\,Return_{i,t}$		−18.86***	−2.44***	−0.11
		(−16.09)	(−4.77)	(−0.27)
$Excess\,Return_{i,t-1}$	−18.82***	−5.94***	−0.55	−0.58
	(−16.22)	(−8.55)	(−1.3)	(−1.07)
$Excess\,Return_{i,t-2}$	−5.3***	−1.93***	−0.93	−0.24
	(−7.44)	(−3.95)	(−1.56)	(−0.52)
$Excess\,Return_{i,t-3}$	−1.72***	−1.13**	−0.12	0.55
	(−3.85)	(−2.15)	(−0.26)	(1.08)
$Beta_{i,t}$	0.07	0.05	0.16	0.09
	(0.55)	(0.42)	(1.21)	(0.67)
$Size_{i,t}$	−0.35***	−0.31***	−0.24***	−0.23***
	(−10.56)	(−10.04)	(−9.08)	(−8.29)
$Momentum_{i,t}$	0	−0.05	−0.08	−0.12**
	(−0.02)	(−0.76)	(−1.35)	(−2.05)
$Volatility_{i,t}$	−2.21	−0.57	−2.98	−0.91
	(−0.18)	(−0.06)	(−0.49)	(−0.14)
$Idio\,Vola_{i,t}$	−7.09	−7.17	−2.72	−4.2
	(−1.44)	(−1.51)	(−0.58)	(−0.91)
$Max\,Ret_{i,t}$	3.74	2.99	2.15	2.5
	(0.78)	(0.78)	(1.36)	(1.17)
$\Delta Volume_{i,t}$	1.8***	0.26***	−0.03	−0.01
	(11.85)	(6.69)	(−1.06)	(−0.4)
$Volume\,Vola_{i,t}$	−0.06	−0.07	−0.08	−0.05
	(−0.88)	(−0.94)	(−1.34)	(−0.83)
$Illiquidity_{i,t}$	−1.23	0.57	1.2	0.79
	(−0.68)	(0.32)	(0.76)	(0.58)
$Skewness_{i,t}$	0	−0.01	0.01	0.02
	(0)	(−0.09)	(0.28)	(0.4)
$Kurtosis_{i,t}$	0	−0.01	−0.01	−0.01
	(−0.3)	(−0.53)	(−1.24)	(−1.39)
$Co\text{-}Skewness_{i,t}$	0	0.01	0.02	−0.01
	(0)	(0.16)	(0.3)	(−0.18)
$Co\text{-}Kurtosis_{i,t}$	0.01	0.01	0.01	0.01
	(0.34)	(0.4)	(0.35)	(0.68)
$VaR_{i,t}$	−2.57	−1.53	0.94	1.71
	(−1.37)	(−1.02)	(0.62)	(1)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243061
Panel B: Filtered ticker				
$Abn\,Attention^{Mention}_{i,t}$	0.32***	0.02	0	−0.01
	(9.84)	(0.76)	(0.03)	(−0.24)
$Abn\,Attention^{Official}_{i,t}$	−0.01	0.04	−0.01	0
	(−0.22)	(0.93)	(−0.41)	(0.13)
$Abn\,Attention^{Ticker}_{i,t}$	1.12***	0.17***	0.03	0.04
	(14.27)	(3.56)	(1.06)	(1.06)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Table III.10: **Investor Attention, Sentiment and Cryptocurrency Returns, 2018-2022**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 165 (103) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	**(2)**	**(3)**	**(4)**
	$Daily\,Excess\,Return_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\,Attention_{i,t}^{All}$	0.82***	0.11***	0.01	0.04
	(17.02)	(2.92)	(0.52)	(1.42)
$Sentiment_{i,t}^{All}$	3.84***	1.25***	0.28	−0.56*
	(9.69)	(4.07)	(0.91)	(−1.73)
$Abn\,Attention_{i,t}^{All}\,X$	1.39***	0.19**	−0.06	−0.05
$Popular\,Tweet_{i,t}^{All}$	(8.92)	(2.19)	(−0.86)	(−0.83)
$Popular\,Tweet_{i,t}^{All}$	0.32***	0.13***	0.02	−0.02
	(5.24)	(2.68)	(0.44)	(−0.57)
$Lexicon_{i,t}^{All}$	2.15***	0.41***	0.06	−0.13
	(13.92)	(3.27)	(0.56)	(−1.06)
$Excess\,Return_{i,t}$		−19.02***	−2.65***	−0.16
		(−16.96)	(−5.24)	(−0.39)
$Excess\,Return_{i,t-1}$	−18.89***	−6.18***	−0.62	−0.44
	(−17.08)	(−9.26)	(−1.48)	(−0.84)
$Excess\,Return_{i,t-2}$	−5.53***	−1.98***	−0.79	−0.24
	(−8.15)	(−4.14)	(−1.38)	(−0.54)
$Excess\,Return_{i,t-3}$	−1.8***	−1.07**	−0.18	0.72
	(−4.17)	(−2.11)	(−0.41)	(1.46)
$Beta_{i,t}$	0.02	0.04	0.15	0.08
	(0.14)	(0.32)	(1.18)	(0.63)
$Size_{i,t}$	−0.36***	−0.31***	−0.23***	−0.21***
	(−10.55)	(−10.43)	(−9.01)	(−8.04)
$Momentum_{i,t}$	−0.02	−0.06	−0.08	−0.12**
	(−0.31)	(−0.93)	(−1.41)	(−2.07)
$Volatility_{i,t}$	1.51	0.24	−3.24	−0.92
	(0.13)	(0.02)	(−0.55)	(−0.14)
$Idio\,Vola_{i,t}$	−9.53**	−7.66*	−2.95	−4.52
	(−1.96)	(−1.65)	(−0.64)	(−1)
$Max\,Ret_{i,t}$	3.14	2.68	2.1	2.39
	(0.68)	(0.72)	(1.36)	(1.15)
$\Delta Volume_{i,t}$	1.78***	0.25***	−0.04	−0.01
	(12.44)	(6.62)	(−1.19)	(−0.29)
$Volume\,Vola_{i,t}$	−0.07	−0.08	−0.08	−0.05
	(−1.04)	(−1.17)	(−1.41)	(−0.78)
$Illiquidity_{i,t}$	−0.38	1.32	1.87	1.33
	(−0.27)	(0.87)	(1.52)	(1.17)
$Skewness_{i,t}$	0.02	0.01	0.02	0.02
	(0.2)	(0.09)	(0.42)	(0.47)
$Kurtosis_{i,t}$	−0.01	−0.01	−0.01	−0.01
	(−0.55)	(−0.64)	(−1.18)	(−1.36)
$Co\text{-}Skewness_{i,t}$	−0.02	−0.01	0.01	−0.05
	(−0.24)	(−0.13)	(0.15)	(−0.59)
$Co\text{-}Kurtosis_{i,t}$	0.02	0.02	0.01	0.02
	(0.65)	(0.67)	(0.56)	(0.72)
$VaR_{i,t}$	−2.42	−1.56	0.66	1.35
	(−1.35)	(−1.07)	(0.44)	(0.8)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	255684	254850	254431	254206
Panel B: Filtered ticker				
$Abn\,Attention_{i,t}^{All}$	0.82***	0.08*	0.04	0.06
	(14.21)	(1.76)	(1.21)	(1.63)
$Sentiment_{i,t}^{All}$	3.73***	1.22***	0.08	−0.72*
	(7.95)	(3.17)	(0.22)	(−1.82)
$Abn\,Attention_{i,t}^{All}\,X$	1.54***	0.24**	−0.07	−0.04
$Popular\,Tweet_{i,t}^{All}$	(8.23)	(2.2)	(−0.82)	(−0.64)
$Popular\,Tweet_{i,t}^{All}$	0.32***	0.14***	0.04	−0.04
	(5.17)	(2.83)	(0.94)	(−0.61)
$Lexicon_{i,t}^{All}$	2.19***	0.43***	0.11	−0.09
	(11.26)	(2.59)	(0.87)	(−0.63)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	160712	160246	159991	159846

Table III.11: **Refined Investor Attention and GitHub Commits**

The dependent variables are daily log-changes in the number of commits ($\Delta Commit_t$). The regression spans the period from 2018 to 2022 for a (filtered) sample of 136 (86) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Standard errors are clustered along weeks and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	**(2)**	**(3)**	**(4)**
	$\Delta\, Commit_i$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention^{Mention}_{i,t}$	−0.33	−0.41	−0.08	−0.12
	(−1.07)	(−1.04)	(−0.24)	(−0.71)
$Abn\ Attention^{Official}_{i,t}$	5.61***	−6.62***	−5.71***	−0.81*
	(5.37)	(−6.5)	(−5.28)	(−1.7)
$Abn\ Attention^{Ticker}_{i,t}$	0.06	−0.04	−0.49	0.13
	(0.16)	(−0.12)	(−1.43)	(0.59)
$Excess\ Return_{i,t}$	1.84	−0.6	0.93	1.95
	(0.8)	(−0.25)	(0.43)	(0.68)
$Excess\ Return_{i,t-1}$	−1.71	0.05	3.45	−3.33
	(−0.75)	(0.02)	(1.27)	(−1.59)
$Excess\ Return_{i,t-2}$	−1.26	3.69	−3.8*	−0.34
	(−0.58)	(1.36)	(−1.75)	(−0.15)
$Excess\ Return_{i,t-3}$	2.91	−3.3	−0.73	0.98
	(1.06)	(−1.55)	(−0.3)	(0.48)
$Beta_{i,t}$	0.17	0.13	−0.01	−0.19
	(0.27)	(0.2)	(−0.02)	(−0.29)
$Size_{i,t}$	0.07	−0.19	−0.14	−0.07
	(0.23)	(−0.59)	(−0.45)	(−0.23)
$Momentum_{i,t}$	−0.07	0.19	0.26	0.06
	(−0.2)	(0.71)	(0.9)	(0.23)
$Volatility_{i,t}$	1.83	−4.38	−1.2	11.64
	(0.06)	(−0.14)	(−0.04)	(0.35)
$Idio\,Vola_{i,t}$	8.54	3.06	−10.48	−17.62
	(0.29)	(0.11)	(−0.36)	(−0.59)
$Max\ Ret_{i,t}$	−3	−2.42	1.24	0.31
	(−1.26)	(−0.96)	(0.6)	(0.1)
$\Delta Volume_{i,t}$	0.34	0.31	−0.54*	0.16
	(0.98)	(0.95)	(−1.77)	(0.64)
$Volume\,Vola_{i,t}$	−0.21	−0.09	−0.15	−0.15
	(−1.09)	(−0.44)	(−0.73)	(−0.75)
$Illiquidity_{i,t}$	4.73	2.19	−2.11	−4.26
	(0.72)	(0.36)	(−0.32)	(−0.66)
$Skewness_{i,t}$	−0.04	−0.02	−0.09	−0.01
	(−0.27)	(−0.19)	(−0.65)	(−0.07)
$Kurtosis_{i,t}$	−0.02	0.01	0.02	0.02
	(−0.33)	(0.2)	(0.42)	(0.61)
$Co\text{-}Skewness_{i,t}$	0.34	0.56**	0.42	0.21
	(0.85)	(1.97)	(1.25)	(0.81)
$Co\text{-}Kurtosis_{i,t}$	−0.02	−0.04	−0.04	−0.04
	(−0.11)	(−0.23)	(−0.22)	(−0.23)
$VaR_{i,t}$	2.28	−2.68	−5.37	−4.34
	(0.55)	(−0.56)	(−1.25)	(−1)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	218873	218769	218668	218568
Panel B: Filtered ticker				
$Abn\ Attention^{Mention}_{i,t}$	−0.34	−0.93*	−0.21	−0.12
	(−1.02)	(−1.69)	(−0.47)	(−0.48)
$Abn\ Attention^{Official}_{i,t}$	6.56***	−7.47***	−6.78***	−1.14*
	(5.06)	(−5.9)	(−4.62)	(−1.79)
$Abn\ Attention^{Ticker}_{i,t}$	−0.13	0.14	−0.6*	0.19
	(−0.34)	(0.38)	(−1.67)	(0.63)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	139313	139244	139178	139113

Table III.12: **Robustness Checks**

The dependent variables are daily excess returns ($Excess\,Return_t$). The regression spans the time period 2018 to 2022 for a sample of 165 cryptocurrencies. Control variables are defined in Panel A of Table III.3. The first model (1) uses raw attention instead of abnormal change as the main independent variable. Models (2) and (3) restrict the sample to coins only or tokens only, respectively. Models (4) and (5) restrict the sample respectively to observations occurring before or after the GameStop short-squeeze. For models (6) and (7), we restrict the sample to assets with a market cap below or above the median, respectively. Finally, model (8) excludes pump and dump events from the sample. Regression coefficients are reported in percentage points. Standard errors are clustered along weeks and cryptocurrencies. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
	Raw Attention	*Coins Only*	*Tokens Only*	*Before GameStop*	*After GameStop*	*Large Only*	*Small Only*	*No P&D Events*
Panel A: Contemporaneous relationship								
	$Daily\,Excess\,Return_{i,t}$							
$Abn\,Attention^{All}_{i,t}$	0.58***	1.02***	1.09***	1.11***	0.88***	1.49***	0.62***	1.05***
	(10.29)	(14.18)	(12.25)	(17.19)	(11.31)	(19.41)	(11.64)	(18.38)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	255684	147494	108190	165359	79856	129704	125160	252490
Panel B: Predictive relationship								
	$Daily\,Excess\,Return_{i,t+1}$							
$Abn\,Attention^{All}_{i,t}$	0.11***	0.15***	0.15***	0.17***	0.13**	0.03	0.13***	0.17***
	(3.68)	(3.19)	(2.58)	(3.84)	(2.41)	(0.68)	(2.89)	(4.42)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	254850	147023	107827	164984	79440	129570	124461	251659

Table III.13: **Refined Investor Attention and Changes in Volume**

The dependent variables are daily logarithmic changes in trading volume. The regression spans the period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Standard errors are clustered along weeks and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	**(2)**	**(3)**	**(4)**
		Daily Δ *Volume*$_i$		
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention^{Mention}_{i,t}$	1.26***	−1.35***	−0.73***	−0.23
	(5.99)	(−7.23)	(−3.8)	(−1.52)
$Abn\ Attention^{Official}_{i,t}$	0.31	0.84***	−0.54*	0.47*
	(1.03)	(2.89)	(−1.75)	(1.73)
$Abn\ Attention^{Ticker}_{i,t}$	7.31***	−3.42***	−1.8***	−0.65***
	(13.03)	(−9.97)	(−6.96)	(−3.34)
$Excess\ Return_{i,t}$	169.03***	44.07***	−46.82***	−13.45***
	(10.53)	(9.95)	(−9.64)	(−4.44)
$Excess\ Return_{i,t-1}$	3.93	−57.62***	−26.09***	−5.53*
	(0.86)	(−9.25)	(−8.9)	(−1.67)
$Excess\ Return_{i,t-2}$	−60.01***	−46.92***	−12.48***	−7.42***
	(−9.76)	(−12.07)	(−3.32)	(−3)
$Excess\ Return_{i,t-3}$	−29.21***	−14.8***	−6.87***	−1.37
	(−9.31)	(−4.8)	(−2.67)	(−0.51)
$Beta_{i,t}$	−1.34***	−1.79***	−0.63	−0.87**
	(−2.94)	(−3.01)	(−1.34)	(−2.28)
$Size_{i,t}$	−0.2	−0.79***	−0.71***	−0.4***
	(−1.54)	(−5.28)	(−5.06)	(−3.3)
$Momentum_{i,t}$	−1.28***	−1.14***	−0.56***	−0.56***
	(−5.28)	(−4.26)	(−2.77)	(−3.11)
$Volatility_{i,t}$	44.57	75.76**	26.27	21.26
	(1.31)	(2.21)	(0.91)	(0.77)
$Idio\,Vola_{i,t}$	−82.84***	−113.26***	−47.35**	−39.21**
	(−4.09)	(−4.75)	(−2.39)	(−2.45)
$Max\ Ret_{i,t}$	12.13	7.97	1.78	1.53
	(1.04)	(0.83)	(0.28)	(0.21)
$\Delta Volume_{i,t}$		−34.06***	−8.12***	−1.92***
		(−67.28)	(−19.27)	(−4.22)
$Volume\,Vola_{i,t}$	0.89**	0.98**	0.49	0.49
	(2.36)	(2.27)	(1.49)	(1.55)
$Illiquidity_{i,t}$	35.42***	36.73***	20.1**	−0.17
	(3.47)	(2.82)	(2)	(−0.02)
$Skewness_{i,t}$	−0.09	0.1	0.14	0.2
	(−0.35)	(0.37)	(0.74)	(1.03)
$Kurtosis_{i,t}$	−0.07*	−0.14***	−0.08**	−0.07*
	(−1.69)	(−3.08)	(−2.25)	(−1.74)
$Co\text{-}Skewness_{i,t}$	0.29	0.48	0.4	−0.13
	(0.82)	(1.05)	(1.12)	(−0.4)
$Co\text{-}Kurtosis_{i,t}$	0.13	0.33**	0.22	0.17
	(1.08)	(2)	(1.53)	(1.53)
$VaR_{i,t}$	7.08	2.76	−3.95	−6.53
	(1.16)	(0.3)	(−0.66)	(−1.5)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243970	243571	243362
Panel B: Filtered ticker				
$Abn\ Attention^{Mention}_{i,t}$	1.16***	−1.62***	−0.55**	−0.34*
	(4.3)	(−7)	(−2.26)	(−1.72)
$Abn\ Attention^{Official}_{i,t}$	0.39	1***	−0.75**	0.81**
	(1.02)	(2.59)	(−2.08)	(2.32)
$Abn\ Attention^{Ticker}_{i,t}$	7.48***	−3.91***	−1.7***	−0.57**
	(10.75)	(−9.31)	(−5.32)	(−2.45)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156868	156616	156475

Table III.14: **Refined Investor Attention and Squared Returns**

The dependent variables are daily squared excess returns. The regression spans the time period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Standard errors are clustered along week and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	**(2)**	**(3)**	**(4)**
		Daily Squared R_i^e		
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention_{i,t}^{Mention}$	0.06***	0.03***	0.02**	0.01
	(6.28)	(3.49)	(2.26)	(1.5)
$Abn\ Attention_{i,t}^{Official}$	−0.01	−0.02*	−0.01	0.02
	(−1.16)	(−1.73)	(−0.77)	(1.24)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.09***	0.08***	0.06***
	(8.35)	(5.68)	(4.8)	(3.44)
$Excess\ Return_{i,t}$	12.98***	0.28	0.58***	0.34
	(17.16)	(0.87)	(2.96)	(1.54)
$Excess\ Return_{i,t-1}$	2.84***	0.7***	0.51**	0.34
	(9.01)	(3.13)	(2.09)	(1.21)
$Excess\ Return_{i,t-2}$	1.6***	0.59**	0.36	0.21
	(7.67)	(2.5)	(1.37)	(1.44)
$Excess\ Return_{i,t-3}$	0.74***	0.18	0.04	0.22
	(4.05)	(0.9)	(0.29)	(1.23)
$Beta_{i,t}$	−0.21***	−0.23***	−0.23***	−0.21***
	(−2.91)	(−3)	(−2.8)	(−2.83)
$Size_{i,t}$	0.01	−0.05***	−0.05***	−0.06***
	(0.67)	(−3.87)	(−3.43)	(−4.04)
$Momentum_{i,t}$	0.07**	0.07**	0.02	0.01
	(2.33)	(2.29)	(0.65)	(0.29)
$Volatility_{i,t}$	16.71***	16.28***	11.03**	9.46***
	(4.62)	(4.86)	(2.5)	(2.66)
$Idio\,Vola_{i,t}$	9.66***	6.39***	5.26**	6***
	(4.18)	(2.94)	(2.37)	(2.96)
$Max\ Ret_{i,t}$	−6.31***	−5.08***	−2.68*	−2.56**
	(−4.65)	(−4.15)	(−1.73)	(−2.34)
$\Delta Volume_{i,t}$	0.24***	0.08***	0.02	−0.01
	(10.35)	(5.77)	(1.44)	(−0.83)
$Volume\,Vola_{i,t}$	0.02	0.04	0.04	0.05
	(0.6)	(0.9)	(0.93)	(1.14)
$Illiquidity_{i,t}$	1	0.04	0.75	0.63
	(0.63)	(0.03)	(0.49)	(0.48)
$Skewness_{i,t}$	0.04	0.05	0	0
	(1.26)	(1.59)	(0.11)	(−0.17)
$Kurtosis_{i,t}$	−0.02***	−0.03***	−0.02***	−0.02***
	(−4.08)	(−5.66)	(−3.63)	(−4.06)
$Co\text{-}Skewness_{i,t}$	0.09*	0.11**	0.09*	0.12**
	(1.85)	(2.35)	(1.92)	(2.52)
$Co\text{-}Kurtosis_{i,t}$	0.06***	0.07***	0.06***	0.06***
	(3.86)	(4.33)	(3.59)	(3.81)
$VaR_{i,t}$	−3.28***	−3.22***	−3.89***	−3.95***
	(−2.59)	(−2.72)	(−2.75)	(−3.06)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243061
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	0.05***	0.02	0.02	0.01
	(4.23)	(1.51)	(1.41)	(0.88)
$Abn\ Attention_{i,t}^{Official}$	−0.01	−0.01	−0.01	0.02
	(−1.01)	(−0.58)	(−0.94)	(1.27)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.1***	0.09***	0.06***
	(7.73)	(5.14)	(4.94)	(3.32)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Table III.15: **Refined Investor Attention and Idiosyncratic Squared Returns**

The dependent variables are daily idiosyncratic squared excess returns. The regression spans the time period 2018 to 2022 for a (filtered) sample of 154 (99) cryptocurrencies. Control variables are defined in Panel A of Table III.3. Standard errors are clustered along week and cryptocurrencies. Regression coefficients are reported in percentage points. The t-statistics are given in parentheses below the coefficients. ***, **, * indicate statistical significance at the 1%, 5% and the 10%, respectively.

	(1)	(2)	(3)	(4)
	Daily *Idiosyncratic* $(R_i^e)^2$			
	t	$t+1$	$t+2$	$t+3$
Panel A: All cryptocurrencies				
$Abn\ Attention_{i,t}^{Mention}$	0.07***	0.03***	0.03***	0.02**
	(6.9)	(4.1)	(2.84)	(2.07)
$Abn\ Attention_{i,t}^{Official}$	−0.02	−0.02**	−0.01	0.02
	(−1.51)	(−1.97)	(−0.72)	(1.47)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.09***	0.08***	0.06***
	(8.32)	(5.57)	(4.85)	(3.55)
$Excess\ Return_{i,t}$	12.45***	0.41	0.62***	0.28
	(17.86)	(1.32)	(3.06)	(1.25)
$Excess\ Return_{i,t-1}$	2.88***	0.81***	0.49**	0.4
	(9.29)	(3.71)	(2.04)	(1.45)
$Excess\ Return_{i,t-2}$	1.65***	0.6***	0.42	0.27*
	(7.89)	(2.58)	(1.6)	(1.72)
$Excess\ Return_{i,t-3}$	0.73***	0.22	0.1	0.27
	(3.91)	(1.08)	(0.63)	(1.57)
$Beta_{i,t}$	−0.27***	−0.28***	−0.29***	−0.27***
	(−3.69)	(−3.74)	(−3.47)	(−3.55)
$Size_{i,t}$	−0.01	−0.07***	−0.07***	−0.08***
	(−0.7)	(−5.24)	(−4.62)	(−5.32)
$Momentum_{i,t}$	0.07**	0.06**	0.01	0
	(2.2)	(2.06)	(0.27)	(−0.1)
$Volatility_{i,t}$	14.65***	14.79***	10.41**	9.29***
	(4.08)	(4.54)	(2.46)	(2.7)
$Idio\,Vola_{i,t}$	11.16***	7.59***	6.42***	7.12***
	(4.9)	(3.46)	(2.88)	(3.39)
$Max\ Ret_{i,t}$	−6.06***	−4.9***	−2.75*	−2.78***
	(−4.56)	(−4.09)	(−1.82)	(−2.6)
$\Delta Volume_{i,t}$	0.23***	0.07***	0.02	−0.01
	(9.97)	(5.43)	(1.33)	(−0.61)
$Volume\,Vola_{i,t}$	0.02	0.04	0.05	0.05
	(0.69)	(1.04)	(1.09)	(1.27)
$Illiquidity_{i,t}$	1.03	−0.12	0.54	0.42
	(0.63)	(−0.07)	(0.36)	(0.32)
$Skewness_{i,t}$	0.04	0.04	0	0
	(1.11)	(1.44)	(0.03)	(−0.11)
$Kurtosis_{i,t}$	−0.02***	−0.03***	−0.02***	−0.02***
	(−4.06)	(−5.71)	(−3.71)	(−4.36)
$Co\text{-}Skewness_{i,t}$	0.11**	0.12***	0.11**	0.14***
	(2.39)	(2.75)	(2.28)	(2.82)
$Co\text{-}Kurtosis_{i,t}$	0.07***	0.08***	0.06***	0.06***
	(4.03)	(4.54)	(3.78)	(4.08)
$VaR_{i,t}$	−3.4***	−3.27***	−3.8***	−3.85***
	(−2.81)	(−2.88)	(−2.71)	(−2.94)
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	244395	243633	243260	243060
Panel B: Filtered ticker				
$Abn\ Attention_{i,t}^{Mention}$	0.06***	0.02*	0.02*	0.01
	(4.89)	(1.92)	(1.65)	(1.27)
$Abn\ Attention_{i,t}^{Official}$	−0.02*	−0.01	−0.01	0.02
	(−1.72)	(−1.1)	(−0.78)	(1.21)
$Abn\ Attention_{i,t}^{Ticker}$	0.16***	0.1***	0.08***	0.07***
	(7.85)	(4.99)	(4.9)	(3.66)
Controls	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Cryptocurrency FE	Yes	Yes	Yes	Yes
Observations	157131	156682	156442	156304

Bibliography

Agarwal, V., S. Ruenzi, and F. Weigert. 2024. Unobserved performance of hedge funds. *The Journal of Finance* 79:3203–59.

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5:31–56.

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

Ammann, M., T. Burdorf, L. Liebi, and S. Stöckl. 2022. Survivorship and delisting bias in cryptocurrency markets. Unpublished Working Paper.

Andrei, D., and M. Hasler. 2015. Investor attention and stock market volatility. *Review of Financial Studies* 28:33–72.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. 2006. The cross-section of volatility and expected returns. *Journal of Finance* 61:259–99.

Ardia, D., and K. Bluteau. 2024. Twitter and cryptocurrency pump-and-dumps. *International Review of Financial Analysis* 95:103479–.

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

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen. 2013. Value and momentum everywhere. *Journal of Finance* 68:929–85.

Bailey, M., R. Cao, T. Kuchler, and J. Stroebel. 2018. The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy* 126:2224–76.

Bali, T. G., N. Cakici, and R. F. Whitelaw. 2011. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99:427–46.

Barber, B. M., X. Huang, T. Odean, and C. Schwarz. 2022. Attention-induced trading and returns: Evidence from robinhood users. *Journal of Finance* 77:3141–90.

Barber, B. M., and T. Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies* 21:785–818.

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.

Barillas, F., and J. Shanken. 2017. Which alpha? *Review of Financial Studies* 30:1316–38.

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.

Bartov, E., L. Faurel, and P. S. Mohanram. 2018. Can twitter help predict firm-level earnings and stock returns? *Accounting Review* 93:25–57.

Benedetti, H., and L. Kostovetsky. 2021. Digital tulips? returns to investors in initial coin offerings. *Journal of Corporate Finance* 66:101786–.

Benetton, M., W. Mullins, M. Niessner, and J. Toczynski. 2024. Celebrity influencers: This is not financial advice. Unpublished Working Paper.

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.

Borri, N., D. Massacci, M. Rubin, and D. Ruzzi. 2022. Crypto risk premia. Unpublished Working Paper.

Bose, D., H. Cordes, S. Nolte, J. C. Schneider, and C. F. Camerer. 2022. Decision weights for experimental asset prices based on visual salience. *Review of Financial Studies* 35:5094–126.

Broadstock, D. C., and D. Zhang. 2019. Social-media and intraday stock returns: The pricing power of sentiment. *Finance Research Letters* 30:116–23.

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., P. De, Y. J. Hu, and B.-H. Hwang. 2014. Wisdom of crowds: The value of stock opinions transmitted through social media. *Review of Financial Studies* 27:1367–403.

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.

Chen, M., M. X. Hanauer, and T. Kalsbach. 2024. Design choices, machine learning, and the cross-section of stock returns. *Available at SSRN 5031755* .

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

Chordia, T., A. Subrahmanyam, and V. R. Anshuman. 2001. Trading activity and expected stock returns. *Journal of financial Economics* 59:3–32.

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.

Cong, L. W., Y. Li, and N. Wang. 2021. Tokenomics: Dynamic adoption and valuation. *Review of Financial Studies* 34:1105–55.

Cookson, J. A., J. E. Engelberg, and W. Mullins. 2023. Echo chambers. *Review of Financial Studies* 36:450–500.

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.

Da, Z., J. Engelberg, and P. Gao. 2011. In search of attention. *Journal of Finance* 66:1461–99.

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

Daniel, K., D. Hirshleifer, and A. Subrahmanyam. 1998. Investor psychology and security market under-and overreactions. *Journal of Finance* 53:1839–85.

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.

Fang, H., and T.-Y. Lai. 1997. Co-kurtosis and capital asset pricing. *Financial Review* 32:293–307.

Fausch, J., M. Frigg, S. Ruenzi, and F. Weigert. 2024. Machine learning mutual fund flows. *Available at SSRN 4812038* .

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.

Fieberg, C., S. Günther, T. Poddig, and A. Zaremba. 2024. Non-standard errors in the cryptocurrency world. *International Review of Financial Analysis* 92:103106–26.

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

Fong, W. M., and B. Toh. 2014. Investor sentiment and the max effect. *Journal of Banking & Finance* 46:190–201.

Frazzini, A., J. Friedman, and L. Pomorski. 2016. Deactivating active share. *Financial Analysts Journal* 72:14–21.

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

Frydman, C., and B. Wang. 2020. The impact of salience on investor behavior: Evidence from a natural experiment. *Journal of Finance* 75:229–76.

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.

Gu, C., and A. Kurov. 2020. Informational role of social media: Evidence from twitter sentiment. *Journal of Banking & Finance* 121:105969–.

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.

Harvey, C. R., and A. Siddique. 2000. Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263–95.

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.

Hillert, A., H. Jacobs, and S. Müller. 2014. Media makes momentum. *Review of Financial Studies* 27:3467–501.

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

Hong, H., and J. C. Stein. 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance* 54:2143–84.

Hosseini, A., G. Jostova, A. Philipov, and R. Savickas. 2020. The social media risk premium. Unpublished Working Paper.

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.

Jiang, H., P. Liu, A. F. Roch, and X. Zhou. 2023. Social media sentiment and bitcoin price dynamics: Evidence from 36 million postings. Unpublished Working Paper.

Jondeau, E., and M. Rockinger. 2006. The copula-garch model of conditional dependencies: An international stock market application. *Journal of international money and finance* 25:827–53.

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.

Kraus, A., and R. H. Litzenberger. 1976. Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085–100.

Kumar, A., S. Ruenzi, and M. Ungeheuer. 2021. Daily winners and losers. Unpublished Working Paper.

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

Li, B., and A. G. Rossi. 2020. Design choices, machine learning, and the cross-section of stock returns. *Available at SSRN 3737667* .

Liu, Y., J. Sheng, and W. Wang. 2022. Technology and cryptocurrency valuation: Evidence from machine learning. Unpublished Working Paper.

Liu, Y., and A. Tsyvinski. 2021. Risks and returns of cryptocurrency. *Review of Financial Studies* 34:2689–727.

Liu, Y., A. Tsyvinski, and X. Wu. 2022. Common risk factors in cryptocurrency. *Journal of Finance* 77:1133–77.

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

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66:35–65.

Lyandres, E., B. Palazzo, and D. Rabetti. 2022. Initial coin offering (ico) success and post-ico performance. *Management Science* 68:8658–79.

Lyócsa, Š., E. Baumöhl, and T. Vỳrost. 2022. Yolo trading: Riding with the herd during the gamestop episode. *Finance Research Letters* 46:102359–.

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.

Menkveld, A. J., A. Dreber, F. Holzmeister, J. Huber, M. Johannesson, M. Kirchler, M. Razen, U. Weitzel, D. Abad, M. M. Abudy, T. Adrian, Y. Ait-Sahalia, O. Akmansoy, J. Alcock, V. Alexeev, A. Aloosh, L. Amato, D. Amaya, J. Angel, A. Bach, E. Baidoo, G. Bakalli, A. Barbon, O. Bashchenko, P. C. Bindra, G. H. Bjonnes, J. Black, B. S. Black, S. Bohorquez, O. Bondarenko, C. S. Bos, C. Bosch-Rosa, E. Bouri, C. T. Brownlees, A. Calamia, V. N. Cao, G. Capelle-Blancard, L. Capera, M. Caporin, A. Carrion, T. Caskurlu, B. Chakrabarty, M. Chernov, W. M. Cheung, L. B. Chincarini, T. Chordia, S. C. Chow, B. Clapham, J.-E. Colliard, C. Comerton-Forde, E. Curran, T. Dao, W. Dare, R. J. Davies, R. De Blasis, G. De Nard, F. Declerck, O. Deev, H. Degryse, S. Deku, C. Desagre, M. A. van Dijk, C. Dim, T. Dimpfl, Y. Dong, P. Drummond, T. L. Dudda, A. Dumitrescu, T. Dyakov, A. H. Dyhrberg, M. Dzieliński, A. Eksi, I. El Kalak, S. ter Ellen, N. Eugster, M. D. D. Evans, M. Farrell, E. Félez-Viñas, G. Ferrara, E. M. Ferrouhi, A. Flori, J. Fluharty-Jaidee, S. Foley, K. Y. L. Fong, T. Foucault, T. Franus, F. A. Franzoni, B. Frijns, M. Frömmel, S. Fu, S. Füllbrunn, B. Gan, T. Gehrig, D. Gerritsen, J. Gil-Bazo, L. R. Glosten, T. Gomez, A. Gorbenko, U. Güçbilmez, J. Grammig, V. Gregoire, B. Hagströmer, J. Hambuckers, E. Hapnes, J. H. Harris, L. Harris, S. Hartmann, J.-B. Hasse, N. Hautsch, X. He, D. Heath, S. Hediger, T. Hendershott, A. M. Hibbert, E. Hjalmarsson, S. A. Hoelscher, P. Hoffmann, C. W. Holden, A. R. Horenstein, W. Huang, D. Huang, C. Hurlin, A. Ivashchenko, S. R. Iyer, H. Jahanshahloo, N. Jalkh, C. M. Jones, S. Jurkatis, P. Jylha, A. Kaeck, G. Kaiser, A. Karam, E. Karmaziene, B. Kassner, M. Kaustia, E. Kazak, F. Kearney, V. van Kervel, S. Khan, M. Khomyn, T. Klein, O. Klein, A. Klos, M. Koetter, J. P. Krahnen, A. Kolokolov, R. A. Korajczyk, R. Kozhan, A. Kwan, Q. Lajaunie, F. E. Lam, M. Lambert, H. Langlois, J. Lausen, T. Lauter, M. Leippold, V. Levin, Y. Li, M. H. Li, C. Y. Liew, T. Lindner, O. B. Linton, J. Liu, A. Liu, G. Llorente, M. Lof, A. Lohr, F. A. Longstaff, A. Lopez-Lira, S. Mankad, N. Mano, A. Marchal, C. Martineau, F. Mazzola, D. Meloso, R. Mihet, V. Mohan, S. Moinas, D. Moore, L. Mu, D. Muravyev, D. Murphy, G. Neszveda, C. Neumeier, U. Nielsson, M. Nimalendran, S. Nolte, L. L. Norden, P. O'Neill, K. Obaid, B. A. Ødegaard, P. Östberg, M. Painter, S. Palan, I. Palit, A. Park, R. Pascual, P. Pasquariello, L. Pastor, V. Patel, A. J. Patton, N. D. Pearson, L. Pelizzon, M. Pelster, C. Pérignon, C. Pfiffer, R. Philip, T. Plíhal, P. Prakash, O.-A. Press, T. Prodromou, T. J. Putniņš, G. Raizada, D. A. Rakowski, A. Ranaldo, L. Regis, S. Reitz, T. Renault, R. W. Renjie, R. Renò, S. Riddiough, K. Rinne, P. Rintamäki, R. Riordan, T. Rittmannsberger, I. Rodríguez-Longarela, D. Rösch, L. Rognone, B. Roseman, I. Rosu, S. Roy, N. Rudolf, S. Rush, K. Rzayev, A. Rzeźnik, A. Sanford, H. Sankaran, A. Sarkar, L. Sarno, O. Scaillet, S. Scharnowski, K. R. Schenk-Hoppé, A. Schertler, M. Schneider, F. Schroeder, N. Schuerhoff, P. Schuster, M. A. Schwarz, M. S. Seasholes, N. Seeger, O. Shachar, A. Shkilko, J. Shui, M. Sikic, G. Simion, L. A. Smales, P. Söderlind, E. Sojli, K. Sokolov, L. Spokeviciute, D. Stefanova, M. G. Subrahmanyam, S. Neusüss, B. Szaszi, O. Talavera, Y. Tang, N. Taylor, W. W. Tham, E. Theissen, J. Thimme, I. Tonks, H. Tran, L. Trapin, A. B. Trolle, G. Valente, R. A. Van Ness, A. Vasquez, T. Verousis, P. Verwijmeren, A. Vilhelmsson, G. Vilkov, V. Vladimirov, S. Vogel, S. Voigt, W. Wagner, T. Walther, P. Weiss, M. van der Wel, I. M. Werner, P. J. Westerholm, C. Westheide, E. Wipplinger, M. Wolf, C. C. P. Wolff, L. Wolk, W.-K. Wong, J. Wrampelmeyer, S. Xia, D. Xiu, K. Xu, C. Xu, P. K. Yadav, J. Yagüe, C. Yan, A. Yang, W. Yoo, W. Yu, S. Yu, B. Z. Yueshen, D. Yuferova, M. Zamojski, A. Zareei, S. Zeisberger, S. S. Zhang, X. Zhang, Z. Zhong, Z. I. Zhou, C. Zhou, S. Zhu, M. Zoican, R. C. J. Zwinkels, J. Chen, T. Duevski, G. Gao, R. Gemayel, D. Gilder, P. Kuhle, E. Pagnotta, M. Pelli, J. Sönksen, L. Zhang, K. Ilczuk, D. Bogoev, Y. Qian, H. C. Wika, Y. Yu, L. Zhao, M. Mi, L. Bao, A. Vaduva, M. Prokopczuk, A. Avetikian, and Z.-X. Wu. 2024. Nonstandard errors. *Journal of Finance* 79:2339–90.

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

Müller, S., N. Pugachyov, and F. Weigert. 2024. Forecasting mutual fund performance – combining return-based with portfolio holdings-based predictors. *Available at SSRN 4887900* .

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.

———. 2022. Game on: Social networks and markets. *Journal of Financial Economics* 146:1097–119.

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

Renault, T. 2017. Intraday online investor sentiment and return patterns in the us stock market. *Journal of Banking & Finance* 84:25–40.

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

Scheuch, C., S. Voigt, and P. Weiss. 2023. *Tidy Finance with R.* Chapman and Hall/CRC.

Schmeling, M., A. Schrimpf, and K. Todorov. 2023. Crypto carry. Unpublished Working Paper.

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.

Sockin, M., and W. Xiong. 2023. A model of cryptocurrencies. *Management Science* 69:6684–707.

Soebhag, A., B. Van Vliet, and P. Verwijmeren. 2024. Non-standard errors in asset pricing: Mind your sorts. *Journal of Empirical Finance* 78:517–35.

Solomon, D. H., E. Soltes, and D. Sosyura. 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics* 113:53–72.

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

Tan, H.-T., E. Y. Wang, and B. Zhou. 2015. How does readability influence investors' judgments? consistency of benchmark performance matters. *The Accounting Review* 90:371–93.

Walter, D., R. Weber, and P. Weiss. 2024. Methodological uncertainty in portfolio sorts. *Available at SSRN 4164117* .

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.