Essays on Mutual Fund Activeness and Sustainability as a Flow Determinant

D I S S E R T A T I O N
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences
and International Affairs
to obtain the title of
Doctor of Philosophy in Finance

submitted by

Sebastian Fischer

from

Germany

Approved on the application of

Prof. Dr. Manuel Ammann

and

Prof. Dr. Markus Schmid

Dissertation no. 4885

Difo-Druck GmbH, Untersiemau 2019

The University of St.Gallen, School of Management, Economics, Law, Social Sciences and International Affairs hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, May 21, 2019

The President:

Prof. Dr. Thomas Bieger

Für Mama und Papa.

Danke.

Acknowledgements

I am very grateful for the continuous and great support that I have received throughout my PhD studies. I want to thank my supervisor Prof. Dr. Manuel Ammann for his patient guidance, his helpful advice and in particular his great empathy. Thank you, Manuel. I am thankful to my co-supervisor Prof. Dr. Markus Schmid for his valuable feedback and to my co-authors for the excellent collaboration. I am particularly grateful to Prof. Dr. Florian Weigert, who was not only a co-author, but became a teacher and friend. Thank you, Florian. Furthermore, I want to thank my outstanding colleagues and friends Alexander Cochardt, Alexander Feser, Stephan Heller, Mathis Mörke, Jan-Philip Schade and Simon Straumann for their invaluable advice, honest feedback and for countless insightful discussions. Most importantly, I would like to thank my family and Sarah for their enduring support and encouragement.

St. Gallen, February 10, 2019
Sebastian Fischer

Table of Content

List of Tables

List of Figures

Summary

This dissertation contributes to two recent debates in the mutual fund literature: The impact of sustainability on mutual fund flows and the connection between fund activeness and mutual fund performance.

In March 2016, Morningstar, one of the leading information providers in the mutual fund industry, introduced its mutual fund Sustainability Rating. The Rating provides investors with an easy-to-understand measure to identify funds that invest in accordance with high environmental, social, and governance standards. Chapter 1 investigates the effect of this Rating on mutual fund flows. An average high-rated retail fund receives up to USD 10.1 million higher net flows and an average low-rated retail fund suffers from up to USD 3.5 million lower net flows than an average-rated fund during the first year after the publication of the Rating. This result stresses the importance of sustainability as an investment criterion and the impact of the Sustainability Rating as a source of information to private investors.

Chapters 2 through 4 examine whether the trading activity of a fund manager or fund activeness, that is the deviation of a fund portfolio from its benchmark, is linked to future performance. The fund literature has identified various activity measures that can predict fund returns. Chapter 2 shows that two of the most important measures, Active Share and the R^2 selectivity measure, have not been good predictors after 2003 when controlling for different benchmark indices and alternative risk factors. Chapter 3 examines the investment performance of funds whose exposures to the risk factors of the Carhart model vary significantly over time. The analysis shows that funds with volatile factor weights achieve on average lower returns than funds with stable factor exposures. After testing for alternative explanations, this result provides evidence that fund managers fail to time risk factors. This finding also contributes to the current debate on whether risk factors can be timed. Chapter 4 addresses the question whether fund managers trade more in times of large market mispricing and, therefore, whether fund turnover is positively correlated to the subsequent fund performance. The results confirm respective findings from earlier research for an international mutual fund sample. They additionally show that this turnover-performance relationship is particularly strong in countries with highly skilled fund managers, who trade more in times of high market opportunities. Furthermore, the effect is stronger in markets with a low performance persistence.

Zusammenfassung

Die vorliegende Dissertation widmet sich zwei aktuellen Fragestellungen zum Thema Aktienfonds: Der Bedeutung von Nachhaltigkeit für den Zu- und Abfluss von Investorengeldern sowie dem Zusammenhang zwischen der Anlagestrategie und dem Anlageerfolg eines Fonds.

Mit der Einführung eines Nachhaltigkeits-Ratings für Fonds hat der Datenanbieter Morningstar im März 2016 eine einfach zu interpretierende Kennzahl geschaffen, mit deren Hilfe Investoren Aktienfonds mit hoher ökologischer und sozialer Verantwortung identifizieren können. Kapitel 1 untersucht, wie Anleger auf die Erstveröffentlichung dieses Ratings reagieren. Während Retail-Fonds mit einem schlechten Nachhaltigkeits-Rating im ersten Jahr um durchschnittlich bis zu 3.5 Millionen USD höhere Abflüsse als durchschnittlich bewerte Fonds verzeichnen, investieren Privatanlegen um bis zu 10 Millionen USD mehr in gut bewerte Fonds. Dieses Ergebnis belegt die Bedeutung nachhaltiger Investmentprodukte und des Nachhaltigkeitsrating als Informationsquelle für Privatanleger.

Kapitel 2 bis 4 untersuchen, ob die Handelsaktivität eine Fondmanagers oder die Abweichung eines Fondportfolios von seinem Benchmark zukünftige Renditen vorhersagen. Die Fondsliteratur hat diverse Aktivitätsmasse identifiziert, die Fondsrenditen prognostizieren. Kapitel 2 zeigt, dass zwei der auch in der Praxis bedeutendsten Masse, Active Share und das R^2-Selectivity-Mass, bei Berücksichtigung unterschiedlicher Benchmarks sowie alternativer Risikofaktoren nach 2003 keine verlässlichen Renditeindikatoren mehr sind. Kapitel 3 untersucht den Anlageerfolg von Fonds, deren Exposition gegenüber den Risikofaktoren des Carhart-Modells über die Zeit stark schwankt. Die Datenanalyse zeigt, dass Fonds mit volatilen Faktor-Gewichten im Mittel geringere Renditen erzielen als Fonds mit stabilen Faktor-Gewichten. Nach Ausschluss alternativer Erklärungsansätze liefert dieses Ergebnis Evidenz dafür, dass Fondsmanager beim Versuch, Risikofaktoren zu timen, mehrheitlich scheitern. Diese Erkenntnis trägt auch zur aktuell geführten Debatte über das Timing von Risikofaktoren bei. Kapitel 4 widmet sich der Frage, ob Fondsmanager in Zeiten von Fehlbewertungen am Aktienmarkt vermehrt handeln und daher auf Jahre hoher Handelsaktivität Jahre hoher Fondsrenditen folgen. Die Analyse bestätigt entsprechende Resultate früherer Forschung auch für internationale Fondsmärkte und zeigt, dass dieser Turnover-Performance-Zusammenhang vor allem in Ländern mit sehr guten Fondsmanagern besteht. Der Effekt ist zudem in Märkten mit geringer Renditepersistenz besonders ausgeprägt.

Chapter 1

The Impact of the Morningstar Sustainability Rating on Mutual Fund Flows

Manuel Ammann, Christopher Bauer, Sebastian Fischer, Philipp Müller*

Status: Forthcoming in the European Financial Management

Abstract

We examine the effect of the introduction of Morningstar's Sustainability Rating in March 2016 on mutual fund flows. Exploiting this shock to the availability of sustainability information, we find strong evidence that retail investors shift money away from low-rated and into high-rated funds. An average high-rated retail fund receives between \$4.1 million and \$10.1 million higher net flows and an average low-rated retail fund suffers from \$1.0 million to \$3.5 million lower net flows than an average-rated fund during the first year after the publication of the Rating. Institutional investors react much more weakly to the publication of the Rating.

JEL Classification: G11, G14, G23

Keywords: Mutual Fund, Sustainability, Investment Decisions, Information

* We thank John Doukas (the editor) and an anonymous referee for their helpful comments. We also thank Daniel Höchle, Sebastian Utz and Florian Weigert as well as the participants of the 2017 TiF Seminar at the University of St. Gallen and the participants of the SGF Conference 2018 for their helpful remarks. We are grateful to Morningstar Switzerland GmbH for providing access to Morningstar Direct as well as a unique dataset on the entire history of the Morningstar Sustainability Rating.

1.1 Introduction

Academic research has shed light on the empirical relation between fund flows and various performance measures,[1] on the divergence of this performance-flow relationship in different investor clienteles,[2] and on the marginal impact of reduced search costs through increased marketing efforts.[3] Much less notice has been taken of the increasing attention of investors to sustainable investments and its impact on fund flows. While assets under professional management utilizing sustainable investment criteria grew by 33% from 2014 to 2016, more than one in every five dollars under professional management in the United States (approximately \$8.72 trillion) is invested according to sustainable investment strategies.[4] Despite this already substantial market share of sustainable investments, a recent market survey suggests this sector may grow even further, as 75% of investors have expressed interest in sustainable investments.[5] Thus, a new variable is added to the decision-making process of investors.

The role of sustainability in investments is discussed by literature linking stock and fund performance to environmental, social and governance (ESG) criteria.[6] There exists, however, little empirical evidence on the impact of sustainability on fund flows and on the use of sustainability information in investment decisions. Massa (2003) suggests that investors select funds based on performance-related as well as non-performance-related characteristics. Statman (2008) interviews social investors and finds that ethical, societal and religious values influence their investment decisions. He observes that this investor clientele evaluates an investment by combining its social responsibility and return characteristics. Riedl and Smeets (2017) link individual investor data to survey responses and find that the decision to invest in socially responsible mutual funds can be explained by social preferences and social signaling

[1] See, for example, Chevalier and Ellison (1997), Sirri and Tufano (1998), and Ivković and Weisbenner (2009).
[2] Del Guercio and Tkac (2002) show that institutional investors, in contrast to retail investors, punish poorly performing managers by withdrawing assets under management but do not invest in recent winners proportionally.
[3] Sirri and Tufano (1998) show that funds, which receive greater media attention and belong to larger complexes grow more rapidly than other funds. Moreover, they document that the performance−flow relationship is most pronounced for funds with greater marketing efforts.
[4] The US SIF Foundation's "2016 Report on US Sustainable, Responsible and Impact Investing Trends" reports \$40.3 trillion of total assets under professional management in the USA, of which \$8.72 trillion have been invested according to sustainable investment strategies.
[5] The study can be found in Morgan Stanley's 2017 edition of the Sustainable Signals series, "New Data from the Individual Investor."
[6] Filbeck et al. (2009) provide an overview of empirical research investigating the stock performance of sustainable companies. A comprehensive literature overview on the performance of socially responsible funds can be found in Renneboog et al. (2008) and Capelle-Blancard and Monjon (2014).

rather than financial motives. Bollen (2007) argues that investors have a multi-attribute utility function and therefore profit from owning socially responsible investments. His findings, especially a lower volatility of fund flows and a lower (higher) sensitivity of flows to negative (positive) past performance for socially responsible investment (SRI) funds compared to convenient funds, support this framework.[7] Benson and Humphrey (2008) and El Ghoul and Karoui (2017) find a weaker performance-flow relationship for sustainable funds. Renneboog et al. (2011) agree with this result for all SRI funds but funds with environmental screens, for which fund flows react more sensitively to past performance. Whereas those prior results provide indirect evidence that investors appreciate sustainability, a causal relationship between sustainability and fund flows has not been established so far. To test for such a relationship, we make use of an exogenous shock to the availability of sustainability information: the launch of Morningstar's Sustainability Rating in March 2016.

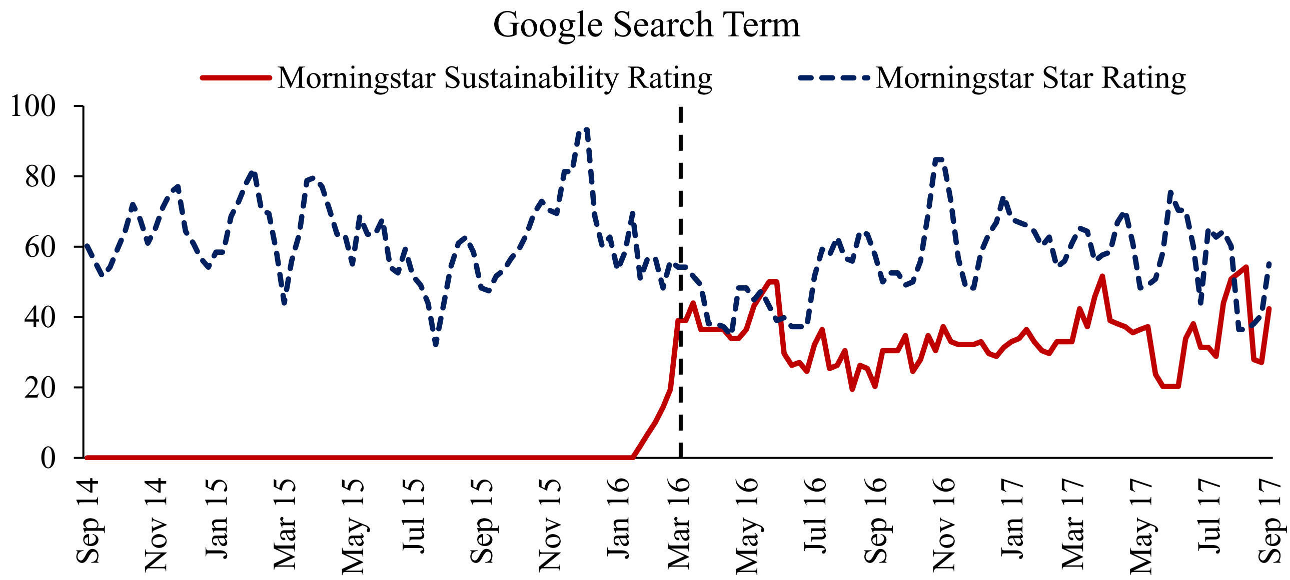
Economists widely recognize the complexity of consumers' purchasing decisions in the mutual fund marketplace by means of costly search. Retail investors face thousands of choices and often lack access to up-to-date information on potential fund investments or are unable to process sophisticated information. Del Guercio and Tkac (2002), Evans and Fahlenbrach (2012) and Salganik-Shoshan (2016) all show that retail investors react to simple return measures like past raw returns, whereas institutional investors pay attention to more sophisticated performance measures such as multi-factor alphas. As a result, the academic literature has documented the substantial impact of information intermediaries who provide free access to clearly displayed information.[8] Due to the ongoing debate about the definition of sustainability and the fact that information on sustainability has been available at company level only, we expect a freely accessible rating on sustainability to have such an impact, too. If investors have a multi-attribute utility function, as proposed by Bollen (2007), but cannot assess a fund's level of sustainability, they will rely on a third-party judgment in order to align their investments to their preferences. Prior to March 2016 there was no such freely accessible and reliable information.

[7] As pointed out by Barnett and Salomon (2006), there is a heterogeneity within SRI funds concerning their type of social screening. Throughout this paper we will refer to SRI funds and other funds complying with ESG criteria as sustainable funds.

[8] Del Guercio and Tkac (2002) provide robust empirical evidence that a package of fund quality information embodied in the Morningstar Star Rating affects investor flow independently of the influence of other common measures of fund performance.

Figure 1.1: Google search interest in Morningstar Ratings

This graph shows the four-week moving average of the relative Google search interest in the search terms "Morningstar Sustainability Rating" and "Morningstar Star Rating." The dashed vertical line indicates the initial publication date of the Sustainability Rating



With the publication of its Sustainability Rating, Morningstar, one of the leading information providers in the mutual fund industry, has transformed sustainability from a difficult-to-grasp characteristic into an easy-to-understand figure. Morningstar's Sustainability Rating measures a fund's conformity to ESG criteria and assigns each mutual fund share class to a rating category between 1 (low sustainability) and 5 (high sustainability). Specifically, funds among the top 10% are assigned a Sustainability Rating of 5, whereas the bottom 10% of funds receive a rating of 1. An analysis of the relative Google search interest, displayed in Figure 1.1, reveals great attention to the Morningstar Sustainability Rating, not only upon its public launch but also in subsequent months. In spring 2016, the term "Morningstar Sustainability Rating" was about as popular as the well-established "Morningstar Star Rating" and remained so during the next year.

The introduction of the Sustainability Rating in March 2016 constitutes a shock to investors' investment decisions as it provides them with freely accessible information on the sustainability of a majority of US equity mutual funds. We expect investors to adjust their investments in response to the additional information in order to align them to their preference for sustainable investments. We particularly expect the Rating to be informative to retail investors due to their limited informational resources and stronger interest in sustainable investments.[9] Whereas institutional investors already had access to databases providing both fund holding data and company-level information on sustainability prior to the launch of Morningstar's Sustainability Rating, it is unlikely that retail investors had access to this data.

To examine our hypotheses, we would ideally compare funds with a published Sustainability Rating to comparable funds with the same but unpublished rating. Morningstar, however, simultaneously launched its Sustainability Rating for the vast majority of mutual funds on February 29, 2016 (available for Morningstar Direct users only) and March 17, 2016 (publicly available without cost). Funds that did not receive a rating during those months cannot serve as a valid control group as Morningstar selected funds that received a Sustainability Rating in early 2016 based on size and the availability of holding data. Therefore, the two groups of funds cannot be compared. To derive a sound estimate of the effect of the Morningstar Sustainability Rating on fund flows, we therefore employ three different empirical methodologies: panel regressions, propensity score matching, and an event study.

In a first step we use panel regressions to measure the impact of rating categories on fund flows. We have unpublished data on the funds' Sustainability Rating prior to the launch of the Morningstar Sustainability Rating and can therefore observe the effect of a high or low Sustainability Rating before and after its first-time publication. We find that a published Sustainability Rating has a strong impact on fund flows. The effect is only significant for retail share classes and does not appear prior to the launch of the Rating. A high-rated retail fund receives a 0.78 percentage points per month higher net flow than a low-rated fund.

To ensure that our panel regression results are not driven by the comparison of heterogeneous groups of funds and to also account for the nonlinear relationships between fund characteristics and fund flows−as argued in the case of performance by, for example, Chevalier and Ellison (1997), Sirri and Tufano (1998) and Ivković and Weisbenner (2009)−we apply propensity score matching. We match retail funds with a high and funds with a low Sustainability Rating to comparable funds with an average rating. This leaves us with two groups of funds that do not differ significantly in any relevant fund characteristics other than the

[9] A compilation of survey evidence indicates that retail investors display a substantially stronger interest in sustainable investment strategies than institutional investors. The December/January 2016 issue of the *Morningstar* magazine provides an overview of existing studies.

Sustainability Rating. The comparison of flows between these funds confirms the results of panel regressions. Retail investors react to a high Sustainability Rating by a 9.41 percentage points higher net inflow during the first year after the launch. Low-rated funds receive a 6.80 percentage points lower net flow during the 12 months following the initial rating publication.

The first two methods, panel regressions and propensity score matching, provide insights from a cross-sectional comparison of funds with different Sustainability Ratings. In contrast, the third methodology used, an event study as proposed by Del Guercio and Tkac (2008), measures the effect of the initial publication of the Morningstar Sustainability Rating from the time series of single fund data. We use fund characteristics and past flows to estimate a fund's monthly expected net flows around the launch of the Morningstar Sustainability Rating and compare them to the actual flows. We then examine the difference (i.e. abnormal flows) for different rating categories. Again, the results suggest a strong relationship between the Sustainability Rating and mutual fund flows after the initial publication. High-rated funds receive an abnormal flow of 1.83% during the first 6 months after the ranking was published. Low-rated funds suffer from an abnormal flow of –1.01% during the same half-year period.

We calculate the economic value of a high and low Sustainability Rating (compared to an average Rating) from all three approaches. We find that an average high-rated retail fund receives between \$4.1 million and \$10.1 million higher inflows per year than would have been expected for an average rating. A low-rated retail fund suffers from a \$1.0 million to \$3.5 million lower net flow per year compared to an average-rated fund. We find that the impact of Morningstar's Sustainability Rating is much weaker for institutional share classes.

To better understand the drivers of these results we apply panel regressions at a fund portfolio level, for which we can split net flows into gross inflows and gross outflows. We find that both inflows and outflows are affected by the launch of the Rating, but the effect is significantly stronger for gross inflows and we only observe the asymmetrical impact of the Rating for gross inflows where the high rated funds receive disproportionally larger amounts of new investments. An additional analysis adds measures of the Rating distribution within a fund family to the panel regression. Nanda et al. (2004) show positive spillover effects of good past performance within fund families. We do not find such a positive spillover for the Sustainability Rating but observe a negative effect of a high Sustainability Rating of other funds within the same fund family. This result would be consistent with fund families focusing their marketing activities on funds with a high Sustainability Rating.

Our study of the Morningstar Sustainability Rating and its marginal impact on mutual fund flows provides new insights into the decision-making of investors and contributes to the existing mutual fund literature in four ways. First, our research provides empirical evidence for retail investors' strong interest in sustainable investment strategies, which has so far only been documented in qualitative market surveys. By showing that retail investors invest in funds with the highest Sustainability Rating while withdrawing money from lower-rated funds, our paper establishes a causal link between sustainability and mutual fund flows and supports a model in which investors have multi-attribute utility functions.[10] By showing that investors react disproportionately strongly to a high Rating and divest from funds with the lowest Rating at a much lower rate, we provide additional evidence on investors' sensitivity to levels of sustainability. Second, given the crucial role of information intermediaries throughout an investor's purchasing process, our paper complements the existing literature by demonstrating a significant marginal impact of condensed and clearly displayed sustainability information. Consistent with the findings of Del Guercio and Tkac (2008), who demonstrate a significant investors' reaction to quality information, we provide robust evidence that the aggregated sustainability information incorporated in the Morningstar Sustainability Rating affects mutual fund flows independently of the impact of other factors. Third, we show that retail investors are much more sensitive to the publication of the Rating than institutional investors. This provides evidence that in particular retail investors react to unsophisticated information. This result has previously been shown for performance measures, for example by Del Guercio and Tkac (2002) and Evans and Fahlenbrach (2012). Our paper expands its validity beyond the performance dimension. Finally, the approaches employed in our paper allow us to estimate the economic magnitude of the demonstrated Sustainability Rating effect in terms of additional dollar flows allocated by mutual fund investors. Our findings document a substantial economic impact of the recently launched Sustainability Rating.

The remainder of this paper is structured as follows: Section 1.2 introduces the Morningstar Sustainability Rating in greater detail and describes our data set. Section 1.3 contains our

---

[10] As pointed out by an anonymous referee, there might also be non-altruistic reasons for investors' preference toward sustainable investment opportunities. Besides social preferences and social signaling as suggested by Statman (2008) and Riedl and Smeets (2017) neither the majority of previous studies nor our work can rule out that investors may interpret sustainability as a synonym for favorable risk-return characteristics, due, for example, to higher quality or better governance. Literature on the financial value of social responsible investments is inconclusive; see, for example, Fabozzi et al. (2008) who find sin stocks to outperform and Kempf and Osthoff (2007) who find stocks with high socially responsible ratings to outperform.

results for panel regressions (Section 1.3.1), propensity score matching (Section 1.3.2) and the event study setting (Section 1.3.3). Section 1.3.4 analyses the Rating's gross effects on fund inflows and outflows and Section 1.3.5 investigates the role of fund families for the sustainability-flow relationship. Section 1.4 concludes.

1.2 Data and summary statistics

1.2.1 Background on the Morningstar Sustainability Rating

The Morningstar Sustainability Rating indicates to what extent a fund holds securities whose issuers are successfully managing environmental, social and governance risks and opportunities. It evaluates a fund's level of sustainability relative to funds of the same Morningstar Category. The Rating is a holding-based measure and is calculated from companies' ESG and controversy scores provided by Sustainalytics, which evaluate companies relative to their global industry peers.

First, Morningstar derives an aggregate portfolio sustainability score, based on latest holdings data and the holdings' ESG and controversy scores. Specifically, the portfolio sustainability score is calculated as the difference between the asset-weighted average of normalized company-level ESG scores (0–100) and the asset-weighted average of controversy scores (0–20). Within each global industry group, the company-level ESG scores are normalized to have a mean of 50 and standard deviation of 10 to make them comparable across industry peer groups, which is essential in a portfolio context. A fund is only evaluated if at least 50% of its assets are covered by a company ESG and controversy score, whereby only equities and corporate debt are considered. Finally, within each Morningstar Category the funds with the 10% highest (lowest) portfolio sustainability scores receive a Sustainability Rating of high (low). The next top and bottom 22.5% are rated above and below average and the middle 35% are categorized as average. Portfolios receive a rating 1 month and 6 business days after their reported as-of date, and funds are ranked relative to peers on the same 1 month and 6 business day lag.[11]

[11] For a full explanation of the Morningstar Sustainability Rating methodology refer to https://corporate1.morningstar.com/Morningstar-Sustainability-Rating-Methodology–2.

The rating was first launched on February 29, 2016, initially only available for institutional investors via subscription-based Morningstar Direct. On March 17, 2016 the public launch followed, making the Sustainability Rating accessible on Morningstar's website without registration and free of charge. For both launch dates the published Sustainability Ratings were based on end-of-December 2015 portfolio data. Subsequent rating updates are issued monthly based on the most recent company and holdings data.

1.2.2 Sample selection

For our empirical analysis we merge the CRSP Survivor-Bias-Free Mutual Fund Database with the Morningstar Direct database. Additionally, besides the monthly time series of published Sustainability Ratings from March 2016 to March 2017, we obtain unpublished Sustainability Ratings from November 2015 to February 2016 on a monthly basis directly from Morningstar.[12] We extrapolate the Sustainability Ratings from November 2015 to October and September 2015, so that our sample extends to six full months before the Sustainability Rating publication.[13] To justify the extrapolation, we compute monthly transition probabilities between Sustainability Rating categories. Share classes remain in the same rating category with a probability of more than 80%. Thus, extending our sample of prepublication months by 50% outweighs the minor approximation error caused by extrapolation.

Our sample focuses on actively managed domestic US equity mutual funds. Due to the limited availability of Sustainability Ratings, especially in the pre-publication period, we eliminate balanced, bond, index, international, and sector funds. Specifically, we exclude all funds not assigned to the Morningstar Global Categories *US Equity Large Cap Value, US Equity Large Cap Blend, US Equity Large Cap Growth, US Equity Mid Cap* or *US Equity Small Cap* and focus our analysis on share classes explicitly marked either as retail or institutional.[14] We exclude share classes closed to investors and share classes with total net assets below $1million. We also delete all observations that coincide with a fund merger, since the flows are likely to be distorted. As flows of different share classes may not be closely related, we consider each share class of a fund to be a distinct fund. Thus, in contrast to studies of fund

---

[12] We thank Morningstar for providing a unique dataset on the historical Morningstar Sustainability Rating from November 2015 to March 2017.
[13] Our results remain unchanged when repeating the analysis without extrapolating unpublished data.
[14] CRSP variables RETAIL_FUND and INST_FUND

performance, we are not double-counting observations using individual share classes. Table 1.1 shows the final number of distinct funds and share classes in our sample from September 2015 to March 2017. A share class is included in the sample whenever a Sustainability Rating is available. The number of funds and share classes remains close to unchanged from September 2015 to September 2016, with about 1000 distinct funds and 2900 share classes. In October 2016, a sharp rise in sample size to more than 1300 funds and over 3700 share classes is observed. This is due to an improved Sustainability Rating coverage. After October 2016, the sample size remains relatively constant. In total, our final dataset contains 60,644 fund-month observations with up to 3804 distinct share classes per month. The fraction of institutional share classes remains constant at about 49% over the whole period. In the following, all analyses are based on share classes and we refer to them as funds.

Monthly data on total net assets (MTNA), returns (MRET), expense ratios (EXP_RATIO), turnover ratios (TURN_RATIO), and the fund inception date (FIRST_OFFER_DT) are collected from the CRSP Mutual Fund Database. Missing values for expense ratios and turnover ratios are supplemented with data from the Morningstar Direct database. Further, we obtain monthly Morningstar Star Ratings from Morningstar Direct. A fund's style is determined by its Morningstar Global Category.

Table 1.1: Sample size by month

This table lists the monthly number of distinct US domestic equity mutual funds and share classes from September 2015 to March 2017. Only share classes classified as retail or institutional are considered. The monthly ratio of retail share classes is reported in the bottom row. A share class is included in the sample whenever a Sustainability Rating is available for the respective month.

Month	**Sep 15**	**Oct 15**	**Nov 15**	**Dec 15**	**Jan 16**	**Feb 16**	**Mar 16**	**Apr 16**	**May 16**	**Jun 16**
Funds	994	994	994	994	994	997	998	1001	1006	1007
Share classes	2914	2914	2914	2915	2915	2913	2914	2923	2939	2939
Retail (%)	48.7	48.7	48.7	48.6	48.6	48.8	48.7	48.7	48.7	48.7
Month	**Jul 16**	**Aug 16**	**Sep 16**	**Oct 16**	**Nov 16**	**Dec 16**	**Jan 17**	**Feb 17**	**Mar 17**	
Funds	1011	1012	1002	1312	1348	1345	1343	1337	1330	
Share classes	2945	2950	2922	3711	3804	3797	3794	3778	3743	
Retail (%)	48.7	48.8	48.9	49.2	49.0	49.0	49.0	49.2	49.5	

1.2.3 Summary statistics and variable definitions

By grouping funds according to their Sustainability Rating, we analyze the relationship between flows and the Sustainability Rating before and after its launch, and thus reveal first insights into investors' reaction to the new rating. We define monthly relative net flow as the net growth in fund assets beyond reinvested returns. Formally, it is calculated as

$$\mathrm{FLOW}_{i,t} = \frac{\mathrm{TNA}_{i,t} - \mathrm{TNA}_{i,t-1}(1+R_{i,t})}{\mathrm{TNA}_{i,t-1}}, \tag{1}$$

where $TNA_{i,t}$ are total net assets of fund i at the end of month t, and $R_{i,t}$ is the return of fund i during that month. This measure reflects the percentage growth of a funds' assets under management in excess of the growth that would have occurred if no new funds had flowed in and all dividends had been reinvested. To mitigate the influence of extreme outliers, flows are winsorized at the 1% and 99% levels. Additionally, we report data on various performance measures, return volatility as a measure of risk, fund size, expenses, turnover and the fund age in years, as these characteristics are identified as major drivers of fund flows in the mutual fund flow literature.

Specifically, Del Guercio and Tkac (2008) report a strong positive and convex relationship between the Morningstar Star Rating and mutual fund flows. They conclude that the Star Rating not only captures the general nonlinear relationship between performance and fund flows (see, for example, Ippolito, 1992; Gruber, 1996; Chevalier and Ellison, 1997; Goetzmann and Peles, 1997; Sirri and Tufano, 1998), but also has substantial independent influence on mutual fund investors' investment decisions, and consequently on fund flows. The Star Rating measures historical performance with respect to both return and risk relative to its peer group. Load-adjusted returns are used to compute three-, five- and ten-year risk-adjusted performance measures for each fund. Further, Ivković and Weisbenner (2009) and Sirri and Tufano (1998) point out the strong predictive power of alpha and raw returns for mutual fund flows. Therefore, to account for mutual fund investors' performance chasing behavior, we include the Morningstar Star Rating (or "Performance Rating") as a widely respected medium- to long-term performance measure, supplemented by monthly raw returns and 12-month Carhart (1997) four-factor alphas to cover shorter-term performance measures as well. We use monthly Fama and French (1993) as well as momentum risk factors from Kenneth R. French's website.[15]

Moreover, Ivković and Weisbenner (2009) and Sirri and Tufano (1998) find that mutual fund investors prefer funds with lower expense ratios, leading to a negative fee-fund-flow relationship. Further, Chevalier and Ellison (1997) and Huang et al. (2007) report that the level of flows is lower for older funds. Among others, Sirri and Tufano (1998) and Huang et al. (2007) find funds with a higher return volatility to receive fewer inflows and we therefore add the 12-month return volatility as a measure of risk. The literature on the turnover-flow relationship is sparse and mostly inconclusive. However, as the more sustainable funds in our sample have on average substantially lower turnover, we control for this characteristic to rule out any potential influence on our results. Finally, we measure the size of a fund by its total net assets under management, reflecting the fact that an equal dollar flow will have a larger percentage impact on smaller funds.

Panel A of Table 1.2 provides summary statistics for these characteristics for all funds in our sample from September 2015 to February 2016, that is, prior to the launch of the Sustainability Rating. In the 6-month period, funds on average experienced outflows (−0.24%) and had negative average monthly returns (−0.84%) and negative annualized 12-month four-factor alphas (−1.51%). Grouping funds by their Sustainability Rating cannot reveal any specific pattern in relative net flows. The difference in flows between funds within the highest and lowest Sustainability Rating category is slightly negative (−0.24%) but nonsignificant. The least outflows can be observed for funds with a below average Sustainability Rating (−0.03%) and the highest outflows for funds with the highest Sustainability Rating (−0.61%). Regarding the control variables, funds in the highest rating category tend to have significantly superior performance measures, a lower return volatility, twice as much in assets under management and significantly lower expense and turnover ratios compared to funds in the lowest rating category. The funds in the two rating categories are on average of the same age.

Panels B and C of Table 1.2 describe retail and institutional funds separately from September 2015 to February 2016, that is, prior to the publication of the Sustainability Rating. Institutional funds on average suffer from stronger outflows, have a higher Performance Rating and

[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/dat_library.html

less negative 12-month four-factor alphas, are about 30% smaller, 5 years younger and have lower expense ratios than retail funds. Comparing relative net flows between the highest and lowest Sustainability Rating category, noticeable differences between retail and institutional funds exist. Retail funds of both rating categories have similar net flows during this time period. In contrast, institutional funds in the lowest rating category have the weakest net outflows (−0.03%) compared to the highest rating category which observes the strongest net outflows (−0.68%).

Remarkably, after the launch of the Sustainability Rating the sustainability-flow relationship changed substantially, as shown by Panel D of Table 1.2, displaying mean characteristics for all funds after the launch of the Sustainability Rating from March 2016 to March 2017. In contrast to the pre-launch period, a distinct pattern can be identified. Flows are strictly increasing in the Sustainability Rating, leading to a highly significant monthly flow differential between the highest and lowest rating category of 0.93%.

Further, in line with our conjecture that different investor clienteles appreciate the new sustainability measure unequally, Panels E and F of Table 1.2 reveal considerable differences in the flow-sustainability relationship between retail and institutional funds. In fact, flows are strictly increasing in the Sustainability Rating for both share class types. However, the monthly flow differential for retail funds (1.28%) is much more pronounced than that for institutional funds (0.49%). Moreover, there is a sharp increase in flows (0.56%) moving from retail funds rated as above average to funds rated with the highest Sustainability Rating. Similarly, a marked drop in flows comparing below-average-rated to low-rated retail funds (−0.40%) arises. In contrast, the flows to the three middle rating categories are relatively similar. For institutional funds a comparable pattern cannot be observed, with flows much more equally dispersed over the five rating categories.

Thus, first insights favor our hypothesis that investors, especially retail investors, react to the sustainability information that becomes public with Morningstar's Sustainability Rating. However, some of the control variables−especially the performance measures−differ significantly between the highest and the lowest Rating category. For retail as well as for institutional funds, past performance is increasing in the Sustainability Rating. Specifically, funds with the highest Sustainability Rating category have a significantly better Performance Rating and 1-year alpha than funds with the lowest Sustainability Rating. A similar result holds for

Table 1.2: Mean fund characteristics sorted by the Sustainability Rating before and after the initial rating publication

This table presents the mean values of fund characteristics of our sample. All values are reported at a share class level. Panel A shows the mean characteristics for the months between September 2015 and February 2016, i.e. for the 6 months prior to the public launch of the Sustainability Rating. The share classes are sorted according to their most recent Sustainability Rating. Panels B and C display mean values for the subsample of retail and institutional share classes, respectively. Panels D-F present the same mean characteristics for the months from March 2016 to March 2017, i.e. after the public launch of the Sustainability Rating. Panel D comprises all share classes included in the sample. Panels E and F split the sample into retail and institutional share classes, respectively. A share class is included in the sample whenever a Sustainability Rating is available. Relative net flow is defined, as in Sirri and Tufano (1998), as $[\mathrm{TNA_t} - (1 + R_t)\mathrm{TNA}_{t-1}]/\mathrm{TNA}_{t-1}$. Monthly returns are obtained from the MRET variable in the CRSP mutual fund database. The Performance Rating is the one- to five-star Morningstar Star Rating provided by Morningstar. The Carhart (1997) four-factor alpha is calculated using monthly returns over the prior 12 months. The return volatility is the standard deviation of monthly net returns over the previous 12 months. The fund age is the number of years that the fund has been in existence up to the initial publication of the Sustainability Rating in March 2016 and is calculated using the FIRST_OFFER_DT variable in the CRSP mutual fund database. Total net assets, turnover ratios and expense ratios are respectively obtained from the MTNA, TURN_RATIO and EXP_RATIO variables in the CRSP mutual fund database. The column headed "(5)−(1)" presents difference-in-means tests for mean characteristics of high- and low-rated share classes. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively.

Panel A: Mean fund characteristics for retail and institutional funds (09/2015–02/2016)

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	(5)-(1)	Total
Relative net flow (%)	–0.36	–0.03	–0.35	–0.12	–0.61	–0.24	–0.24
Monthly return (%)	–1.27	–0.96	–0.90	–0.62	–0.47	0.69***	–0.84
Performance Rating	2.86	3.19	3.07	3.11	3.18	0.31***	3.09
12-month alpha (%)	–2.17	–1.03	–1.49	–1.59	–1.81	0.32**	–1.51
12-month volatility (%)	4.08	3.95	3.90	3.79	3.68	–0.40***	3.88
Total net assets ($m)	428.93	770.15	944.21	704.18	991.19	484.67***	794.51
Fund age in years	14.33	14.12	15.02	14.54	14.31	0.04	14.57
Turnover ratio (%)	56.33	62.51	57.95	57.67	42.00	–14.45***	57.44
Expense ratio (%)	1.27	1.17	1.15	1.19	1.11	–0.14***	1.17
Retail share classes (%)	58.10	49.36	50.19	53.13	57.61	–1.49	52.18
Number of observations	1568	3264	5099	3813	1189		14933
	(10.5%)	(21.86%)	(34.15%)	(25.53%)	(7.96%)		(100%)

(continued)

(continued)

Panel B: Mean fund characteristics for retail funds (09/2015–02/2016)

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	(5)-(1)	Total
Relative net flow (%)	–0.60	0.20	–0.30	0.01	–0.56	0.05	–0.17
Monthly return (%)	–1.28	–0.92	–0.94	–0.65	–0.50	0.71 ***	–0.86
Performance Rating	2.59	2.89	2.81	2.86	3.01	0.42 ***	2.83
12-month alpha (%)	–2.61	–1.24	–1.89	–1.93	–1.85	0.64 ***	–1.84
12-month volatility (%)	4.09	3.95	3.90	3.81	3.68	–0.41 ***	3.89
Total net assets ($m)	502.37	878.54	1108.81	815.45	1206.07	644.61 ***	922.57
Fund age in years	16.94	16.38	18.05	17.16	17.42	0.66	17.29
Turnover ratio (%)	58.57	63.03	57.62	56.55	41.17	–17.18 ***	57.12
Expense ratio (%)	1.43	1.40	1.36	1.42	1.27	–0.13 ***	1.38
Number of observations	911	1611	2559	2026	685		7792
	(11.69%)	(20.68%)	(32.84%)	(26%)	(8.79%)		(100%)

Panel C: Mean fund characteristics for institutional funds (09/2015–02/2016)

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	(5)-(1)	Total
Relative net flow (%)	–0.03	–0.24	–0.40	–0.27	–0.68	–0.65	–0.32
Monthly return (%)	–1.24	–0.99	–0.85	–0.59	–0.45	0.67 ***	–0.83
Performance Rating	3.24	3.47	3.33	3.40	3.40	0.15 ***	3.38
12-month alpha (%)	–1.55	–0.82	–1.08	–1.21	–1.74	–0.08	–1.14
12-month volatility (%)	4.07	3.95	3.90	3.77	3.68	–0.39 ***	3.88
Total net assets ($m)	327.09	664.52	778.39	578.03	699.15	301.77 ***	654.78
Fund age in years	10.71	11.92	11.96	11.56	10.09	–0.49	11.60
Turnover ratio (%)	53.23	62.01	58.29	58.94	43.13	–11.11 ***	57.78
Expense ratio (%)	1.03	0.94	0.93	0.94	0.89	–0.12 ***	0.94
Number of observations	657	1653	2540	1787	504		7141
	(9.20%)	(23.15%)	(35.57%)	(25.02%)	(7.06%)		(100%)

Panel D: Mean fund characteristics for retail and institutional funds (03/2016–03/2017)

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	(5)-(1)	Total
Relative net flow (%)	–1.24	–0.93	–0.77	–0.65	–0.33	0.93 ***	–0.78
Monthly return (%)	1.73	1.74	1.75	1.69	1.65	–0.07	1.72
Performance Rating	2.76	3.00	3.03	3.12	3.06	0.30 ***	3.02
12-month alpha (%)	–4.56	–3.47	–3.20	–2.63	–1.89	2.86 ***	–3.13
12-month volatility (%)	4.19	3.91	3.87	3.74	3.60	–0.59 ***	3.85
Total net assets ($m)	393.14	920.36	647.51	774.74	852.28	407.85 ***	733.48
Fund age in years	13.29	13.49	13.82	14.61	13.43	0.12	13.86
Turnover ratio (%)	65.58	61.20	60.13	56.22	46.06	–18.21 ***	58.67
Expense ratio (%)	1.25	1.16	1.18	1.19	1.16	–0.08 ***	1.18
Retail share classes (%)	55.52	49.07	50.38	52.42	55.06	0.44	51.52
Number of observations	3723	8476	12973	9639	3371		38182
	(9.75%)	(22.2%)	(33.98%)	(25.24%)	(8.83%)		(100%)

(continued)

(continued)

Panel E: Mean fund characteristics for retail funds (03/2016–03/2017)

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	(5)-(1)	Total
Relative net flow (%)	–1.57	–1.17	–1.13	–0.85	–0.29	1.28 ***	–1.03
Monthly return (%)	1.75	1.69	1.72	1.70	1.67	–0.08	1.71
Performance Rating	2.51	2.77	2.82	2.90	2.94	0.42 ***	2.81
12-month alpha (%)	–4.82	–3.73	–3.49	–2.99	–2.16	2.65 ***	–3.43
12-month volatility (%)	4.19	3.90	3.88	3.77	3.63	–0.57 ***	3.86
Total net assets ($m)	403.89	1084.35	691.14	970.32	1062.95	595.17 ***	850.89
Fund age in years	15.42	16.44	16.49	17.57	16.00	0.28	16.60
Turnover ratio (%)	64.26	62.45	60.77	55.36	45.24	–16.26 ***	58.64
Expense ratio (%)	1.45	1.40	1.40	1.41	1.34	–0.09 ***	1.40
Number of observations	2067	4159	6536	5053	1856		19671
	(10.51%)	(21.14%)	(33.23%)	(25.69%)	(9.44%)		(100%)

Panel F: Mean fund characteristics for institutional funds (03/2016–03/2017)

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	(5)-(1)	Total
Relative net flow (%)	–0.84	–0.70	–0.41	–0.44	–0.37	0.49 **	–0.52
Monthly return (%)	1.70	1.79	1.78	1.68	1.63	–0.05	1.74
Performance Rating	3.06	3.22	3.24	3.35	3.22	0.15 ***	3.24
12-month alpha (%)	–4.24	–3.22	–2.90	–2.23	–1.57	2.71 ***	–2.82
12-month volatility (%)	4.19	3.93	3.86	3.70	3.57	–0.63 ***	3.84
Total net assets ($m)	379.72	762.37	603.22	559.24	594.20	190.58 ***	608.71
Fund age in years	10.64	10.66	11.10	11.35	10.28	–0.11	10.95
Turnover ratio (%)	67.23	60.00	59.49	57.16	47.05	–20.41 ***	58.71
Expense ratio (%)	1.01	0.94	0.95	0.95	0.94	–0.07 ***	0.95
Number of observations	1656	4317	6437	4586	1515		18511
	(8.95%)	(23.32%)	(34.77%)	(24.77%)	(8.18%)		(100%)

the turnover and the expense ratio, both decreasing in the Sustainability Rating, revealing significant differences between funds of the top and bottom rating categories. Therefore, it is crucial to control for the potential influence of these disparities in characteristics on our results, and thus to disentangle the flow-sustainability relationship from other effects, to infer a marginal impact of the Sustainability Rating on flows. To this end, we proceed with three empirical methodologies: panel regressions, propensity score matching, and an event study.

1.3 Results

1.3.1 Panel regression

To study investors' reaction to the publication of Morningstar's Sustainability Rating we would ideally like to compare a group of funds for which the Sustainability Rating was made public to comparable funds with an identical but unpublished Sustainability Rating. Since we cannot observe comparable funds for which the rating was not published, we rely on a comparison of funds with different Sustainability Ratings. Instead of measuring the effect of the publication on single funds we measure the difference between the publication effects for funds with different rating classes. As a first approach, we use panel regressions to determine the impact of the Sustainability Ratings along with other control variables on fund flows. We regress the monthly net flow $F_{i,t}$ on the fund's Sustainability Rating. We treat the Morningstar Sustainability Rating with its five rating classes, 1 (low) to 5 (high), as a categorical variable since we do not expect the effect to be linear.

We select a wide range of control variables that have been found to influence mutual fund flows, in particular the 1-year alpha calculated from a Carhart (1997) model, the 1-month raw return and the categorical Morningstar Performance Rating. Using all three performance measures, we control for short-, mid- and long-term performance effects. We additionally control for the return volatility as a measure of risk, for fund size by including the logarithm of a fund's total net assets, for fund expenses estimated by the fund's net expense ratio and its turnover ratio, and for fund age.

Because we want to measure investors' reaction to the Sustainability Rating we use the most recent rating as of the beginning of the month and also lag all control variables, that is, we use fund characteristics measured at the end of the previous month. We additionally include month-style fixed effects to account for time-varying overall flows into and out of the mutual fund industry and for flows between different investment styles.

We want to examine whether investors react to information from Morningstar's Sustainability Rating. If so, we expect investors to buy funds with a high Sustainability Rating and sell funds with a low Rating. This effect, however, should not occur prior to the publication of the first Morningstar Sustainability Rating, or else the effect might just be due to an overall higher popularity of sustainable funds and not to the publication of the Rating. We therefore split our sample in two parts, one covering the months September 2015 to February 2016, the time prior to the public launch, and one covering March 2016 to March 2017, when the Rating was publicly available. We are able to conduct this analysis because we received unpublished Sustainability Ratings for the months prior to the public launch directly from Morningstar. Our results are displayed in Table 1.3. Whereas we do not find any significant relationship between a fund's unpublished Sustainability Rating during the months prior to the public launch (column 1), investors clearly react to the Rating after its publication (column 2). Low-rated funds receive a 0.23 percentage points per month lower net flow than average-rated funds, and high-rated funds a 0.29 percentage points per month higher net flow than average-rated funds. Those differences are statistically significant, and flows are monotonically increasing in the Sustainability Rating.

Investors are expected to react to the availability of the Sustainability Rating only if they consider ESG criteria during their asset allocation process and if the Rating reveals information they did not have access to before. We therefore expect different results for institutional and retail investors. Qualitative studies indicate that institutional US investors have a much lower interest in sustainable investments.[16] Even for those institutional investors who consider ESG criteria, the Sustainability Rating does not contain as much information as for retail investors. All the information that is used to calculate a fund's Sustainability Rating has been available to professional investors before. For example, holding data can be obtained from quarterly SEC filings and firms' ESG scores are available from data providers such as Bloomberg. We therefore split the data sample for the months after the Rating's public launch into a subsample of institutional funds and a subsample of retail funds and repeat ordinary least squares regressions for both. As reported in columns 3 and 4 of Table 1.3, we do not find any significant effect of the Sustainability Rating for institutional funds. The effect, on the other hand, is even stronger for retail funds. A low-rated (below-average-rated) retail fund receives a 0.33 percentage points (0.12 percentage points) per month lower net flow than an average-rated fund. A high-rated (above-average-rated) fund receives a 0.45 percentage points (0.11 percentage points) per month higher net flow. Given those significant results for retail funds only, we will focus all further analyses on retail funds. Our results suggest that retail investors gain information from the publication of Morningstar's Sustainability Rating and adapt their investments to that information. Given the median size of a retail share class

[16] The December/January 2016 issue of the *Morningstar* magazine provides an overview of existing studies.

Table 1.3: Relative net flow-sustainability regression

This table reports the results from ordinary least squares panel regressions of monthly fund flows on the Morningstar Sustainability Rating and other fund characteristics. Monthly net flows are defined, as in Sirri and Tufano (1998), as $[\mathrm{TNA}_t - (1 + R_t)\mathrm{TNA}_{t-1}]/\mathrm{TNA}_{t-1}$. The Sustainability Rating is as of the beginning of the month and all other fund characteristics are as of the end of the previous month. Monthly returns are obtained from the CRSP mutual fund database (MRET). The Carhart four-factor alpha and return volatility are calculated using monthly returns over the prior 12 months. The Performance Rating is the one- to five-star Morningstar Star Rating provided by Morningstar. Fund age is the number of years since the inception date (FIRST_OFFER_DT from CRSP). Total net assets, turnover ratios and expense ratios are also obtained from the CRSP mutual fund database (MTNA, TURN_RATIO and EXP_RATIO). The sample is constructed as described in Section 1.2 with single observations for each share class and month. We split the sample into multiple subsamples and each column refers to one of these subsamples: Column 1 reports results for all observations between September 2015 and February 2016, column 2 for all observations between March 2016 and March 2017. Column 3 refers to retail share classes during March 2016 to March 2017, column 4 to institutional share classes during that time (identified by the CRSP variables RETAIL_FUND and INST_FUND). Column 5 refers to funds that did not receive an initial Sustainability Rating in March 2016 but later and the sample contains all fund–month observations from March 2016 on and prior to the fund's initial Rating publication. Standard errors are double clustered on month and share class level and t-values are reported in parentheses. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	Dependent variable: Monthly relative net flow (%)				
	(1) Before launch (All)	**(2) After launch (All)**	**(3) After launch (Retail)**	**(4) After launch (Inst.)**	**(5) Unrated after launch (All)**
1-month lagged sustainability rating					
Low	–0.272	–0.231 *	–0.334 **	–0.134	0.665
	(–1.32)	(–1.94)	(–2.52)	(–0.63)	(1.52)
Below average	–0.039	–0.134	–0.118	–0.152	1.016 *
	(–0.30)	(–1.78)	(–1.30)	(–1.18)	(1.98)
Above average	0.113	0.031	0.108	–0.042	0.449
	(0.75)	(0.37)	(1.03)	(–0.27)	(1.26)
High	–0.100	0.293 **	0.445 ***	0.141	–0.060
	(–0.42)	(2.43)	(3.50)	(0.77)	(–0.06)
1-month lagged performance rating					
Low	–0.354	–0.365	–0.129	–1.228 ***	–1.331
	(–1.63)	(–1.42)	(–0.63)	(–3.15)	(–1.79)
Below average	–0.522 **	–0.604 ***	–0.462 ***	–0.926 ***	–1.149 **
	(–3.72)	(–4.43)	(–5.08)	(–3.71)	(–3.43)
Above average	1.071 ***	1.039 ***	1.030 ***	1.075 ***	1.834 ***
	(6.83)	(10.31)	(6.97)	(8.12)	(4.27)
High	3.265 ***	3.213 ***	3.348 ***	3.174 ***	3.175 ***
	(6.77)	(11.14)	(7.61)	(9.26)	(4.82)

(continued)

(continued)

	Dependent variable: Monthly relative net flow (%)				
	(1) Before launch (All)	**(2) After launch (All)**	**(3) After launch (Retail)**	**(4) After launch (Inst.)**	**(5) Unrated after launch (All)**
1-month lagged 12-month alpha (%)	0.240 ***	0.156 ***	0.151 ***	0.161 ***	0.179 ***
	(6.70)	(5.50)	(5.64)	(4.54)	(4.60)
1-month lagged monthly return (%)	–0.071	0.074	0.146 ***	–0.029	–0.185 **
	(–1.78)	(1.57)	(3.52)	(–0.48)	(–3.76)
1-month lagged 12-month return volatility (%)	–0.346 *	0.158	0.229	–0.010	0.418
	(–2.24)	(0.99)	(0.84)	(1.02)	(1.29)
1-month lagged turnover ratio (%)	–0.000	–0.001	–0.000	–0.002	0.007
	(–0.12)	(–1.02)	(–0.23)	(–1.20)	(1.30)
1-month lagged expense ratio (%)	0.581	–0.378 **	–0.176	–1.105 ***	–0.171
	(1.90)	(–2.18)	(–1.34)	(–3.66)	(–0.27)
Log 1-month lagged tna ($m)	–0.203 **	–0.250 ***	–0.240 ***	–0.303 ***	–0.197
	(–3.01)	(–8.59)	(–6.02)	(–6.75)	(–1.75)
Fund age in years	–0.002	–0.003	0.006	–0.036 ***	–0.026
	(–0.38)	(–1.08)	(1.34)	(–3.22)	(–0.72)
Month-style fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	14945	38202	19676	18526	2418
Adjusted R2	0.08	0.07	0.09	0.07	0.09

of $89.4 million, an average fund receives $4.83 million more inflows over one year if it obtains a high instead of an average Sustainability Rating. On the other hand, it receives $3.54 million lower net flows if it is rated low instead of average.

We are not aware of any confounding events that might have caused an increased interest in sustainable investments among retail investors during the period from March 2016 to March 2017. The underperformance of the Dow Jones Sustainability World Index versus the Dow Jones Industrial Average Index during those months does not support such an alternative explanation. To further increase the credibility of the causal relationship between Morningstar's Sustainability Rating and fund flows we repeat the panel regression for a smaller subsample of mainly US Small Cap funds which did not receive a Sustainability Rating in March 2016, but from a later point in time on. We only include observations after March 2016, but prior to the month when a fund received its first Rating, and we further assume that the fund would have had its initial rating during previous months. Using this sample, we do not find any relationship that would support the alternative explanation of an overall increased interest in sustainable investments (column 5).

Altogether, panel regressions strongly suggest that retail investors react to the launch of the Morningstar Sustainability Rating by adjusting their investment decisions and investing in higher-rated funds.

1.3.2 Matching

As we compare the impact of published Sustainability Ratings on fund flows across rating categories we need to consider that the Sustainability Rating is not independent of other fund characteristics such as size or past performance and that those differences might cause differences in fund flows. Adding a series of control variables to panel regressions corrects for linear relationship between fund characteristics and fund flows. To account for differences in fund characteristics, for which a linear relationship (or log-linear in the case of size) might not be appropriate, we additionally apply nearest-neighbor matching. In particular, Chevalier and Ellison (1997) and Sirri and Tufano (1998) have shown that the relationship between past performance and fund flows is convex, and Ivković and Weisbenner (2009) show that relative past performance induces inflows but not outflows.

We would like to compare two groups of funds with very similar fund characteristics but different Sustainability Ratings and therefore use propensity score matching to construct such groups. To ensure that we compare funds that received their initial rating during the same month and because the majority of funds received a Sustainability Rating during its launch in March 2016 we restrict our sample to those funds for the matching. Because we expect the effect of the Sustainability Rating to occur only for retail funds we drop all other funds from the sample. We match funds based on their initial rating (i.e. the Rating as of March 17, 2016) and on fund characteristics at the end of February 2016 and keep the matched sample constant over time. Our matching procedure consists of three steps. First, we define the treatment group and the population of funds from which we will construct the control group. We examine three settings: funds with a high Sustainability Rating versus a matched control sample of funds with an average Rating (“high versus average”); funds with a low versus a matched control sample of funds with an average Rating (“low versus average”); and funds with a high Rating versus a matched control sample of funds with a low Rating (“high versus low”).

In a second step, we run a logit regression on all funds in our treatment group and on those funds the control group is selected from (e.g. all funds with a high or average rating for the high versus average setting) to estimate propensity scores. The logit model regresses dummy variables indicating whether a fund is in the treatment group or not on the 1-year alpha from a Carhart (1997) model, the fund's Morningstar Performance Rating, its size (log of total net assets), age, expense ratio and turnover ratio. Propensity scores are the fitted values from this model and can be interpreted as the probability of being a fund of our treatment group. Unreported results show that funds are more likely to have a better Sustainability Rating if they have a higher past alpha and a better Performance Rating. Funds with a high Sustainability Rating are smaller and have lower expense and turnover rations than funds with an average rating but those three characteristics are not significant for the logit models in the "low versus average" and "high versus low" settings.

In a third step, we match each fund of the treatment group to the fund in the group of potential control funds with the closest propensity score. If the propensity scores of the matched pair differ by more than 0.025 we drop the pair from our sample. We allow the same control fund to be matched multiple times. In doing so, we end up with the treatment group and a matched control group with similar fund characteristics. Table 1.4 shows that there is no significant difference with respect to most matching variables between funds in the treatment and the control group which indicates that our match is of high quality. In the "high versus average" setting we observe a significant difference only in the median fund age. For the "low versus average" no mean or median matching variable differs between the treatment and the control group. There are, however, significant differences between the turnover and the expense ratios in the "high versus low" setting. We know from panel regressions that the expense ratio has a negative impact on fund flows, and since the expense ratio is higher for the funds in our treatment group ("high"), this difference will, if at all, induce a downward bias in our results. Since turnover ratios are not found to influence fund flows significantly, neither difference harms the quality of our matching.

Given this matching, we observe the difference in fund flows during the 6 months prior and the 12 months subsequent to the launch of the Morningstar Sustainability Rating. Table 1.5 displays the results. During the 3- and 6-month period prior to the launch of the Morningstar Sustainability Rating, that is, December 2015 to February 2016 and September 2015 to February 2016, there is no significant difference in fund flows between our treatment and the

Table 1.4: Propensity score matching−sample comparison

This table reports the mean and median values of all variables included in the propensity score matching for the treatment and matched control groups. Only retail share classes with a published Sustainability Rating in March 2016 are considered for the matching. Funds that cannot be matched are excluded from the sample. All fund characteristics are as of end of February 2016. Panel A shows the values for the "high versus average" setting, where the treatment group consists of high-rated share classes according to the Morningstar Sustainability Rating in March 2016 and the control group is selected from average-rated funds. Panel B reports values for the "low versus average" setting, with low-rated funds forming the treatment group and the control group being selected from average-rated funds. Panel C shows characteristics for the high versus low setting, where high-rated funds form the treatment group and the control group is selected from low-rated funds. The Carhart four-factor alpha is calculated using monthly returns (MRET from CRSP) over the prior 12 months. The Performance Rating is the one- to five-star Morningstar Star Rating provided by Morningstar. Fund age is the number of years since the inception date (FIRST_OFFER_DT from CRSP). Total net assets, turnover ratios and expense ratios are also obtained from the CRSP mutual fund database (MTNA, TURN_ RATIO and EXP_RATIO). The last two columns report the p-values from a t-test and a Mood's median test to test for differences between the mean and median values in the treatment and the control group. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	Treatment		**Control**		**Diff. test *p*-value**	
	Mean	**Median**	**Mean**	**Median**	**Mean**	**Median**
Panel A: Matched sample comparison-model: high vs. average						
12-month alpha (%)	–1.57	–1.36	–1.92	–1.68	0.41	0.49
Performance Rating	3.01	3.00	2.90	3.00	0.45	0.49
Log total net assets ($m)	4.67	4.68	4.78	4.34	0.71	0.68
Fund age in years	16.63	13.01	17.97	15.01	0.46	0.04 **
Expense ratio (%)	1.30	1.23	1.26	1.22	0.47	0.89
Turnover ratio (%)	48.29	35.00	43.21	40.00	0.28	0.68
Panel B: Matched sample comparison-model: low vs. average						
12-month alpha (%)	–3.52	–2.93	–3.46	–3.24	0.89	0.29
Performance Rating	2.47	3.00	2.49	2.00	0.83	0.41
Log total net assets ($m)	4.43	4.29	4.72	4.59	0.21	0.20
Fund age in years	16.55	14.68	16.28	14.97	0.85	0.56
Expense ratio (%)	1.41	1.32	1.46	1.45	0.36	0.29
Turnover ratio (%)	56.28	52.00	51.47	49.00	0.27	0.56
Panel C: Matched sample comparison-model: high vs. low						
12-month alpha (%)	–2.10	–1.91	–1.80	–1.97	0.48	0.88
Performance Rating	2.81	3.00	2.75	3.00	0.65	0.18
Log total net assets ($m)	4.39	4.17	4.61	4.13	0.48	0.88
Fund age in years	16.75	12.84	16.24	14.61	0.78	0.47
Expense ratio (%)	1.34	1.24	1.25	1.12	0.10 *	0.01 ***
Turnover ratio (%)	50.25	39.00	55.25	56.00	0.32	0.00 ***

Table 1.5: Propensity score matching−treatment effects

This table presents the average treatment effect on the treated from a nearest-neighbor propensity score matching. Funds are matched based on their characteristics (12-month alpha, Performance Rating, size, fund age, expense ratio and turnover ratio) as of the end of February 2016 and differences in net flows between the treatment and the control group are measured. The results are reported for three settings: For "high versus average" and "low versus average", the treatment group consists of all high- or low-rated funds according to the Morningstar Sustainability Rating in March 2016, while the control group is selected from all average-rated funds. For "high versus low", the treatment group consists of all high-rated funds and the control group is selected from the low-rated funds. Only retail share classes with a Sustainability Rating published in March 2016 are included in the sample. The differences in net flows are calculated over six different time intervals: 6 months prior (−6 to 0), and 3 months prior to March 2016 (−3 to 0) as well as for the 3, 6, 9, and 12 months starting from March 2016 (0 to 3, …, 0 to 12). Net flows are defined, as in Sirri and Tufano (1998), as $[\mathrm{TNA}_t - (1 + R_t)\mathrm{TNA}_{t-1}]/\mathrm{TNA}_{t-1}$. Differences in flows are measured in percent. t-statistics are reported in parentheses using robust Abadie and Imbens (2016) standard errors. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

		Months					
	N	**–6 to 0**	**–3 to 0**	**0 to 3**	**0 to 6**	**0 to 9**	**0 to 12**
High vs. average	105	–1.35	1.91	3.01 *	6.50 ***	8.16 ***	9.41 **
		(–0.48)	(1.03)	(1.87)	(2.97)	(2.89)	(2.32)
Low vs. average	145	–1.86	–0.49	–2.45 **	–2.92 *	–4.81 **	–6.80 **
		(–1.29)	(–0.56)	(–2.38)	(–1.72)	(–2.37)	(–2.34)
High vs. low	95	1.31	1.36	3.45 **	4.89 **	10.00 ***	12.04 ***
		(0.59)	(1.06)	(2.26)	(2.33)	(2.89)	(2.73)

matched control group. Mutual funds with a high Sustainability Rating get neither significantly higher nor significantly lower net flows than average-rated or low-rated funds during that time. Low-rated funds are not subject to significantly lower flows than average-rated funds. The impact of the Sustainability Rating, however, rises after its publication.

During the months March to May 2016, high-rated funds receive a 3.01 percentage points higher net flow than average-rated funds and 3.45 percentage points higher net flows than low-rated funds. Low-rated funds have a 2.45 percentage points lower net flow than average-rated funds during that time period.[17] The effect of higher net flows into high-rated funds and out of low-rated funds continues during the subsequent months and totals 9.41 percentage

[17] We should not expect the high versus low effect to equal the sum of the high versus average and low versus average effects since the fund characteristics of the treatment group in the low versus average setting differ significantly from the treatment groups of the other settings and since we are measuring an average effect on the treated.

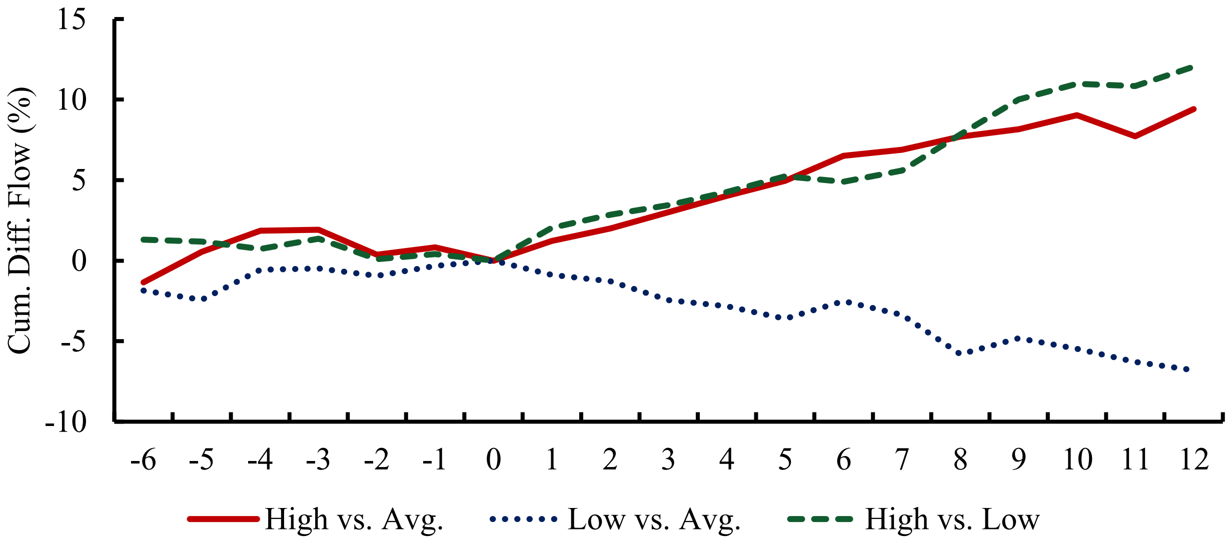
points during the first 12 months for high-rated versus average-rated funds, and 12.04 percentage points for high-rated versus low-rated funds. Low-rated funds receive 6.80 percentage points lower net flows than average-rated funds during the first year after the launch of Morningstar's Sustainability Rating. Figure 1.2 1.2 visualizes these results, displaying the cumulative difference in flows during the months prior to and after the launch. Differences prior to the launch are close to zero and increase or decrease in line with the reported values. The almost monotonically increasing or decreasing cumulative difference in flows between March 2016 and February 2017 indicates that funds with a higher (lower) Sustainability Rating received higher (lower) net flows during almost every month subsequent to the publication of the Rating.

Since the median size of a retail share class in our matching with a high (low) Sustainability Rating is \$108 million (\$65 million), the effect of a 9.41 percentage point (6.80 percentage point) flow difference should be interpreted as follows. A high-rated fund according to Morningstar's Sustainability Rating receives on average \$10.1 million higher net inflows during March 2016 to February 2017 than would have been expected if it had received an average Sustainability Rating. In contrast, an average low-rated fund receives about \$2.1 million lower net flows than expected if it were given an average Sustainability Rating, thus indicating that the effect of the Morningstar Sustainability Rating is both statistically and economically significant. The effect can only be observed after the initial Rating publication, indicating that retail investors react to the new information.

As a robustness test, we alternate the matching. We select alternative matching variables, especially different measures of past performance, and do not consider the expense ratio and the turnover ratio. We also change the distance thresholds from 0.025 to values between 0.01 and 0.03. This does not qualitatively change our results. We also look at a setting where funds with a low Rating are our treatment group and funds with a high Rating are matched ("low versus high"). Results confirm our finds from "high versus low". We repeat the matching procedure for the sample of institutional funds, but in line with our predictions and the results from panel regressions we do not find significant differences in flows between high- and average-rated and between high- and low-rated funds subsequent to the launch of Morningstar's Sustainability Rating (unreported results). We find weakly significant evidence for a rating effect when matching low- to average-rated funds.

Figure 1.2: Propensity score matching−cumulative differences in flows

This figure visualizes the average treatment effect on the treated from a nearest-neighbor propensity score matching. It shows the cumulative flow differences between the treatment and the control group for the 6 months prior to and the 12 months after the launch of the Morningstar Sustainability Rating. Funds are matched based on their characteristics (12-month alpha, Performance Rating, size, fund age, expense ratio and turnover ratio) as of the end of February 2016 and differences in net flows between the treatment and the control group are measured. The results are displayed for three settings: for "high versus average" and "low versus average", the treatment group consists of all high- or low-rated funds according to the Morningstar Sustainability Rating in March 2016, while the control group is selected from all average-rated funds; for "high versus low", the treatment group consists of all high-rated funds and the control group is selected from the low-rated funds. Only retail share classes with a Sustainability Rating published in March 2016 are included in the sample. Net flows are defined, as in Sirri and Tufano (1998), as $[\text{TNA}_t - (1 + R_t)\text{TNA}_{t-1}]/\text{TNA}_{t-1}$. We define $t = 0$ as the beginning of the month the Sustainability Rating was launched in (i.e. March 1, 2016). For the time period prior to March 2016 (i.e. $t < 0$), cumulative differences in flows are calculated as the sum of the monthly differences in average flows of the treatment and the control group during the months t to 0 (e.g. $t = -3$ shows the sum of flow differences for the months from December 2015 to February 2016). For the time period after the launch (i.e. $t > 0$), cumulative differences in flows are calculated as the sum of the monthly differences in average flows of the treatment and the control group during the months 0 to t (e.g. $t = 3$ shows the sum of flow differences for the months from March to May 2016).



1.3.3 Event study

The panel regression and propensity score matching approaches both allow insights into the flow-sustainability relationship by comparing the flows between funds in different Sustainability Rating categories. In contrast, the following event study approach provides an opportunity to directly examine the investors' flow response to the launch of the Sustainability Rating on a fund-by-fund basis. Specifically, each fund flow following the introduction of the

Rating is benchmarked to a model of its expected flow to obtain an estimate of the abnormal flow triggered by the publication of Morningstar's Sustainability Rating. This subsection describes the details of how the event study approach is implemented and the diagnostic tests employed to assess the reliability of the benchmark model.

1.3.3.1 Methodology

To evaluate the marginal effect of the Sustainability Rating on fund flows, the benchmark model essentially needs to disentangle the independent "sustainability effect" from other performance and non-performance characteristics, potentially influencing fund flows. Therefore, the inherent principles of the method employed follow a traditional event study approach as described by Campbell et al. (1997) and adjust flows for influences other than the Sustainability Rating.

To estimate a time-series benchmark regression for each individual fund i, we choose our model in accordance with the discussions of Del Guercio and Tkac (2008), who employ an event study approach to analyze the independent effect of Morningstar's Star Rating on fund flows. Thus, given the model's viability,[18] we compute the expected flows by estimating the following time-series benchmark regression for each individual fund i:

$$F_t^i = \gamma^i + \beta_1^i \mathrm{SPCF}_t^i + \beta_2^i F_{t-1}^i + \beta_3^i R_{t-1}^i + \beta_4^i \Delta\alpha_{t-1}^i + \beta_5^i (\Delta\alpha_{t-1}^i)^2 + e_t^i, \qquad (2)$$

where F_t^i is the relative net flow to fund i at time t, F_{t-1}^i is the lagged relative net flow to fund i at time $t-1$, and $SPCF_t^i$ is the average relative net flow at time t to all funds in the same style and same lagged performance category as fund i. R_{t-1}^i is the monthly net return of fund i observed in the preceding month. $\Delta\alpha_{t-1}^i$ represents the changes in fund i's Carhart (1997) four-factor alpha from time $t-2$ to $t-1$ calculated over a 12-month period using monthly data of the Carhart four-factor model premiums. Thus, our benchmark regression model includes variables found to be important predictors of fund flows in the mutual fund literature.

---

[18] Analyses of the model's fit in Del Guercio and Tkac (2008) indicate a strong case for the model's reliability. Moreover, analyzing the benchmark model by means of observed relative net flows and our normal flow estimates grouped by the funds' Performance Rating, Figure 1.B1 in Appendix 1.B shows the model's capability of replicating the convex flow-performance relationship. Table 1.A1 in Appendix 1.A provides further evidence that the model successfully adjusts fund flows for influences different from the Sustainability Rating by assessing the statistical significance of average abnormal fund flows for the 6 months after initial rating grouped by the funds' respective Performance Rating.

To account for potential multicollinearity issues due to the high correlation among the factors used in the time-series regressions, we use the change in alpha, instead of the absolute measure, as suggested by Del Guercio and Tkac (2008). Following their approach further, we additionally include a squared term of the change in a fund's alpha to control for the convex flow-performance relationship. Different from the model in Del Guercio and Tkac (2008), we choose a proxy for the average relative net flow in the same style and lagged Performance Rating category, to account not only for the overall attractiveness of a fund's investment style, but also for marginal effects of the Performance Rating.

For the event study, we define the beginning of the month during which the Sustainability Rating was made accessible to retail investors as the event date (t = 0). To estimate the coefficients of the benchmark model, the estimation window was chosen to have a length of 24 months[19], and was set to capture the months directly preceding the event month, i.e. t = [−24; 0]. Applying the estimated benchmark model parameters to the event window, we measure the abnormal flows for a given fund i in each month around the time of the event:

$$\mathrm{AF}_t^i = F_t^i - \gamma^i - \beta_1^i \mathrm{SPCF}_t^i - \beta_2^i F_{t-1}^i - \beta_3^i R_{t-1}^i - \beta_4^i \Delta\alpha_{t-1}^i - \beta_5^i \left(\Delta\alpha_{t-1}^i\right)^2. \qquad (3)$$

Thus, AF_t^i represents the relative abnormal net fund flow at time t to the individual fund i, which is equal to the actual observed relative net fund flow F_t^i minus the expected relative net fund flow, the latter being calculated by the style-performance-category flow, the lagged fund flow, the lagged raw return and the lagged change in alpha as well as the square of the lagged change in alpha. The term γ^i represents the fund-specific average abnormal flow, which is expected to cover predictors that are roughly constant over time, such as the fund age, turnover or expenses, and are therefore not included in a fund-wise time-series regression.

To identify funds' flow response to the introduction of the Sustainability Rating, we group funds according to their corresponding rating and report abnormal flow statistics over the event window t = [0; 6] in the main test of this section. For the subsequent analysis, this paper applies the event study methods described in Dodd and Warner (1983) and Patell (1976) to the abnormal flow estimates. Particularly, the relative abnormal flow estimates for each fund

[19] Analyses of the models fit and the benchmark residuals by means of RMSE suggest that an estimation window length of 24 months provides the highest level of in-sample precision. However, varying the window lengths does not change the outcomes qualitatively.

i at time t within the event window are standardized by the square root of their estimated forecast variance (root mean square error, RMSE) and an additional forecasting correction term to form the standardized relative abnormal flow SAF_t^i. This standardization procedure leads to a different weighting scheme in the observations. Thus, by standardizing the relative abnormal flow, funds showing higher precision in their estimates are implicitly assigned more weight in the calculation of the average relative abnormal flow across funds in each event month. To assess the temporal perspective of the event, the cumulative standardized relative abnormal flows $CSAF_t^i$ are calculated for each fund i by summing the standardized relative abnormal flows of each fund over the specified event window. By forming the average across funds for each Sustainability Rating category, the average standardized relative abnormal fund flow $\overline{\mathrm{SAF}_t}$ and the average cumulative standardized relative abnormal fund flow $\overline{CSAF_t}$ are obtained.

To assess the statistical significance of the $\overline{SAF_t}$ and $\overline{CSAF_t}$, we employ the cross-sectional test introduced by Boehmer et al. (1991), which divides standardized abnormal flows and cumulative standardized abnormal flows by the contemporaneous standard errors. To account for a potential change in the variance of the standardized abnormal flow in the event window relative to the estimation window period, this approach calculates the standard errors from the event-period abnormal flows. If, for example, the abnormal flows are exposed to higher volatility in the event window, the standard errors used by Boehmer et al. (1991) lead to reduced t-statistics. In turn, this approach mitigates the risk of biased inferences. The abnormal flow statistics reported in this section also contain test statistics for the nonparametric sign test as in Campbell et al. (1997).

1.3.3.2 Results

Tables 1.6 and 1.7 contain the average standardized relative abnormal flow $\overline{SAF_t}$ and the average cumulative standardized relative abnormal flow $\overline{CSAF_t}$, *r*espectively, for all Sustainability Rating categories for each event month from 1 to 6. Given the observed unequal reaction of different investor clienteles to the Rating, we focus on the abnormal flow response on all retail funds obtaining a Sustainability Rating.

Table 1.6 reports the results for the average standardized relative abnormal fund flow $\overline{SAF_t}$. For the highest sustainability category, $\overline{SAF_t}$ is significantly positive for the majority of months. Thus, given the relative consistency in the significance level throughout the event window, the results indicate that some retail investors immediately respond to the publication of the highest Sustainability Rating, whereas others respond with a significant lag. In contrast, funds in the lowest sustainability category have a significantly negative $\overline{SAF_t}$ for most event months. Standardized abnormal flows to the lowest sustainability category are nonsignificantly positive only in event month 4. For funds being assigned an intermediate Sustainability Rating (below average, average, above average), we cannot derive a statistically significant effect from the outcome presented in Table 1.6. We observe a retail investors' reaction to both the highest and lowest Sustainability Rating. The only partly significant $\overline{SAF_t}$ for the lowest rating category, however, suggests that investors react more strongly to the highest Sustainability Rating.

The results are confirmed by the average cumulative standardized relative abnormal flows $\overline{CSAF_t}$ reported in Table 1.7. Both the $\overline{CSAF_t}$ to the highest and to the lowest Sustainability Rating are highly significant, with twice the cumulative effect for the highest Rating category. By employing a nonparametric sign test, the results of the average cumulative standardized relative abnormal fund flow $\overline{CSAF_t}$ confirm the pattern of abnormal flow directions by yielding significant test statistics for the clear majority of the observations. For an additional verification of this result, a difference-in-means test is applied to assess the statistical difference between the low and high end of the sustainability spectrum. Indeed, the corresponding test statistics reported for $\overline{SAF_t}$ and $\overline{CSAF_t}$ provide strong evidence for a significant difference in the flow responses and therefore provide further indication of investors' reaction to the Sustainability Rating.

Overall, the results on the sustainability-flow relationship paint an intriguing picture of how retail investors respond to different rating categories. Given the significantly stronger reaction to superior ratings when compared to bottom-end ratings, the data provide supportive evidence that retail investors flock to funds with the highest Sustainability Rating while not fleeing from lower-rated funds at the same rate. Thus, the results indicate that the well-documented convex relationship between past performance and net flow also holds for the sustainability-flow relationship. Similar to the findings of Chevalier and Ellison (1997) and Sirri and Tufano (1998), showing that poorly performing funds only suffer modest losses and top performers accumulate large inflows, we find a high standardized relative abnormal flow for the most sustainable funds.

Table 1.6: Standardized relative abnormal flow after the launch of the Sustainability Rating

Panel A reports the average standardized relative abnormal fund flows $\overline{SAF_t}$ grouped according to the corresponding Sustainability Rating for the 6 months after a fund was initially rated (t = [0; 6]). We define the standardized relative abnormal fund flow at time t as the actual relative net flow minus the expected flow standardized by the forecast correct standard error, as described in Dodd and Warner (1983). To calculate the expected fund flow, we estimate the coefficients of the subsequent benchmark model for each fund individually over an estimation window of 24 months (t = [−24; 0]). Specifically, we regress the fund's monthly relative flow on the average relative flow at time t to funds in the same style and lagged Performance Rating category, its time $t-1$ flow, its time $t-1$ raw return, its change in the Carhart four-factor alpha from $t-2$ to $t-1$, and its change in alpha from $t-2$ to $t-1$ squared. Further, Panel A reports nonparametric sign tests under the null hypothesis that it is equally probable that sample funds have positive or negative standardized abnormal flows, and a difference-in-means test, showing the discrepancy in the average standardized abnormal flows between high and low Sustainability Ratings. Panel B reports the t-statistics of the average standardized relative abnormal flows grouped by the assigned Sustainability Rating. To account for a potential shift in the variance of the $\overline{\mathrm{SAF}_t}$ in the event window relative to the estimation window, the standard errors of the $\overline{\mathrm{SAF}_t}$ are calculated from the event window, as in Boehmer et al. (1991). *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

Event month	Low		Below Avg.		Average		Above Avg.		High		High-Low
	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$
Panel A: Average standardized relative abnormal flow by Sustainability Rating and event month (coefficients)											
1	–0.041	45.65	0.014	48.96	0.067	50.45	0.013	50.24	0.287***	59.58**	0.329***
2	–0.207***	37.50***	–0.090*	48.52	–0.024	48.89	0.090**	52.41	0.232***	56.84*	0.440***
3	–0.185***	41.30**	–0.043	44.70*	0.042	47.05	0.076*	51.58	0.090	54.10	0.275***
4	0.070	58.69**	0.033	51.78	0.080*	52.47	–0.063	47.18	0.252***	61.80***	0.182*
5	–0.105*	47.28	–0.096**	42.55***	–0.049	49.17	0.105**	52.65	0.354***	61.11***	0.459***
6	–0.208***	42.39**	–0.137***	43.32**	–0.001	47.88	–0.129***	44.55**	0.320***	66.20***	0.529***
Panel B: Average standardized relative abnormal flow by Sustainability Rating and event month (t-statistics)											
1	–0.62	–1.17	0.34	–0.38	1.06	0.21	0.30	0.09	3.14***	2.31**	2.66***
2	–3.25***	–3.39***	–1.92*	–0.54	–0.73	–0.51	2.24**	0.98	2.71***	1.65*	3.79***
3	–2.91***	–2.35**	–1.12	–1.95*	0.56	–1.37	1.79*	0.64	1.21	0.99	2.59***
4	1.37	2.35**	0.71	0.65	1.65*	1.15	–1.37	–1.13	3.60***	2.83***	1.92*
5	–1.64*	–0.73	–2.09**	–2.72***	–1.40	–0.38	2.01**	1.08	4.54***	2.66***	4.18***
6	–3.57***	–2.06**	–3.35***	–2.45**	–0.02	–0.98	–3.13***	–2.21**	4.15***	3.90***	5.02***

Table 1.7: Cumulative standardized relative abnormal flow after the launch of the Sustainability Rating

Panel A reports the average cumulative standardized relative abnormal fund flows $\overline{CSAF_t}$ grouped according to the corresponding Sustainability Rating for the 6 months after a fund was initially rated (t = [0; 6]). We compute the cumulative standardized relative abnormal flow for each fund by summing the standardized relative abnormal flow from event month 1 to 6. We then form the average of the $\overline{CSAF_t}$ across all funds for each Sustainability Rating category. Further, Panel A reports nonparametric sign tests under the null hypothesis that it is equally probable that sample funds have positive or negative cumulative standardized abnormal flows, and a difference-in-means test, showing the discrepancy in the average cumulative standardized relative abnormal flows between high and low Sustainability Ratings. Panel B reports the t-statistics of the average cumulative standardized abnormal flows grouped by the assigned Sustainability Rating. To account for a potential shift in the variance of the $\overline{SAF_t}$ in the event window relative to the estimation window, the standard errors of the $\overline{SAF_t}$ are calculated from the event window, as in Boehmer et al. (1991). *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

	Low		Below Avg.		Average		Above Avg.		High		High-Low
Event month	$\overline{CSAF_t}$	% > 0	$\overline{CSAF_t}$	% > 0	$\overline{CSAF_t}$	% > 0	$\overline{CSAF_t}$	% > 0	$\overline{CSAF_t}$	% > 0	$\overline{CSAF_t}$
Panel A: Average cumulative standardized relative abnormal flow by Sustainability Rating and event month (coefficients)											
1	–0.041	45.65	0.014	48.96	0.067	50.45	0.013	50.24	0.287***	59.58**	0.329***
2	–0.240**	41.84**	–0.090	45.29*	0.062	50.00	0.114*	50.48	0.511***	57.53*	0.751***
3	–0.388***	39.13***	–0.125	45.88	0.098	48.16	0.226***	55.74**	0.559***	59.58**	0.948***
4	–0.344**	42.39**	–0.084	47.02	0.113	49.17	0.130	52.07	0.871***	58.33**	1.216***
5	–0.382**	42.39**	–0.181	45.53	0.187	48.25	0.211	50.72	1.278***	61.80***	1.660***
6	–0.620***	40.21***	–0.298*	43.91**	0.166	46.42	* 0.101	49.63	1.484***	62.06***	2.105***
Panel B: Average cumulative standardized relative abnormal flow by Sustainability Rating and event month (t-statistics)											
1	–0.62	–1.17	0.34	–0.38	1.06	0.21	0.30	0.09	3.14***	2.31**	2.66***
2	–2.35**	–2.21**	–1.28	–1.73*	0.81	0.00	1.88*	0.19	3.72***	1.82*	4.04***
3	–2.89***	–2.94***	–1.40	–1.51	0.84	–0.85	2.79***	2.32**	3.37***	2.31**	4.09***
4	–2.21**	–2.06**	–0.71	–1.09	0.97	–0.38	1.23	0.84	4.33***	2.00**	4.39***
5	–1.96**	–2.06**	–1.31	–1.63	1.26	–0.81	1.58	0.29	5.42***	2.83***	4.98***
6	–2.75***	–2.65***	–1.85*	–2.23**	0.98	–1.67	* 0.64	–0.14	5.24***	2.90***	5.35***

The crucial prerequisite for the interpretation of the "sustainability effect", however, is the benchmark model's capability to successfully isolate the effect of the Sustainability Rating on fund flows. If the average abnormal flow is triggered by the introduction of the Sustainability Rating, we do not expect to observe any significant abnormal flow in months preceding the event date. To implement a corresponding test, the event window and the estimation window are shifted by 6 months, that is, the event window then spans the period of 6 months prior to the initially defined event month (t = [−6; 0]) and the estimation window captures the 24 months preceding this period (t = [−30; −6]). To group the funds by their corresponding Sustainability Rating we use unpublished Sustainability Ratings obtained directly from Morningstar (going back to November 2015) and extrapolated values for September and October 2015. For this test, we group the funds according to their monthly Sustainability Rating and calculate the average standardized relative abnormal flow $\overline{SAF_t}$, reported in Table 1.8. Generally, we are not able to derive a consistent pattern from the average standardized relative abnormal flows $\overline{SAF_t}$ in the event window t = [−6; 0] and observe statistically nonsignificant results for the overwhelming majority of event months and rating categories. Moreover, the difference-in-means test for the outcomes of the highest and lowest Sustainability Rating shows insignificant results with reversed signs for the t-statistics, compared to the results following the launch of the Sustainability Rating. Overall, the results reported in Table 1.8 indicate that the "sustainability effect" is not driven by a general attractiveness. Thus, we conclude that the effect is due to the launch of Morningstar's Sustainability Rating.

Finally, to allow for an economic interpretation, we repeat the event study, but without the standardization procedure. By doing so, we find that the positive abnormal flow response identified for funds in the highest rating category sums to a total of 1.83 percentage points during the first 6 months after the publication of the Sustainability Rating. In contrast, low-rated funds receive 1.01 percentage points lower net flows than they should have expected without the attainment of a rating during the first two quarters after the launch of the Rating. Figure 1.3 visualizes these results by presenting the average cumulative abnormal flows for all rating categories for a 6-month period prior and subsequent to the launch of the

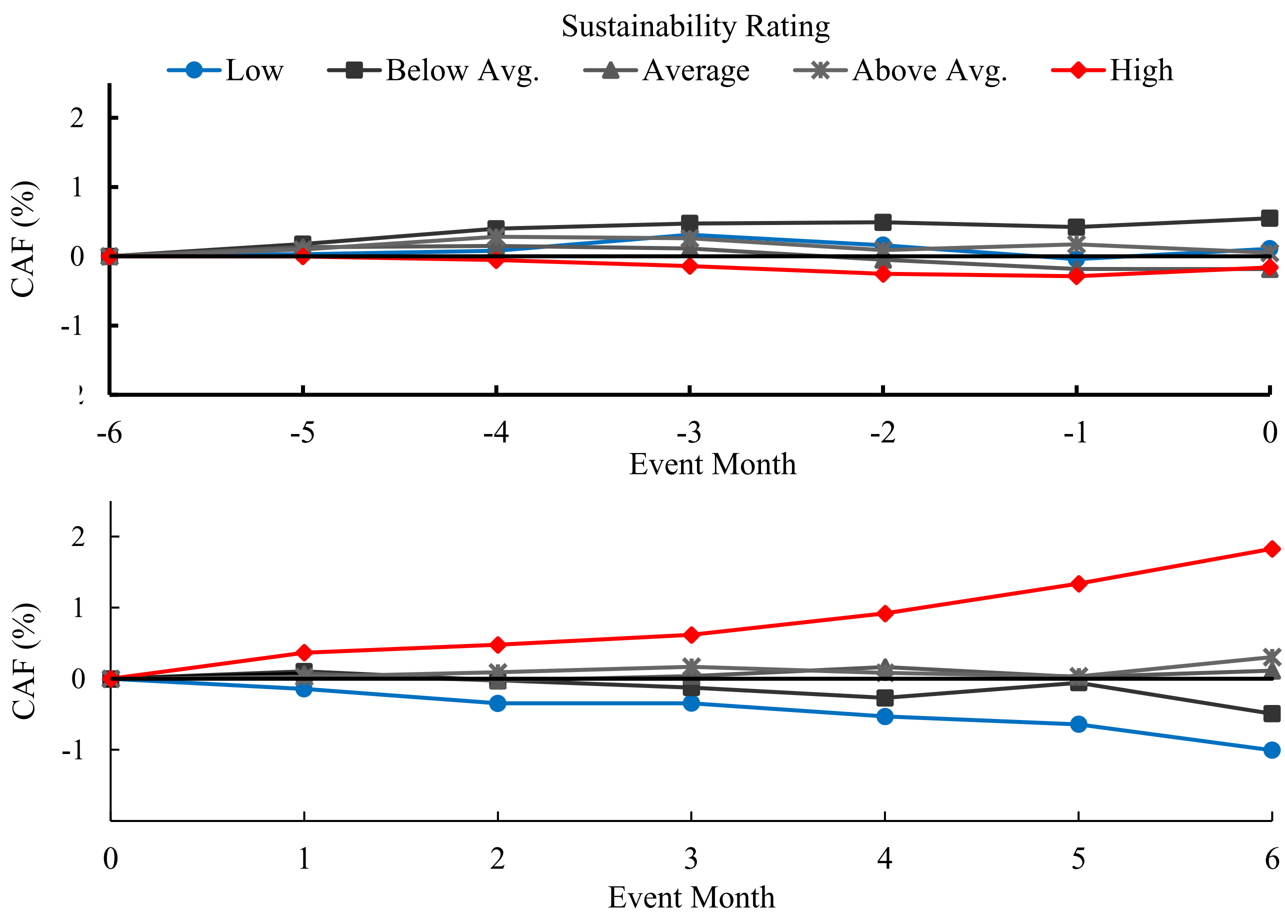
Table 1.8: Placebo event study−standardized relative abnormal flow before the launch of the Sustainability Rating

Panel A reports the average standardized relative abnormal fund flows $\overline{SAF_t}$ grouped according to the corresponding Sustainability Rating for the 6 months before a fund was initially rated (t = [−6; 0]). To ensure the availability of the Sustainability Rating for the full sample of 6 months, we extrapolate the November 2015 Rating to October and September 2015. We define the standardized relative abnormal fund flow at time t as the actual relative net flow minus the normal, or expected, flow standardized by the forecast correct standard error, as described in Dodd and Warner (1983). To calculate the normal fund flow, we estimate the loadings of the subsequent benchmark model for each fund individually over an estimation window of 24 months (t = [−30; −6]). Specifically, we regress the fund's monthly relative flow on the average relative flow at time t to funds in the same style and lagged Performance Rating category, its time $t-1$ flow, its time $t-1$ raw return, its change in the Carhart four-factor alpha from $t-2$ to $t-1$, and its change in alpha from $t-2$ to $t-1$squared. Further, Panel A reports nonparametric sign tests under the null hypothesis that it is equally probable that sample funds have positive or negative standardized abnormal flows, and a difference-in means-test, showing the discrepancy in the average standardized relative abnormal flows between high and low Sustainability Ratings. Panel B reports the t-statistics of the average standardized abnormal fund flows grouped by the assigned Sustainability Rating. To account for a potential shift in the variance of the $\overline{SAF_t}$ in the event window relative to the estimation window, the standard errors of the $\overline{SAF_t}$ are calculated from the event window, as in Boehmer et al. (1991). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Event month	Low		Below Avg.		Average		Above Avg.		High		High-Low
	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$
Panel A: Average cumulative standardized relative abnormal flow by Sustainability Rating and event month (coefficients)											
–6	0.023	52.44	0.170***	55.59*	0.131*	58.92***	0.099**	56.70**	0.005	57.79	–0.017
–5	0.064	52.48	0.222***	59.92***	–0.021	51.07	0.168	50.30	–0.057	47.66	–0.122
–4	0.211	59.57**	0.077	52.40	–0.010	52.06	0.008	49.07	–0.289**	39.25**	–0.500***
–3	–0.103	55.10	–0.051	47.69	–0.230*	40.93***	–0.049	48.23	–0.358**	39.47**	–0.255
–2	–0.222***	43.42	–0.088	53.20	0.054	53.05	0.044	55.87**	–0.066	43.36	0.155
–1	0.054	47.29	0.125	55.43*	–0.049	46.69	–0.039	51.77	–0.091	43.51	–0.145
Panel B: Average cumulative standardized relative abnormal flow by Sustainability Rating and event month (t-statistics)											
–6	0.34	0.58	3.17***	1.80*	1.72*	3.68***	1.99**	2.42**	0.07	1.62	–0.16
–5	0.87	0.58	3.12***	3.14***	–0.43	0.43	1.55	0.11	–0.43	–0.48	–0.75
–4	1.56	2.27**	1.03	0.75	–0.18	0.83	0.15	–0.33	–2.49**	–2.22**	–2.59***
–3	–0.78	1.23	–0.43	–0.74	–1.87*	–3.66***	–0.55	–0.62	–2.33**	–2.24**	–1.18
–2	–2.82***	–1.62	–1.08	1.01	1.06	1.23	0.74	2.08**	–0.63	–1.41	1.09
–1	0.73	–0.65	1.60	1.77*	–0.90	–1.33	–0.63	0.62	–0.78	–1.34	–0.97

Figure 1.3: Cumulative relative abnormal flow before and after the launch of the Sustainability Rating

The figure visualizes the average cumulative relative abnormal fund flows $\overline{CAF_t}$ of each Sustainability Rating category for the 6-month periods before and after a fund initially received a Sustainability Rating. To ensure the availability of the Sustainability Rating for the full sample of 6 months, we extrapolate the November 2015 Rating to October and September 2015. We compute the cumulative relative abnormal flow for each fund by summing the relative abnormal flow for the respective event period and form the average of the $\overline{CAF_t}$ across all funds for each Sustainability Rating category. To allow for an economic interpretation, we define the relative abnormal fund flow at time t as the actual relative net flow minus the normal, or expected, flow, but omit the standardization procedure employed in the main test. Expected flow is based on the benchmark regression, whereby the fund's monthly relative flow is regressed on the average relative flow at time t to funds in the same style and lagged Performance Rating category, its time $t-1$ flow, its time $t-1$ raw return, its change in the Carhart four-factor alpha from $t-2$ to $t-1$, and its change in alpha from $t-2$ to $t-1$ squared. For the calculation of the $\overline{CAF_t}$ after the Rating's launch, we estimate the benchmark model parameters for each fund individually over an estimation window of 24 months (t = [−24; 0]), while shifting the estimation window by 6 months (t = [−30; −6]) to calculate the $\overline{CAF_t}$ prior to the introduction



Sustainability Rating.[20] Since the median size of a retail share class with a high (low) Sustainability Rating is \$111 million (\$48 million), the flow response can be interpreted as follows. A fund that is assigned the highest Sustainability Rating experiences on average \$4.05 million per year higher net inflows than expected without the introduction of the Rating. In contrast, the launch of the Rating leads on average to a \$0.98 million per year lower net flow for a typical low-rated fund. In accordance with previous sections, the findings indicate that the effect of the Sustainability Rating is both statistically and economically significant.

1.3.4 Fund inflows and outflows

The results from Sections 1.3.1–1.3.3 provide robust evidence that the introduction of the Morningstar Sustainability Rating has a significant impact on mutual fund net flows. As net flows are calculated from the change in total net assets there might be two potential sources of this result. Investors might either invest more new money into funds with a high Sustainability Rating than into funds with a low Rating or sell more existing positions of low-rated funds than of high-rated funds as a response to the publication of the Rating. We therefore separate gross inflows from gross outflows and analyze the effect of the Morningstar Sustainability Rating on both. Prior research has made this distinction between inflows and outflows in investigating the performance-flow relationship. Bergstresser and Poterba (2002) find that gross inflows react more sensitively to past performance than gross outflows. O'Neal (2004) highlights that both gross inflows and gross outflows react to past performance. This result is also supported by Cashman et al. (2012). Both papers agree on an asymmetric effect on inflows, with the best performing funds receiving disproportionately high inflows. Ivković and Weisbenner (2009) refine those results by showing that inflows react to relative performance measures whereas outflows are driven by a fund's absolute performance.

To examine the effect of the Morningstar Sustainability Rating on gross inflows and gross outflows we use data from the SEC form N-SAR.[21] All registered investment companies are required to semi-annually report monthly sales and redemptions on this form and we collect

---

[20] For the calculation of the cumulative abnormal flows prior to the launch of the Sustainability Rating, we shift the event window and the estimation window by 6 months and sum the relative abnormal flow derived thereby for each fund from $t = -6$ to $t = 0$.We then average the cumulative abnormal flows across funds for each Sustainability Rating category.

[21] These data were previously used by, for example, Edelen (1999), Bergstresser and Poterba (2002), and Cashman et al. (2012).

those data from the Morningstar Direct database. Because the N-SAR reports are filed at portfolio and not at share class level, we aggregate fund data at the portfolio level. A fund's total net assets are added across share classes. We calculate portfolio flows, returns, turnover and expense ratios as the weighted average of single share classes' characteristics using the lagged size of each share class as the respective weight. We also use portfolio-level alphas and return volatilities. To indicate that variables refer to a portfolio and not to single share classes, we superscript variables with p, for example $TNA^{p}_{i,t}$ for the total net assets of portfolio i at time t. In some cases, multiple share classes of the same portfolio have different Performance Ratings, where the difference is usually just one Rating category. We then calculate a portfolio's Performance Rating as the weighted mean of the single share classes' Performance Rating using 1 to 5 for the five Rating categories and rounding the resulting average Rating to the nearest integer. Monthly gross inflows and outflows are scaled by a fund's total net assets at the beginning of the month, that is,

$$\mathrm{INFLOW}^{p}_{i,t} = \frac{\mathrm{SALES}^{p}_{i,t}}{\mathrm{TNA}^{p}_{i,t-1}} \tag{4}$$

$$\mathrm{OUTFLOW}^{p}_{i,t} = \frac{\mathrm{REDEMPTIONS}^{p}_{i,t}}{\mathrm{TNA}^{p}_{i,t-1}} \tag{5}$$

where $SALES^{p}_{i,t}$ are the sum of new sales and other sales and $REDEMPTIONS^{p}_{i,t}$ are the redemptions of portfolio i during month t as reported in the NSAR form. Inflows and outflows are winsorized at the 1% and 99% levels.

We repeat the panel regressions from Section 1.3.1 using $SALES^{p}_{i,t}$ and $REDEMPTIONS^{p}_{i,t}$ as the dependent variables to examine the effect of the Morningstar Sustainability Rating on gross inflows and outflows. Table 1.9 displays the results for the time period after the Rating was made public (March 2016 to March 2017). Regressions coefficients indicate that both gross inflows and gross outflows are affected by the Sustainability Rating, but the effect is stronger for inflows. We find the sustainability-inflow relationship to be convex. Mutual funds with the highest (lowest) Sustainability Rating receive inflows 0.40 (0.14) percentage points per month higher (lower) than average rated funds. The relationship is rather symmetric for fund outflows. Low (high) rated funds are subject to 0.11 (0.14) percentage points per month higher (lower) outflows than average rated funds. Despite the economic significance of these results, only the higher inflow into high-rated funds is statistically significant.

Table 1.9: Relative net flow−sustainability regression

This table reports the results from ordinary least squares panel regressions of monthly fund inflows and outflows on the Morningstar Sustainability Rating and other fund characteristics during March 2016 to March 2017. Monthly fund inflows and outflows are defined as $INFLOW_t^p = SALES_t^p / TNA_{t-1}^p$ and $OUTFLOW_t^p = REDEMPTIONS_t^p / TNA_{t-1}^p$, where $SALES_t^p$ are the sum of new sales and other sales and $REDEMPTIONS_t^p$ are the redemptions of a fund as reported in the SEC form N-SAR. The Sustainability Rating is as of the beginning of the month and all other fund characteristics are as of the end of the previous month. Monthly returns are obtained from the CRSP mutual fund database (MRET). The Carhart four-factor alpha and return volatility are calculated using monthly returns over the prior 12 months. The Performance Rating is the one- to five-star Morningstar Star Rating provided by Morningstar. Fund age is the number of years since the inception date (FIRST_OFFER_DT from CRSP). Total net assets, turnover ratios and expense ratios are also obtained from the CRSP mutual fund database (MTNA, TURN_RATIO and EXP_RATIO). Fund characteristics are aggregated across all share classes of the same fund as described in Section 1.3.4. We regress monthly inflows and outflows on fund characteristics between March 2016 and March 2017. Column 1 reports the results for gross inflows, column 2 for gross outflows. Standard errors are double clustered on month and fund level and t-values are reported in parentheses. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	(1) INFLOWᵖ	**(2) OUTFLOWᵖ**
1-month lagged Sustainability Rating		
Low	–0.138	0.110
	(–0.85)	(0.63)
Below average	0.075	0.079
	(0.58)	(0.79)
Above Average	–0.046	–0.146
	(–0.39)	(–1.53)
High	0.403 *	–0.144
	(1.89)	(–0.85)
1-month lagged Performance Rating		
Low	–0.763 ***	–0.288
	(–3.15)	(–1.22)
Below average	–0.289 **	0.115
	(–2.19)	(0.82)
Above average	0.908 ***	0.024
	(7.30)	(0.28)
High	2.676 ***	0.039
	(7.23)	(0.24)

(continued)

(continued)

	(1) INFLOW^p	(2) OUTFLOW^p
1-month lagged 12-month alpha (%)	0.077 ***	–0.041 **
	(4.42)	(–2.98)
1-month lagged monthly return (%)	–0.005	–0.031
	(–0.14)	(–0.87)
1-month lagged 12-month return volatility (%)	0.327 *	0.096
	(2.05)	(0.88)
1-month lagged turnover ratio (%)	0.003 **	0.005 ***
	(2.33)	(3.44)
1-month lagged expense ratio (%)	0.311	0.355
	(1.67)	(1.77)
Log 1-month lagged total net assets ($m)	–0.064	0.103 **
	(–1.47)	(2.36)
Fund age in years	–0.018 ***	–0.017 ***
	(–4.51)	(–4.58)
Month-style fixed effects	Yes	Yes
Number of observations	7727	7736
Adjusted R^2	0.16	0.04

The results indicate that the effect of the Morningstar Sustainability Rating is most pronounced for the inflow into high-rated funds. Investors pay high attention to the Rating when deciding where to invest new money.

1.3.5 Fund families

Roughly 80% of the funds in our sample are members of a fund family and those fund families consist on average of 4.4 fund portfolios and 9.4 share classes. As pointed out by Nanda et al. (2004), fund families offer advantages like greater flexibility in the allocation of human resources. Investors might attribute reputation to fund families rather than single funds which might induce spillover effects within fund families. Nanda et al. (2004) find that a fund's outstanding past performance results in increased cash inflows not only to the respective fund but also to other funds of the same family. The same effect might occur for a high ESG rating. On the other hand, fund families might utilize the great attention to Morningstar's

Sustainability Rating and concentrate their marketing activities to the most sustainable fund, which might lead to negative spillover effects. To investigate the spillover effects within a fund family, we repeat the panel regressions of Section 1.3.1 but add additional explanatory variables. We add the standard deviation of Sustainability Ratings within a fund family as well as the mean Sustainability Rating of all funds within a family but the fund itself. Following Nanda et al. (2004) we also add two dummy variables, *SUSTAINABLE* and *UNSUSTAINABLE*. *SUSTAINABLE* (*UNSUSTAINABLE*) is set to 1 if a fund does not have a high (low) Sustainability Rating but there is a fund within the same fund family that has a high (low) rating; otherwise it is set to 0.

Table 1.10 reports the results. A higher mean family Rating results in a lower fund flow as indicated by the negative coefficient of this variable. The effect, however, only becomes statistically and economically significant if we focus on extreme cases using the *SUSTAINABLE* and *UNSUSTAINABLE* dummy variables. Funds that belong to a fund family with a low-rated fund but do not have a low rating themselves have a 0.11 percentage points per month higher net flow than they would have without a low-rated fund in their family. Funds that are not rated high but belong to a fund family with a high-rated fund receive a 0.17 percentage points per month lower flow than without belonging to such a fund family. The latter effect is statistically significant. These results indicate negative spillover effects within fund families. Contrary to what is documented by Nanda et al. (2004) with respect to the Performance Rating, funds do not profit from high Sustainability Ratings within their fund family, but investors shift money within fund families toward the most sustainable funds. To provide further evidence to this result we add month-family fixed effects to the panel regression (column 4 of Table 1.10).

Even after controlling for monthly family-level flows the effect of the Morningstar Sustainability Rating remains strong, thus indicating that the effect is not driven by aggregated flows at a family level. Our analysis does not identify the channel behind this result, but the results are in line with fund families focusing their marketing activities on members with the highest Sustainability Rating. We also add the standard deviation of Sustainability Ratings within a fund family as an explanatory variable. Results indicate that funds that belong to families with a wide dispersion of Sustainability Rating suffer from significantly lower net flows subsequent to the publication of the Rating.

Table 1.10: Relative net flow−sustainability regression including family characteristics

This table reports the results from ordinary least squares panel regressions of monthly fund flows on the Morningstar Sustainability Rating, other fund characteristics and fund family characteristics during March 2016 to March 2017. Monthly net flows are defined, as in Sirri and Tufano (1998), as $[\mathrm{TNA}_t - (1 + R_t) * \mathrm{TNA}_{t-1}]/\mathrm{TNA}_{t-1}$. The Sustainability Rating and fund family characteristics are as of the beginning of the month and all other fund characteristics are as of the end of the previous month. *SUSTAINABLE* (*UNSUSTAINABLE*) equals 1 if a fund does not have a high (low) Sustainability Rating but there is a fund within the same fund family that has a high (low) rating, and it is set to 0 otherwise. Mean family Rating is the average of the Sustainability Rating of all other funds of the same fund family and family Rating standard deviation is the standard deviation of all the funds' Sustainability Ratings within a fund family. Lagged fund-level control variables include the Morningstar Performance Rating, the 12-month alpha and monthly return, the 12-month return volatility, the turnover ratio, total expense ratio, fund size and age as described in Section 1.2.2. The sample is constructed as described in Section 1.2 with single observations for each share class and month. Funds that do not belong to a fund family are excluded. Standard errors are double clustered on month and share class level and t-values are reported in parentheses. Significance at the 10%, 5%, and 1% level is denoted by *, **, and ***, respectively.

	Dependent variable: Monthly relative net flow (%)			
	(1)	**(2)**	**(3)**	**(4)**
1-month lagged Sustainability Rating				
Low	–0.296 **	–0.197	–0.189	–0.402 **
	(–2.26)	(–1.62)	(–1.55)	(–2.71)
Below average	–0.177 *	–0.149 *	–0.132	–0.234 **
	(–2.12)	(–1.84)	(–1.64)	(–2.28)
Above average	0.064	0.035	0.039	0.021
	(0.69)	(0.39)	(0.43)	(0.21)
High	0.345 **	0.226 *	0.293 **	0.294
	(2.65)	(1.81)	(2.41)	(1.76)
1-month lagged mean family Rating	–0.095			
	(–1.58)			
1-month lagged *SUSTAINABLE*		–0.174 *		
		(–1.89)		
1-month lagged *UNSUSTAINABLE*		0.111		
		(1.37)		
1-month lagged std. dev. Of family Rating			–0.154 *	
			(–1.78)	
Fund-level control variables	Yes	Yes	Yes	Yes
Month-style fixed effects	Yes	Yes	Yes	Yes
Month-family fixed effects	No	No	No	Yes
Number of observations	36476	36476	36476	36380
Adjusted R^2	0.08	0.08	0.07	0.09

1.4 Conclusion

In this paper we study investors' behavior by mutual fund flows that occur subsequent to the introduction of the Morningstar Sustainability Rating in March 2016. The Rating provides investors with a publicly available and an easy-to-grasp indicator of a fund's conformity to ESG criteria. Based on surveys that show investors' preference for sustainable investment opportunities, we hypothesize that the Morningstar Sustainability Rating helps retail investors to overcome the obstacle of identifying sustainable investment funds. We therefore expect retail investors to react to the publication of the Sustainability Rating by withdrawing money from low-rated funds and investing it in high-rated funds. We apply three methodologies to examine the causal relationship between the publication of the Sustainability Rating and subsequent fund flows: panel regressions, a nearest-neighbor propensity score matching, and an event study approach. We find that retail investors indeed react to the availability of the Morningstar Sustainability Rating. Panel regressions show that, among retail share classes, the highest-rated funds receive a 0.45 percentage points per month higher net flow than average-rated funds and a 0.78 percentage points per month higher net flow than low-rated funds. We do not find such a relationship prior to the launch of the Sustainability Rating or for funds with an unpublished rating, suggesting that the results are driven by the publication of the Rating. The matching confirms these results for retail funds. During the first year after the launch, net flows to high-rated funds are 9.41 percentage points higher than to comparable average-rated funds. During the same time low-rated funds suffer from a 6.80 percentage points lower net flow than a comparable average-rated fund.

Our event study finds that high-rated retail funds receive an abnormal inflow of 1.83% during the first 6 months after the initial rating publication whereas low-rated retail funds receive an abnormal flow of –1.01% during the same half-year period. For an average fund, these findings translate into \$4.1 million to \$10.1 million higher inflows for high-rated funds and \$1.0 million to \$3.5 million higher outflows from low-rated funds compared to an average-rated fund during the first year after the publication of the Rating. As expected, the effect of the publication of Morningstar's Sustainability Rating is much lower for institutional investors.

Given the recency of the introduction of the Morningstar Sustainability Rating and the limited data history, it remains an open question how the Rating will affect retail and institutional investors' behavior in the long run. Observing the first year after the launch of the Rating, our results suggest that retail investors value sustainable investments and that the Morningstar

Sustainability Rating enables them to incorporate this preference into their fund selection process. Retail investors appreciate the condensation of public but hard-to-grasp information on sustainability into an easy-to-read figure and shift their investments accordingly.

References

Abadie, A., Imbens, G. W., 2016. Matching on the estimated propensity score. Econometrica, 84, 781−807.

Barnett, M. L., Salomon, R. M., 2006. Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance. Strategic Management Journal, 27, 1101−1122.

Benson, K. L., Humphrey, J. E., 2008. Socially responsible investment funds: Investor reaction to current and past returns. Journal of Banking & Finance, 32, 1850−1859.

Bergstresser, D., Poterba, J., 2002. Do after-tax returns affect mutual fund inflows?. Journal of Financial Economics, 63, 381−414.

Boehmer, E., Masumeci, J., Poulsen, A. B., 1991. Event-study methodology under conditions of event-induced variance. Journal of Financial Economics, 30, 253−272.

Bollen, N. P. B., 2007. Mutual fund attributes and investor behavior. Journal of Financial and Quantitative Analysis, 42, 683−708.

Campbell, J. Y., Lo, A. W., MacKinlay, A. C., 1997. Event-Study Analysis. 149−180 (Princeton University Press, Princeton, NJ).

Capelle-Blancard, G., Monjon, S., 2014. The performance of socially responsible funds: Does the screening process matter? European Financial Management, 20, 494−520.

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

Cashman, G. D., Deli, D. N, Nardari, F.,Villupuram, S., 2012. Investors do respond to poor mutual fund performance: Evidence from inflows and outflows. Financial Review, 47, 719−739.

Chevalier, J., Ellison, G., 1997. Risk taking by mutual funds as a response to incentives. Journal of Political Economy, 105, 1167−1200.

Del Guercio, D., Tkac, P. A., 2002. The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds. Journal of Financial and Quantitative Analysis, 37, 523−557.

Del Guercio, D., Tkac, P. A., 2008. Star power: The effect of Morningstar ratings on mutual fund flow. Journal of Financial and Quantitative Analysis, 43, 907−936.

Dodd, P., Warner, J. B., 1983. On corporate governance: A study of proxy contests. Journal of Financial Economics, 11, 401−438.

Edelen, R. M., 1999. Investor flows and the assessed performance of open-end mutual funds. Journal of Financial Economics, 53, 439−466.

El Ghoul, S., Karoui, A., 2017. Does corporate social responsibility affect mutual fund performance and flows?. Journal of Banking & Finance, 77, 53−63.

Evans, R. B, Fahlenbrach, R., 2012. Institutional investors and mutual fund governance: Evidence from retail− institutional fund twins. Review of Financial Studies, 25, 3530−3571.

Fabozzi, F. J., Ma, K.C., Oliphant, B. J., 2008. Sin stock returns. Journal of Portfolio Management, 35, 82.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33, 3−56.

Filbeck, G., Gorman, R., Zhao, X., 2009. The “best corporate citizens”: Are they good for their shareholders?. Financial Review, 44, 239−262.

Goetzmann, W. N., Peles, N., 1997. Cognitive dissonance and mutual fund investors. Journal of Financial Research, 20, 145−158.

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

Huang, J., Wei, K.D., Yan, H., 2007. Participation costs and the sensitivity of fund flows to past performance. Journal of Finance, 62, 1273−1311.

Ippolito, R. A., 1992. Consumer reaction to measures of poor quality: Evidence from the mutual fund industry. Journal of Law and Economics, 35, 45−70.

Ivković, Z., Weisbenner, S., 2009. Individual investor mutual fund flows. Journal of Financial Economics, 92, 223−237.

Kempf, A., Osthoff, P., 2007. The effect of socially responsible investing on portfolio performance. European Financial Management, 13, 908−922.

Massa, M., 2003. How do family strategies affect fund performance? When performance-maximization is not the only game in town. Journal of Financial Economics, 67, 249−304.

Nanda, V., Wang, Z. J., Zheng, L., 2004. Family values and the star phenomenon: Strategies of mutual fund families. Review of Financial Studies, 17, 667−698.

O'Neal, E. S., 2004. Purchase and redemption patterns of US equity mutual funds. Financial Management, 63−90.

Patell, J. M., 1976. Corporate forecasts of earnings per share and stock price behavior: Empirical test. Journal of Accounting Research, 14, 246−276.

Renneboog, L., Ter Horst, J., Zhang, C., 2008. The price of ethics and stakeholder governance: The performance of socially responsible mutual funds. Journal of Corporate Finance, 14, 302−322.

Renneboog, L., Ter Horst, J., Zhang, C., 2011. Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds. Journal of Financial Intermediation, 20, 562−588.

Riedl, A., Smeets, P., 2017. Why do investors hold socially responsible mutual funds? Journal of Finance, 72, 2505–2550.

Salganik-Shoshan, G., 2016. Investment flows: Retail versus institutional mutual funds. Journal of Asset Management, 17, 34−44.

Sirri, E. R., Tufano, P., 1998. Costly search and mutual fund flows. Journal of Finance, 53, 1589−1622.

Appendix 1.A: Event study–Benchmark diagnostics

Table 1.A1: Benchmark diagnostics−standardized relative abnormal flow sorted by the Performance Rating in the event window

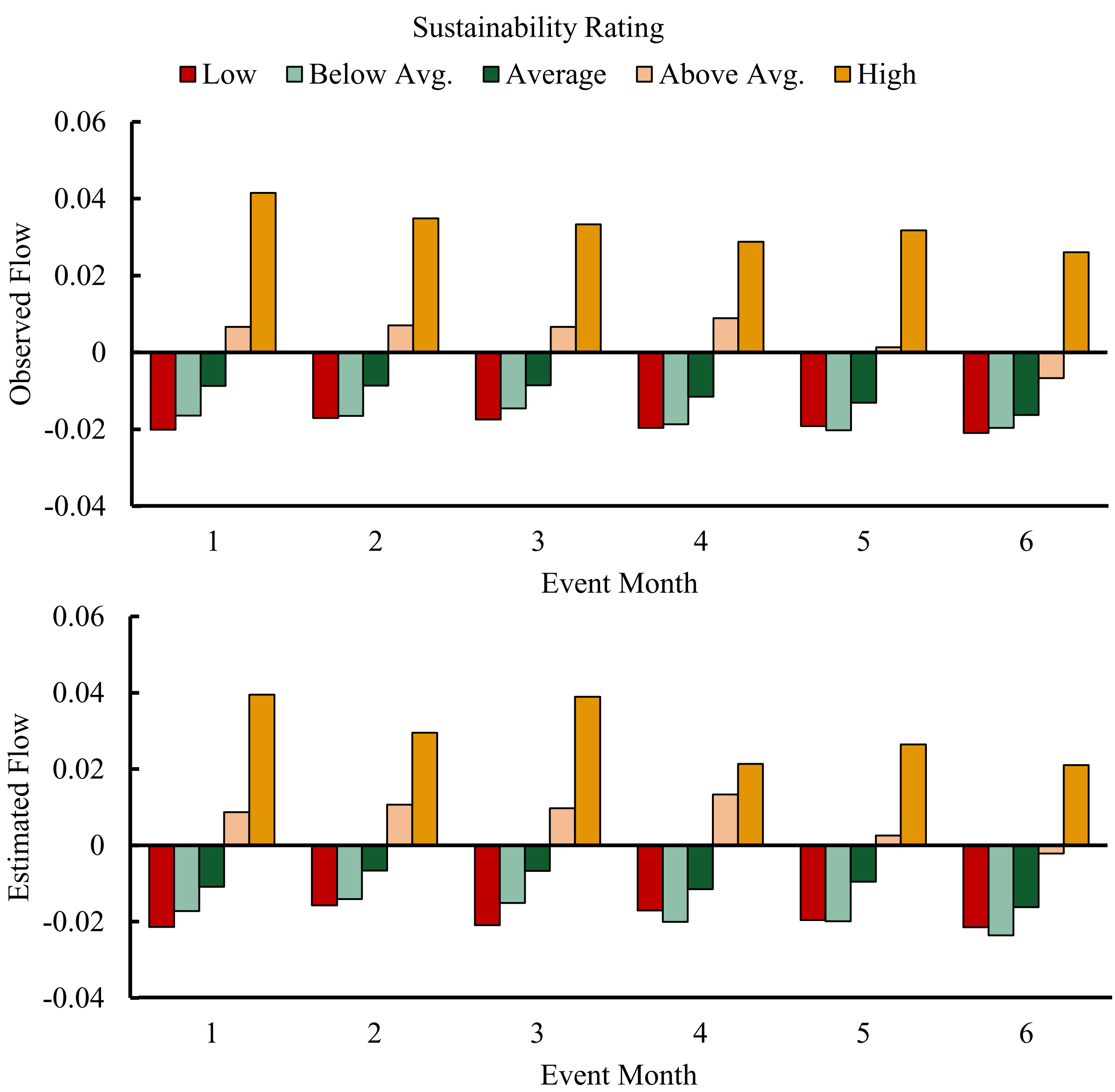
Panel A reports the average standardized relative abnormal fund flows $\overline{SAF_t}$ grouped according to the corresponding Performance Rating for months 1 to 6 after a fund was initially rated. We define the standardized relative abnormal fund flow at time t as the actual relative net flow minus the normal, or expected, flow standardized by the forecast correct standard error, as described in Dodd and Warner (1983). To calculate the normal fund flow, we estimate the loadings of the subsequent benchmark model for each fund individually over an estimation window of 24 months (t = [−24; 0]). Specifically, we regress the fund's monthly relative flow on the average relative flow at time t to funds in the same style and lagged Performance Rating category, its time $t-1$ flow, its time $t-1$ raw return, its change in the Carhart four-factor alpha from t–2 to t−1, and its change in alpha from t–2 to t−1 squared. Further, Panel A reports nonparametric sign tests under the null hypothesis that it is equally probable that sample funds have positive or negative standardized abnormal flows and a difference-in-means test, showing the discrepancy in the average standardized relative abnormal flows between high and low Performance Ratings. Panel B reports the t-statistics of the average standardized abnormal flows grouped by the assigned Performance Rating. To account for a potential shift in the variance of the $\overline{SAF_t}$ in the event window relative to the estimation window, the standard errors of the $\overline{SAF_t}$ are calculated from the event window, as in Boehmer et al. (1991). *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Event month	Low		Below Avg.		Average		Above Avg.		High		High-Low
	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$	% > 0	$\overline{SAF_t}$
Panel A: Average standardized relative abnormal flow by Performance Rating and event month (coefficients)											
1	0.110	50.00	0.080*	55.04**	0.070	53.82 **	0.011	44.40*	0.334**	57.62	0.223
2	–0.043	44.26	0.069	54.12*	–0.058	45.99 **	–0.060	43.94**	0.042	42.85	0.086
3	–0.022	47.50	0.012	47.15	–0.003	47.34	–0.026	45.74	0.017	45.00	0.039
4	0.022	50.86	0.089**	58.35***	0.052	52.94	–0.089	41.95***	–0.092	39.62	–0.115
5	0.060	51.16	0.132***	56.44***	–0.037	48.89	–0.111**	44.55*	–0.162	38.70*	–0.222
6	–0.192**	46.40	0.053	52.11	–0.067**	46.57 *	–0.036	46.06	–0.208	43.10	–0.016
Panel B: Average standardized relative abnormal flow by Performance Rating and event month (t-statistics)											
1	1.08	0.00	1.87*	2.20**	1.59	1.97 **	0.20	–1.89*	2.17**	1.17	1.17
2	–0.53	–1.26	1.63	1.79*	–1.60	–2.06 **	–1.00	–2.05**	0.29	–1.06	0.49
3	–0.24	–0.54	0.33	–1.23	–0.09	–1.36	–0.42	–1.42	0.11	–0.77	0.22
4	0.29	0.18	2.23**	3.67***	1.47	1.49	–1.41	–2.72***	–0.67	–1.51	–0.69
5	0.85	0.26	2.97***	2.80***	–1.10	–0.55	–2.09**	–1.86*	–1.19	–1.77*	–1.37
6	–2.21**	–0.80	1.30	0.94	–1.96**	–1.73 *	–0.66	–1.28	–1.29	–1.05	–0.08

Appendix 1.B: Event study – Observed versus estimated flows

Figure 1.B1: Observed versus estimated flow sorted by Performance Rating in the event window

The figure shows the average normal flow estimates together with the average observed flows grouped according to the corresponding Performance Rating for the 6 months after the launch of the Sustainability Rating (the period from March 2016 to August 2016). To calculate the normal fund flow, we estimate the loadings of the subsequent benchmark model for each fund individually over an estimation window of 24 months ($t = [-24; 0]$). Specifically, we regress the fund's monthly relative flow on the average relative flow at time t to funds in the same style and lagged Performance Rating category, its time $t-1$ flow, its time $t-1$ raw return, its change in the Carhart four-factor alpha from $t-2$ to $t-1$, and its change in alpha from $t-2$ to $t-1$ squared.



Chapter 2

Is Fund Activeness Really a Predictor of Performance?

Sebastian Fischer*

Status: Working Paper (February 2019)

Abstract

Measures of fund management activeness such as the R^2 from a risk factor regression and Active Share are claimed to predict future fund returns. I show that neither Active Share nor R^2 have been good predictors in recent times when controlling for different benchmark indices. R^2 only performs well during the years 2000−2002. I find that the most actively managed portfolios load on alternative risk factors such as profitability, quality, and the low-beta anomaly. Adding those risk factors can explain part of the return predictive effects of activeness reported in prior research.

JEL Classification: G11, G14, G20, G23
Keywords: Mutual Fund, Active Share

* I thank Florian Weigert and Markus Schmid for helpful comments and Morningstar for providing access to Morningstar Direct.

2.1 Introduction

The academic debate on the value of active fund management has been going on since Jensen (1968) first introduced his well-known alpha measure. Meanwhile there is a widely accepted consensus that, net of fees, actively managed funds *on average* underperform a risk-adjusted benchmark. Among others, Malkiel (1995), Gruber (1996), Carhart (1997), as well as Fama and French (2010) support this view. Studies of Grinblatt and Titman (1989, 1993) and Daniel et al. (1997) show that gross of fees mutual funds generate positive abnormal returns. Wermers (2000) combines both views as he finds mutual fund managers to exhibit positive stock picking ability but he finds funds to have negative abnormal net returns. His analysis shows that trading expenses, non-equity holdings and management fees melt down the value added by fund managers' stock selection and lead to the negative shareholder value.

Since investors nevertheless select rather expensive actively managed mutual funds, they expect those products to beat passive benchmarks. A requirement for this outperformance to happen and to compensate for management fees is that fund managers trade on their own investment ideas and therefore actively deviate from passive benchmarks. Research, however, has shown that many mutual funds stick closely to their benchmark and are therefore called closet indexers (Cremers et al. (2016)). This evidence has attracted attention for two reasons. First, selling a (close to) passive investment strategy labeled as an active mutual fund is mislabeling and has led to investigations by investor protection authorities such as ESMA (ESMA (2016)). Second, academic research documents that closet indexing is linked to underperformance (Petajisto (2013)). Studies even suggest that a fund's activeness, which is some form of deviation from a benchmark, can predict future fund returns. For example, Wermers (2003) finds a high tracking error volatility to be a good indicator for future out-performance. Cremers and Petajisto (2009) as well as Petajisto (2013) measure the deviation of a fund's portfolio holdings from the constituents of its self-reported benchmark index (or another index if the fund's holdings are more related to any) and call it "Active Share". The former find that funds in the highest Active Share quintile beat their benchmark index by 1.05% p.a. after controlling for additional risk factor exposure. Amihud and Goyenko (2013) use the R^2 from a regression of a fund's past returns on risk factors as a proxy for fund activeness. They find that funds with a low R^2 and therefore a high stock picking activity, which they call selectivity, generate abnormal net returns of 0.60% per year and even 3.80% for funds with highest past alphas. Doshi et al. (2015) measure activeness as the difference between a fund's actual portfolio weighting and a hypothetical cap-weighting of the fund's holdings. They find that funds that (almost) cap-weight their holdings underperform those with most active portfolio weighting decisions by 2.35% p.a. (Carhart (1997) alphas of net returns) with the most active funds generating a positive abnormal net return of 0.68% per year.

The rationale behind those findings on activeness is as follows. Fund managers make active trading decisions based on skill and superior information. The more skill or information a fund manager has, the more promising investment opportunities she will find. Once a fund manager has realized all the identified investment opportunities she will invest the remaining capital according to some benchmark. Therefore, activeness is an indicator of skill and information, which lead to future abnormal returns. This interpretation is backed by the theoretical work of Berk and Green (2004), who argue that fund managers can only identify a limited number of investment opportunities.

Despite the high popularity of activeness measures, there is some doubt about the predictive power of activeness for future fund performance. In the case of Active Share this led to a vigorous debate between Frazzini et al. (2016) and Cremers (2015). The former claim that much of the return predictive power of Active Share is due to differences in the average Active Share of funds with different benchmarks and the fact that benchmark indices themselves have an alpha when measured in the conventional Fama and French (1993) and Carhart (1997) framework (Cremers et al. (2013)). It remains unclear to what extend this heterogeneity across benchmark indices affects the return predictive power of other activeness measures. Additionally, Cremers et al. (2016) raise doubt about the relationship between the R^2-based selectivity measure and performance as they do not find such a relationship in their international fund sample.

My paper seeks to validate the performance of activeness measures, especially Active Share and Amihud's and Goyenko's (2013) R^2-based selectivity measure, as a predictor of future fund returns. Furthermore, I analyze the investment patterns of the most active fund managers. I use two data sets: The original Cremers and Petajisto (2009) data set including almost all US equity mutual funds between 1980 and 2003 and a second data set covering the large subsample of US mutual funds that report the S&P 500 as their primary benchmark during the years 2004 to 2015. I first analyze the first dataset and its subsample of S&P 500 benchmarked funds during the 1980–2003 time period and then turn to more recent years.

My results confirm the return predictive power of Amihud's and Goyenko's R^2 selectivity measure during the 1980–2003 time period. High selectivity funds outperform the least active counterparts by 1.08% p.a. I do, however, not find this predictive power of selectivity in recent years, but instead observe a negative relationship between selectivity and future fund returns. Looking more closely into the return time series, I find that selectivity has only predicted performance well during the post dot-com bubble years 2000−2002. As to Active Share, the predictive power strongly decreases from 2.64% p.a. to 0.31% p.a. when controlling for the different benchmark indices and becomes statistically nonsignificant in both samples. This is in line with the findings of Frazzini et al. (2016). My results suggest that activeness is not a reliable concept to distinguish skillful fund managers from underperformers.

In addition to analyzing the return predictive power of activeness measures, I measure the characteristics of active fund management and examine the channels through which truly active fund managers deviate from their passive counterparts. I find that prior to 2004 active fund managers did profit from an exposure to risk factors not controlled for by in the Carhart (1997) four-factor model, namely the Fama and French (2015) profitability and investment and – closely linked – the Asness et al. (2014) quality factors. The exposure of the most actively managed funds to those non-traditional risk factors, however, disappears and even turns partially negative after 2003. The more active fund portfolios are instead exposed to the Frazzini and Pedersen (2014) betting-against-beta factor afterwards, whereas they contain less liquidity risk as measured by the Amihud and Noh (2016) illiquid-minus-liquid factor. Altogether, the inclusion of additional risk factors lowers the performance predictive power of activeness measures. I therefore claim that neither Active Share nor the R^2-based selectivity measure reliably predict outperformance. They do, instead, in large parts measure the (positive or negative) exposure to risks not controlled for by the Carhart (1997) four-factor model. This finding is closely related to the result of Cremers and Pareek (2016), who find that the outperformance of high Active Share funds during 1990–2013 is reduced by roughly 2/3 when the Asness et al. (2014) quality-minus-junk factor is added to the factor model used to evaluate fund alphas.

The contribution of my paper is twofold. First and foremost, I add to the literature on fund activeness and its performance predictive power by conducting a broad analysis on different data samples. I especially add to the debate between Frazzini et al. (2016) and Cremers (2015) on the informational value of Active Share. Second, I add to the literature on mutual fund performance evaluation by providing evidence that the inclusion of the profitability, investment, illiquidity, and betting-against-beta factors helps explain return differences between the more and the less active funds.

The remainder of this paper is structured as follows. Section 2.2 introduces the measures of activeness I consider in my analyses. Sections 2.3 and 2.4 describe my sample selection and calculation procedure as well as my final data sets. Section 2.5 reports my findings regarding the validity of activeness as a predictor of future fund performance. Section 2.6 analyzes the characteristics of truly active fund managers that explain parts of the return predictability. Section 2.7 concludes.

2.2 Measures of fund activeness

I focus on two different measures of fund activeness that have been developed in prior research: The R^2-based selectivity measure and Active Share. Both measures differ in the underlying idea of active portfolio management as well as the required data to calculate these measures. I also calculate each fund's tracking error as a measure of activeness but will pay less attention to it since my analysis shows that tracking error does not have a significant performance predictive power.

2.2.1 Amihud's and Goyenko's (2013) selectivity measure

Amihud and Goyenko (2013) propose to measure fund activeness as the R^2 from a time-series regression of a fund's returns on risk factors. As proposed in the original paper, each month I regress monthly fund net returns from the past two years on a Carhart (1997) four-factor model. The resulting R^2 indicates to what extent a fund's return volatility can be explained by the volatility of risk factors and thus measures how much a fund is involved in investment strategies that go beyond factor loading, e.g. stock picking or short-term factor timing. Since a high R^2 is equivalent to low activeness, Amihud and Goyenko (2013) call $1–R^2$ the selectivity measure. Selectivity is easy to calculate since it does not require any holding data or any information on a fund's benchmark and it can be updated on a monthly basis. It captures simple passive strategies as well as strategies that load on risk factors or any static mixture of passive strategies (e.g. 50% market portfolio, 50% small cap portfolio). Stock picking will typically reduce the R^2 and therefore increase the selectivity of a fund. Short-term factor timing (i.e. within 24 months timing activities) will lead to a higher selectivity whereas long-time timing effects will increase selectivity only slightly as the measure results from a regression during a 24-month period.

2.2.2 Active Share

Active Share is the deviation of actual fund holdings from an explicit benchmark. At each time *t* a fund's Active Share is calculated as:

$$Active\ Share_t = \frac{1}{2}\sum_{i=1}^{N}|w_{f,i} - w_{index,i}|$$

with $i=1,\ldots N$ being all assets available, $w_{f,i}$ being the fund's portfolio weight of asset *i,* and $w_{index,i}$ being the weight of asset *i* in the benchmark index. Following Cremers and Petajisto (2009), I only consider a fund's equity holdings because equity mutual funds usually hold very few derivatives or other non-equity assets. Following Petajisto (2003), I use a fund's self-declared benchmark index to calculate Active Share. Active Share evaluates fund managers' activeness by a snapshot of fund holding. Since most funds only report their holdings on a quarterly basis and with a time lag, Active Share cannot be updated monthly. Unlike selectivity, however, Active Share does not require any fund history. Active Share does not distinguish funds which deviate in a systematic way from those which actually do stock picking. An S&P 500 benchmarked fund whose manager passively invests 60% into the S&P 500 and 40% in the S&P 400 would have an Active Share of 0.4 and thus the same as a fund whose manager passively invests 60% in the S&P 500 and actively picks non-US or small-cap stocks to constitute the remaining 40% of the portfolio.

2.2.3 Tracking Error Volatility

Tracking error volatility, also known as tracking error, is defined as the volatility of the fund's return in excess of a benchmark. I consider two different benchmark definitions: The fund's explicit self-declared benchmark following Wermers (2003) and an implicit CAPM benchmark following Petajisto (2013), i.e. $R_f + \beta R_m$ with R_f and R_m being the risk-free rate and the market excess return respectively and β being the fund's market beta from a one-factor model. I call the resulting measures of tracking error $TE_{explicit}$ and $TE_{implicit}$, respectively. Following Wermers (2003), I use the standard deviation of monthly fund excess returns over the past 36 months to determine tracking error volatilities. A fund's beta is also calculated from time-series regressions over the past 36 months. Tracking error measures the risk that comes with any systematic und unsystematic deviation from the benchmark. Just as selectivity, its calculation does not require any holding data, but relies on historical return information.

2.3 Data and the calculation of activity measures

My analysis builds on two data sets. First, I use the data that was originally used by Cremers and Petajisto (2009) and later extended, reused and publicly provided by Petajisto (2013). It consists of funds from the CRSP mutual fund universe which were identified as equity funds by investment objective codes from Wiesenberg, ICDI, and Spectrum as well as a code from Thomson Reuters and which are reported to on average invest at least 70% of their portfolio in equities. Cremers and Petajisto (2009) used MFLINKS to match those data to the Thomson Reuters database to obtain fund holdings which are based on mandatory quarterly SEC filings. Funds for which less than 60% of their investments could be identified via CRSP as well as sector funds and funds with total net assets below USD 10 million were dropped from the sample. For the remaining sample, Petajisto provides quarterly data on Active Share between 1980 and 2009 as well as the primary benchmark collected from Morningstar on his personal website.[1] I limit the dataset to the years 1980–2003 as in Cremers and Petajisto (2009). I additionally drop index funds and enhanced index funds. I neither recalculate Active Share nor systematically collect the funds' benchmarks but use that information as provided by Petajisto. I combine the data with monthly fund returns and monthly assets under management data from the CRSP survivorship-bias free mutual fund database. For funds with multiple share classes, I aggregate assets under management and calculate monthly returns as the mean of single share classes' returns weighted by total assets. Whereas the selectivity measure can be computed monthly, I do not have Active Share data on such a regular basis.[2] For each fund, I therefore use the most recent available data on Active Share if this is not older than 6 months. I end up with a total of 154'016 fund-month observations for 1,749 distinct funds between 03/1980 and 12/2003.

[1] http://www.petajisto.net/data.html
[2] I require return data for at least 16 of the previous 24 months to calculate the R^2-based selectivity measure.

The second data set considered in my analysis combines fund holdings, fund returns and fund characteristics from the CRSP Survivorship-Bias Free Mutual Fund Database with additional information from Morningstar Direct. I start out considering all mutual funds included in the CRSP database during the 2004–2015 time period.[3] For each fund, I get monthly return data as well as data on total net assets from CRSP. Using CUSIP, I match the fund data with data from Morningstar Direct as of 03/06/2016 to obtain each fund's primary prospectus benchmark and the Morningstar US category group. For funds with multiple share classes, I aggregate the data at a portfolio level as done in the first data sample. I only consider funds whose investment focus is "US Equities" as indicated by the Morningstar US category group. I exclude index funds, balanced funds, international funds, and sector funds by the CRSP Index Fund Flag, the CRSP Objective Code and by screening fund names for key terms such as "balanced" or "index". I only include funds whose total net assets reach USD 10 million at some time during my sample period. Frazzini et al. (2016) claim that the return predictive power of Active Share is due to different Active Share values across groups of funds with different benchmark indices and the abnormal return of benchmark indices themselves (as noted by Cremers et al. (2013)). Considering this concern, I reduce my sample to those funds that declare the S&P 500 to be their primary benchmark. The S&P 500 is by far the most popular benchmark and thus leaves me with 1,563 mutual funds.[4] For 1,455 of those funds I get complete holding information from the CRSP mutual fund database. I drop all holdings but stocks, e.g. derivatives, bonds and loans, funds, cash positions as well as other balance sheet items (account receivables etc.). I identify those holdings either by CUSIP or by their security name. For example, I drop all holdings whose names include one of the strings "Swap", "Forward" or "ETF". Holdings without a CUSIP as well as all short positions are also dropped. This leaves me with 3,832,883 fund-stock-month observations.

I use historical S&P 500 constituents and each constituent's market capitalization, both of which are taken from CRSP. I retrieve index returns from Thomson Datastream and download monthly Fama and French (1993) as well as momentum risk factors from Kenneth R. French's website.[5] Monthly quality-minus-junk as well as betting-against-beta risk factors are provided

[3] CRSP database contains comprehensive fund holding information from 2003 on. I start my sample in 2004 to consider two disjoint data samples.
[4] Second most popular is the Russell 1000 Growth index, which is benchmarked by only 593 funds throughout the sample period.
[5] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

by AQR Capital Management, LLC, via their website.[6] I construct the illiquid-minus-liquid risk factor, as described by Amihud and Noh (2016) and Amihud (2002), using return, volume and price data from CRSP.

Using the holding data, I calculate Active Share. I first aggregate holdings of several share classes of the same firm. This becomes relevant from 2014 on, when for few companies such as Alphabet Inc. the S&P 500 includes several share classes. It should not be considered a sign of activeness if a fund manager decides to hold only one of the share classes to substitute the other one. I match the fund holding data with S&P 500 constituents in three steps. If a holding can be matched to an index constituent by its historical CUSIP, this is the preferred way. I also use the "CRSP Link" linking table in a second step to match remaining fund holdings to firms that are listed in the S&P 500. Last but not least, I match holding data manually if the security name, ticker data and CUSIP given in the holding data unequivocally identify the security as a stock of an S&P 500 listed company. My values of Active Share are very close to those of Petajisto (2013). Among the 5'992 fund-month observations for which both data sets provide Active Share data, the correlation of Active Share is 0.97 and even increases to 0.99 if I drop funds for which Petajisto (2013) finds a benchmark other than the S&P 500 and therefore calculates Active Share relative to a different index.[7] Just as in the first data set, I consider values on Active Share that were measured at the end of a month for the subsequent 6 months unless new data becomes available. I drop fund-month observations if assets under management are below USD 10 million or if my equity holding data covers less than 60% of the fund's assets. The R^2-based selectivity measure is calculated as in the first data set. I end up with a total of 71,063 fund-month observations of 1,488 distinct funds between 01/2004 and 12/2015.

2.4 Descriptive statistics

My first dataset (1980–2003) consists of 154,016 monthly observations of 1,749 funds. Descriptive statistics are reported in Table 2.1. The average fund has about USD 1.1 billion assets under management which is driven by a few very large funds. The median fund size is USD 222 million. For the vast majority of those observations, I obtain all four measures of

[6] https://www.aqr.com/Insights/Datasets

[7] I attribute this difference to the fact that benchmark indices are determined from snapshots that were taken at different points in time in both samples. I checked a subsample of the respective funds and found that my benchmark indication is correct in the vast majority of cases.

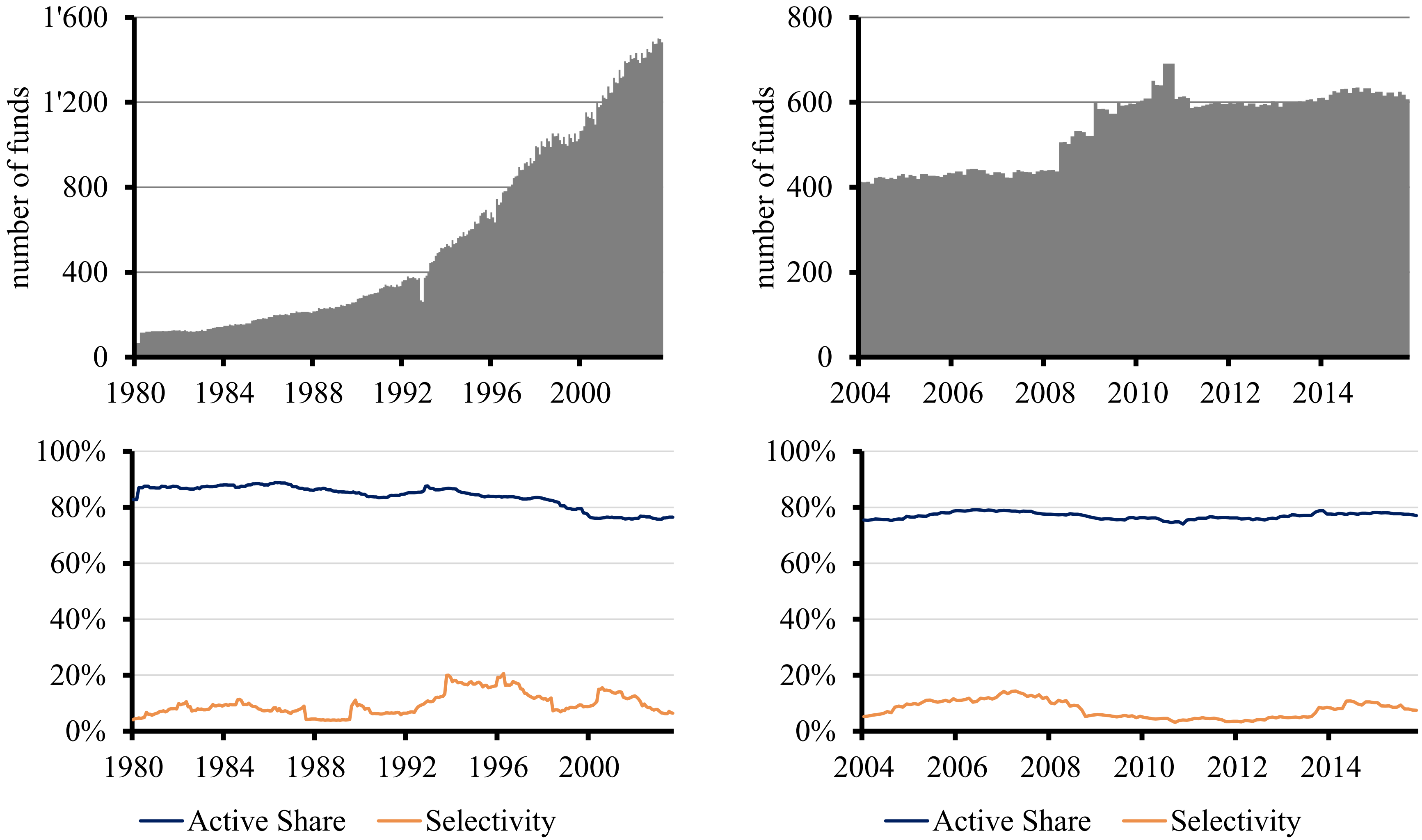
Table 2.1: Descriptive data on sample size and activeness measures

This table provides a descriptive overview over the sample size, the fund's total net assets and values of the activeness measures. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. $TE_{explicit}$ and $TE_{implicit}$ are a fund's tracking error volatility with respect to either the explicit benchmark named in a fund's prospectus or the implicit benchmark return calculated from a one-factor model. The mean as well as the 1%, 25%, 50%, 75% and 99% quantile values are reported. Panel A refers to the 1980–2003 mutual fund dataset. Panel B refers to the 2004–2015 dataset of S&P 500 benchmarked mutual funds.

	Number of observations	Mean	1%	25%	50%	75%	99%
			Panel A: Data set 1, 1980–2003				
Number of funds	1,749						
Number of fund-month-observations	154,016						
Total assets (in mn. USD)	154,016	1,093.0	11.7	71.4	221.8	721.1	16,441.3
Selectivity	146,783	0.11	0.01	0.05	0.08	0.14	0.48
Active Share	154,016	0.81	0.36	0.72	0.85	0.94	1.00
$TE_{explicit}$	138,730	0.026	0.008	0.016	0.023	0.032	0.079
$TE_{implicit}$	139,676	0.026	0.007	0.014	0.022	0.033	0.085
			Panel B: Data set 2, 2004–2015				
Number of funds	1,488						
Number of fund-month-observations	71,063						
Total assets (in mn. USD)	64,758	2,483.3	11.0	65.9	265.0	1,277.8	49,492.6
Selectivity	60,382	0.07	0.00	0.02	0.05	0.09	0.38
Active Share	64,230	0.76	0.32	0.67	0.78	0.88	1.00
$TE_{explicit}$	55,166	0.011	0.015	0.020	0.028	0.028	0.052
$TE_{implicit}$	55,166	0.012	0.004	0.008	0.010	0.015	0.039

Figure 2.1: Time series of observations and activeness measures

This figure depicts the number of observations and the average value of Active Share and selectivity over time. The two graphs on the left show values for the 1980–2003 data set, the graphs on the right refer to the 2004–2015 data set. The number of observations are shown in the upper graphs, average values of the active measures in the bottom graphs.



activeness. Average selectivity is 0.11 representing an R^2 of 0.89. The top 1% of all fund-month observations has an R^2 of 0.52 or below. Active Share reaches from 0.36 to 1.00 (1% and 99% quintile) with a mean of 0.81. Active Share and selectivity have a cross-sectional correlation of 0.35. Even though my sample reaches from 1980 to 2003 there are just 66 funds during early months and up to 1,501 funds in late 2003 as Figure 2.1 shows. Whereas Active Share slightly decreases over time, there is a non-monotonic variation in average selectivity reaching from 4% in the early 1980s to over 20% in mid 1990s. My first dataset combines funds with different investment styles and 19 different benchmark indices. Table 2.2 reports average values of Active Share and selectivity by benchmark index. Notably, almost one third of all funds and over 40% of all fund-month observations report the S&P 500 as their primary prospectus benchmark. There is a tendency of small cap funds to have higher values of Active Share and selectivity. Large cap funds tend to have lower activeness values.

The second dataset (2004–2015) consists of 1,488 funds and 71,063 fund-month observations. The average fund size is USD 2.48 billion and thus more than twice as large as in the first sample. This effect is driven by more very large funds. Levels of selectivity, Active Share as well as tracking error are below those in the first sample. This is in line with my expectation since my second sample consists of S&P 500 benchmarked funds only and the separation of the first sample into index groups has shown that large cap funds tend to have lower values of activeness. The number of funds increases during the 2004–2015 time period starting with 413 observations in 01/2004 and growing to more than 600 observations by the end of 2015. During 2010, the sample size even peaks at 691. The number of funds increases most during the post-crises years 2009 and 2010. Neither measure of activeness follows a clear trend but selectivity peaks during 2006 and 2007. As in the first sample Active Share and selectivity are positively correlated in the cross-section (ρ=0.46).

2.5 Fund activeness as a predictor of performance

Prior research has shown that more active funds, as defined by selectivity or Active Share, outperform their less active counterparts. In the light of these findings, it seems that activeness is a universal predictor of future fund returns. Using both of my data sets, I test this claim and evaluate the predictive power of both measures. I also conduct the analysis for the tracking error measures. In a first step, I use the 1980–2003 data that were originally used by Cremers and Petajisto (2009). It is plausible to assume that my results not only resemble those of

Table 2.2: Observations and activeness measures by benchmark index

This table summarizes the number of funds and observations as well as the average selectivity and Active Share values by benchmark index. The benchmark index is determined by a fund's prospectus. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. $TE_{explicit}$ and $TE_{implicit}$ are a fund's tracking error volatility with respect to either the explicit benchmark named in a fund's prospectus or the implicit benchmark return calculated from a one-factor model. All data refer to the 1980–2003 data set. Mean selectivity and Active Share are calculated as the mean of panel data.

Index	Cap. focus	Distinct funds	Fund-month-obs.	Mean selectivity	Mean Active Share	Mean $TE_{explicit}$	Mean $TE_{implicit}$
Russell 1000	Large	31	2,508	0.09	0.73	0.017	0.014
Russell 1000 Growth	Large	184	17,119	0.09	0.75	0.024	0.022
Russell 1000 Value	Large	141	11,361	0.10	0.77	0.018	0.022
Russell 2000	Small	163	13,979	0.14	0.95	0.031	0.034
Russell 2000 Growth	Small	128	8,813	0.11	0.91	0.035	0.051
Russell 2000 Value	Small	77	5,046	0.15	0.91	0.024	0.033
Russell 3000	All	22	1,938	0.11	0.86	0.022	0.021
Russell 3000 Growth	All	15	1,333	0.11	0.80	0.032	0.031
Russell 3000 Value	All	14	652	0.11	0.80	0.017	0.023
Russell Midcap	Midcap	22	2,047	0.13	0.92	0.025	0.026
Russell Midcap Growth	Midcap	119	8,070	0.13	0.85	0.038	0.044
Russell Midcap Value	Midcap	43	2,231	0.15	0.90	0.027	0.030
S&P 400	Midcap	53	4,000	0.12	0.87	0.033	0.030
S&P 500	Large	564	62,139	0.10	0.78	0.025	0.020
S&P 500 Growth	Large	93	6,685	0.11	0.71	0.026	0.019
S&P 500 Value	Large	33	2,044	0.14	0.78	0.022	0.021
S&P 600	Small	30	1,768	0.13	0.92	0.038	0.040
Wilshire 4500	Small and Midcap	6	970	0.09	0.92	0.027	0.030
Wilshire 5000	All	11	1,313	0.08	0.97	0.029	0.028
TOTAL		**1,749**	**154,016**	**0.11**	**0.81**	**0.026**	**0.026**

Cremers and Petajisto (2009) but should also be similar to the results of Amihud and Goyenko (2013), who use 1988–2010 data of a comparable fund sample. In a second step, I conduct the analysis for the more homogenous subsample consisting of those funds that benchmark the S&P 500. Finally, I turn to my second data set and analyze the performance predictive power of fund activeness for S&P 500 benchmarked funds during the years 2004–2015.

I use a portfolio formation approach as proposed by Cremers and Petajisto (2009) and Amihud and Goyenko (2013). Each month, I sort all funds into quintile portfolios by either selectivity, Active Share, or tracking error. I then calculate each portfolio's next month's return as the equal-weighted mean of the returns of all funds in the respective portfolio. As I do this over the entire sample period, I end up with five quintile return time series per measure. Notably, this approach does not require any forward-looking techniques as all activeness measures are calculated using historical data. I regress those return series on the Fama and French (1993) and Carhart (1997) risk factors. Table 2.3 shows the mean returns as well as the regression results for the five portfolio return time-series (low selectivity to high selectivity) where selectivity serves as the measure of activeness. The annualized mean return displayed in Panel A ranges from 13.66% for funds with low selectivity to 16.94% for funds with high selectivity, monotonically increasing in selectivity. Regressing the excess returns of each quintile portfolio on the Carhart (1997) model results in an annualized alpha between −0.69% for low selectivity and 0.39%, for high selectivity funds as displayed in Panel B of Table 2.3. Thus, the 20% most active funds generate a positive abnormal return and outperform their 20% least active counterparts by 3.28% p.a. in terms of simple returns and 1.08% p.a. after risk-adjustment. The effect is not as large as the risk-adjusted outperformance of 2.05% reported by Amihud and Goyenko (2013), yet economically (and for simple returns also statistically) significant. The risk factor loadings provide strong evidence that funds that are more active tend to have higher exposures towards the size, value and momentum factor as all three regression coefficients monotonically increase in selectivity.

I conduct the same analyses for Active Share and both measures of tracking error. Table 2.4 reports the results. Simple returns of the more active quintile portfolios are higher for all the activeness measures and returns are monotonically increasing in activeness for all quintile portfolios but the most active quintile when sorted by $TE_{implicit}$ (Panel A). When adjusting for the Fama and French (1993) and Carhart (1997) risk factor exposures, however, the outperformance of funds with high Active Share and $TE_{explicit}$ decreases to only 0.03% and 0.73%

Table 2.3: Returns and factor loadings of selectivity quintile portfolios (1980–2003)

This table reports regression results based on the 1980–2003 data. Each month, funds are sorted into quintiles by the selectivity measure, which is 1 minus the R^2 of a time-series regression of fund returns during the past 24 months on a Carhart (1997) model. The following month's fund returns are averaged within each quintile. Panel A reports annualized mean returns of the resulting time series of quintile portfolios between 1980 and 2003. Panel B reports the annualized alphas, factor loadings and R^2 of a Fama/French (1993) and Carhart (1997) time-series regression of each quintile's excess returns. The first column reports the respective values for an equal-weighted portfolio, the last column the difference between portfolios of highest and lowest selectivity. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	All funds	Low selectivity	(2)	(3)	(4)	High selectivity	High-low selectivity
	Panel A: Simple returns, 1980–2003						
Annualized Mean Return	15.09*** (4.31)	13.66*** (3.95)	14.05*** (3.95)	14.94*** (4.16)	15.85*** (4.42)	16.94*** (5.00)	3.28*** (3.02)
	Panel B: Fama/French and Carhart regressions, 1980–2003						
Annualized α	−0.49 (−0.94)	−0.69* (−1.65)	−0.82 (−1.60)	−0.75 (−1.30)	−0.57 (−0.82)	0.39 (0.56)	1.08 (1.63)
Mkt-Rf	0.98*** (95.66)	0.97*** (113.21)	0.98*** (93.74)	0.99*** (83.91)	0.99*** (69.00)	0.94*** (66.75)	−0.03** (−2.21)
SMB	0.18*** (14.00)	−0.02* (−1.82)	0.13*** (9.63)	0.21*** (14.37)	0.28*** (15.44)	0.29*** (16.23)	0.31*** (18.02)
HML	0.02 (1.25)	−0.05*** (−4.18)	−0.03** (−2.17)	0.01 (0.72)	0.07*** (3.32)	0.12*** (5.42)	0.17*** (8.28)
Mom	0.01 (1.55)	−0.02** (−2.18)	−0.01 (−1.41)	0.01 (1.34)	0.04*** (3.03)	0.05*** (4.29)	0.07*** (5.84)
R^2	0.98	0.99	0.98	0.98	0.96	0.96	0.59

(Panel B). Funds with a high $TE_{implicit}$ even underperform those with a low $TE_{implicit}$. None of the portfolios sorted on those three activeness measures has a positive alpha. Cremers and Petajisto (2009) find a performance predictive effect of Active Share only when measuring performance by the Carhart (1997) alpha of benchmark-adjusted returns.[8] I follow this procedure in Panel C of Table 2.4 and find a statistically significant outperformance of high Active Share funds of 2.64% p.a. over the low Active Share funds and still, but statistically not

[8] Cremers and Petajisto (2009) find a positive but statistically nonsignificant difference in net return alphas of 0.23% p.a. between high Active Share and low Active Share funds.

significant, 1.07% p.a. over the implicit Carhart model benchmark. Those results are comparable to the original paper.[9] Sorting on selectivity or either tracking error measure results in a significantly negative alpha of benchmark-adjusted returns for the least active funds and a negative but statistically not significant alpha for the most active funds. The outperformance of the most active over the least active funds increases to 1.61% p.a. for selectivity and 0.88% p.a. and 1.49% p.a. for the tracking error measures.

The use of benchmark-adjusted returns as presented in Panel C of Table 2.4 has raised a vigorous debate between Frazzini et al. (2016) and Cremers (2015). The former claims that funds whose benchmark index has a negative Carhart (1997) alpha tend to have a high Active Share and vice versa. Deducting benchmark returns when measuring the performance might drive the result. Therefore, they claim that the performance predictive power of Active Share is an effect of different benchmarks. For a first indication, whether this might be the case in my sample, I plot the average Active Share per benchmark index against the 1980–2003 alpha of the respective index returns. Figure 2.2 displays this graph. There indeed seems to be a negative relationship between index return alphas and the funds' Active Share as indicated by the dotted regression line.

To account for this concern, I use a different portfolio sort approach, which has also been suggested by Frazzini et al. (2016). Instead of assigning quintiles over the complete sample, I form quintiles separately for each benchmark index. I eventually still aggregate funds of all benchmarks when averaging portfolio returns but assigning quintiles by benchmark assures that the funds of each distinct benchmark are equally distributed across all five quintile portfolios.[10] Therefore, the effect stressed by Frazzini et al. (2016) is the same for all five quintile portfolios. I report results based on this alternative sorting procedure in Table 2.5. The risk-adjusted outperformance of funds with high selectivity slightly changes to 1.11% (statistically significant) and funds with a high Active Share outperform those with low Active Share by only 0.24% (excess return alpha) and 0.31% (benchmark-adjusted alpha).[11] I therefore

[9] Cremers/Petajisto (2009) report an outperformance of high Active Share funds over low Active Share funds by 2.98% and an alpha of benchmark-adjusted returns of 1.05%.

[10] Cremers (2015) argues that a within-benchmark sort changes the definition of Active Share for exactly this reason. He claims that "Active Share is basically a measure of the amount of stock picking, and small cap managers have generally more stock picking opportunities..." Therefore, he claims, there should be more small cap funds in the high Active Share quintiles.

[11] Small differences between alphas of excess returns and benchmark-adjusted returns occur when the number of funds per benchmark is not a multiple of five and therefore the number of funds per benchmark slightly varies across quintile portfolios.

Table 2.4: Returns of quintile portfolios for all activeness measures (1980–2003)

This table reports results based on the 1980–2003 data. Each month, funds are sorted into quintiles by one of four different activeness measures as indicated in the first column. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. $TE_{explicit}$ and $TE_{implicit}$ are a fund's tracking error volatility with respect to either the explicit benchmark named in a fund's prospectus or the implicit benchmark return calculated from a one-factor model. The following month's fund returns are averaged within each quintile. Panel A reports annualized mean returns of the resulting time series of quintile portfolios. Panel B reports the annualized alphas of a Fama/French (1993) and Carhart (1997) time-series regression of each quintile's excess returns. The first column reports the respective values for an equal-weighted portfolio, the last column For Panel C regressions, monthly quintile portfolio returns are constructed by averaging the funds' excess returns over the individual benchmark return and returns are regressed on Fama/French (1993) and Carhart (1997) factors. Annualized alphas are reported. The last column contains the difference between the portfolios of highest and lowest activeness. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Low activeness	(2)	(3)	(4)	High activeness	High-low activeness
Panel A: Simple returns						
Selectivity	13.66*** (3.95)	14.05*** (3.95)	14.94*** (4.16)	15.85*** (4.42)	16.94*** (5.00)	3.28*** (3.02)
Active Share	14.15*** (4.32)	14.15*** (4.23)	14.68*** (4.22)	15.76*** (4.20)	16.66*** (4.26)	2.51 (1.36)
$TE_{explicit}$	14.12*** (4.43)	14.87*** (4.53)	14.89*** (4.39)	15.26*** (4.30)	15.96*** (3.84)	1.83 (1.00)
$TE_{implicit}$	13.22*** (4.23)	14.23*** (4.45)	15.85*** (4.78)	16.33*** (4.58)	15.52*** (3.36)	2.31 (0.91)
Panel B: Fama/French and Carhart alphas						
Selectivity	−0.69* (−1.65)	−0.82 (−1.60)	−0.75 (−1.30)	−0.57 (−0.82)	0.39 (0.56)	1.08 (1.63)
Active Share	−0.17 (−0.50)	−0.58 (−1.50)	−0.86 (−1.56)	−0.60 (−0.83)	−0.14 (−0.15)	0.03 (0.03)
$TE_{explicit}$	−0.92** (−2.25)	−0.37 (−0.80)	−0.50 (−0.93)	−0.61 (−0.98)	−0.19 (−0.24)	0.73 (1.00)
$TE_{implicit}$	−0.83 (−2.66)	−0.47 (−1.13)	−0.05 (−0.08)	−0.06 (−0.07)	−1.17 (−1.13)	−0.34 (−0.39)
Panel C: Fama/French and Carhart alphas of benchmark-adjusted returns						
Selectivity	−1.70*** (−2.64)	−1.34** (−2.05)	−0.87 (−1.27)	−0.98 (−1.29)	−0.09 (−0.11)	1.61*** (2.78)
Active Share	−1.57** (−2.39)	−1.79*** (−2.71)	−1.90*** (−2.61)	−0.60 (−0.81)	1.07 (1.09)	2.64*** (2.74)
$TE_{explicit}$	−1.42*** (−2.94)	−1.14* (−1.77)	−1.24* (−1.76)	−1.01 (−1.27)	−0.54 (−0.59)	0.88 (1.28)
$TE_{implicit}$	−1.99*** (−2.85)	−1.71** (−2.54)	−1.00 (−1.37)	−0.06 (−0.07)	−0.51 (−0.59)	1.49* (1.82)

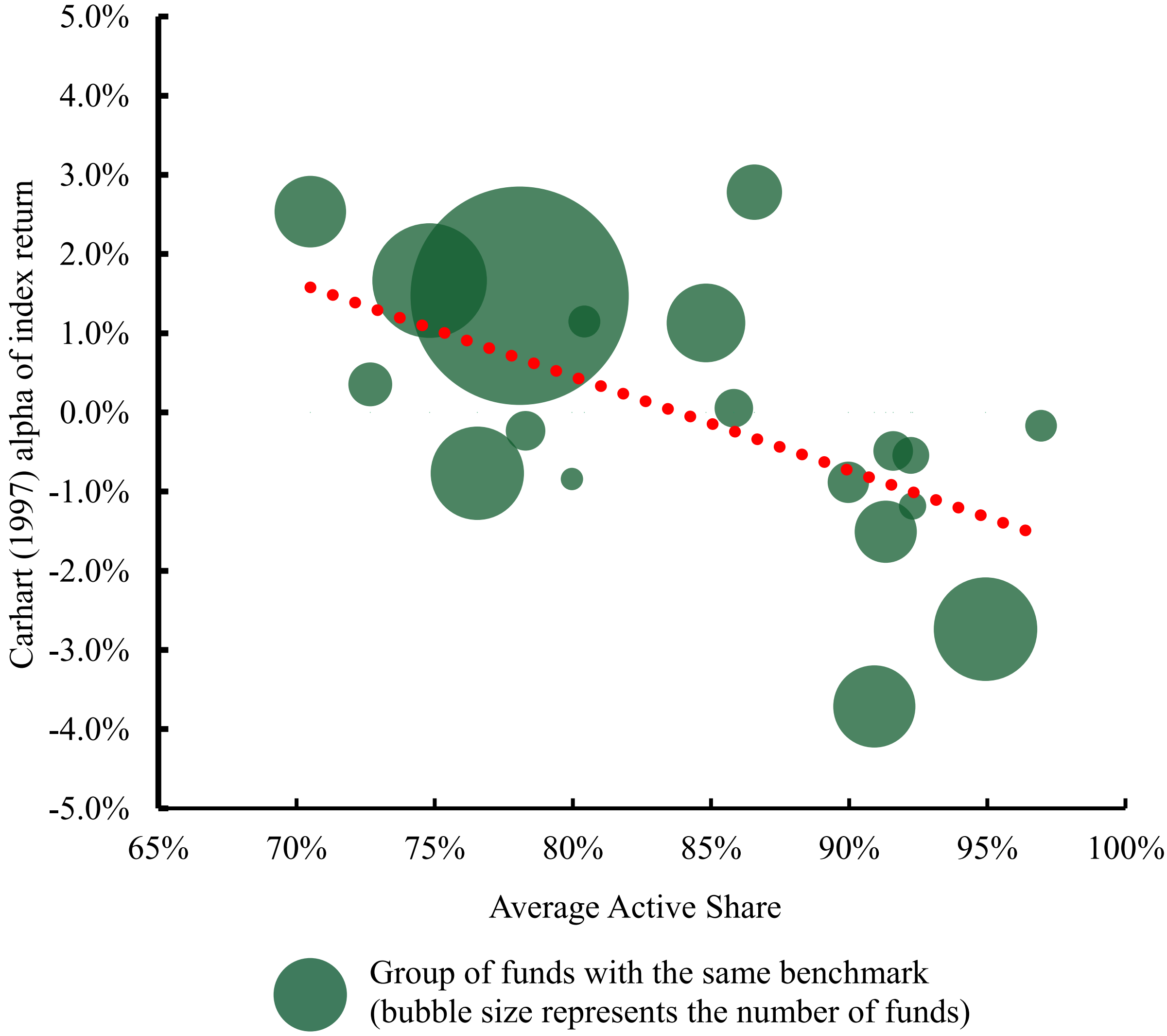
Table 2.5: Returns of benchmark-specific sorted quintile portfolios for all activeness measures (1980–2003)

This table reports results based on the 1980–2003 data. Each month, funds are sorted into quintiles by one of four different activeness measures as indicated in the first column. For each benchmark index the sorting is executed separately across funds with this benchmark. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. $TE_{explicit}$ and $TE_{implicit}$ are a fund's tracking error volatility with respect to either the explicit benchmark named in a fund's prospectus or the implicit benchmark return calculated from a one-factor model. The following month's fund returns are averaged within each quintile. Panel A reports the annualized alphas of a Fama/French (1993) and Carhart (1997) time-series regression of each quintile's excess returns over the risk-free rate. The first column reports the respective values for an equal-weighted portfolio, the last column For Panel B regressions, monthly quintile portfolio returns are constructed by averaging the funds' excess returns over the individual benchmark return and returns are regressed on Fama/French (1993) and Carhart (1997) factors. Annualized alphas are reported. The last column contains the difference between the portfolios of highest and lowest activeness. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Low activeness	(2)	(3)	(4)	High activeness	High-low activeness
Panel A: Fama/French and Carhart alphas						
Selectivity	−0.97**	−0.50	−0.63	−0.53	0.14	1.11**
	(−2.27)	(−1.02)	(−1.15)	(−0.82)	(0.21)	(2.03)
Active Share	−0.41	−0.83*	−0.85	−0.12	−0.17	0.24
	(−1.04)	(−1.88)	(−1.60)	(−0.18)	(−0.22)	(0.34)
$TE_{explicit}$	−0.93**	−0.42	−0.53	−0.40	−0.34	0.60
	(−2.41)	(−0.92)	(−0.94)	(−0.66)	(−0.47)	(1.00)
$TE_{implicit}$	−0.56*	−0.66*	−0.58	−0.39	−0.38	0.18
	(−1.49)	(−1.46)	(−1.14)	(−0.61)	(−0.49)	(0.27)
Panel B: **Fama/French and Carhart alphas of benchmark-adjusted returns**						
Selectivity	−1.51***	−1.04	−1.16*	−1.08	−0.38	1.13**
	(−2.59)	(−1.62)	(−1.69)	(−1.34)	(−0.48)	(2.09)
Active Share	−0.94*	−1.33**	−1.38**	−0.65	−0.63	0.31
	(−1.74)	(−2.26)	(−2.04)	(−0.82)	(−0.67)	(0.44)
$TE_{explicit}$	−1.54***	−0.99	−1.08	−0.96*	−0.87	0.67
	(−2.86)	(−1.62)	(−1.51)	(−1.24)	(−1.00)	(1.16)
$TE_{implicit}$	−1.16**	−1.19*	−1.13*	−0.95	−0.90	0.26
	(−2.07)	(−1.95)	(−1.68)	(−1.19)	(−0.99)	(0.40)

Figure 2.2: Active Share by benchmark index and benchmark index alpha (1980–2003)

This graph plots the average Active Share of all funds from my 1980–2003 data sample with the same benchmark index and the alpha from a regression of the respective benchmark returns on the Fama/French (1993) and Carhart (1997) factors during 1980–2003. Each dot represents a group of funds with the same benchmark index. The size of each circle represents the number of funds with the specific benchmark. The dotted line represents the results of an (unweighted) regression of index alpha on average Active Share.



conclude that approximately 90% of the effect of Active Share can be attributed to the effect of different benchmarks. Notably, after applying the benchmark-specific sorting procedure no single quintile portfolio has a positive alpha. Whereas this result is in line with the findings of Frazzini et al. (2016), who report the return difference of high Active Share funds to become nonsignificant under this sorting procedure, the within-benchmark sorting procedure does not harm the return predictive power of selectivity. The relationship between selectivity and fund performance does not seem to be driven by different benchmark indices.

So far, my data contains a variety of funds with different investment styles and benchmark indices. As pointed out above, about 40% of all observations refer to funds that are benchmarked against the S&P 500. To diminish the heterogeneity of funds in my sample and at the same time retain a large sample size, I redo the analysis for the subsample of all those funds, whose prospectus names the S&P 500 as the primary benchmark. As discussed by Cremers (2017), one might expect that the performance difference between high active share and low active share funds is more pronounced among small cap funds if there is less price efficiency in low cap markets. Cremers's (2017) empirical evidence, however, does not confirm this hypothesis as he finds a stronger performance difference among large cap funds. This additionally legitimates the reduction of the sample to S&P 500 benchmarked funds. I repeat the quintile portfolio sorting approach and calculate alphas for both the excess returns over the risk-free rate and over the benchmark returns. Table 2.6 summarizes the results. The most active funds outperform the least active funds by 0.62% p.a. after risk-adjustment when sorted by selectivity and 0.38% p.a. when sorted by Active Share. Tracking error based sorting even induces an outperformance of the most active funds by up to 0.94%. Differences between two quintile portfolios are by construction the same for excess returns and benchmark-adjusted returns. Although none of those differences is statistically significant, the most active funds outperform their least active counterparty when sorted by either measure of activeness. Furthermore, the portfolios' alphas increase monotonically for most quintiles.

To test whether this observation holds during later time periods, I consider my second dataset consisting of S&P 500 benchmarked funds during the 2004–2015 time period. I follow the sorting procedure described above and report alphas in Table 2.7. Most prominently, I find that selectivity loses its return predictive effect and the most active funds even underperform the least active funds by 0.92% p.a. after risk-adjustment (statistically not significant). Lagged $TE_{explicit}$ and $TE_{implicit}$ also have a negative effect on fund returns (statistically significant for $TE_{explicit}$). Only the effect of Active Share remains positive but statistically nonsignificant at 0.62% p.a. The risk-adjusted return of quintile portfolios does, however, not increase monotonically in Active Share. I also merge the two samples of S&P 500 benchmarked funds covering the years 1980–2003 and 2004–2015. Unreported results show that the most active funds outperform the least active funds by 0.40% and 0.76% over the entire 36 years when sorted by selectivity and Active Share, respectively (neither difference is statistically significant).

Table 2.6: Returns of quintile portfolios of S&P 500 benchmarked funds for all activeness measures (1980–2003)

This table reports results on a subsample of the 1980–2003 data consisting of all S&P 500 benchmarked funds. Each month, funds are sorted into quintiles by one of four different activeness measures as indicated in the first column. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. $TE_{explicit}$ and $TE_{implicit}$ are a fund's tracking error volatility with respect to either the explicit benchmark named in a fund's prospectus or the implicit benchmark return calculated from a one-factor model. The following month's fund returns are averaged within each quintile. Panel A reports the annualized alphas of a Fama/French (1993) and Carhart (1997) time-series regression of each quintile's excess returns over the risk-free rate. For Panel B regressions, the return of the S&P 500 is subtracted from monthly quintile portfolio returns and the differences are regressed on Fama/French (1993) and Carhart (1997) factors. Annualized alphas are reported. The last column contains the difference between the portfolios of highest and lowest activeness. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Low activeness	(2)	(3)	(4)	High activeness	High-low activeness
Panel A: Fama/French and Carhart Alphas						
Selectivity	−0.62* (−1.71)	−0.14 (−0.33)	−0.33 (−0.68)	−0.10 (−0.15)	0.00 (0.00)	0.62 (0.92)
Active Share	−0.04 (−0.13)	−0.70** (−2.02)	−0.69 (−1.43)	0.04 (0.06)	0.33 (0.36)	0.38 (0.41)
$TE_{explicit}$	−0.88** (−2.35)	−0.09 (−0.18)	−0.22 (−0.36)	−0.21 (−0.32)	−0.11 (−0.15)	0.77 (0.97)
$TE_{implicit}$	−0.72** (−2.25)	−0.36* (−1.00)	−0.48 (−0.91)	−0.01 (−0.01)	0.21 (0.25)	0.94 (1.08)
Panel B: Fama/French and Carhart Alphas of Benchmark-Adjusted Returns						
Selectivity	−2.07*** (−2.29)	−1.59* (−1.67)	−1.78* (−1.77)	−1.56* (−1.40)	−1.45 (−1.24)	0.62 (0.92)
Active Share	−1.50* (−1.67)	−2.14** (−2.30)	−2.13** (−2.19)	−1.42 (−1.30)	−1.12 (−0.86)	0.38 (0.41)
$TE_{explicit}$	−2.33*** (−2.65)	−1.54* (−1.61)	−1.67* (−1.61)	−1.66* (−1.55)	−1.56 (−1.24)	0.77 (0.97)
$TE_{implicit}$	−2.17** (−2.39)	−1.81** (−1.97)	−1.93** (−1.96)	−1.46 (−1.31)	−1.24 (−0.95)	0.94 (1.08)

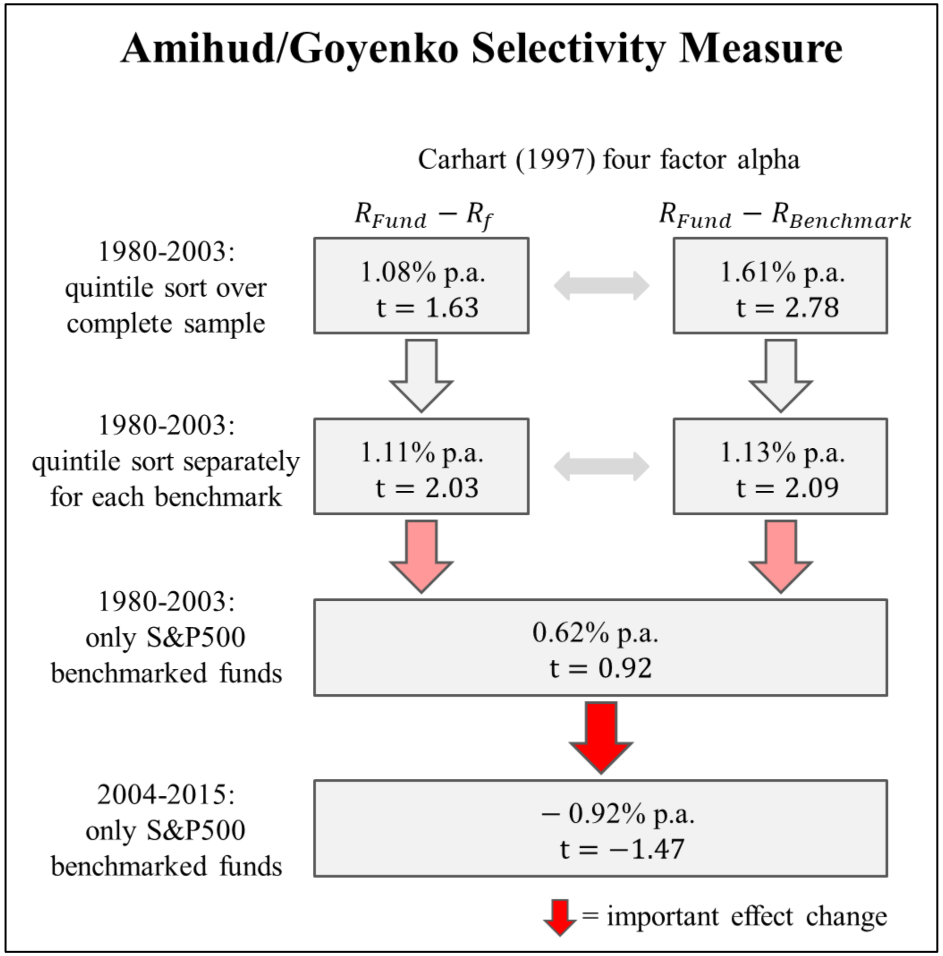
Table 2.7: Returns of quintile portfolios of S&P 500 benchmarked funds for all activeness measures (2004–2015)

This table reports results on a sample of 2004–2015 data containing S&P 500 benchmarked funds. Each month, funds are sorted into quintiles by one of four different activeness measures as indicated in the first column. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. $TE_{explicit}$ and $TE_{implicit}$ are a fund's tracking error volatility with respect to either the explicit benchmark named in a fund's prospectus or the implicit benchmark return calculated from a one-factor model. The following month's fund returns are averaged within each quintile. Panel A reports the annualized alphas of a Fama/French (1993) and Carhart (1997) time-series regression of each quintile's excess returns over the risk-free rate. For Panel B regressions, the return of the S&P 500 is subtracted from monthly quintile portfolio returns and the differences are regressed on Fama/French (1993) and Carhart (1997) factors. Annualized alphas are reported. The last column contains the difference between the portfolios of highest and lowest activeness. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Low activeness	(2)	(3)	(4)	High activeness	High-low activeness
Panel A: **Fama/French and Carhart Alphas**						
Selectivity	−0.96***	−1.16***	−1.04***	−1.12***	−1.88***	−0.92
	(−4.63)	(−4.30)	(−2.93)	(−2.99)	(−2.95)	(−1.47)
Active Share	−1.01***	−1.13***	−1.48***	−1.53***	−0.40	0.62
	(−5.33)	(−4.88)	(−5.62)	(−3.53)	(−0.56)	(0.86)
$TE_{explicit}$	−0.35	−0.80***	−1.45***	−1.38***	−1.88**	−1.53*
	(−1.17)	(−3.08)	(−5.17)	(−3.57)	(−2.19)	(−1.67)
$TE_{implicit}$	−0.86***	−1.12***	−1.05***	−1.38***	−1.47*	−0.61
	(−4.27)	(−3.96)	(−4.01)	(−3.33)	(−1.72)	(−0.71)
Panel B: **Fama/French and Carhart Alphas of Benchmark-Adjusted Returns**						
Selectivity	−0.84	−1.04	−0.92	−1.00	−1.76	−0.92
	(−0.44)	(−0.55)	(−0.47)	(−0.51)	(−0.84)	(−1.47)
Active Share	−0.89	−0.79	−1.36	−1.41	−0.27	0.62
	(−0.48)	(−0.41)	(−0.71)	(−0.72)	(−0.13)	(0.86)
$TE_{explicit}$	−0.33	−0.79	−1.43	−1.36	−1.87	−1.53*
	(−0.16)	(−0.39)	(−0.72)	(−0.67)	(−0.83)	(−1.67)
$TE_{implicit}$	−0.85	−1.10	−1.03	−1.36	−1.45	−0.61
	(−0.43)	(−0.56)	(−0.51)	(−0.67)	(−0.64)	(−0.71)

Figure 2.3: Summary of the difference of the four-factor alpha of the most active funds over the least active funds

These schemes display the alphas and t-values resulting from Fama/French (1993) and Carhart (1997) four-factor regressions following the quintile portfolio sorts based on Active Share and selectivity. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. Alphas of the return differences of most active and least active fund quintiles are shown for the entire 1980–2003 data set sorted over the entire sample and separately by benchmark index, the subsample of S&P 500 benchmarked funds as well as the funds from my 2004–2015 data sample. The results are shown for regression of excess returns over the risk-free rate as well as excess return over the respective benchmark index.



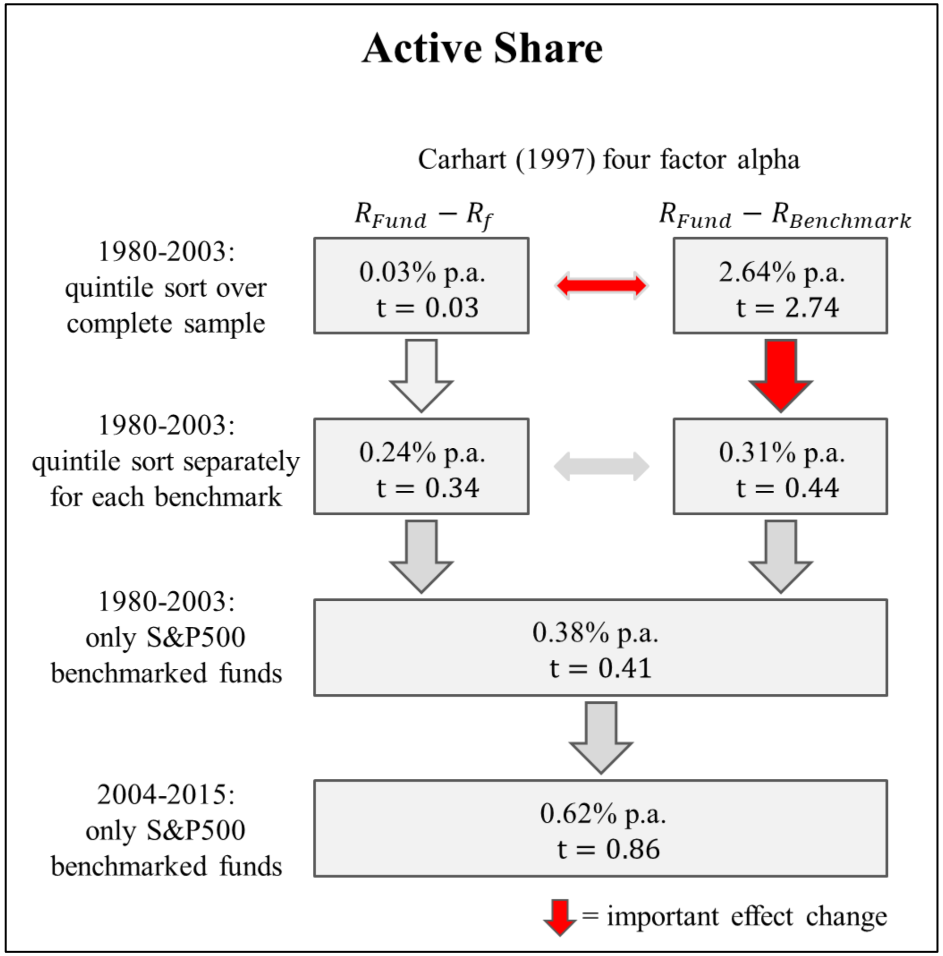
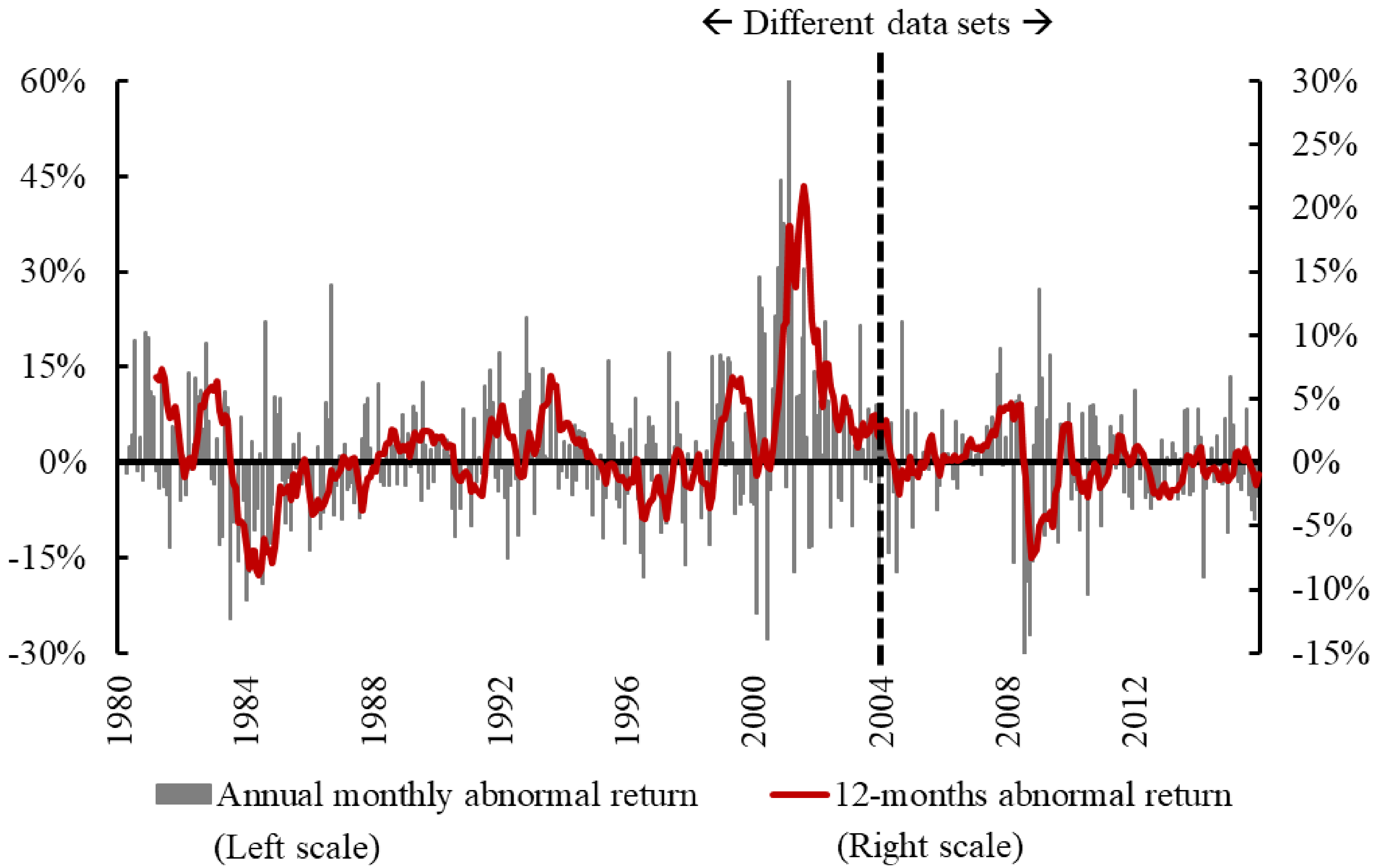


Figure 2.4: Risk-adjusted return difference between high selectivity and low selectivity funds over time (1980–2015)

This graph displays the monthly abnormal return differences between high selectivity and low selectivity funds as well as their 12-month moving average. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. The monthly abnormal return difference is the sum of the monthly residual and the constant alpha from a time-series regression of the return differences between high selectivity (top quintile) and low selectivity (lowest quintile) funds on the Fama/French (1993) and Carhart (1997) factors. Grey bars show the annualized value per month, the red line shows the 12-month rolling regression value. The data is combined from my two data sets (1980–2003 and 2004–2015) and covers all S&P 500 benchmarked funds.



The results so far question the predictive power of activeness. Active Share loses much of its power when controlling for different benchmark indices. The return predictive power of selectivity, $TE_{explicit}$ and $TE_{implicit}$ even changes signs in recent years. I summarize the results for selectivity and Active Share in Figure 2.3. In the case of the selectivity measure, I have discovered a major change of the predictive power of high selectivity in recent years. I use the regression results of the merged 1980–2015 sample of S&P 500 benchmarked funds to visualize the time series of abnormal returns as follows. Each month, I add the residual to the regression alpha to determine a monthly abnormal return. I plot those values as well as the 12-month moving average in Figure 2.4. The graph indicates that large parts of the outperformance of high selectivity funds occurred during the early 2000s. In unreported results, I regress the return difference of the most and the least active quintile portfolios during the 1980–2015 time series on the Fama and French (1993) and Carhart (1997) risk factors but leave out the years 2000–2002. The alpha is 0.04% p.a. I conclude that there is no outperformance apart from the years 2000–2002. The same regression only during the years 2000−2002 returns an alpha of 3.13% p.a.

2.6 Channels of activity of the most active funds

Having analyzed the return predictive power of Active Share and selectivity, I subsequently study how the most active funds invest in order to beat their benchmark. I investigate whether exposures to other, less established risk factors can help explain return differences between the most and the least active funds.

So far, I have controlled for the Fama and French (1993) and Carhart (1997) risk factors to distinguish abnormal returns from simple risk factor loading strategies. Academic research has proposed a variety of alternative risk factors. I apply some of those alternative risk factor models, namely the Fama and French (2015) five factor model, the model of Asness et al. (2014) including a quality factor, the model of Frazzini and Pedersen (2014) including a betting-against-beta factor and a model controlling for an illiquidity risk factor (Amihud and Noh, 2016). I measure the factor exposures using those models to investigate the investment strategies of active fund managers. In a first step, I use the merged sample from the 1980–2003 and the 2004–2015 datasets containing all S&P 500 benchmarked funds. Table 2.8 reports the regression results for the portfolio buying the 20% most active funds and short selling the 20% least active funds by either selectivity or Active Share. I find that high selectivity funds have a significantly higher exposure to quality factors, either measured by the Fama and French (2015) profitability factor or the Asness et al. (2014) quality factor. They also invest more into conservative investment stocks. The positive loading on the betting-against-beta factor together with the negative market factor loading indicates that most active fund managers invest into small-beta stocks. All those factor exposures persist when I control for all factors at a time (I leave out the quality-minus-junk factor as it is highly correlated to the profitability factor), and I additionally measure a slightly negative exposure towards the illiquidity risk factor, that is high selectivity funds hold more liquid stocks than their less active counterparts. Those results are similar when I apply the risk factor models on the Active Share quintile portfolios. Loadings on profitability, quality, and the low-beta anomaly remain positive, investment and liquidity loadings are close to zero. For both activeness measures the model's R^2 increases noticeably when adding the Fama and French (2015) profitability factor, thus suggesting that investments into stocks with robust profitability help explain the time-series return variance. Adding the profitability factor also eliminates the (statistically not significant) outperformance of the most active quintile portfolios over the least active quintile.

Based on my observation that more actively managed funds are exposed to additional risk factors, I investigate whether this effect can explain the return differences I discovered in Section 2.5, especially the return predictive power of selectivity during the pre−2004 period, its opposite effect after 2003 and the outperformance of high Active Share funds when measuring benchmark-adjusted alphas. I therefore repeat all the portfolio sorts from above, that is on the 1980–2003 sample of funds with heterogeneous benchmarks and the 2004–2015 sample of S&P 500 benchmarked funds. Funds are sorted by either selectivity or Active Share and for the first data sample I apply the sorting over the entire sample as well as separately per benchmark index. Unlike before, I regress the portfolios' excess return over the risk-free rate and over the benchmark return on an eight-factor model including the Fama and French (1993) and Carhart (1997) risk factors as well as the profitability, investment, betting-against-beta and illiquidity risk factors. The results are illustrated in Figure 2.5. The 20% funds with the highest selectivity, i.e. the lowest R^2, still outperform the 20% funds with the lowest selectivity, i.e. the highest R^2, during the 1980–2003 time period. This result holds when funds are sorted by selectivity either over the entire sample or benchmark-wise and when the return is either benchmark-adjusted or not. The outperformance, however, is not statistically significant in any setting, and drops below 1.0% p.a.[12]

Table 2.8: Multi-factor alpha of activeness quintile portfolios of S&P 500 benchmarked funds (1980–2015)

This table reports regression results based on a dataset of S&P 500 benchmarked funds merged from my 1980–2003 and 2004–2015 datasets. Each month, funds are sorted into quintiles by either selectivity (panel A) or Active Share (panel B). Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. The following month's fund returns are averaged within each quintile. This table reports the alphas of return differences between the 5th and the 1st quintile (high minus low activeness). Besides the Fama/French (1997) and Carhart (1997) risk factors, the alpha is calculated using the Fama/French (2015) five-factor model, the Asness et al. (2014) quality-minus-junk factor, the Frazzini/Pedersen (2014) betting-against-beta factor and the Amihud/Noh (2016) illiquidity factor. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High Selectivity (Q5) minus Low Selectivity (Q1) Funds						
Annualized	0.74	−0.33	−0.06	0.17	0.54	−0.33
α	(1.14)	(−0.52)	(−0.09)	(0.26)	(0.82)	(−0.52)
Mkt-Rf	−0.05***	−0.04***	−0.04***	−0.05***	−0.05***	−0.05***
	(−5.10)	(−3.61)	(−3.25)	(−5.31)	(−4.31)	(−4.29)
SMB	0.16***	0.19***	0.18***	0.16***	0.14***	0.23***
	(11.26)	(12.62)	(10.68)	(11.38)	(5.95)	(8.60)
HML	0.12***	0.08***	0.13***	0.09***	0.12***	0.07***
	(8.37)	(4.09)	(8.64)	(5.53)	(6.85)	(3.13)
Mom	0.01		0.01	−0.00	0.01	−0.01
	(1.30)		(0.69)	(−0.25)	(1.36)	(−1.32)
CMA		0.07**				0.08**
		(2.59)				(2.57)
RMW		0.11***				0.10***
		(5.57)				(4.75)
QMJ			0.05**			
			(2.08)			
BAB				0.06***		0.05***
				(4.44)		(3.59)
IML					0.04	−0.10*
					(0.92)	(−1.94)
R^2	0.32	0.37	0.33	0.35	0.32	0.39

(continued)

[12] As presented above, the outperformance is statistically significant (up to the 1% level) and lies between 1.08% p.a. and 1.61% p.a. when using only the Fama and French (1993) and Carhart (1997) risk factors.

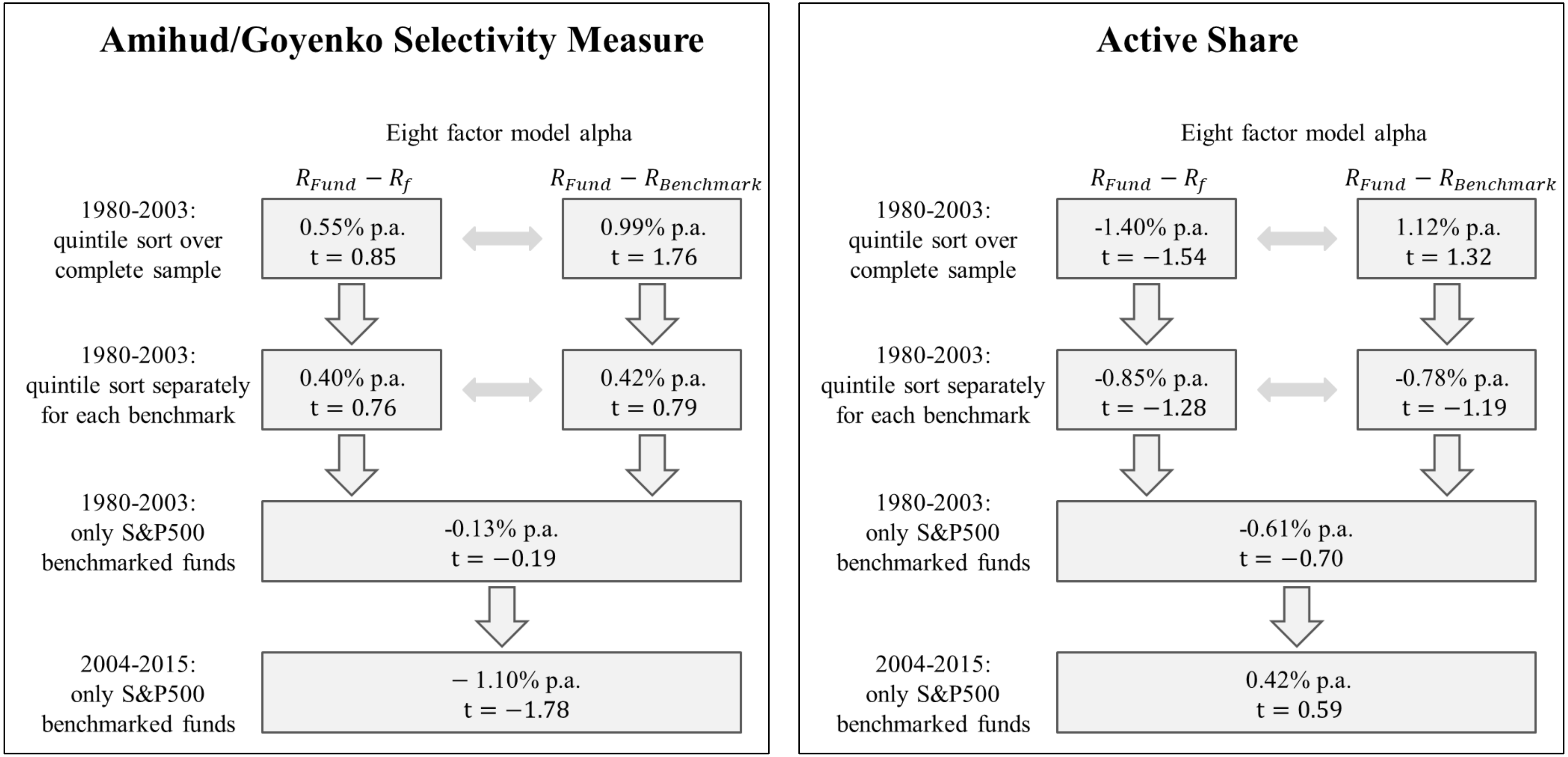
(continued)

Panel B: High Active Share (Q5) minus Low Active Share (Q1) Funds						
Annualized α	0.75 (1.15)	−0.32 (−0.51)	−0.06 (−0.08)	0.18 (0.27)	0.54 (0.82)	−0.32 (−0.51)
Mkt-Rf	0.03* (2.06)	0.05*** (3.57)	0.05*** (3.57)	0.03* (2.01)	0.04** (2.53)	0.03** (2.34)
SMB	0.37*** (20.31)	0.40*** (22.30)	0.40*** (19.25)	0.36*** (20.64)	0.32*** (10.70)	0.46*** (13.93)
HML	0.07*** (3.84)	0.07*** (3.04)	0.09*** (4.46)	0.03 (1.39)	0.06** (2.49)	0.03 (1.29)
Mom	−0.02 (−1.61)		−0.03** (−2.51)	−0.04*** (−3.08)	−0.02 (−1.51)	−0.05*** (−3.84)
CMA		−0.02 (−0.49)				0.00 (0.01)
RMW		0.17*** (6.95)				0.17*** (6.63)
QMJ			0.10*** (3.41)			
BAB				0.07*** (4.51)		0.05*** (3.01)
IML					0.10 (1.74)	−0.08 (−1.30)
R^2	0.51	0.56	0.53	0.54	0.52	0.58

For the sample of S&P 500 benchmarked funds, the more active funds (by selectivity) underperform the less actively managed funds before and after 2003 and, as before, this underperformance is larger during recent years. If I measure activeness by a fund's Active Share, the annual performance difference between the most and the least active fund quintiles of the 1980–2003 sample is reduced by 1.0−1.5 percentage points when applying the eight-factor model. Thus, the most active funds underperform the least active funds in terms of the eight-factor alpha when sorted by Active Share within each benchmark index group or when comparing excess returns over the risk-free rate. Only when sorting funds over the entire sample and using benchmark-adjusted returns, the most active funds still beat the least active funds when applying the eight-factor model. None of these differences, however, remains statistically significant at the 10% level. For the samples of S&P 500 benchmarked funds, the performance difference is −0.61% p.a. between 1980 and 2003 and 0.42% p.a. between 2004 and 2015, neither of which is statistically significant.

Figure 2.5: Summary of the difference of the eight-factor alpha of the most active funds over the least active funds

These schemes display the alphas and t-values resulting from regressions following the quintile portfolio sorts based on Active Share and selectivity. Selectivity is 1 minus the R^2 of a time-series regression of fund returns on a Carhart (1997) model. Active Share is the deviation of portfolio holdings from the fund's benchmark as described in Section 2.2.2. Alphas of the return differences of most active and least active fund quintiles are shown for the entire 1980–2003 data set sorted over the entire sample and separately by benchmark index, the subsample of S&P 500 benchmarked funds as well as the funds from my 2004–2015 data sample. The alphas are calculated using the five Fama/French (2015) factors, the Carhart (1997) momentum factor, the Frazzini/Pedersen (2014) betting-against-beta factor and the Amihud/Noh (2016) illiquidity factor. The results are shown for regression of excess returns over the risk-free rate as well as excess return over the respective benchmark index.



Overall, the addition of the profitability, investment, betting-against-beta and illiquidity risk factors lowers the performance difference between the most and the least active funds. This change mostly results from a significantly higher exposure of the most active funds to the profitability risk factor prior to 2004 (unreported results). This higher exposure cannot be observed in the 2004–2015 sample but a significant exposure of the more active funds to the betting-against-beta factor is measured during this time period. Additionally, the more active funds have a negative exposure to the illiquidity factor during 2004–2015. I can thus conclude that high Active Share and high selectivity funds deviate from the benchmark by loading on additional risk factors, in particular on profitability during the 1980–2003 period and on betting-against-beta as well as negatively on illiquidity afterwards. The additional risk factor exposure causes large parts of the abnormal returns of the most active funds that are reported in earlier research.

2.7 Conclusions

Prior research claims that measures of fund activeness can predict future fund returns. I challenge those findings by eliminating the fund heterogeneity in the data sample and investigating the effect in a more recent data set. My findings suggest that fund activeness is not a reliable indicator of superior future fund performance. My findings on Active Share support the claim of Frazzini et al. (2016) that large parts of the positive effects of Active Share are due to different benchmark indices in a mixed fund data sample. When controlling for these different benchmark indices, the return predictive power of Active Share becomes statistically nonsignificant and no single quintile portfolio realizes a positive net-of-fee abnormal return.

With respect to the selectivity measure, I find a return predictive power prior to 2003 but a statistically nonsignificant negative relationship afterwards. My further analyses suggest that high selectivity funds have performed remarkably well during the burst of the dot-com bubble, that is 2000−2002, but rather poorly throughout all the other years. Considering the by far largest subsample of S&P 500 benchmarked funds over the entire sample period, I find a positive, yet statistically nonsignificant outperformance of high selectivity as well high Active Share funds over their least active counterparts.

I study the investment behavior of the most active funds and find that highly active funds by either selectivity or Active Share are exposed to alternative risk factors, especially profitability and betting-against-beta. This additional result can be interpreted in two ways. Active fund managers either generate value from market anomalies or expose their portfolio to additional risk, which the investor is compensated for. Depending on this judgement one might hope to achieve a marginally better performance by choosing funds with high measures of activeness. The outperformance of the most active funds as discovered by prior research, however, does not withstand my analyses throughout different fund universes and time periods.

References

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

Amihud Y., Goyenko R., 2013. Mutual fund R^2 as predictor of performance. Review of Financial Studies, 2, 667–694.

Amihud, Y., Noh, J., 2016. The pricing of the illiquidity factor's systematic risk. Working paper.

Asness, C. S., Frazzini, A., Pedersen, L. H., 2014. Quality minus junk. Working paper.

Berk J., Green R., 2004. Mutual fund flows and performance in rational markets. Journal of Political Economy, 112, 1269–1295.

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

Cremers, M., 2015. AQR in wonderland: Down the rabbit hole of 'Deactivating Active Share' (and back out again?). Working Paper.

Cremers, M., 2017. Active Share and the three pillars of active management: skill, conviction, and opportunity. Financial Analysts Journal, 73(2), 61−79.

Cremers, M., Ferreira, M., Matos, P., Starks, L., 2016. Indexing and active fund management: International evidence. Journal of Financial Economics, 120, 539−560.

Cremers, M., Pareek, A., 2016. Patient capital outperformance: The investment skill of high active share managers who trade infrequently. Journal of Financial Economics, 122, 288−306.

Cremers, M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies, 22, 3329–3365.

Cremers, M., Petajisto, A., Zitzewitz, E., 2013. Should benchmark indices have alpha? Revisiting performance evaluation. Critical Finance Review, 2, 1–48.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance, 52, 1035–1058.

Doshi, H., Elkamhi, R., Simutin, M., 2015. Managerial activeness and mutual fund performance. Review of Asset Pricing Studies, 5, 156−184.

ESMA (2016). Supervisory work on potential closet index tracking. ESMA, 2016/65.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics, 33, 3−56.

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

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics, 116, 1−22.

Frazzini, A., Friedman, J., Pomorski, L., 2016. Deactivating Active Share. Financial Analysts Journal, 72(2), 14−21.

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

Grinblatt, M., Titman, S., 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. Journal of Business, 62, 393–416.

Grinblatt, M., Titman, S., 1993. Performance measurement without benchmarks: An examination of mutual fund returns. Journal of Business, 66, 47–68.

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

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

Malkiel, B. G., 1995. Returns from investing in equity mutual funds 1971 to 1991. Journal of Finance, 50, 77–91.

Petajisto, A., 2013. Active Share and mutual fund performance. Financial Analysts Journal, 69(4), 73−93.

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

Wermers, R., 2003. Are mutual fund shareholders compensated for active management "bets"?. Working Paper.

Chapter 3

Risk Factor Exposure Variation and Mutual Fund Performance

Manuel Ammann, Sebastian Fischer and Florian Weigert*

Status: Working Paper (February 2019)

Abstract

We investigate the relationship between a mutual fund's variation in systematic risk factor exposures and its future performance. Using a dynamic state space version of Carhart's (1997) four-factor model to capture risk factor variation, we find that funds with volatile risk factor exposures underperform funds with stable risk factor exposures by 147 basis points p.a. This underperformance is neither explained by volatile risk factor loadings of a fund's equity holdings nor driven by a fund's forced trading through investor flows. We conclude that fund managers voluntarily attempt to time risk factors, but are unsuccessful at doing so.

JEL Classification: G11, G14, G20, G23

Keywords: Mutual Fund, Market Timing, Factor Timing, Kalman Filter

* We thank Yakov Amihud, Lars Kaiser, Stefan Ruenzi, Sebastian Stöckl, Erik Theissen, Marco Wilkens, the participants of the Annual Meeting of the German Finance Association 2017, the Cologne Colloquium on Financial Markets 2018, the Lancaster Frontiers of Factor Investing Conference 2018, the participants at the research seminars of the University of Liechtenstein, the Universities of Constance and the University of St. Gallen for their helpful comments. All errors are our own.

3.1 Introduction

Among academics, there is a widely accepted consensus that mutual funds, on average, generate small positive abnormal gross returns, but fail to beat a risk-adjusted benchmark net of fees.[1] Therefore, the focus of the academic mutual fund literature has moved to the question which investment and fund characteristics lead to future abnormal returns and whether there are indicators that identify top-performers ex ante. To achieve the goal of future benchmark-adjusted outperformance, a fund manager can generally pursue three different investment approaches. First, she can expose the fund to alternative risk factors, such as liquidity risk (Pástor and Stambough, 2003, and Dong et al., 2017), volatility risk (Ang et al., 2006), or tail risk (Kelly and Jiang, 2014, and Chabi-Yo et al., 2018) to earn the associated risk premium.[2] Second, she can deviate from the benchmark portfolio and engage in stock picking, i.e., tilt her portfolio towards stocks that are likely to outperform in the future (see Wermer, 2000, and Cremers and Petajisto, 2009). Third, the fund manager can vary her exposure to systematic risk factors, i.e., increase (decrease) her exposure to a risk factor when it is likely to pay a high (low) premium in the future. Our paper is concerned with the latter investment approach and establishes a comprehensive framework to study the relationship between the volatility of a fund's exposure to different systematic risk factors and future performance.

To measure a mutual fund's variation in systematic risk factor exposures, we propose to apply the Carhart (1997) four-factor model with time-varying exposures that follow a mean-reverting process. We choose to apply the Carhart (1997) model because it is widely used to measure mutual fund performance.[3] Its systematic risk factors are the market return (*MKT*) factor, the size (*SMB*) factor, the book-to-market (*HML*) factor, and the momentum (*UMD*) factor. To estimate our model we use a Kalman filter and Kalman smoother technique. We apply the model to a period of 3 years of weekly return data in a rolling manner and measure the volatility of the factor loadings during this estimation period to the *MKT* factor, the *SMB*

---

[1] Among others, Malkiel (1995), Gruber (1996), Carhart (1997), and Fama and French (2010) document that mutual funds underperform their respective benchmark net of fees. Using detailed portfolio holdings data, Grinblatt and Titman (1989, 1993) and Daniel et al. (1997) observe that gross of fees, mutual funds generate positive abnormal returns. Wermers (2000) combines both views and shows that mutual funds exhibit positive stock picking ability, which is – however – too low to cover expenses and transaction costs.

[2] Of course, this approach is only suitable if the mutual fund's benchmark does not account for these alternative risk factors.

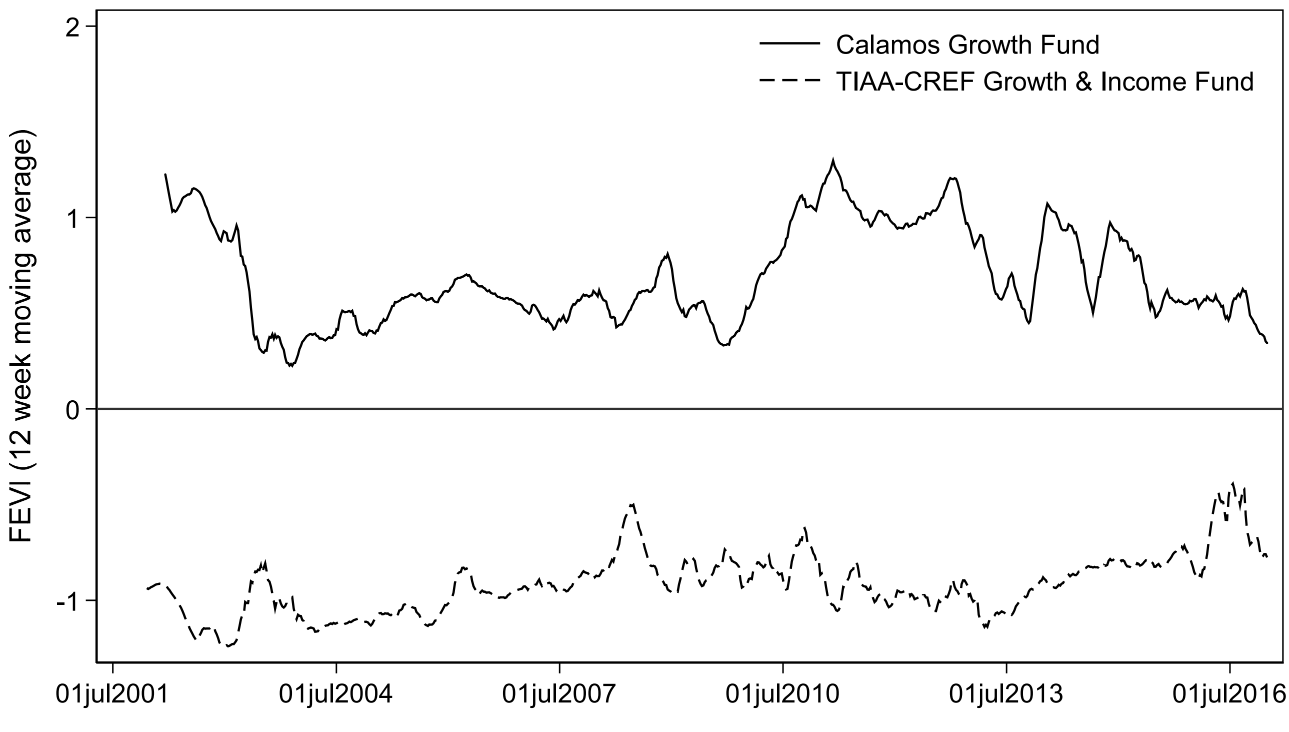
[3] See, e.g, Berk and van Binsbergen (2015) and Barber et al. (2016). Our results are stable for alternative factor models such as the Fama and French (1993) three-factor model and the Fama and French (2015) five-factor model.

factor, the *HML* factor, and the *UMD* factor. To express a fund's total level of factor exposure variation, we compute an overall *Factor Exposure Volatility Indicator (FEVI)* by averaging and standardizing the individual market, size, value, and momentum volatility measures.

To illustrate the concept of systematic risk factor variation and the relevance of the *FEVI* measure, we provide an example of two large and well-established equity mutual funds, the TIAA-CREF Growth & Income Fund and the Calamos Growth Fund, in the time period from 2002 to 2016 in Figure 3.1.[4] Both funds follow a similar investment style and Morningstar classifies them as US Large Cap Growth Equity funds. However, comparing the two funds' factor loading volatilities reveals significant differences: Whereas the TIAA-CREF Growth & Income Fund's market beta measured within single calendar years between 2002 and 2016

Figure 3.1: FEVI of Calamos Growth Fund and TIAA-CREF Growth & Income Fund

This figure plots the 2002−2016 time series of the *FEVI* for the Calamos Growth Fund and the TIAA-CREF Growth & Income Fund. We calculate the *FEVI* from the volatilities of factor exposure obtained from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The *FEVI* is calculated from the past three years of weekly return data. A *FEVI* >0 indicates an above average factor exposure volatility.



[4] The TIAA-CREF Growth & Income Fund was incepted in 1997, the Calamos Growth Fund in 1990. At the end of 2016, USD 5.6 bn. were invested in the TIAA-CREF Growth & Income Fund; the Calamos Growth Fund has total net assets of USD 1.8 bn.

varies between 0.95 to 1.09 (Δ=0.14), Calamos Growth Fund's market beta fluctuates between 0.86 and 1.25 (Δ=0.39) during the same time. The exposures to the *SMB* factor (−0.15 to 0.03 for the TIAA-CREF Growth & Income Fund vs. −0.12 to 0.67 for the Calamos Growth Fund), the *HML* factor (−0.19 to 0.05 vs. −0.80 to 0.42) and the *UMD* factor (−0.02 to 0.13 vs. −0.20 to 0.56) support the impression that the Calamos Growth Fund has more volatile risk factor exposures than the TIAA-CREF Growth & Income Fund. Figure 3.1 plots the two funds' *FEVI*s throughout our sample period with positive (negative) values indicating an above (below)-average of overall factor exposure volatility.

In addition, we observe that differences in *FEVI* can be traced to differences in fund characteristics: In particular, the Calamos Growth Fund has a higher turnover ratio than the TIAA-CREF Growth & Income Fund (90% vs. 83% in 2016) and a less diversified portfolio (79 vs. 189 stock holdings as of the end of 2016). Finally, when comparing the performance of both funds, we observe that the TIAA-CREF Growth & Income Fund, which has stable risk factor loadings, outperformed the volatile Calamos Growth Fund by 1.6% per year between 2002 and 2016.[5]

In this paper, we investigate whether performance differences between funds with high risk factor exposure volatility and those with low risk factor exposure volatility are systematic in a large sample of US equity mutual funds in the time period from the late 2000 to 2016. We first show that factor exposure volatility is a persistent fund characteristic, i.e., funds that are sorted into decile portfolios with the lowest (highest) factor loading volatility in year t have a likelihood of 79% (75%) to remain in the lowest (highest) three deciles in year $t+3$. Second, and most importantly, we find that risk factor exposure volatility is associated with future fund underperformance. A portfolio of the 20% funds with the highest *FEVI* underperforms the 20% funds with the lowest *FEVI* by 147 basis points p.a. at 1% statistical significance when we adjust the returns by the risk factors of the Carhart (1997) four-factor model. Similarly, sorting funds on individual *MKT*-, *HML*-, or *UMD*-factor exposure volatilities, results

[5] The average yearly performance of the primary share classes was 7.2% for the TIAA-CREF Growth & Income Fund and 5.6% for the Calamos Growth Fund in our sample period.

in underperformance of the most volatile funds by 102, 82, and 120 basis points p.a., respectively, with statistical significance at least at the 5% level.6

We check whether the underperformance of funds with high FEVI can be rationalized by their return exposure to other asset pricing models and/or the impact of correlated fund characteristics. For this purpose, we risk-adjust the return spread between funds with high FEVI and funds with low FEVI using different asset pricing models, such as the one-factor CAPM model, the Fama and French (1993) three-factor model, the Fama andFrench (1993) three-factor model extended by a short and long term reversal factor, as well as the Carhart (1997) model extended by the Frazzini and Pedersen (2014) betting-against-beta factor, the Baker and Wurgler (2006) sentiment factor, and the Pástor and Stambaugh (2003) liquidity factor. We find that—in all cases—the underperformance of the high FEVI funds remains statistically and economically significant when accounting for these additional asset pricing models.7 In a similar vein, we observe that the relationship between FEVI in month t and risk-adjusted returns in month t+1 is significantly negative when we account for different fund characteristics, such as fund size, fund age, manager tenure, expenses, turnover, past performance, and fund flows in multivariate Fama and MacBeth regressions. We also confirm that the association between FEVI and future risk-adjusted returns remains negative within different sub-periods and when we alter the setting of our empirical analysis in different robustness checks.

Factor exposure volatility of a fund is conceptually related to two activeness measures that have been shown to affect fund performance in the cross-section: the R^2 measure of Amihud and Goyenko (2013) and the risk-shifting (*RS*) measure of Huang, Sialm, and Zhang (2011). We show that the predictability of *FEVI* for future fund returns is not subsumed by these other activeness measures. In particular, when explicitly controlling for R^2 (*RS*) in portfolio double-sorts, the impact of *FEVI* on future risk-adjusted returns remains negative with economically significant −179 (−79) basis points p.a. Hence, we find compelling evidence that funds with stable exposures to systematic risk factor have higher future risk-adjusted returns than funds

[6] Funds with high *SMB*-factor loading volatility underperform funds with low *SMB*-factor loading volatility by 61 basis points p.a. The performance spread between high and low *SMB*-factor loading volatility funds is statistically significantly indifferent from zero.

[7] When we adjust the high minus low *FEVI* return spread by the Fama/French (2015) five factor model, the statistical significance is slightly above the 10% level. The economic underperformance remains large with a value of −0.85% p.a.

with volatile risk exposures. This effect has not been documented in the academic literature before.

Why do funds with high factor exposure volatility earn low returns in the future and what are the sources for this underperformance? We examine two channels that are potentially related to the underperformance of high *FEVI* funds. First, we investigate whether underperformance is explained by the factor exposure volatility of the funds' long equity portfolio holdings. Armstrong et al. (2013) find that stocks with high risk factor loading uncertainty earn low future returns; hence, the underperformance of high *FEVI* funds could be driven by the low returns of stocks with high risk factor loading uncertainty in their portfolios. We provide evidence—in the line of Armstrong et al. (2013)—that stocks with high risk factor loading uncertainty earn low returns also on the funds' portfolio level; however, we also observe that the underperformance of high *FEVI* funds cannot be solely explained by the risk factor uncertainty of the funds' portfolio holdings.

Second, we analyze whether the underperformance of high *FEVI* funds is related to forced trading of funds due to substantial investor in- and outflows. If investors redeem (or heavily invest) their money from (into) a fund, the fund manager is forced to trade to satisfy investors' liquidity needs (to keep the portfolio invested in equity). Such trades are likely to shift the funds' exposure to systematic risk factors and can also lead to future underperformance (since the manager has to trade quickly and potentially accept disproportionate transaction costs). For this purpose, we estimate the relationship between *FEVI* in month t and risk-adjusted performance in month $t+1$ for funds with different likelihoods of forced trading (approximated by the cumulative amount and cumulative absolute amount of investor flows obtained during the past three years). We do not find evidence that the negative relationship between *FEVI* and future underperformance changes with a fund's likelihood to be affected in forced trading.

Given that both of those explanations fail, our conjecture is that a fund manager voluntarily alters the exposure to systematic risk factors to earn the associated premia in the future, i.e., the manager engages in (unsuccessful) risk factor timing.[8] To check which fund and manager

[8] Our definition of risk factor timing does not distinguish between intended timing attempts (e.g., based on a fund's explicit risk factor timing investment strategy) and unintended, but tolerated portfolio shifts (which nevertheless induce factor exposure volatility in systematic risk factors). We do not differentiate between these

characteristics are associated with risk factor timing, we regress a fund's *FEVI* in month $t+1$ on different observable variables in month t. We find that risk factor timing is particularly prevalent among funds with long management tenure, high turnover, high total expense ratios, and high past fund inflows. These results are in line with previous results from the literature and support the notion that (i) fund manager behavior is influenced by career concerns with young managers having no incentive to expose their portfolios to unsystematic risk (Chevalier and Ellison, 1999), (ii) risk factor timing is an actively enforced and expensive investment strategy, and (iii) risk factor timing is pursued by fund managers who were successful in the past, earn high inflows, and become overconfident in their trading decisions and risk factor forecasts (Puetz and Ruenzi, 2011).

The question whether mutual funds can successfully time risk factor exposures has so far mainly been studied in the context of market timing and produced conflicting results. Whereas the majority of earlier studies, such as Treynor and Mazuy (1966, TM), Henriksson and Merton (1981, HM), Ferson and Schadt (1996) and Kacperczyk and Seru (2007), do not find evidence that fund managers can time the market, more recent studies provide at least some evidence for successful market timing, such as Mamaysky et al. (2008), Jiang et al. (2007), Bollen and Busse (2001), Elton et al. (2012), and Kacperczyk et al. (2014), when applying daily mutual fund data or concentrating on special market situations. Chen and Liang (2007) find that hedge funds, which explicitly claim to time the market, have more favorable risk-return profiles due to successful return and volatility timing. The literature on timing ability beyond the market factor is rather scarce. Kryzanowski et al. (1997) find that only a low proportion of funds attempts to time macro-factors. Investigating changes in fund holdings, Daniel et al. (1997) observe that mutual fund managers do not possess timing abilities with respect to stock characteristics and Benos et al. (2010), who extend the analysis of Bollen and Busse (2001) to a Carhart (1997) model, do not find factor timing abilities either.[9] Busse (1999), Giambona and Golec (2009), and Kim and In (2012) examine volatility timing of mutual funds, while Bodnaruk et al. (2014) document downside risk timing ability of some

approaches since managers usually do not have to report their investment strategy in such a detailed way and even when they do, this description is potentially misleading (see Sensoy, 2009, for the case of deceptive self-designed benchmark indices in the mutual fund industry).

[9] In contrast, Swinkels and Tjong-a-Tjoe (2007) detect positive risk factor timing skills within a very small US fund sample when applying the TM and HM measures to a four-factor model.

fund managers. Finally, Huang et al. (2011) document that funds that intensively shift their total risk exposure over time underperform funds with a stable risk level.

Our research contributes to the mutual fund literature on market and risk factor timing fourfold. First, our proposed *FEVI* measure can directly assess a fund's timing activity whereas earlier models, e.g. TM and HM, only observe performance effects of timing activity. This also enables us to observe high persistence in *FEVI* as an investment characteristic.[10] Second, our model allows us to estimate a fund's factor exposure volatility simultaneously with respect to different risk factors. The vast majority of prior research on timing ability of mutual funds focuses on market timing only. Hence, our result of a negative return effect of *FEVI* goes beyond the most prominent findings of no positive market timing skill. Third, we also contribute to the literature on fund activeness as timing is one element of activeness and is closely linked to—yet not covered by—earlier developed activeness measures such as the Amihud and Goyenko (2013) selectivity measure or the Huang et al. (2011) risk shifting measure. Finally, we contribute to the ongoing debate among academics and investment management practitioners, whether (and how) risk factors can be timed. Numerous papers suggest factor timing strategies, such as Barroso and Santa-Cara (2015) and Moreira and Muir (2017), who show that volatility predicts the momentum and other alternative risk premiums. Among others, Asness et al. (2000) and Arnott et al. (2016) advocate using risk factors' value spread as a signal to time factors. Yet, the question whether those results can be exploited out of sample remains unsolved. Asness (2016) articulates doubts about the performance of risk factor timing. We contribute to this discussion by documenting that professional and sophisticated investors, such as mutual fund managers, are apparently unsuccessful at the timing of risk factors.

The reminder of this paper is structured as follows. Section 3.2 describes the data and introduces our measure of factor timing activity. Section 3.3 links factor timing to mutual fund performance and Section 3.4 examines the drivers of factor timing. Section 3.5 concludes.

[10] Goetzmann et al. (2000) show that the HM parametric test fails to detect timing skill when timing decisions are made at a higher frequency then the observed return intervals. Jiang et al. (2007) note that the measures of TM and HM are subject to artificial timing biases and propose a holding-based measure. Our measure does not require any fund holding data, which might be difficult to access for most investors and which is generally only available on a low, quarterly frequency.

3.2 Data and the factor exposure volatility indicator

In this section we describe the data used in this study and discuss the methodology of the empirical analysis. We also provide summary statistics for the overall *FEVI* and examine its persistence.

3.2.1 Data selection

We investigate the relationship between the volatility of a fund's risk factor exposure and its performance using a sample of actively managed US equity mutual funds. We select our fund universe from the CRSP Survivor-Bias Free Mutual Fund Database and use daily net returns as well as quarterly updated fund characteristics in the empirical analysis. We start our data selection process with all mutual funds included in the CRSP Survivor-Bias Free Mutual Fund Database during the 1998 - 2016 time period. This time window is determined by the availability of daily fund returns. We use Objective Codes from CRSP and Lipper as well as the Strategic Insights Objective Code to determine fund styles and assign each fund to either *Growth and Income, Growth, Income, Hedged, Mid Cap, Small Cap* or *Micro Cap*.[11] Funds that cannot be matched to one of these categories as well as funds with missing fund names are dropped from our sample. We exclude index funds, balanced funds, international funds, and sector funds according to the CRSP Index Fund Flag, CRSP Objective Code and by screening fund names for key terms such as "balanced" or "index". We additionally exclude funds with less than 70% of equity holdings and funds with total net assets of less than USD 15 million. This leaves us with a total number of 3,816 funds in the sample.

We obtain quarterly data on fund age, management tenure, turnover ratio, total expense ratio and total assets under management as well as daily net returns for our sample funds and aggregate those data across all share classes of each fund. Fund age is the age of the oldest share class, total net assets are the sum of the total net assets of all share classes and turnover ratio, total expense ratio and daily returns are the weighted means of single share classes' data, weighted by the share classes' total net assets. We additionally calculate 12-months fund flows for each fund by $flow_t = (tna_t - tna_{t-1year})/(tna_{t-1year} * (1 + ret_{(t-1year,t)}))$, where tna_t are the total net assets at time t and $ret_{(t-1year,t)}$ is the 1-year return (net of fees)

[11] We find that actively managed funds that mainly invest into large caps or whose name contains strings that indicate a large cap investment strategy are mostly classified as *Growth* or *Growth and Income*.

during the past 12 months. We winsorize the data on age, tenure, turnover, expense ratio, flows, and total net assets at the 1%-level. For a sub-analysis in Section 3.4.1 of the paper, we also use equity portfolio holding implied returns. To calculate these returns, we obtain quarterly holding data from CRSP and use the securities' historical CUSIP number to link it to daily stock returns from CRSP.

For our empirical analysis, we aggregate daily returns into weekly as well as monthly data. Following Bollen and Busse (2001), we measure the volatility of risk factor exposures based on weekly returns. Our performance analysis is then based on monthly returns. Since we do not have monthly observations on fund characteristics, we assign the last available data point to each fund if it is not older than 12 months. We calculate weekly and monthly Fama and French (1993) as well as momentum risk factors from daily data, which we obtain from Kenneth R. French's website.[12] We also collect monthly data for the Fama and French (2015) five-factor model as well as a short and long term reversal factor from Kenneth R. French's website. In addition, we gather data on the Frazzini and Pedersen (2014) betting-against-beta factor from AQR, data for the Baker and Wurgler (2006) sentiment factor from Jeffrey Wurgler's website and data for the Pástor and Stambaugh (2003) liquidity factor from WRDS.[13]

3.2.2 Volatility of risk factor exposures in a dynamic factor model

Traditional asset pricing factor models such as the capital asset pricing model (CAPM), the Fama and French (1993) three factor model, and the Carhart (1997) four-factor model assume a linear relationship between an asset's excess return and the respective factor premia. The size of this relationship, represented by β, is traditionally assumed to be constant over time, which allows estimating values of β using an OLS regression framework. Even if this assumption of constant βs holds for single securities it might not be valid for managed portfolios such as mutual funds, as pointed out by Mamaysky et al. (2008), because any varying exposure due to a fund's tactical asset allocation would not be reflected correctly. We

[12] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[13] Data for the betting-against-beta factor is retrieved from https://www.aqr.com/Insights/Datasets/Betting-Against-Beta-Equity-Factors-Monthly and data for market sentiment from http://people.stern.nyu.edu/jwurgler/.

model such time-varying exposures by applying the Carhart (1997) four-factor model with dynamic risk factor loadings β_t, which is represented by the following state space model:

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_{RMRF,i,t} * \left(r_{m,t} - r_{f,t}\right) + \beta_{SMB,i,t} * SMB_t +$$
$$\beta_{HML,i,t} * HML_t + \beta_{UMD,i,t} * UMD_t + \varepsilon_{i,t},$$
$$\beta_{j,i,t} = \beta_{j,i,t-1} + \theta_{j,i}(\mu_{j,i} - \beta_{j,i,t-1}) + \eta_{j,i,t} \quad \text{for} \quad j \in \{RMRF, SMB, HML, UMD\},$$

where $r_{m,t}$ is the market return, $r_{f,t}$ the risk-free rate at time t and SMB_t, HML_t and UMD_t denote the Fama and French (1993) and Carhart (1997) risk factors at time t. The model differs from a classical Carhart (1997) model as it allows the factor loadings to change over time. In our main empirical specification, we assume the factor loadings to follow a mean-reverting process with four time-invariant mean factors μ (one with respect to each risk factor). The four time-invariant values of θ indicate the pace at which the loadings revert to its mean. Those values are unknown and estimated empirically together with the values of β_t. Forcing $\theta = 0$ leads to a model that assumes risk factor loadings to follow a random walk as introduced by Black et al. (1992). In our robustness check, we recalculate our results enforcing this random walk. Our results remain qualitatively unchanged and remain statistically significant. The disturbance terms $\varepsilon_{i,t}$ and $\eta_{j,i,t}$ are normally distributed with zero mean and unknown standard deviations.

For each month we calculate a fund's factor exposure volatility. To do so, we apply the model to the past three years of weekly fund return data and use a Kalman filter and Kalman smoother technique to estimate the dynamics of all unknown parameters.[14] This yields a time series of 156 weekly values of β_{RMRF}, β_{SMB}, β_{HML} and β_{UMD} per fund in the three-year period. For each of the four βs, we compute the standard deviation across time, i.e. $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$. These standard deviations express the volatility of the fund's exposure to the respective risk factors during the three-year period: Generally, a higher $\sigma(\beta)$ indicates a less stable factor exposure with regard to a certain risk factor.[15] To express a fund's overall level of exposure volatility with respect to all the risk factors we aggregate

[14] We shortly describe the Kalman filter and the Kalman smoother technique in the Appendix. Within each three-year window we require funds to have at least 104 weekly return observations.

[15] To prevent outliers influencing our empirical tests, we censor observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ or $\sigma(\beta_{UMD})$ are among the highest 1% of all observations.

the four measures to one overall Factor Exposure Volatility Indicator (*FEVI*). We determine this $FEVI$ as follows: At each point in time, we calculate the cross-sectional mean and standard deviation for each factor exposure volatility measure $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ and standardize all estimated values of $\sigma(\beta)$ by demeaning (using the cross-sectional mean) the estimates and dividing them by the respective cross-sectional standard deviation. Our $FEVI$ is then defined as the average of the four standardized values, i.e.,

$$FEVI = \frac{1}{4}\left(\frac{\sigma(\beta_{RMRF})-\overline{\sigma(\beta_{RMRF})}}{SD(\sigma(\beta_{RMRF}))} + \frac{\sigma(\beta_{SMB})-\overline{\sigma(\beta_{SMB})}}{SD(\sigma(\beta_{SMB}))} + \frac{\sigma(\beta_{HML})-\overline{\sigma(\beta_{HML})}}{SD(\sigma(\beta_{HML}))} + \frac{\sigma(\beta_{UMD})-\overline{\sigma(\beta_{UMD})}}{SD(\sigma(\beta_{UMD}))}\right),$$

where $\overline{\sigma(\beta)}$ is the cross-sectional mean and $SD(\sigma(\beta))$ the cross-sectional standard deviation of $\sigma(\beta)$. Subsequently, we will refer to a fund's risk factor exposure volatility measured over the past three years ending at time *t* as the fund's exposure volatility (or exposure variation or factor loading volatility as synonyms) at time *t*. We will investigate the relationship between future fund performance and a fund's exposure volatility in Section 3.3.

3.2.3 Summary statistics and the persistence of factor exposure volatility

Daily fund returns—and hence, calculated weekly returns for our empirical tests—are available from CRSP by the end of 1998. We calculate our exposure volatility measures from past three years' net returns. If more than two but less than three years of data are available, we calculate factor exposure volatility using the available data. Therefore, our final dataset reaches from the end of 2000 to 2016. It contains 300,519 observations and 3,816 distinct funds. Table 3.1 provides summary statistics for the main variables of the empirical analysis. Average and median fund sizes are USD 1,329 and 324 million, which indicate a skewed distribution of size across funds. On average, the age of a fund is 15.7 years and management has been in office for 7.5 years. The average turnover ratio is 75% per year, but there is a wide variance ranging from 3% to 342%. Total expenses range from 0.14% p.a. to 2.23% p.a. with a mean of 1.15%. The average yearly flow is positive (2.0% of past *TNA*) but its median is at −6.0% suggesting that there are high net inflows into few funds but smaller net outflows from the majority of funds. All four estimated parameters of $\sigma(\beta)$ show a pronounced heterogeneity in factor exposure volatility ranging from a very stable factor exposure ($\sigma(\beta)$<0.0001) to values as large as 4.2 times the average $\sigma(\beta)$.[16] The mean variation in factor loading ($\sigma(\beta)$)

[16] The maximum values of the market, *SMB*, *HML*, and *UMD* exposure volatilities are 0.42, 0.74, 1.00, 0.51.

is highest for the *HML* risk factor, followed by the *SMB*-, the *UMD*-, and the market risk factor, which is in line with results of Engle (2016) who finds betas of industry portfolios to vary over time with the *HML* being the most volatile. As expected and by construction, the average *FEVI* is close to 0, but there are some funds with very volatile factor exposures (maximum *FEVI* = 4.57) and some funds with very stable factor loadings (minimum *FEVI* = −1.80). Panel B reports the estimates of exposure volatility by fund style. Mid Cap, Small Cap, and Micro Cap funds tend to have less stable risk factor exposures than Growth, Growth and Income, and Income funds. The row "other" summarizes very few observations of funds that were classified as large cap funds as well as funds that have been included in our sample but whose assigned styles change during the sample period.

Table 3.2 reports the average cross-sectional correlations between the four measures of exposure volatility (Panel A) as well as between the *FEVI* and fund characteristics (Panel B). The correlation between the volatilities of exposures to single risk factors ranges from 0.20 to 0.33, thus indicating that the factor exposure variation with regard to a single risk factor does not strongly imply exposure volatility with respect to other risk factors. Funds with unstable factor exposures (measured by a high *FEVI*) tend to be smaller, more expensive and show a higher turnover ratio. These results provide first evidence that factor exposure volatility might not be a randomly occurring observation but might be connected to a fund's active trading. We investigate the relationship between exposure variation and fund characteristics more thoroughly in Section 3.4.4 of the paper.

Figure 3.2 plots the time series of equally-weighted average measures of factor exposure volatility over all funds in our sample. Measures of exposure volatility with respect to the market, *SMB* and *UMD* risk factors appear to be relatively stable over time whereas *HML* exposure volatility slightly peeks during the pre-crises years and after 2013. Overall, the variation of factor exposures seems to be prevalent in different market situations and periods of economic booms and recessions.

We also investigate the persistence of factor loading volatility. If factor loading variation is related to mutual fund returns, long-term investors can only profit from this result if factor exposure volatility is a stable fund characteristic rather than a quickly changing investment trend. To study persistence, we sort funds into ten deciles by their *FEVI*. We do so every

Table 3.1: Descriptive statistics and factor exposure volatility by fund style

Panel A of this table provides a descriptive overview over the sample size and fund characteristics. Size, age, management tenure, turnover ratio and total expense ratio are obtained from the CRSP Survivor-Bias Free Mutual Fund Database and relative fund flows are calculated over the past year using $flow_t = (tna_t - tna_{t-1year})/(tna_{t-1year} * (1 + ret_{(t-1year,t)}))$. Fund styles are mainly determined by a fund's CRSP objective code. Funds are aggregated on a portfolio level and size is the sum of all share classes' total assets, fund age is the age of the oldest share class and all other characteristics as well as returns are calculated as the size-weighted mean of all share classes. The measures of factor exposure variation $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are the standard deviation of a fund's weekly factor loading during the past three years. The factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor exposure volatility. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample. Panel B reports the average measures of factor exposure variation and *FEVI* by fund style.

Panel A: Fund characteristics

	# Obs.	Mean	1%	25%	50%	75%	99%
Number of funds	3,816						
Fund-Week-observations	300,519						
Total assets (in mn. USD)	300,519	1,329	19	107	324	1,049	21,268
Fund age (years)	300,361	15.73	2.57	7.65	12.68	19.04	72.42
Manager tenure (years)	241,442	7.51	0.42	3.66	6.33	10.09	25.76
Turnover ratio	265,402	0.75	0.03	0.30	0.58	0.99	3.42
Total expense ratio (in %)	266,142	1.15	0.14	0.92	1.14	1.37	2.23
Relative fund flow	300,311	0.02	−0.59	−0.15	−0.06	0.08	1.90
$\sigma(\beta_{RMRF})$	300,519	0.1220	0.0260	0.0828	0.1105	0.1495	0.3343
$\sigma(\beta_{SMB})$	300,519	0.2179	0.0250	0.1242	0.1936	0.2847	0.6311
$\sigma(\beta_{HML})$	300,519	0.2401	0.0200	0.1295	0.2035	0.3099	0.7831
$\sigma(\beta_{UMD})$	300,519	0.1465	0.0138	0.0814	0.1278	0.1925	0.4273
FEVI	301,908	0.0065	−1.1529	−0.4874	−0.0937	0.3895	2.0284

(continued)

(continued)

Panel B: Mean values of factor exposure volatility by fund style

Fund Style	# Funds / # Obs.	$\sigma(\beta_{RMRF})$	$\sigma(\beta_{SMB})$	$\sigma(\beta_{HML})$	$\sigma(\beta_{UMD})$	$FEVI$
Growth and Income	773 /57,877	0.105	0.173	0.194	0.118	−0.343
Growth	1,636/127,455	0.120	0.212	0.235	0.144	−0.029
Hedged	49 / 2,369	0.132	0.223	0.236	0.127	0.071
Income	202 /14,016	0.108	0.175	0.209	0.130	−0.247
Mid Cap	430 /36,644	0.141	0.263	0.275	0.184	0.365
Small Cap	675 /58,711	0.133	0.249	0.277	0.159	0.217
Micro Cap	45 / 4,275	0.155	0.315	0.321	.0190	0.629
Other	6 / 172	0.105	0.173	0.194	0.118	−0.177

month and leave those decile portfolios unchanged to observe the average value of the $FEVI$ during the 12 months prior and the 72 months after the formation period. Figure 3.3 displays this time series of average $FEVI$ values of funds sorted in decile portfolios. We find that the difference in $FEVI$ becomes smaller during the 12 months before and 36 months after the portfolio formation, but no two decile portfolios cross lines or converge to a common value. Even after 36 months, from where on the calculation window of the $FEVI$ does not overlap with the calculation window of the $FEVI$ at the formation period, the average $FEVI$s of the decile portfolios remain in an unchanged order. Thus, we conclude that the persistence of factor loading volatility remains strong even in the long run (i.e., for a period up to 6 years in the future).

The transition matrix in Table 3.3 underlines this conclusion numerically. It displays the likelihood that a fund sorted in decile portfolio i in year t, appears in decile portfolio j in year t and year $t+3$, respectively. Our results indicate that about 60% of all funds in the decile with most stable (unstable) factor exposures remain in the decile with most (least) stable exposures after one year and over 90% of all funds in the decile with most stable (unstable) exposures remain in the three deciles with most (least) stable exposures after one year. This might partially be by construction since the $FEVI$ has been estimated over a three-year time window. Panel B therefore displays transitions over a period of three years. Results do not change qualitatively. After three years, 45% (41%) of the funds in the lowest (highest) $FEVI$ decile

Table 3.2: Cross-sectional correlations between factor exposure volatility and fund characteristics

Panel A of this table reports the average cross-sectional correlations between measures of factor exposure variation $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ and the $FEVI$. The measures of factor exposure variation are the standard deviation of a fund's weekly factor exposures over the past three years. The factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The $FEVI$ is the mean of the four cross-sectionally standardized factor exposure volatilities. Panel B reports the correlations between fund characteristics and the $FEVI$. Fund size, age, management tenure, turnover ratio and total expense ratio are obtained from the CRSP Survivor-Bias Free Mutual Fund Database and relative fund flows are calculated over the past year using $flow_t = (tna_t - tna_{t-1year})/(tna_{t-1year} * (1 + ret_{(t-1year,t)}))$. Funds are aggregated on a portfolio level and size is the sum of all share classes' total assets, fund age is the age of the oldest share class and all other characteristics as well as returns are calculated as the size-weighted mean of all share classes. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample.

Panel A: Average cross-sectional correlations between measures of factor exposure volatility

	$\sigma(\beta_{RMRF})$	$\sigma(\beta_{SMB})$	$\sigma(\beta_{HML})$	$\sigma(\beta_{UMD})$	$FEVI$
$\sigma(\beta_{RMRF})$	1.00				
$\sigma(\beta_{SMB})$	0.26	1.00			
$\sigma(\beta_{HML})$	0.28	0.20	1.00		
$\sigma(\beta_{UMD})$	0.33	0.31	0.29	1.00	
$FEVI$	0.69	0.65	0.65	0.71	1.00

Panel B: Average cross-sectional correlations between fund characteristics and the FEVI

	$FEVI$	Total exp. ratio	Turnover ratio	Relative fund flow	ln(total assets)	ln(fund age)	ln(tenure)
$FEVI$	1.00						
Total exp. ratio	0.35	1.00					
Turnover ratio	0.22	0.22	1.00				
Relative fund flow	0.00	−0.05	−0.05	1.00			
ln(total assets)	−0.11	−0.32	−0.16	0.07	1.00		
ln(fund age)	−0.01	−0.07	−0.06	−0.19	0.37	1.00	
ln(tenure)	0.04	−0.03	−0.15	−0.02	0.05	0.18	1.00

Figure 3.2: Volatility of risk factor exposures over time

This figure shows the evolution of cross-sectional factor exposure volatilities $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ over time. The volatility measures are the standard deviation of factor loadings during the past three years and factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process.

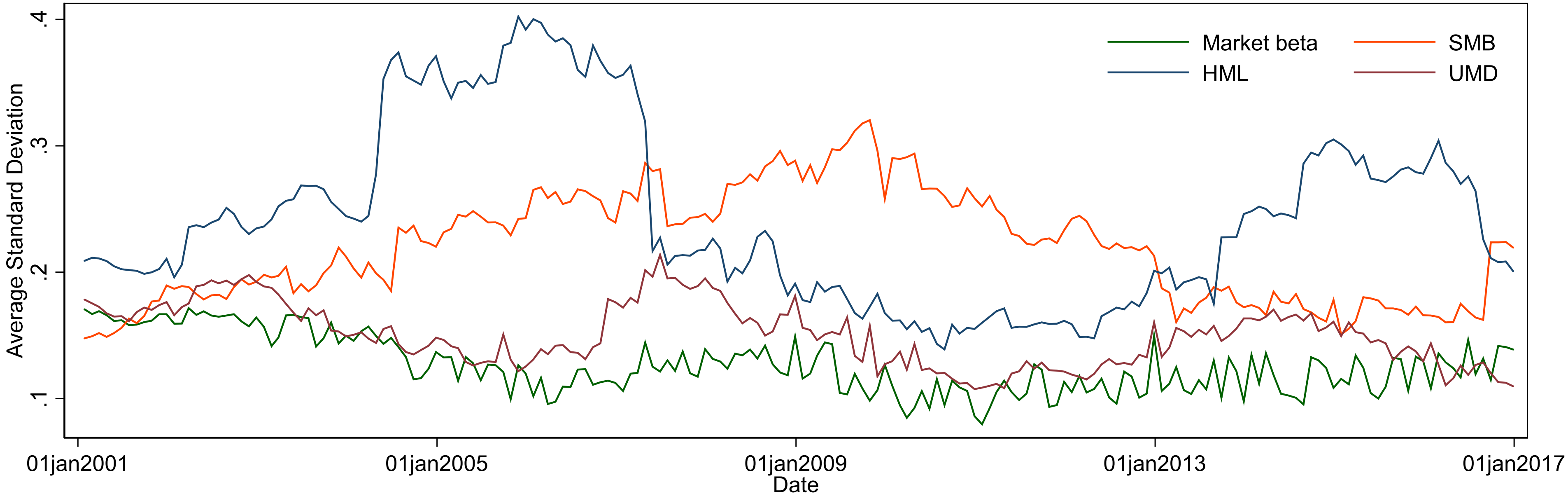


Figure 3.3: FEVI persistence

This figure shows the evolution of the mean *FEVI* (see Section 3.2.2 for the calculation of the *FEVI*) of decile portfolios over time. Each month funds are sorted into ten deciles by the current value of the *FEVI*, which is calculated over the past three years of weekly net returns. The average values of the *FEVI* of those deciles are displayed over time, starting 12 months prior to and ending 72 months after the formation period.

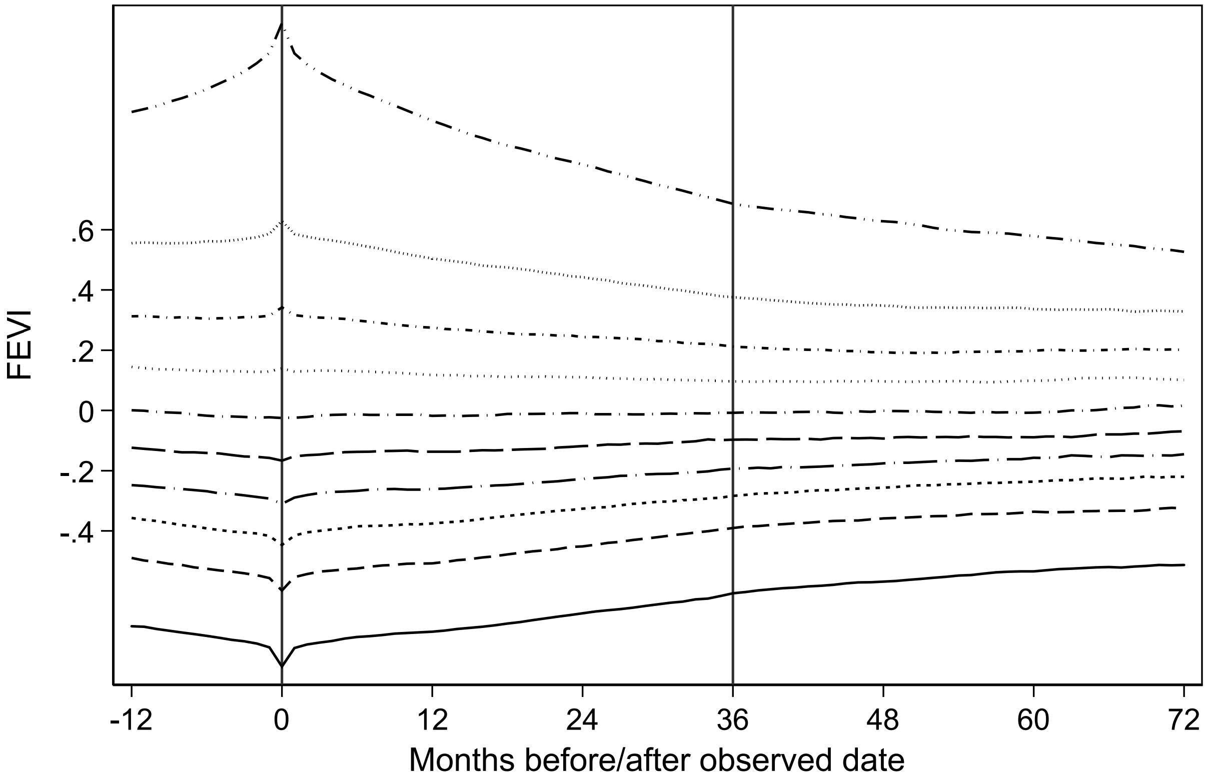


Table 3.3: Factor exposure volatility transition matrix

This table displays a transition matrix of mutual funds between deciles sorted on the $FEVI$ over a period of one year (Panel A) and three years (Panel B). A fund's $FEVI$ is defined as the mean of its cross-sectionally standardized measures of factor loading variation $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$. Measures of factor exposure variation are the weekly standard deviation of a fund's factor loadings obtained from a dynamic version of Carhart's (1997) four-factor model over the previous three years. Each week we sort funds by their $FEVI$. The first column reports the average $FEVI$ of funds within each decile upon its formation as well as one or three years later. The last column reports the percentage of funds within each decile that drop out of our sample within the next year or the next three years, respectively. For all other funds the table reports the transitions between the original decile and the decile funds would have been sorted into if the sorting was done one year or three years later.

Panel A: 1-year transition matrix and attrition rate

Current Decile	Mean initial / final $FEVI$	1	2	3	4	5	6	7	8	9	10	Attrition Rate
1	−0.86 / −0.73	60.47	22.31	8.82	4.15	2.14	0.96	0.58	0.29	0.18	0.09	6.47
2	−0.60 /−0.51	21.02	32.09	22.07	12.27	6.40	3.20	1.53	0.87	0.47	0.08	6.92
3	−0.45 / −0.37	8.62	21.01	25.10	19.59	12.35	6.72	3.62	1.99	0.75	0.24	6.65
4	−0.31 / −0.26	4.52	11.79	18.65	21.98	18.61	11.68	7.10	3.59	1.62	0.46	7.21
5	−0.16 / −0.14	2.27	6.21	11.63	17.55	21.31	17.72	12.18	6.98	3.23	0.92	6.70
6	−0.02 / −0.18	1.45	3.49	6.78	11.47	17.31	20.66	18.45	12.63	6.05	1.73	7.33
7	0.14 / 0.12	0.86	2.05	3.94	6.85	11.77	17.40	22.59	19.55	11.07	3.92	6.73
8	0.34 / 0.27	0.72	1.19	2.11	3.73	7.05	11.90	18.95	24.77	21.05	8.54	7.51
9	0.63 / 0.50	0.41	0.72	1.17	2.01	3.42	6.76	11.55	20.24	31.38	22.34	8.34
10	1.28 / 0.95	0.30	0.39	0.52	0.67	1.36	2.09	4.33	8.91	23.32	58.12	14.24

(continued)

(continued)

Panel B: 3-year transition matrix and attrition rate

Current Decile	Mean initial / final *FEVI*	1	2	3	4	5	6	7	8	9	10	Attrition Rate
1	−0.86 / −0.61	45.34	21.04	12.64	7.70	5.20	3.37	2.03	1.26	0.82	0.59	16.10
2	−0.60 /−0.39	19.82	22.22	17.89	13.17	9.74	6.38	4.42	3.17	2.20	0.99	18.29
3	−0.45 / −0.28	11.59	17.72	17.16	15.24	12.53	9.19	6.84	4.85	3.51	1.37	18.70
4	−0.31 / −0.19	7.79	13.06	15.30	14.79	14.11	11.03	9.45	7.66	4.63	2.17	18.69
5	−0.16 / −0.10	4.91	9.28	12.05	13.91	14.42	13.42	12.35	9.95	6.52	3.19	18.70
6	−0.02 / −0.01	3.43	6.81	10.04	11.49	13.76	13.52	14.35	12.48	8.93	5.18	19.26
7	0.14 / 0.10	2.45	5.09	7.30	9.18	11.15	13.78	15.47	14.69	13.22	7.67	18.69
8	0.34 / 0.21	1.77	3.50	4.64	6.94	9.55	12.08	15.57	16.97	17.24	11.73	19.37
9	0.63 / 0.38	1.16	2.44	3.24	4.88	7.30	9.90	12.80	17.12	21.03	20.12	20.43
10	1.28 / 0.69	0.67	1.18	1.59	2.51	4.22	6.13	8.60	12.84	21.61	40.66	26.22

still remain in this decile and 79% (75%) of all funds in the lowest (highest) $FEVI$ decile remain within the lowest (highest) three deciles after three years. We also provide summary statistics of the attrition rate, that is, the percentage of funds that leave our sample within the following one year or the following three years, respectively. Funds with unstable factor exposures are more likely to drop from our sample within the next years. Only 6% (16%) of all the funds in the lowest $FEVI$ decile leave our sample within the next year (three years), but the probability increases as factor exposure volatility increases and reaches 14% (26%) for the 10% of the funds with the highest exposure volatility.

In summary, Section 3.2 displays summary statistics of the main variables in our study and shows that, overall, the volatility of risk factor exposures is a persistent characteristic of a mutual fund. Moreover, we find that funds with less stable factor exposures are more likely to drop from our sample and that the correlation between individual factor exposure volatility measures is moderate. Hence, high factor exposure volatility to an individual factor does not necessarily imply high factor exposure volatility to another factor.

3.3 Volatility of risk factor exposures and mutual fund performance

This section investigates the relationship between the volatility of risk factor exposures and future mutual fund performance. We examine univariate portfolio sorts in Section 3.3.1, multivariate Fama-MacBeth regressions in Section 3.3.2, and bivariate portfolio sorts in Section 3.3.3. We perform additional robustness checks and document the stability of our main results in Section 3.3.4.

3.3.1 Univariate portfolio sorts

We are interested in the relationship between the volatility of factor exposures, measured by $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ as well as the overall $FEVI$, and the future performance of mutual funds. We start by applying univariate portfolio sorts to investigate this relationship. Each month t, we sort all funds in our sample by the volatility of either a specific risk factor exposure (i.e., by either $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$, $\sigma(\beta_{UMD})$) or by the $FEVI$ and assign them into five quintile portfolios, each portfolio holding one fifth of all funds. As factor loading volatility differs significantly between fund styles, we sort the funds within the same style, thus ensuring that the number of funds of a certain fund style is (almost) the same for all five quintile portfolios. We keep these portfolios unchanged for one month and calculate the quintile portfolio returns in month $t+1$ as the equal-weighted mean of the funds' returns within this portfolio. We resort the portfolios every month by the most recent factor exposure volatility measure and therefore obtain a monthly return time series for each quintile portfolio.

Table 3.4 reports the average abnormal, risk-adjusted returns of these portfolios with each column referring to a specific sorting criterion. As our asset pricing model for the risk-adjustment, we use the Carhart (1997) four-factor factor model. We specifically examine the differences in abnormal returns between funds with a high and low factor exposure volatility, i.e., funds that are sorted in portfolio five and portfolio one according to each measure.

Our results reveal that the risk-adjusted spread between funds with high and low exposure volatility is negative and statistically significant (at least at the 5% significance level) for market, value, and momentum exposure variation as well as for the overall *FEVI*. Funds in the fifth portfolio, i.e. funds with unstable factor exposures, underperform the funds in the first portfolio, i.e. funds with stable factor exposures, in terms of abnormal returns by 102 (market factor), 82 (value factor), 120 (momentum factor) and 147 (overall *FEVI*) basis points p.a., respectively. Furthermore, the abnormal returns decrease monotonically in the market, value, momentum exposure volatility and the overall *FEVI*. The relationship between the exposure volatility to the size factor and abnormal returns is also negative, yet statistically not significant at the 10% level.[17]

To rule out that these results are driven by other risk factors and/or the choice of the factor model, we repeat the portfolio sorts for the *FEVI* and calculate each quintile's abnormal return for different alternative asset pricing models in Table 3.5. Again, we focus to interpret the results of the (5)−(1) difference portfolio between funds with unstable and stable factor exposures. To control for additional risk factors, we use the one-factor CAPM model, the Fama and French (1993) three-factor model, the Fama and French (2015) five-factor model and the Fama and French (1993) model plus a short term and a long term reversal factor provided by Kenneth French's homepage.

[17] Notably, no single quintile portfolio has a positive alpha. This is not surprising as we use net returns and funds are known to show, on average, significantly negative abnormal return after fees. The abnormal return is particularly low for funds in the quintile with the highest exposure volatility when sorted by any measure.

Table 3.4: Abnormal returns of quintile portfolios sorted by factor exposure volatility

This table reports the abnormal returns of fund portfolios sorted on the volatility of factor exposures. Each month we sort funds into five quintiles by either a single factor variation measure $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ or by the *FEVI*. Measures of factor exposure variation are the weekly standard deviation of a fund's factor loadings obtained from a dynamic version of Carhart's (1997) four-factor model over the previous three years as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor loading variation. The sorting is done within each style category, where fund styles are mainly determined by a fund's CRSP objective code. We keep the portfolios constant for one month and calculate the equal weighted portfolio return from funds' net return. Each column represents the sorting by the volatility of the exposure with respect to a distinct risk factor. We report reports Carhart (1997) alphas for each quintile portfolio (Rows 1−5) as well as the difference between the portfolios with most and least volatile factor exposures (High−Low). We regress the return time series on a Carhart (1997) factor model and report the annualized alphas. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1) $\sigma(\beta_{RMRF})$	(2) $\sigma(\beta_{SMB})$	(3) $\sigma(\beta_{HML})$	(4) $\sigma(\beta_{UMD})$	(5) $FEVI$
Low factor exposure volatility	−0.96%*** (−2.91)	−1.17%*** (−4.07)	−1.01%** (−2.38)	−0.91%*** (−2.82)	−0.80%** (−2.56)
(2)	−1.28%*** (−3.54)	−1.39%*** (−4.12)	−1.10%*** (−3.01)	−1.12%*** (−3.17)	−0.97%*** (−2.92)
(3)	−1.33%*** (−3.25)	−1.25%*** (−3.24)	−1.43%*** (−3.77)	−1.19%*** (−3.07)	−1.34%*** (−3.66)
(4)	−1.44%*** (−3.12)	−1.40%*** (−2.91)	−1.59%*** (−4.06)	−1.62%** (−3.48)	−1.58%*** (−3.17)
High factor exposure volatility	−1.98%*** (−3.47)	−1.78%*** (−2.88)	−1.82%*** (−3.38)	−2.11%*** (−3.56)	−2.27%*** (−3.50)
High−Low exposure volatility	−1.02%** (−2.24)	−0.61% (−1.33)	−0.82%** (−2.43)	−1.20%*** (−2.68)	−1.47%*** (−2.76)

We also apply the Carhart (1997) model including, either, the Frazzini and Pedersen (2014) betting-against-beta factor, the Baker and Wurgler (2006) sentiment factor or the Pástor and Stambaugh (2003) liquidity factor in alternative specifications. We find that our results remain qualitatively unchanged and statistically significant for almost all alternative factor models (while getting even more significant for some of the additional models).

.

Table 3.5: Abnormal returns of quintile portfolios sorted by the FEVI under different factor models

This table reports the abnormal returns of fund portfolios sorted by the *FEVI*. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor variation, which are defined as the standard deviation of a fund's weekly factor exposures during the past three years. The factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The sorting is done within each style category, where fund styles are mainly determined by a fund's CRSP objective code. We keep the portfolios constant for one month and calculate the equal weighted portfolio return from funds' net return. We regress each quintile portfolio's return time series on different factor models. Each column refers to one factor model, namely the one-factor model including only the market factor, the Fama/French (1993) three-factor model, the Carhart (1997) four-factor model, the Fama/French (2015) five factor model, a Fama/French (1993) three-factor model extended by a short and long term reversal factor as well as a Carhart (1997) model extended by the Frazzini/Pedersen (2014) betting-against-beta factor, the Baker/Wurgler (2006) sentiment factor or the Pástor/Stambaugh (2003) liquidity factor. We report the annualized alphas for each quintile portfolio (Rows 1−5) as well as the difference between the portfolios with most and least volatile factor exposures (High−Low). T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1) Carhart (1997)	(2) 1-Factor	(3) Fama/French 3 Factors (1993)	(4) Fama/French 5 Factors (2015)	(5) FF3 + Reversal	(6) Carhart + BaB	(7) Carhart + Sentiment	(8) Pástor-Stambaugh
Low factor exposure volatility	−0.80%**	−0.14%	−0.76%**	−1.28%***	−0.81%**	−1.20%***	−0.62%*	−0.94%***
	(−2.56)	(−0.28)	(−2.37)	(−4.40)	(−2.55)	(−4.01)	(−1.89)	(−3.01)
(2)	−0.97%***	−0.37%*	−0.94%***	−1.31%***	−0.98%***	−1.34%***	−0.84%**	−1.20%***
	(−2.92)	(−0.74)	(−2.76)	(−4.06)	(−2.91)	(−4.10)	(−2.36)	(−3.71)
(3)	−1.34%***	−0.76%	−1.30%***	−1.45%***	−1.34%***	−1.70%***	−1.21%***	−1.57%***
	(−3.66)	(−1.43)	(−3.46)	(−3.92)	(−3.67)	(−4.68)	(−3.12)	(−4.44)
(4)	−1.58%***	−0.96%	−1.52%**	−1.71%**	−1.53%***	−1.98%***	−1.41%***	−1.83%***
	(−3.17)	(−1.42)	(−2.93)	(−3.26)	(−2.97)	(−3.94)	(−2.66)	(−3.70)
High factor exposure volatility	−2.27%***	−1.61%*	−2.20%***	−2.14%***	−2.20%***	−2.82%***	−2.12%***	−2.50%***
	(−3.50)	(−1.83)	(−3.27)	(−3.15)	(−3.34)	(−4.32)	(−3.09)	(−3.75)
High−Low exposure volatility	−1.47%***	−1.47%**	−1.43%***	−0.86%	−1.39%***	−1.61%***	−1.50%***	−1.56%***
	(−2.76)	(−2.08)	(−2.65)	(−1.60)	(−2.67)	(−2.93)	(−2.78)	(−2.77)

Solely the Fama and French five-factor model reduces the return difference between funds with high and low values of $FEVI$ to 86 basis points and is borderline significant at the 10% level. We thus conclude that the underperformance of mutual funds with volatile risk factor exposures is not explained by alternative asset pricing risk factors.

3.3.2 Fama-MacBeth regressions

To check whether there is a negative impact of factor exposure volatility on performance when controlling for different fund characteristics at the same time, we proceed to investigate the relationship between factor loading volatility and future fund returns using Fama-MacBeth regressions. We calculate a fund's abnormal return at month t, α_t, as the difference between the actual fund performance during this month and the expected fund performance calculated from a Carhart (1997) model, that is $\alpha_t = r_{i,t} - E[r_{i,t}]$, where

$$E[r_{i,t}] = r_f + \beta_{mkt,i,t} * (r_{m,t} - r_{f,t}) + \beta_{smb,i,t} * SMB_t + \beta_{hml,i,t} * HML_t + \beta_{mom,i,t} * UMD_t,$$

and β_t are estimated by an OLS regression over the previous three years of weekly return data.[18] We conduct Fama-MacBeth regressions with abnormal returns during the next month (or cumulated over the next six and twelve months) as the dependent variable and different fund characteristics as independent variables. As independent variables, we use a fund's *ln(TNA)*, *ln(fund age)*, *ln(manager tenure)*, expenses, turnover, lagged alpha and past fund flows. Table 3.6 reports our results using Newey-West standard errors with a lag of 1 month and style dummies.

Specification (1) reports that the volatility of market, size, value, and momentum exposures have, on average, a negative effect on abnormal returns. This effect is statistically significant for the volatility of market, size and momentum exposures. In specification (2), we pool the individual measures to the $FEVI$ measure. The average effect of the $FEVI$ on future abnormal returns is negative and statistically significant at the 1% level. We also show that this result holds for the six-month and twelve-month abnormal returns in specifications (3) and (4). As a side note, we verify already established relationships between fund characteristics and performance in our multivariate regressions. In particular, we document a significantly negative

[18] We also obtain estimates of β_t when applying the dynamic factor model during the same three-year period. Using those estimates of β_t instead yields qualitatively unchanged results.

Table 3.6: Fama-MacBeth regressions

This table reports the results of Fama-MacBeth regressions of annualized abnormal fund returns on measures of factor exposure volatility and controls. Monthly expected returns are calculated from a Carhart (1997) model where the factor loadings are estimated over the past three years using an OLS regression. Abnormal returns are the differences between monthly returns and the expected returns. Fama-MacBeth regressions are applied on the panel data of monthly abnormal returns. The first two columns report results where the dependent variable is the next month's abnormal return, the last two columns report results where the cumulated abnormal return over the next six or 12 months is regressed on fund characteristics. Fund characteristics are calculated as described in Section 3.2.1. Measures of factor exposure variation are the weekly standard deviation of a fund's factor loadings obtained from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor loading variation. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample. T-statistics are reported in parentheses. We use Newey-West standard errors (lag = 1 month) for the regressions with abnormal returns as the dependent variable. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

Explanatory variables	(1) annualized $alpha_{j,t}$	(2) annualized $alpha_{j,t}$	(3) 6-months CAR	(4) 12-months CAR
$\sigma(\beta_{RMRF})$	−0.060***			
	(−3.33)			
$\sigma(\beta_{SMB})$	−0.032**			
	(−2.15)			
$\sigma(\beta_{HML})$	−0.011			
	(−1.15)			
$\sigma(\beta_{UMD})$	−0.025*			
	(−1.78)			
FEVI		−1.036***	−0.763***	−0.616***
		(−2.78)	(−3.99)	(−5.21)
ln(tna)	−0.158**	−0.160**	−0.157***	−0.135***
	(−2.02)	(−2.03)	(−3.50)	(−4.96)
ln(fund age)	0.144	0.163	0.150**	0.090*
	(1.03)	(1.18)	(2.46)	(1.77)
ln(manager tenure)	0.013	0.006	−0.024	−0.017
	(0.17)	(0.08)	(−0.50)	(−0.56)
Expenses	−0.777***	−0.783***	−0.743***	−0.731***
	(−5.72)	(−5.58)	(−12.63)	(−18.50)
Turnover	−0.207	−0.211	−0.349**	−0.348***
	(−0.96)	(−0.96)	(−2.56)	(−3.16)
Lagged Alpha	0.262***	0.259***	0.189***	0.160***
	(5.86)	(5.70)	(8.11)	(8.70)
Fund Flows	0.181	0.201	0.105	−0.101
	(0.68)	(0.75)	(0.84)	(−1.12)
Style Dummies	YES	YES	YES	YES
Average R^2	0.14	0.13	0.14	0.15

relationship between fund size (expenses) and performance as well as a significantly positive relationship between past performance and performance.

We also analyze the economic impact of our results. The average cross-sectional standard deviation of the volatilities of market, size, value and momentum exposure are 0.06, 0.12, 0.14, and 0.09. Thus, a one standard deviation increase of the volatility of market, size, value, and momentum loadings leads to a decrease of annualized abnormal returns by 35, 38, 15, and 22 basis points p.a. The economic impact of the overall *FEVI* is also substantial: Specification (2) reports that a one standard deviation increase of factor exposure variation reduces abnormal future returns by 71 basis points p.a.

To demonstrate that our results are stable and do not depend on a specific economic environment, we split our sample in different subsets and repeat the Fama-MacBeth regressions as in specification (2) of Table 3.6. We split the 192 sample months by business cycle into 166 months of expansion and 26 months of recession as defined by the NBER. We also split the sample into months with a positive and negative market risk premium and additionally consider a subsample that excludes the months of the financial crises, that is from November 2007 to February 2009. We find that, throughout all subsets, the variation of risk factor exposures as measured by the *FEVI* is associated with lower future abnormal performance, as reported in Table 3.7. In addition to this result, we find that the coefficient estimate of the *FEVI* is more negative during recessions, months with a negative market performance, and less negative for the sample excluding the financial crisis. The level of statistical significance varies across sub-periods, which is partially due to the decreased number of observations within the subsets.

3.3.3 Bivariate portfolio sorts

The volatility of factor exposures is conceptually related to measures of fund manager activeness, which have already been linked to mutual fund performance in prior research. Amihud and Goyenko (2013) show that a low R^2 obtained from an OLS regression of fund returns on a Carhart (1997) model predicts future fund returns. They interpret this low R^2 as selectivity and claim that a higher selectivity might indicate a fund manager's conviction resulting from superior skill. Opposed to that, Huang et al. (2011) find that mutual funds that change their risk levels significantly over time underperform mutual funds with a more stable risk level. The authors suggest that risk shifting might be either an indication of inferior manager ability or a result of agency issues.

Table 3.7: Fama-MacBeth-regressions within sub-periods

This table reports the results of Fama-MacBeth regressions of annualized one-month abnormal fund returns on the *FEVI* and controls for several subperiods. Each month expected returns are calculated from a Carhart (1997) model where the factor loadings are estimated over the past three years of weekly return data from an OLS regression. Abnormal returns are the differences between actual monthly returns and the expected returns. Fama-MacBeth regressions are applied on the panel data of monthly abnormal returns. Funds are aggregated on a portfolio level and fund characteristics are calculated as described in Section 3.2.1. Measures of factor exposure variation are the weekly standard deviation of a fund's factor loadings obtained from a dynamic version of Carhart's (1997) four-factor model over the previous three years as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor loading variation. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample. The regressions are conducted for several sub-periods, that is, expansion and recession months as defined by the NBER, months with a positive and negative market risk premium, as well as during all months besides the 11/07 - 02/09 financial crisis. T-statistics are reported in parentheses. We use Newey-West standard errors (lag = 1 month) for the regressions with abnormal returns as the dependent variable. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1) NBER Expansion	(2) NBER Recession	(3) Positive MRP	(4) Negative MRP	(5) Without crises (11/2007 – 02/2009)
FEVI	−0.736**	−2.953	−0.644	−1.635**	−0.880**
	(−2.33)	(−1.62)	(−1.56)	(−2.52)	(−2.55)
ln(tna)	−0.094*	−0.587	0.085	−0.535***	−0.136*
	(−1.70)	(−1.21)	(1.07)	(−3.70)	(−1.84)
ln(fund age)	0.014	1.115	0.114	0.237	0.074
	(0.13)	(1.50)	(0.78)	(1.06)	(0.57)
ln(manager tenure)	0.025	−0.112	0.120	−0.167	−0.003
	(0.35)	(−0.36)	(1.30)	(−1.16)	(−0.04)
Expenses	−0.729***	−1.128**	−0.605***	−1.055***	−0.793***
	(−5.10)	(−2.19)	(−3.16)	(−4.23)	(−5.43)
Turnover	−0.197	−0.305	0.243	−0.904**	−0.185
	(−1.00)	(−0.29)	(1.11)	(−2.24)	(−0.87)
Lagged Alpha	0.280***	0.126	0.302***	0.194**	0.265***
	(5.75)	(1.00)	(5.14)	(2.58)	(5.64)
Fund Flows	0.026	1.321	0.654*	−0.491	0.283
	(0.10)	(1.09)	(1.85)	(−1.24)	(1.00)
Style Dummies	YES	YES	YES	YES	YES
Months	166	26	116	76	176
Average R^2	0.12	0.16	0.13	0.13	0.13

To investigate whether the negative relationship between factor exposure volatility and mutual fund performance persists beyond those other measures (R^2 and risk shifting), we perform bivariate portfolio sorts based on the *FEVI* and the measures of activeness. We calculate a fund's R^2 following Amihud and Goyenko (2013) from an OLS regression of net returns on a Carhart (1997) model but use three years of weekly return data to comply with the calculation of our *FEVI* measure. For those funds for which holding data are available from CRSP, we also calculate the holding based risk shifting measure following Huang et al. (2011).[19] Whereas the risk shifting measure of Huang et al. (2011) is hardly correlated to the *FEVI* (ρ=0.04), there is a considerable negative correlation between the *FEVI* and the Amihud and Goyenko R^2 (ρ= −0.55).

We perform the bivariate portfolio sorts as follows: Each month we sort the funds by one of the two measures (R^2 and risk shifting) into five quintiles. As before, we define quintiles per style category to ensure an (almost) equal distribution of fund styles within each quintile. Within each of the resulting quintiles we sort funds by their *FEVI* and form quintiles such that we end up with two sets of 5x5 quintile portfolios. We keep the portfolios unchanged over the next month and calculate the respective portfolio returns. For each portfolio we calculate the abnormal return using the Carhart (1997) model. Table 3.8 displays the results where the first sorting is done by either R^2 (Panel A) or the risk shifting measure (Panel B).

Again, we are particularly interested in the difference between the portfolios of high and low factor exposure variation formed within each R^2 or risk shifting quintile, respectively. When looking at the results in Panel A, we observe that the quintiles with the highest *FEVI* underperform the quintiles with the lowest *FEVI* in all cases and the difference is statistically significant at the 1% level for almost all R^2-quintiles and significant at the 10% level for those with the lowest R^2. This result is important, because there is a negative relationship between R^2 and our *FEVI* by construction.[20] By showing that a high *FEVI* is associated with lower abnormal returns even within R^2-quintiles, we provide strong evidence that our *FEVI*

[19] Mutual fund holding data is available from CRSP starting in December 2001 for few funds and from June 2002 for a larger sample of funds. We need three years of data to calculate risk shifting and therefore our subsample for any analysis using holding data and risk shifting starts in July 2004 only. This reduces the overall number of fund-month observations to 233,251.

[20] Funds with a high *FEVI* have volatile loadings on risk factors and thus, a static risk factor model might not explain much of the return volatility and will have a low R^2 when estimated using an OLS regression.

Table 3.8: Portfolio double sorts, activeness and factor exposure volatility

This table reports the results of bivariate portfolio sorts. Funds are first sorted into five quintiles by either the Amihud/Goyenko (2013, Panel A) R^2 measure or the fund's Huang et al. (2011, Panel B) risk shifting measure. Within each quintile funds are sorted into five quintiles by their *FEVI*. R^2 is calculated from a Carhart (1997) model over three years of return data and we follow Huang et al. (2011) to calculate the risk shifting measure. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor exposure variation, which are defined as the standard deviation of a fund's weekly factor exposures during the past three years. The factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The sorting is done within each style category, where fund styles are mainly determined by a fund's CRSP objective code. We keep the portfolios constant for one month and calculate the equal weighted portfolio returns. We regress each quintile portfolio's return time series on the Carhart (1997) four-factor model. Within each panel, we report the annualized alphas for each 5x5 portfolio as well as the difference between the portfolios with most and least volatile factor exposures (High−Low) within each of 5 quintile portfolios from the first sorting step. T-statistics are reported in parentheses. We also report average coefficients and t-statistics from the High−Low returns. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Low *FEVI*	(2)	(3)	(4)	High *FEVI*	High-Minus-Low
Panel A: Sorting on R^2 and FEVI						
Low R^2	0.14%	−0.23%	−0.69%	−0.78%	−1.30%	−1.44%*
	(0.22)	(−0.32)	(−0.82)	(−0.93)	(−1.21)	(−1.68)
(2)	−0.49%	−0.47%	−1.29%**	−0.64%	−2.52%***	−2.04%***
	(−0.83)	(−0.79)	(−2.02)	(−1.01)	(−3.78)	(−2.60)
(3)	−0.71%	−1.07%**	−1.15%**	−2.16%***	−2.40%***	−1.69%***
	(−1.56)	(−2.29)	(−2.34)	(−4.29)	(−3.90)	(−2.64)
(4)	−0.71%*	−1.64%***	−1.76%***	−2.07%***	−2.63%***	−1.92%***
	(−1.77)	(−4.24)	(−4.30)	(−4.49)	(−5.39)	(−3.56)
High R^2	−0.71%***	−1.52%***	−2.10%***	−1.84%***	−2.55%***	−1.84%***
	(−2.93)	(−5.21)	(−5.69)	(−4.63)	(−5.60)	(−4.08)
					Average Coefficient	−1.79%***
Panel B: Sorting on Risk Shifting and FEVI						
Low Risk Shifting	−1.52%***	−1.56%***	−1.65%***	−2.54%***	−3.31%***	−1.79%***
	(−4.52)	(−3.71)	(−3.43)	(−4.43)	(−4.68)	(−3.01)
(2)	−1.52%***	−1.35%***	−1.08%***	−1.23%***	−1.98%***	−0.46%
	(−5.31)	(−3.99)	(−2.90)	(−2.64)	(3.33)	(−0.82)
(3)	−0.79%***	−0.97%***	−1.07%***	−1.21%***	−1.79%***	−1.00%*
	(−2.99)	(−3.05)	(−2.90)	(−2.62)	(−3.10)	(−1.87)
(4)	−0.80%***	−0.79%**	−1.09%***	−0.49%	−1.02%*	−0.21%
	(−2.71)	(−2.44)	(−2.75)	(−1.00)	(−1.67)	(−0.42)
High Risk Shifting	−0.83%**	−0.42%	−1.23%**	−1.25%**	−1.34%*	−0.50%
	(−2.48)	(−0.95)	(−2.55)	(−2.20)	(−1.68)	(−0.80)
					Average Coefficient	−0.79%

measures a return pattern not captured by the R^2 measure. Sorting on risk shifting and factor exposure volatility in Panel B supports our findings from above: Funds with a high *FEVI* underperform funds with the lowest *FEVI* in every risk-shifting quintile. The effect is statistically and economically particularly prevalent among funds with a low risk shifting measure. Hence, the return pattern due to factor exposure volatility is not subsumed by the effect of a fund manager's risk shifting (as measured by Huang et al., 2011).

3.3.4 Robustness tests

We conduct a series of robustness tests to check that the negative relationship between factor exposure volatility and mutual fund performance remains strong when using value-weighted Fama-MacBeth regressions, using alternative performance measures, varying the dynamics of our state space model or adding additional control variables. We adapt the Fama-MacBeth regressions presented in specification (2) of Table 3.6 and display the results of the stability checks in Table 3.9.

In specification (1), we value-weight the funds during the first stage regressions of the Fama-MacBeth procedure. The results remain unchanged. Then, we regress alternative performance measures on the *FEVI* and other fund characteristics. In specification (2), we use the skill measure of Berk and van Binsbergen (2015), which measures the dollar value a fund manager generates, either presenting itself as a management fee or as over- or under-performance to the investor. Therefore, skill is defined as the product of fund size (total net assets) and the fund's gross excess return over the benchmark. We use the expected return from a Carhart (1997) factor model as a benchmark. We also apply a fund's Sharpe ratio and the manipulation-proof performance measure of Goetzmann et al. (2000) calculated from half a year of weekly returns as performance measures in specifications (3) – (5). For the latter, we set $\rho=2$ and $\rho=3$ to alternate the level of risk penalty. The relationship between a fund's factor exposure volatility—notably, measured during the period prior to the half year the performance measures were calculated for—and fund performance remains negative and statistically significant at the 5% level.

Our dynamic factor model relies on an assumption about the underlying process of factor loadings and we assume a mean reverting process that is for each fund *i* at time t:

$$\beta_{j,i,t} = \beta_{j,i,t-1} + \theta_{j,i}\left(\mu_{j,i} - \beta_{j,i,t-1}\right) + \eta_{j,i,t} \text{ for } j \in \{RMRF, SMB, HML, UMD\}.$$

As an additional robustness test, in Specification (6), we restrict this process to a random walk by setting $\theta_{j,i}$ to 0. This yields:

$$\beta_{j,i,t} = \beta_{j,i,t-1} + \eta_{j,i,t} \text{ for } j \in \{RMRF, SMB, HML, UMD\}.$$

We estimate the dynamics assuming this random walk, measure the variation of βs and calculate a corresponding version of the *FEVI* as described in Section 3.2.2. The relationship between factor exposure volatility and mutual fund performance remains negative and economically and statistically significant at the 2%-level when using this alternative approach.

Another methodological alternative only considers the idiosyncratic variation of betas. That is, we estimate the dynamics of βs using a mean-reverting process. Instead of measuring the variation of βs over time we use the standard deviation of ηs as a measure of factor loading variation. For each of the four risk factors, the error term η is normally distributed and we take the standard deviation of these distributions with respect to the four risk factors as the measures of interest. As before, we calculate a version of the *FEVI* as the average of the cross-sectionally standardized measures and use Fama-MacBeth regressions to determine the relationship between factor loading variation and fund performance in specification (7). Again, we find a negative and statistically significant relationship. We additionally test the robustness of our results by adding the measures of fund activeness discussed in Section 3.3.3 as additional control variables in specification (8). As a result, the relationship between factor exposure volatility and fund performance slightly weakens economically when including Amihud and Goyenko (2013)'s R^2 and Huang et al. (2011)'s risk shifting measure as independent variables, but it remains statistically significant at the 10%-level.

Finally, we perform a placebo test to examine the relationship between factor exposure volatility and fund performance for a sample of index funds. For these funds, any variation of risk factor exposure should be coincidental and not influenced by fund managers' trading decisions. If the relationship between factor exposure volatility and fund returns was due to fund managers' actions, we should not expect this relationship for index funds. We exactly follow the data selection procedure from Section 3.2.1 but instead of dropping index funds, we solely keep index funds in our sample. We identify those funds by the index fund flag from CRSP and additionally hand-pick funds whose names include one of the terms "Index", "S&P", "Wilshire", "Dow" or "Russell". This leaves us with 631 index funds and 33,515 fund-month observations. Specification (9) shows the results of the Fama-MacBeth regressions on this

Table 3.9: Robustness checks

The table reports the results of Fama-MacBeth regressions of different performance measures on the *FEVI* and control variables as well as other robustness checks. Column (1) reports results of Fama-MacBeth regressions where the first-step regressions is size-weighted. In columns (2) − (5) the dependent variable consists of an alternative performance measure: the skill measure of Berk/van Binsbergen (2015), the 26-week sharpe ratio, and the 26-week manipulation-proof performance measure of Goetzmann et al. (2007) with ρ=2 and ρ=3. Columns (6) and (7) report results for alternative models of factor exposure variation. For column (6) we assume risk factor exposures to follow a random walk, and for column (7) we measure the standard deviation of the idiosyncratic component of factor loading variation. In Column (8), the Amihud/Goyenko (2013) R^2 measure and the Huang et al. (2011) risk shifting measure are added as controls. Column (9) displays the results for a sample of index funds. *FEVI* is the mean of the four cross-sectionally standardized measures of factor exposure volatility. Age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample. T-statistics are reported in parentheses. We use Newey-West standard errors (lag=1 month). Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Size-weighted	Alternative Performance Measures				Other Models		Additional Control Var.	Index Funds
Explanatory variables	annualized $alpha_{j,t}$	skill	Sharpe ratio (26 weeks)	MPPM (rho=2)	MPPM (rho=3)	Random Walk	Idiosynchratic β-Volatility	annualized $alpha_{j,t}$	annualized $alpha_{j,t}$
FEVI	−1.029**	−977.482**	−0.031***	−0.828**	−1.103***	−0.974**	−0.842**	−0.670*	−0.045
	(−2.38)	(−2.04)	(−2.69)	(−2.40)	(−3.02)	(−2.43)	(−2.54)	(−1.87)	(−0.12)
ln(tna)	−0.132	−13.973**	−0.004	−0.230***	−0.253***	−0.151*	−0.165	−0.119*	−0.040
	(−1.08)	(−2.19)	(−1.42)	(−3.34)	(−3.56)	(−1.68)	(−2.06)	(−1.74)	(−0.88)
ln(fund age)	0.081	0.634	0.009	0.309***	0.336***	0.086	0.079	0.119	0.083
	(0.49)	(0.35)	(1.55)	(3.42)	(3.65)	(0.72)	(0.59)	(1.10)	(0.60)
ln(manager tenure)	−0.053	−1.017	−0.010***	−0.068	−0.063	0.033	0.022	−0.032	0.120
	(−0.33)	(−0.26)	(−3.14)	(−0.89)	(−0.81)	(0.43)	(0.29)	(−0.46)	(1.32)
Expenses	−1.109***	190.744	−4.497***	−0.785***	−0.798***	−0.761***	−0.753***	−1.015***	−1.024***
	(−3.58)	(0.58)	(−7.15)	(−9.83)	(−9.63)	(−7.31)	(−5.46)	(−6.45)	(−3.79)
Turnover	−0.351	−134.092	−0.791	−0.090	−0.136	0.208	−0.216	−0.100	−0.110
	(−1.30)	(−0.51)	(−1.16)	(−0.49)	(−0.74)	(−0.80)	(−0.96)	(−0.46)	(−0.65)
Lagged Alpha	0.248***	233.256***	1.689***	0.213***	0.231***	0.246***	0.261***	0.249***	0.283***
	(4.36)	(3.42)	(10.78)	(7.01)	(7.46)	(5.51)	(5.75)	(5.21)	(3.47)
Fund Flows	0.024	4.698	−0.011	−0.247	−0.221	0.210	0.198	0.005	−0.178
	(0.07)	(1.32)	(−1.28)	(−1.65)	(−1.46)	(0.95)	(0.75)	(0.03)	(−0.74)
R^2								−0.027	
								(−0.93)	
Risk Shifting								0.402**	
								(2.36)	
Style Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Average R^2	0.20	0.08	0.21	0.27	0.27	0.13	0.13	0.14	0.42

index fund sample. As expected, the relationship between the *FEVI* and future abnormal fund performance is close to zero. This result provides evidence that to our main result of a negative relation between the volatility of risk factor exposures and future fund performance stems from fund managers' trading decisions. A deeper analysis on the drivers of our main results is provided in Section 3.4.

To summarize, in this section we document that the relationship between the volatility of risk factor exposures and future risk-adjusted performance is negative in univariate portfolio sorts, multivariate regressions, and bivariate portfolio sorts when explicitly controlling for related measures. We confirm this result in a large battery of robustness checks and show that our results are not sensitive to several choices we make in our empirical analysis.

3.4 Drivers of factor exposure volatility

Our results indicate a strongly negative relationship between the volatility of a fund's factor exposures and future fund performance. First evidence from the descriptive statistics and the index fund robustness test suggest that fund managers' trading activities might cause this relationship. In this section we analyze potential drivers of mutual funds' factor exposure volatility. We look at equity-induced factor loading variation in Section 3.4.1 and investigate whether the volatility of factor exposures is related to fund flows and thus induced by funds' asset fire sales and purchases in Section 3.4.2. As both is not the case, we conclude that our results are driven by fund managers' voluntary, but unsuccessful, attempts to time risk factors. Section 3.4.3 provides supportive in-sample evidence of negative factor timing skills. We finally relate factor timing to correlated fund characteristics in Section 3.4.4

3.4.1 Factor exposure volatility induced by unstable factor loadings of equity holdings

There might be two potential sources of factor exposure volatility measured by our approach. On the one hand, a fund's trading activity might cause the variation of βs if the fund management shifts holdings accordingly, for example between large cap and small cap stocks. Section 3.4.2 will further break down this channel into forced and unforced trading. On the other hand, even a buy-and-hold strategy might have volatile risk factor exposures if the holdings' factor exposures vary over time. Prior research finds evidence consistent with the latter explanation: Armstrong et al. (2013) show that stocks with high risk factor loading uncertainty with respect to the *MKT* factor, the *SMB* factor, the *HML* factor, and the *UMD* factor earn low future returns.[21] This pattern is likely to be present also on the fund level.

We aim to disentangle factor exposure volatility induced by changes in a fund's asset allocation from factor exposure variation caused by the volatility of the holdings' factor loadings. Therefore, we calculate an additional set of factor exposure volatility measures directly imputed from mutual fund equity portfolio holdings. Most funds report holdings at the end of each quarter and we then calculate weekly returns during a quarter q as the weighted average stock returns during this week, weighted by the fund's portfolio weights as of the end of quarter q−1.[22] The underlying assumption of constant portfolio weights between reporting dates is, among others, in line with the implicit assumptions of the holding-based market timing approach of Jiang et al. (2007). This yields a return series, where short-term investment decisions and the timing of trades of the fund manager remain unconsidered. As in Section 3.2.2, we apply the Kalman filter and smoother to estimate our dynamic version of the Carhart model for this holding-based return series instead of actual fund returns. As before, we compute the volatility of factor loadings with regard to *MKT*, *SMB*, *HML*, and *UMD* factor over a period of 156 weeks and form an *FEVI* by averaging the standardized values of these factor exposure volatility measures. We then investigate whether this overall *FEVI* calculated from fund holdings is also related to future abnormal returns of the fund using Fama-MacBeth regressions in Table 3.10.

Specification (1) repeats the baseline regression setup of specification (2) in Table 3.6 for a comparison of coefficients. In specification (2), we report the results of the relationship between factor exposure variation based on equity holdings data and fund performance. In line with the results of Armstrong et al. (2013), we find that the association between the holdings-based *FEVI* and future abnormal returns is significantly negative. However, we also observe that the coefficient estimate of the *FEVI* decreases by more than 30% in comparison to the holding-based *FEVI* based on actual net returns; if we use both indicators as explanatory variables in regression (3), we document that only the coefficient of the *FEVI* calculated from

[21] Opposed to this view, Lenz (2017) finds that stocks with more volatile market betas earn systematically higher returns than stocks with persistent market risk exposures.
[22] We do not consider a fund whenever the most recent holdings were reported more than one year ago and are missing in the upcoming quarters.

actual net returns remains statistically significant and is 3.4 times as large as the coefficient on the holding-based *FEVI*.

Altogether, these results indicate that the holding-based *FEVI* relates negatively to future abnormal returns; however, it cannot explain the negative association between the *FEVI* calculated from actual net returns and future performance. Hence, we conclude that the factor exposure volatility induced by fund managers' active trading decisions (as opposed to the volatility of fund holdings' risk factor exposures) is the main driver of a fund's underperformance. Section 3.4.2 investigates whether this result might be driven by forced trading due to inflows and outflows rather than strategic or tactical asset allocation decisions.

3.4.2 Forced versus unforced trading

Section 3.4.1 shows that the negative relationship between factor loading volatility and future fund performance is not explained by the underlying portfolio holdings' exposure volatility and hence must be due to factor loading volatility induced by fund managers' trading. This section aims to distinguish between unsolicited trading and forced trading, i.e., trading that is required according to a fund's investor flows. If investors withdraw large amounts from a fund (or invest new money into the fund), the fund management will be forced to sell (or buy) assets and the risk factor exposure might vary as a result of this forced trading.[23] If the negative relationship between the volatility of factor exposures and fund performance is stronger and only present among funds that experience large inflows or outflows, it might not be due to fund managers' unfavorable trading decisions but due to investors' flows and resulting asset sales and purchases. We measure the volatility of factor exposures over a three-year period and investigate the impact of contemporaneous flows, observed over the identical period. Hence, we compute a fund's three-year flow as the sum of yearly flows as described in Section 3.2.1. To detect the impact of fund inflows and outflows on the *FEVI*-performance relationship we construct three subsamples. One subsample consists of all fund-month observations for which the three-year flow lies below the 30% quantile of three-year flows during the same time period. A second subsample consists of all observations with a three-year flow above the 70% quantile. All remainder funds, those with a medium three-year flow between

[23] Coval and Stafford (2007) discuss the phenomena of asset fire sales and purchases for mutual funds. They show that, among others, funds experiencing large outflows tend to decrease existing positions, which creates price pressure in the underlying securities held by the fund.

Table 3.10: Equity portfolio holdings

This table reports the results of Fama-MacBeth regressions of annualized one-month abnormal fund returns on measures of factor exposure volatility and controls. Expected returns are calculated from a OLS regression of a Carhart (1997) model. Abnormal returns are the differences between actual monthly returns and the expected returns. Fama-MacBeth regressions are applied on the panel data of monthly abnormal returns. Funds are aggregated on a portfolio level and fund characteristics are calculated as describes in Section 3.2.1. The measures of factor loading variation $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are the standard deviations of a fund's weekly factor exposures during the past three years. The factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. We assume risk factor exposures to follow a mean-reverting process. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor exposure volatility. Besides the *FEVI* calculated from funds' net returns we calculate a second *FEVI* from funds' equity portfolio holdings. Portfolio holding are reported on a quarterly basis and we assume that between those reporting dates a fund held constant portfolio weights. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ for either the net return based or the holding based approach are amongst the highest 1% are dropped from the sample. T-statistics are reported in parentheses. We use Newey-West standard errors (lag = 1 month) for the regressions with abnormal returns. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1)	(2)	(3)
FEVI (based on returns)	−1.036*** (−2.78)		−0.640* (−1.90)
FEVI (based on holdings)		−0.725* (−2.11)	−0.186 (−0.99)
ln(tna)	−0.160** (−2.03)	−0.075 (−0.97)	−0.073 (−0.97)
ln(fund age)	0.163 (1.18)	0.110 (0.92)	0.103 (0.85)
ln(manager tenure)	0.006 (0.08)	−0.018 (−0.28)	−0.016 (−0.25)
Expenses	−0.783*** (−5.58)	−0.900*** (−4.25)	−0.869*** (−4.22)
Turnover	−0.211 (−0.96)	−0.180 (−0.66)	−0.166 (−0.63)
Lagged Alpha	0.259*** (5.70)	0.271*** (5.07)	0.272*** (5.18)
Fund Flows	0.201 (0.75)	0.093 (0.41)	0.090 (0.40)
Style Dummies	YES	YES	YES
Average R^2	0.13	0.12	0.13

Table 3.11: Fama-MacBeth regressions by flow

This table reports the results of Fama-MacBeth regressions of annualized abnormal fund returns on the *FEVI* and controls within fund subsamples. Funds are sorted into subsamples by either the past 3-year flow (columns 1–3) or the past 3-year absolute flow (columns 4–6). Yearly flows are calculated as $flow_t = (tna_t - tna_{t-1year})/(tna_{t-1year} * (1 + ret_{(t-1year,t)}))$. The 3-year flow is the sum of the year flows during the most recent three years. The 3-year absolute flow is calculated as the sum of the absolute values of yearly flows during the previous three years. Within each subsample we apply the Fama-MacBeth regression as follows. Each month expected returns are calculated from a Carhart (1997) model where the factor loadings are estimated over the past three years of weekly return data from an OLS regression. Abnormal returns are the differences between actual monthly returns and the expected returns. We regress the abnormal returns on the *FEVI* and further control variables. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor exposure volatility. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample. T-statistics are reported in parentheses. We use Newey-West standard errors (lag = 1 month) for the regressions with abnormal returns as the dependent variable. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1)	(2)	(3)	(4)	(5)	(6)
	30% lowest past flows	Medium flows	30% highest past flows	30% lowest absolute flows	Medium absolute flows	30% highest absolute flows
FEVI	−0.668* (−1.89)	−1.117** (−2.55)	−0.967*** (−2.62)	−0.984** (−2.32)	−1.036** (−2.44)	−0.943*** (−2.69)
ln(tna)	−0.069 (−0.72)	−0.134* (−1.79)	−0.191* (−1.77)	−0.140 (−1.50)	−0.181* (−1.91)	−0.218** (−1.98)
ln(fund age)	−0.266 (−1.22)	0.139 (1.00)	−0.077 (−0.41)	0.028 (0.14)	0.206 (1.05)	0.034 (0.17)
ln(manager tenure)	0.106 (1.01)	−0.136 (−0.89)	−0.135 (−0.61)	0.159** (1.97)	−0.099 (−0.74)	−0.169 (−0.77)
Expenses	−0.929** (−2.40)	−0.960*** (−4.35)	−0.524** (−2.03)	−0.980*** (−2.99)	−0.642*** (−2.83)	−0.969*** (−3.47)
Turnover	−0.529 (−1.43)	−0.255 (−1.00)	−0.115 (−0.52)	−0.512 (−1.44)	−0.253 (−0.97)	−0.052 (−0.24)
Lagged Alpha	0.250*** (4.72)	0.312*** (5.78)	0.239*** (3.93)	0.317*** (5.14)	0.244*** (4.73)	0.248*** (4.59)
Fund Flows	1.809 (1.82)	0.405 (0.70)	0.356 (1.43)	−0.351 (−0.35)	0.512 (0.96)	0.223 (0.85)
Style Dummies	YES	YES	YES	YES	YES	YES
Average R^2	0.16	0.17	0.17	0.17	0.16	0.16

the 30% and 70% quantile constitute a third subsample. Within each subsample we repeat the Fama-MacBeth regression. Table 3.11 displays the results. If high outflows or inflows were driving our results, we would expect the relationship between factor exposure volatility and future fund performance to be particularly large for funds with negative or high three-year-flows. The empirical results do not support this idea. In fact, the coefficient of the *FEVI* is lowest for funds in the medium-flow sample, that is, for funds with moderate flows.

Summing up flows over the previous three years might disguise cases where funds had to react to large inflows in one year and to large outflows during another year, i.e., these flows could level out each other. We therefore calculate an absolute flow measure as the sum of the absolute flow values during the three years.[24] We repeat the subsample analysis using this absolute flow measure. We consider subsamples with the 30% highest, 30% lowest and 40% median absolute flows. If large inflows or outflows drove our results, we would expect the *FEVI*-return-relationship to be stronger for funds with a higher absolute flow measure and would disappear for all other funds. However, the results in Panel B of Table 3.11 do not support this expectation and the most negative coefficient estimate of the *FEVI* can be seen for funds with median absolute flows. We thus conclude that the negative relationship between risk factor exposure variation and future fund performance cannot be explained by fire sales and is mainly due to unsolicited trading decisions.

3.4.3 Factor timing (in)ability

Neither volatile factor loadings of the portfolio holdings nor fund flows can explain the underperformance of funds with volatile risk factor exposures. We thus conclude that this underperformance stems from fund management’s unforced trading decisions and we call these trading activities that lead to time-varying risk factor exposures “factor timing”. This definition of factor timing does not require fund managers to vary their risk factor exposures intentionally but also includes any unintended, but tolerated, volatility of risk factor exposures (e.g., as a side effect of stock picking). Portfolio managers do not disclose their investment strategy detailed enough to distinguish both sources. This definition of factor timing is also in line with the prior literature on market and factor timing that does not take into account

[24] As an example, a fund with yearly flows of +50%, −50% and +20% would have an absolute flow measure of 120%.

whether a fund managers intents to time risk factors[25]. Most important, from an investor's perspective, it appears inessential whether a fund manager actually tries to time risk factors. The Carhart (1997) four-factor model has become a state of the art model in the financial industry and thus fund managers should be fully aware of their risk factor exposures. If fund managers allow these exposures to vary, factor timing stems from a neglect to manage stable exposures. So far, our paper detects a negative relationship between factor timing and future, that is, out-of-sample, fund performance.[26] Our approach, however, also can be adapted to observe mutual fund manager's timing ability in an in-sample setting. For this purpose, we estimate the model as presented in Section 3.2.2. Instead of using a rolling 156-week window we estimate the dynamics of a fund's factor loadings over the entire sample period, that is from late 1998 or the fund's inception date (whichever is later) until the end of 2016 or the fund's termination day (whichever is later). For each fund, this yields four time series of factor loadings with respect to the market risk, *SMB*, *HML* and *UMD* risk factor. A fund manager who wants to time a risk factor will increase her risk factor exposure just before she expected a risk factor to pay a high premium and decrease her exposure when she expected a low premium. If the manger was skilled in timing risk factors, we should observe a positive correlation between factor exposures measured from our dynamic Carhart model and the risk premia during the subsequent month. We thus measure this correlation for each fund and each risk factor. Table 3.12 provides an overview over the distribution of these correlations.

The cross-sectional distribution of the correlations does not suggest positive timing skill with respect to either risk factor. Instead, the mean and median correlations are slightly negative (between −0.01 for momentum timing and −0.06 for market timing). Although these results do not have a strong statistical significance (mean values are less than one standard deviation below 0), the in-sample analysis rather supports our main finding of unsuccessful factor timing.

[25] Neither return-based measures such as TM and HM nor holding-based approaches like Jiang et al. (2007) make such a distinction.

[26] This investigation of the return predictive power of factor timing activity is unique and provides insights beyond the results of earlier studies. Other methods, e.g. HM and TM, but also holding-based approaches like Jiang et al. (2007), measure the success of factor timing from an ex post perspective.

Table 3.12: Correlations between factor exposures and future risk premia

This table provides a descriptive overview over the correlations between mutual funds' risk factor exposures and the respective risk premia. We apply the dynamic version of the Carhart (1997) model as described in Section 3.2.2 to each fund over the entire sample period or the sub-period during which the fund was alive. We then calculate the correlation between risk factor exposures and risk premia during the subsequent month.

	Mean	5%	10%	25%	Median	75%	90%	95%	Standard Deviation	Skewness	Kurtosis
$\rho(\beta_{RMRF}, r_{mkt} - r_f)$	−0.064	−0.165	−0.154	−0.120	−0.065	−0.002	0.030	0.040	0.068	0.089	1.917
$\rho(\beta_{SMB}, SMB)$	−0.030	−0.075	−0.067	−0.051	−0.031	−0.014	0.013	0.021	0.031	0.582	4.758
$\rho(\beta_{HML}, HML)$	−0.030	−0.106	−0.086	−0.057	−0.021	0.000	0.019	0.032	0.044	−0.779	3.827
$\rho(\beta_{UMD}, UMD)$	−0.012	−0.081	−0.068	−0.041	−0.012	0.014	0.041	0.075	0.044	0.318	3.345

3.4.4 Fund determinants

To understand which funds are pursuing factor timing, we study the relationship between fund characteristics and the individual measures of exposure volatility with regard to the *MKT*, *SMB*, *HML*, and *UMD* risk factors as well as the overall *FEVI*. Since factor exposure volatilities are estimated using 3-year time windows during our 09/1998−12/2016 sample period, we split our sample into six non-overlapping sub-periods, namely 1999−2001, 2002−2004, 2005−2007, 2008−2010, 2011−2013, and 2014−2016. We regress the measures of factor exposure volatility during those periods on the fund characteristics at the beginning of these periods to observe the relationship between ex-ante fund characteristics and timing activity. Table 3.13 reports the results of the multivariate regressions. Specifications (1) – (4) show the results with the individual exposure volatilities as dependent variables, while specification (5) adapts the *FEVI* as the dependent variables. We focus to interpret the results of regression (5) which documents a significant relationship between three sets of fund characteristics and the *FEVI*.

First, we observe that risk factor timing is most common among funds that are old and that are managed by fund managers with long manager tenure. The result is in line with predictions of Chevalier and Ellison (1999) who suggest that manager's behavior is influenced by career concerns and that younger managers have an incentive to not expose their portfolios to unsystematic risk and hold more conventional portfolios. Second, risk factor timing is positively related to a fund's expenses and portfolio turnover. This finding is in line with Huang et al. (2011) as well as Amihud and Goyenko (2013), who document a positive relationship between a fund's expense ratio and turnover as well as their measures of fund activity. These relationships also confirm our results from Sections 3.4.1 and 3.4.2 that our *FEVI* captures an intended actively implemented investment strategy rather than a coincidental return series characteristic. Furthermore, the positive relationship between a fund's expense ratio and timing activity is either due to additional trading costs for the fund manager's trading strategy (e.g., due to high trading costs or research efforts) or it might indicate investors' willingness to pay for factor timing activity.[27] Finally, we observe that risk factor timing is pursued by fund managers who were successful in the past and have earned high inflows into their strategies due to (i) the availability of cash for new investment strategies, and (ii) changes in the

[27] Amihud/Goyenko (2013) make this argument in the context of selectivity.

Table 3.13: Determinants of factor exposure variation

This table reports the results of multivariate regressions of factor exposure volatility on lagged fund characteristics. We split our sample into non-overlapping 3-year subperiods (1999−2001, 2002−2004, etc.) and regress the measures of factor exposure volatility estimated from the dynamic factor model during those periods on the fund characteristics measured at the beginning of these periods. Fund size, age, management tenure, turnover ratio and total expense ratio are obtained from the CRSP Survivor-Bias Free Mutual Fund Database and fund flows are calculated over the past year. Fund styles are determined by a fund's CRSP objective code. Funds are aggregated on a portfolio level. The measures of factor exposure volatility $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are the standard deviations of a fund's weekly factor exposures during the past three years. The factor loadings are estimated from a dynamic version of Carhart's (1997) four-factor model as introduced in Section 3.2.2. The *FEVI* is the mean of the four cross-sectionally standardized measures of factor exposure volatility. Data on age, tenure, turnover, expense ratio, flows, and total net assets are winsorized at the 1%-level and observations for which the estimated values of $\sigma(\beta_{RMRF})$, $\sigma(\beta_{SMB})$, $\sigma(\beta_{HML})$ and $\sigma(\beta_{UMD})$ are amongst the highest 1% are dropped from the sample. Standard errors are double clustered on fund level and time period. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	(1) $\sigma(\beta_{RMRF})$	(2) $\sigma(\beta_{SMB})$	(3) $\sigma(\beta_{HML})$	(4) $\sigma(\beta_{UMD})$	(5) *FEVI*
ln(tna)	−2.89e−4 (−0.56)	−1.47e−4 (−0.15)	−3.02e−3 (−1.27)	2.12e−3 (1.38)	1.06e−3 (0.10)
ln(fund age)	2.18e−3 (1.47)	6.27e−4 (0.18)	6.67e−3* (1.86)	2.14e−3 (1.11)	3.43e−2 (1.55)
ln(manager tenure)	2.55e−3** (2.45)	8.77e−3 (4.57)	7.28e−3*** (2.86)	6.44e−3*** (7.46)	6.37e−2*** (6.51)
Expenses (in %)	2.78e−2*** (6.78)	6.61e−2*** (18.35)	5.99e−2*** (8.25)	3.99e−2*** (8.46)	5.14e+1*** (9.42)
Turnover ratio	5.19e−3*** (3.09)	1.39e−2*** (5.82)	1.23e−2*** (3.05)	1.56e−2*** (11.71)	1.21e−1*** (8.01)
Past Alpha	7.11e−3 (0.17)	−3.31e−2 (−0.69)	1.10e−2 (0.40)	4.93e−3 (0.18)	5.71e−2 (0.18)
Fund Flows	1.98e−6*** (3.26)	2.41e−3 (1.19)	1.50e−2** (2.52)	2.53e−3** (2.13)	4.75e−2*** (4.86)
Style dummy variables					
Growth and Income	–	–	–	–	–
Growth	0.013*** (4.75)	0.030*** (4.64)	0.033*** (6.33)	0.017*** (4.24)	0.239*** (9.05)
Hedged	0.032** (2.49)	0.052** (2.24)	0.061* (1.71)	0.004 (0.44)	0.388*** (3.48)
Income	0.003* (1.85)	−0.007 (−0.51)	0.007 (−0.83)	0.002 (0.32)	0.012 (0.21)
Micro	0.036*** (6.35)	0.080*** (4.36)	0.079*** (6.24)	0.063*** (5.02)	0.701*** (7.87)
Mid	0.029*** (3.25)	0.066*** (8.31)	0.065*** (3.53)	0.053*** (6.56)	0.560*** (6.85)
Small	0.021*** (2.97)	0.057*** (7.55)	0.062*** (2.73)	0.034*** (5.11)	0.432*** (6.53)
R^2	0.24	0.17	0.23	0.26	0.24

mindset of (successful) managers who become overconfident and spend their money in costly active trading strategies (see Puetz and Ruenzi, 2011).[28]

To summarize, our results reveal that a part of the negative relationship between the volatility of risk factor exposures and future performance is due to factor-loading uncertainty of funds' stock holdings (see Armstrong et al., 2013). However, this effect only partly explains the negative association between our main *FEVI* and future fund performance. Fund flows and thus asset fire sales and purchases cannot explain our results either. We thus conclude that unsuccessful factor timing is the main driver of our results and provide supportive in-sample evidence. Finally, we show that this *FEVI* is strongly correlated to certain fund characteristics, such as fund manager tenure, fund's expenses and portfolio turnover, and past flows.

3.5 Conclusion

Mutual fund managers vary their exposure to risk factors over time. To measure this investment pattern, we propose a new measure of factor loading variation based on a dynamic version of the Carhart (1997) four-factor model. Using this measure, we investigate whether a variation in factor exposure is linked to fund performance within a sample of US mutual funds during the time period from the late 2000 up to 2016.

We find that the volatility of factor exposures is a persistent fund characteristic and associated with future underperformance. A portfolio of the 20% funds with the highest *FEVI* underperforms the 20% funds with the lowest *FEVI* by risk-adjusted 147 basis points p.a. with statistical significance at the 1% level. Similarly, sorting funds on the volatility of individual *MKT*-, *HML*-, or *UMD*-exposures, results in underperformance of the funds with the most unstable factor loadings by 102, 82, and 120 basis points p.a., respectively, with statistical significance at least at the 5% level. We also show that the underperformance is not explained by different risk factors, fund characteristics, or similar activeness measures, such as the R^2 selectivity measure by Amihud and Goyenko (2013) or the Huang et al. (2011) risk shifting measure.

[28] We also include style dummies in our regressions and find that factor exposure volatility is higher for growth funds as well as mid, small and especially micro-cap funds.

Our results also provide evidence that the relationship between factor exposure volatility and performance is mainly driven by fund managers' active trading decisions and less so by the variation of single stocks' factor exposures. Moreover, it is not driven by asset sales and purchases in response to investment flows into or out of the fund. We conclude that unsuccessful factor timing leads to an underperformance and show that risk factor timing is particularly prevalent among funds with long management tenure, high turnover and total expense ratio, and high past fund inflows. Our results do not support the hypothesis that deviations in risk factor exposures are a signal of skill and we recommend that investors should resist the temptation to invest in funds that intentionally or coincidentally vary their exposure to risk factors over time.

References

Amihud, Y., Goyenko, R., 2013. Mutual fund R^2 as predictor of performance. Review of Financial Studies, 26, 667–694.

Armstrong, C. S., Banerjee, S., Corona, C., 2013. Factor-loading uncertainty and expected returns. The Review of Financial Studies, 26, 158–207.

Arnott, R. D., Beck, N., Kalesnik, V., West, J., 2016. How can 'smart beta' go horribly wrong? Working Paper.

Asness, C. S., 2016. The siren song of factor timing aka "smart beta timing" aka "style timing". The Journal of Portfolio Management, 42, 1–6.

Asness, C. S., Friedman, J. A., Krail, R. J., Liew, J. M., 2000. Style timing: Value versus growth. The Journal of Portfolio Management, 26, 50–60.

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

Barber, B. M., Huang, X., Odean, T., 2016. Which factors matter to investors? Evidence from mutual fund flows. The Review of Financial Studies, 29, 2600–2642.

Barroso, P., Santa-Clara, P., 2015. Momentum has its moments. Journal of Financial Economics, 116, 111–120.

Benos, E., Jochec, M., Nyekel, V., 2010. Can mutual funds time risk factors? The Quarterly Review of Economics and Finance, 50, 509–514.

Black, A., Fraser, P., Power, D., 1992. UK unit trust performance 1980–1989: a passive time-varying approach. Journal of Banking & Finance, 16, 1015–1033.

Blake D., Lehmann B., Timmermann, A., 1999. Asset allocation dynamics and pension fund performance. Journal of Business, 72, 429–461.

Bodnaruk, A., Chokaev, B., Simonov, A., 2014. Downside risk timing by mutual funds. Working Paper.

Bollen, N. P., Busse, J. A., 2001. On the timing ability of mutual fund managers. The Journal of Finance, 56, 1075–1094.

Brunnermeier, K., Nagel, S., 2004. Hedge Funds and the Technology Bubble. The Journal of Finance, 59, 2013–2040.

Busse, J. A., 1999. Volatility timing in mutual funds: Evidence from daily returns. Review of Financial Studies, 12, 1009–1041.

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

Chabi-Yo, F., Ruenzi, S., Weigert, F., 2018. Crash sensitivity and the cross-section of expected stock returns. Journal of Financial and Quantitative Analysis, 53, 1059–1100.

Chen, Y., Liang, B., 2007. Do market timing hedge funds time the market? Journal of Financial and Quantitative Analysis, 42, 827−856.

Chevalier, J., Ellison, G., 1999. Career concerns of mutual fund managers. Quarterly Journal of Economics, 114, 389–432.

Coval, J., Stafford, E., 2007: Asset fire sales (and purchases) in equity markets. Journal of Financial Economics, 86, 479−512.

Cremers, K. J. M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. Review of Financial Studies, 22, 3329–3365.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. The Journal of Finance, 52, 1035–1058.

Dong, X., Feng, S., Sadka, R., 2017. Liquidity Risk and Mutual Fund Performance. Management Science. Forthcoming.

Elton, E. J., Gruber, J. M., Blake, C. R., 2012. An examination of mutual fund timing ability using monthly holdings data. Review of Finance, 16, 619–645.

Engle, R. F., 2016. Dynamic conditional beta. Journal of Financial Econometrics, 14, 643–667.

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

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

Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics, 116, 1–22.

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

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

Giambona, E., Golec, J., 2009. Mutual fund volatility and management fees. Journal of Banking and Finance, 33, 589–599.

Goetzmann, W. N., Ingersoll, J., Ivković, Z., 2000. Monthly measurement of daily timers. Journal of Financial and Quantitative Analysis 35, 257−290.

Grinblatt, M., Titman, S., 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. Journal of Business, 62, 393–416.

Grinblatt, M., Titman, S., 1993. Performance measurement without benchmarks: An examination of mutual fund returns. Journal of Business, 66, 47–68.

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

Helske, J., 2016. KFAS: Exponential family state space models in R. Working Paper.

Henriksson, R. D., Merton, R. C., 1981. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. Journal of Business, 54, 513–533.

Hollstein, F., Prokopczuk, M. (2016). Estimating beta. Journal of Financial and Quantitative Analysis, 51, 1437−1466.

Huang, J., Sialm, C., Zhang, H., 2011. Risk shifting and mutual fund performance. Review of Financial Studies, 24, 2575–2616.

Jiang, G. J., Yao, T., Yu, T., 2007. Do mutual funds time the market? Evidence from portfolio holdings. Journal of Financial Economics, 86, 724–758.

Jostova, G., Philipov, A., 2005. Bayesian analysis of stochastic betas. Journal of Financial and Quantitative Analysis, 40, 747–778.

Kacperczyk, M., Nieuwerburgh, S. V., Veldkamp, L., 2014. Time-varying fund manager skill. The Journal of Finance, 69, 1455–1484.

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

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 35–45.

Kelly, B., Jiang, H., 2014. Tail risk and asset prices. Review of Financial Studies, 27, 2841–2871.

Kim, S., In F., 2012. False discoveries in volatility timing of mutual funds. Journal of Banking and Finance, 36, 2083–2094.

Kryzanowski, L., Lalancette, S., To, M. C., 1997. Performance attribution using an APT with prespecified macrofactors and time-varying risk premia and betas. Journal of Financial and Quantitative Analysis, 32, 205−224.

Lenz, M., 2017. Factor loading persistence and expected returns. Working paper.

Malkiel, B. G., 1995. Returns from investing in equity mutual funds 1971 to 1991. The Journal of Finance, 50, 77–91.

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

Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. Review of Financial Studies, 21, 233–264.

Moreira, A., Muir, T., 2017. Volatility-managed portfolios. The Journal of Finance, 72, 1611–1644.

Pástor, L., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political economy, 111, 642–685.

Puetz, A., Ruenzi, S., 2011. Overconfidence among professional investors: Evidence from mutual fund managers. Journal of Business Finance & Accounting, 38, 684–712.

Rachev, S. T., Mittnik, S., Fabozzi, S. M., 2007. Financial Econometrics: From basics to advanced modeling techniques. Wiley.

Racicot, F. É., Théore, R., 2009. Modeling hedge fund returns using the Kalman filter: An errors-in-variables perspective. Working Paper.

Sensoy, B.A., 2009, Performance evaluation and self-designated benchmark indexes in the mutual fund industry. Journal of Financial Economics, Journal of Financial Economics, 92, 25−39.

Swinkels, L., Tjong-A-Tjoe, L., 2007. Can mutual funds time investment styles? Journal of Asset Management, 8, 123–132.

Swinkels, L., Van Der Sluis, P. J., 2006. Return-based style analysis with time-varying exposures. European Journal of Finance, 12, 529–552.

Treynor, J., Mazuy, K., 1966. Can mutual funds outguess the market? Harvard Business Review 44, 131–136.

Wells, C., 1994. Variable betas on the Stockholm exchange 1971−1989. Applied Financial Economics, 4, 75–92.

Wermers, R., 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. The Journal of Finance, 55, 1655–95.

Appendix 3.A: Kalman Filter

Kalman filtering was introduced to engineering in 1960[29]. The algorithm derives estimates of unobservable state variables from a time series of observable variables that contains statistical noise. In our case, the unobservable state variables are the risk factor loadings, which are estimated from a return time series. The Kalman filter requires a mathematical model that describes the dynamics of the unobservable state variables. In our main specification, we assume the factor loadings to follow a mean-reverting process.

The optimization follows a recursive two-step process. At each time t, the Kalman filter uses information up to time t to estimate the current state variables (i.e., factor loadings) as well as their uncertainties. It then uses the observed noisy measurement (i.e., the fund return) to update the estimate using a weighted average forecast. The algorithm gives more weight to estimates with lower uncertainty. In addition to the Kalman *filter* technique*,* we also apply a Kalman *smoother* in our estimations. The Kalman smoother additionally contains a backward procedure that utilizes observations that occur after time t to estimate state variables at time t. The Kalman smoother is more suitable to estimate the factor loading dynamics from an ex-post perspective. Rachev et al. (2007) provide an introduction to the Kalman filter and its application in finance. Racicot and Théore (2009) provide an overview over the historical use of Kalman filters in finance, which started in the 1980s. Black, Fraser, and Power (1992) have been the first to measure time-varying factor exposures via the Kalman filter and similar approaches have later been used e.g. by Wells (1994), Brunnermeier and Nagel (2004), Jostova and Philipov (2005), Swinkels and Van Der Sluis (2006), Mamaysky, et al. (2007) and Mamaysky, et al. (2008). Hollstein and Prokopczuk (2016) find that a Kalman Filter model outperforms other methods in estimating market betas of single stocks. An important difference among earlier studies and our paper is the assumed process of factor loadings. Whereas some papers assume a random walk, we follow Wells (1994), Jostova and Philipov (2005) who assume a mean-reverting process. This is also in line with the findings of Blake et al. (1999) who document a mean reversion in funds' portfolio weights within a sample of U.K. pension funds. We execute the Kalman filter using adapted functions from the Jouni Helske's KFAS package (Helske, 2016) in the software environment R.

[29] See Kalman (1960).

Chapter 4

International Evidence on Mutual Fund Turnover as a Predictor of Fund Performance

Sebastian Fischer

Status: Working Paper (February 2019)

Abstract

Pástor, Stambaugh and Taylor (2017) model fund turnover as a function of market mispricing and derive a positive time-series relationship between a fund's turnover and its subsequent performance. To provide out-of-sample evidence, I investigate this relationship in an international fund sample between 1992 and 2018. I find fund turnover to be strongly related to future performance in most countries. Furthermore, I establish a positive link between management skill as well as the extent to which fund managers trade on mispricing and the turnover-performance relationship, and I find market momentum to be negatively related to this relationship. My results strongly support the main idea of the original model and enable a better understanding of the circumstances under which fund turnover predicts future fund performance.

JEL Classification: G10, G20

Keywords: Mutual Fund, Turnover, Performance, International Evidence

4.1 Introduction

The ongoing debate whether mutual fund managers possess the skill to outguess the equity market has shaped the mutual fund literature throughout the past decades. The growing size of the global equity mutual fund market with $22.1 trillion invested in almost 39,000 equity funds at the end of 2017 reveals investors' trust in such an ability.[1] The empirical evidence, however, is mixed. An extensive body of literature shows that equity funds underperform their benchmark net of fees.[2] The negative net-of-fee alpha is important for an investor but does not necessarily indicate a lack of skill. Management fees might transfer abnormal returns achieved by skilled fund managers from the investors to the managers. If investors allocate money to the most skilled managers, decreasing returns to scale might erode the abnormal performance even in the presence of fund manager skill.[3] Literature has studied the nature of fund manager skill and found fund managers' stock-picking ability (Chen et al. (2000) and Wermers (2000)), their ability to forecast earnings-related fundamentals (Cohen et al. (2005) and Baker et al. (2010)) and a general market timing ability (Bollen and Busse (2001), Jiang et al. (2007), Mamaysky et al. (2008)).[4] In their 2017 Journal of Finance paper, Pástor, Stambaugh and Taylor (PST hereafter) present a model that transfers fund managers' ability to detect profit opportunities into a positive turnover-performance relationship. Their model and supportive empirical evidence shed further light on the nature of fund managers' skill as it shows that fund managers, on average, have the ability to trade on mispricing. These trades result in a higher performance during a subsequent period, when the mispricing disappears, and thus lead to a positive time-series relationship between fund turnover and future performance. This main result of PST is surprising for two reasons: Firstly, if turnover is a result of market mispricing and fund managers' ability to detect it, then more skilled fund managers should have a higher level of turnover and this would imply a positive cross-sectional relationship. Empirical measures of this relationship should additionally be upward-

[1] See Investment Company Institute (2018). In 2017, total assets managed by global equity mutual funds account for 27.5% of the market capitalization of global equities (in 2010: 23.1%).
[2] See, for example, Jensen (1968), Malkiel (1995), Gruber (1996), Carhart (1997), and Fama and French (2010) on the performance of domestic US equity funds. See Cumby and Glen (1990) on international funds.
[3] Grinblatt and Titman (1989, 1993) and Daniel et al. (1997) find positive gross-of-fees abnormal returns. Berk and Green (2004) and Pástor and Stambaugh (2012) model decreasing returns to scale on a fund-level and industry-level and argue that it might lead to the coexistence of skill and non-positive alphas. Berk and van Binsbergen (2015) and Pástor et al. (2015) provide empirical evidence for this coexistence.
[4] The market timing ability of mutual fund managers is controversial as discussed in Chapter 3 of this dissertation.

biased as some of the time-series relationship will be measured as a contemporaneous relationship due to the crudeness of the lag in the turnover measure.[5] The empirical literature, however, has shown mixed evidence on the cross-sectional relationship between fund turnover and performance.[6] Secondly, the model shows that turnover predicts subsequent performance for single investors. As each trade requires a counterpart, mispricing should not only lead to higher turnover of skilled investors, but there must be unskilled investors who buy (sell) the overpriced (underpriced) assets from (to) the skilled investors. Therefore, mispricing must also increase the turnover of unskilled investors. As a result, turnover should predict future underperformance of unskilled investors. It remains unclear who those investors are and why they make a larger number of unfavorable trades in times of higher market opportunities. Given the strong but surprising results of PST, an out-of-sample study of the turnover-performance relationship can help to either confirm the model's validity or to reject its general validity. Therefore, I apply the empirical study of PST to a broad international fund sample covering 13,378 distinct equity mutual funds that are managed from 40 different countries and invest into 41 different regional markets. Besides an out-of-sample test, the international structure of the fund management companies and the investment areas allows me to investigate differences across countries and markets and link them to differences in the turnover-performance relationship. I particularly link measures of country-specific fund management skill and market-specific return momentum to the international results on the time-series relationship between fund turnover and performance. The results help to understand the requirements for and drivers of the PST model.

In greater detail, the economic intuition behind the model of PST reads as follows. By assumption, there exists mispricing in the (equity) market, which bears profit opportunities and will deliver positive excess returns (over the benchmark) in the future. In order to realize those profit opportunities in some period $t+1$, an investor (or fund manager) has to shift her assets and thus actively trade during period t. This leads to a positive relationship between the optimal portfolio turnover at t and profit at $t+1$. This relationship is concave since an investor

[5] In their empirical study, PST regress monthly fund returns on the turnover during the prior fiscal year. By that, they implicitly assume that any mispricing funds trade on is corrected during the subsequent fiscal year. If, however, some of the mispricing disappears earlier, part of the time-series relationship predicted by PST should empirically be captured by the cross-sectional relationship.

[6] Among others, Elton et al. (1993) and Carhart (1997) find a negative and Dahlquist et al. (2000) and Chen et al. (2000) a positive relationship. Wermers (2000), Kacperczyk et al. (2005) and Edelen et al. (2007) report no significant relationship.

will primarily select stocks with the highest profit opportunities and select stocks with lower profit opportunities at a lower priority. If portfolio turnover is associated with (almost) proportional costs, i.e. the relationship between turnover at *t* and costs at *t* is (close to) linear, there is a level of optimal portfolio turnover. For a skilled investor, whose trading is only driven by such mispricing, this leads to a positive linear time-series relationship between the expected benchmark-adjusted portfolio return and lagged portfolio turnover.[7] This result holds even if an investor trades suboptimal and cannot realize the full profit opportunity, but only a fraction $0 < \delta < 1$ of it. Specifically, the slope *b* of a time-series regression of benchmark-adjusted returns on lagged turnover is determined by four characteristics: It increases in the trading costs *c* associated with portfolio turnover and the skill indicator δ (i.e. the fraction of profit potential an investor can realize), but decreases with the average auto-correlation of the portfolio turnover. Finally, the curvature of the relationship between optimal turnover and future performance influences the regression slope. Besides this main prediction of a positive time-series regression slope (hypothesis 1), PST derive multiple further predictions from their model: The cross-sectional relationship between portfolio turnover and the expected future excess return is also positive (2) but lower than the time-series relationship (3). It is stronger for funds that trade less liquid stocks (4) and among investors that possess greater skill (5). Finally, they predict a positive correlation between turnover across investors (6), a positive correlation between average turnover and proxies for stock mispricing (7) and a predictive power of the average portfolio turnover for single portfolios' performance (8). PST use a sample of domestic US equity mutual funds to provide evidence for those predictions.[8] They operationalize the discrete periods of their model by regressing monthly benchmark-adjusted fund returns (a fund's gross return minus the return of the benchmark index as assigned by Morningstar) on the fund's turnover during the previous fiscal year (i.e. the most recent fiscal year that ended prior to the respective month). Despite the crudeness of this lagged turnover measure, they find convincing evidence for all their predictions. In this paper, I test all eight model predictions for an international fund sample covering funds managed in 40 countries and investing into 41 different equity markets. My results strongly support the main idea of the PST model. The positive within-fund time-series relationship between fund turnover and subsequent performance (hypothesis 1) strongly persists outside

---

[7] As mispricing might occur within industries, across industries or on a market-wide level, the results of PST might be driven be either stock picking, industry selection or market timing ability.
[8] PST use a fund sample merged from Morningstar Direct and CRSP as in Pástor et al. (2015).

the US and can be observed within 18 of the 20 most common investment areas in my fund sample. I do, however, not find support for a cross-sectional relationship outside the US (hypothesis 2) and as a result find strong support for hypothesis 3. In line with PST, I find a strong positive correlation between turnover across funds (hypothesis 6) and a positive correlation between the average portfolio turnover and single portfolios' subsequent performance (hypothesis 8). My results on the hypotheses 4 and 5, a stronger turnover-performance relationship among funds that trade less liquid stocks or whose managers possess greater skill, as well as the relationship between mispricing measures and average turnover (hypothesis 7) are mixed.

International evidence does not only provide an out-of-sample tests to uncover coincidental results, but literature has shown marked differences between global fund markets and performance characteristics. E.g., Ferreira et al. (2012) show that investors in more developed countries react more sensitively to mutual fund performance, which leads to differences in fund managers' risk taking. Ferreira et al. (2013) point out international differences in the determinants of fund performance. They show that the well-established diminishing returns to scale effect in the US fund market does not hold internationally and that funds perform better in countries with stronger legal institutions and more liquid stock markets. Cremers et al. (2016) show that funds perform better in countries with more explicit index funds and fewer closet indexers. In the same vein, international differences in the turnover-performance relationship might be explained by differences in international fund markets and international equity markets and I identify three main determinants. Firstly, funds managed in countries with higher management skill, either measured by the average Berk and van Binsbergen (2015) skill measure or the countries' IMF Financial Development Index, have a stronger turnover-performance relationship. Secondly, this relationship is stronger among funds that trade more heavily during times when mispricing is more likely, that is in times of higher return dispersion, lower market liquidity or a higher market sentiment. Lastly, the relationship is stronger for funds that invest in markets with a low return momentum, which is in markets where an optimal portfolio strategy requires more turnover. Of these three determinants, solely the impact of skill is formally predicted by the PST model. Nevertheless, the latter two determinants also fit the model's economic idea well. Fund managers who trade more when mispricing is more prevalent, are more likely to invest in undervalued stocks than fund managers who trade when there is little mispricing. In markets with a strong momentum, an optimal portfolio strategy requires lower fund turnover and thus high levels of turnover are more likely to be driven by noise trading rather than beneficial investment decisions.

My results help to better understand the motives of mutual funds' trading activity and provide evidence on the circumstances under which fund turnover is beneficial for a mutual fund's performance. By that, my research is not only related to the PST paper, but also contributes to the streams of literature, PST relate to. The PST model indicates that fund managers spot time-varying profit opportunities and exploit them by selling overpriced and purchasing underpriced stocks. By doing so, turnover as a dimension of fund activeness leads to future fund outperformance.

Prior literature has identified measures of fund activeness that are a cross-sectional predictor of fund performance. Kacperczyk et al. (2005) find that funds with more concentrated industry exposures outperform funds with a lower industry concentration. Cremers and Petajisto (2009) show that active share, a measure of a fund's portfolio deviation from its benchmark, predicts a fund's benchmark-adjusted performance. Amihud and Goyenko (2013) use the R^2 of risk factor regression of fund returns on a Carhart (1997) model as a measure of fund activeness and provide empirical evidence of its connection to future returns.[9] In contrast, Huang et al. (2011) find that risk shifting, that is the change of portfolio risk over time, is a negative predictor of future fund performance. Whereas all these measures of fund activeness are positively correlated to a fund's turnover ratio as reported in the original papers, PST are the first who relate fund turnover to fund performance in a time-series regression. My research adds to this literature by showing that turnover as a measure of activeness is a particularly strong predictor of fund performance when management skill is high, when trading is motivated by market mispricing and in markets with a low return momentum. These results contribute to the debate on fund activeness as they show that activeness might only be a good performance predictor among a subset of funds.

The model of PST also adds to the literature on time-varying mutual fund performance. Among others, Kosowski (2011) and Kacperczyk et al. (2014) empirically detect time-varying fund manager skill.[10] My results further refine the contribution of PST by showing

[9] A high R^2 indicates a low level of activeness and is associated with future underperformance.
[10] A more extensive discussion of this literature is provided in the original PST paper.

that not all fund managers trade on market mispricing, but those that do, perform better during periods following times of larger mispricing and thus after periods when turnover was high.

Finally, PST relate to the broad literature on "herding", that is correlated trading. A large body of literature, e.g. Grinblatt et al. (1995) and Wermers (1999), detects herding behavior among funds.[11] Sias (2004) provides evidence that herding is a result of institutional investors inferring information from each other's trades. Walter and Weber (2006) identify changes in benchmark index compositions to cause herding. Brown et al. (2013) argue that due to career concerns, mutual funds herd in response to sell-side analyst reports. Among others, Nofsinger and Sias (1999) and Sias et al. (2001) find herding to impact asset prices. Dasgupta et al. (2011) provide a theoretical framework that explains this price impact. PST contribute to this literature by showing significant commonalities among mutual funds and I add to this contribution by showing that the commonality is not a national phenomenon as it is not stronger within single fund management countries or among funds that share a common regional investment area.

The reminder is structured as follows. Sections 4.2 describes the data sources and provides descriptive statistics on my international fund sample. Section 4.3 investigates the time-series and cross-sectional relationship between fund turnover and mutual fund performance. Section 4.4 links fund characteristics to the turnover-performance relationship and Section 4.5 observes international comovements in fund turnover and structural drivers of turnover. Section 4.6 provides explanations for the international differences in the turnover-performance relationship. Section 4.7 concludes.

4.2 Data selection and descriptive statistics

I collect international fund data from the Morningstar Direct[12] (MD) mutual fund database, which has previously been used by e.g. Busse et al. (2013), Cremers and Pareek (2016), Patel and Sarkissian (2017). Morningstar collects information on mutual funds from numerous countries and also provides some proprietary information such as the Morningstar Category, a holding-based fund classification, and a corresponding Morningstar Index that serves as a benchmark for funds of that Morningstar Category. Specifically, I define the fund universe

[11] The broader herding literature is discussed in the original PST paper as well as by Wermers (2002).
[12] I thank Morningstar Switzerland GmbH for providing access to Morningstar Direct.

used in this paper by all equity funds (as indicated by the Global Broad Category Group in MD) for which any return information between January 1992 and August 2018 is available from MD. To prevent a survivorship bias, I also include defunct funds. For all funds, I collect monthly gross returns and net returns, monthly data on total net assets and the Morningstar Category a fund is assigned to as well as the fund's monthly equity allocation (in % of total assets). Throughout the paper, I use USD as the base currency. For funds that do not report USD values as their base currency, MD calculates USD returns and converts the fund size into USD.[13] I also collect annual data on a fund's net expense ratio and its annual portfolio turnover starting in 1991. The turnover ratio from MD is either calculated in accordance with the regulatory requirements of the US SEC or the requirements of the UCITS. For Taiwanese funds, its calculation follows the Taiwan FSC requirements. The US SEC method divides the lesser of a fund's purchases and sales by its average monthly net assets. The UCITS method takes the sum of purchases and sales (both in absolute values) minus the sum of investors' inflows and outflows (both in absolute values) and divides this difference by the fund's average monthly net assets. For Taiwanese funds, sales turnover and purchases turnover are calculated separately and only the lower ratio is reported. Due to the different calculation methods, a comparison of turnover ratios across countries has to be interpreted with caution, although all three methods should yield similar results. For the main focus of this paper, the within-fund turnover-performance relationship, however, the different calculation methods do not matter as the method remains constant over time within a fund domicile and therefore within each fund. I aggregate funds across share classes by matching funds with a common *fundid* in MD. A fund's size is the sum of the total assets of its share classes, its fees and returns are calculated as the weighted average of single share classes' fees and returns with the weights being given by the share classes' total net assets at the end of the previous month. To prevent a survivorship bias stemming from reporting conventions, I follow Elton et al. (1996) and drop funds with total net assets (*TNA*) below USD 15 million.[14] I exclude funds that were identified as index or enhanced index funds by MD.[15]

---

[13] Following PST, I consider excess returns over a fund benchmark index. This excess return is independent of the currency. Therefore the choice of USD as the single base currency does not influence my results.
[14] The same threshold has been used by PST and previously by e.g. Chen et al. (2004), Yan (2008), Amihud and Goyenko (2013) and others.
[15] According to MD's definition, enhanced index funds "attempt to match an index's performance. Unlike index funds, however, enhanced index funds attempt to better the index by either adding value or reducing volatility through selective stock-picking."

As in PST, I use excess gross returns, i.e. the difference between fund returns before fees and the return of the benchmark index. If a fund's monthly gross return is missing in MD, I calculate it from monthly net return data by adding 1/12 of the ex post reported annual expense ratio. For the benchmark return, I use the monthly USD return of the Morningstar Index which MD assigns as a benchmark to the respective Morningstar Category.

The international evidence in this paper will not only serve as an out-of-sample test of PST's results, but I will relate country-specific results to country and market characteristics. I therefore classify each fund in two geographic dimensions, its investment area, i.e. the country or region a fund primarily invests in, and the location of its fund management.[16] To determine the former, I collect multiple indicators of a fund's geographic investment area from MD: A fund's Global Category, its investment area according to MD, the Morningstar Category, Morningstar Index, and its MPT index. I assign this information to countries and regions (e.g. Latin America) and determine a fund's investment area by its prospectus if the five indicators yield contradictory results. To determine the fund management location, I use the fund management's phone number and address from MD. I match phone numbers to countries by the country code (if given) and in some cases by the structure of the phone number. For 54% of all funds, the address and phone number point to the same country. For 35% of all funds I can only match one of the two criteria unequivocally to a country. In 8% of all cases the two country indicators are conflicting. I then assign the fund to the country the address corresponds to. For the remaining 3% I cannot determine a fund management country and I drop these funds from the sample. Finally, I drop funds for which I have fund turnover data for only one year or equal turnover values for all years. The main result of PST is a within-fund time-series relationship between fund turnover and subsequent fund performance. To measure such a relationship, I require heterogeneity in turnover of each single fund.

Applying all abovementioned filters leaves me with 1,129,597 fund-month observations of 13,378 distinct mutual funds. These funds are managed in 40 different countries and invest into 41 different investment areas, including single countries as well as broader regions (e.g.

[16] Besides the fund's management country or its investment area a fund could also be classified by its legal domicile or the country a fund is available for sale in. Both dimensions are used by e.g. Cremers et al. (2016). A fund's country available for sale, however, is not unique and is therefore not an appropriate identifier to separate funds into non-overlapping subsamples. A fund's domicile is determined by regulatory considerations. Therefore, I expect the funds' management countries to be a better indicator of commonalities than their domiciles.

"Middle East and Africa" or "Emerging Markets"). Table 4.1 displays the number of distinct funds of the 20 most common fund management countries and investment areas. Roughly 25% of all funds in my sample are domestic US equity funds, the subset of funds investigated by PST.[17] Besides the US, most of the funds in my sample invest in global or European equity markets. Most funds are managed from the US, Luxembourg, Canada or the UK. My sample misses funds from some important fund markets, in particular Japan. This is due to missing data for these funds in MD. This selection bias should, however, not influence my results quantitatively as there is no plausible connection between a fund's turnover-performance relationship and its data availability in MD. Table 4.2 provides descriptive statistics on US managed and non-US managed funds in the final sample. Roughly 46% of all observations in my sample refer to non-US managed funds. On average, these funds have a higher turnover (92% versus 78%)[18], are significantly smaller (USD 352 million versus USD 1,167 million) and charge higher annual fees (1.64% versus 1.21%) than US managed funds. The average monthly gross returns of US managed and non-US managed funds are similar at about 0.81% per month and both classes of funds on average slightly outperform their benchmark (0.02% per month for non-US managed funds and 0.06% per month for US managed funds).

My subsequent analysis also uses the monthly cross-sectional standard deviation of stock return as well as the autocorrelation of stock returns for international equity markets. To calculate these measures, I follow Griffin et al. (2010) to construct a sample of international equity returns. For the US, I use common stocks (share code 10 or 11) traded at the NYSE, AMEX, and NASDAQ from CRSP. For all other countries, I collect equities that are traded on domestic exchanges from Datastream. I follow the multi-step filtering described in the Appendix of Griffin et al. (2010) to drop non-common equities from the sample.[19] Specifically I search the asset's name for numerous global and country-specific terms that indicate other assets than common equities (e.g. "Schuldschein" in the German-speaking area). To deal with data errors, I drop monthly returns >500%, and drop any two subsequent returns if one of them is >100% with the two-month return being <20%. I additionally drop any two subsequent returns of 0.

[17] My sample covers the 1992−2018 time series whereas PST observe funds during 1979−2011.
[18] Notably, fund turnover can be negative when calculated according to the UCITS requirements.
[19] Since I use monthly instead of daily returns I modify the return filtering of Griffin et al. (2010) by setting higher cut-off values for extreme returns.

Table 4.1: Mutual fund sample by management country and investment area

This table displays the number of distinct funds in my sample by the funds' management countries and investment areas. The 20 most common regions are displayed, all others are summarized as «other». The sample includes all equity mutual funds for which return and turnover data during the 1992−2018 time period is available from MD. Funds with less than USD 15 million assets under management and funds without variation in fund turnover are dropped from the sample. A fund's investment area is determined by its Global Category, its investment area according to MD, the Morningstar Category, Morningstar Index, and its MPT index. The fund management country is determined by the management company's address and phone number.

	Investment area																					
Country of fund management	Asia Pacific	Canada	China	Emerging Markets	Europe	France	Global	India	Japan	Latin America	Malaysia	Mexico	Norway	Russia	Sweden	Switzer-land	Taiwan	Thailand	UK	USA	other	**Total**
Austria	11	-	-	10	45	-	92	1	5	1	-	-	-	3	-	1	-	-	-	12	15	196
Belgium	4	-	-	4	40	-	45	-	3	-	-	-	-	-	-	-	-	-	-	10	18	124
Canada	26	458	11	45	37	-	481	-	-	-	-	-	-	-	-	-	-	-	-	281	4	1343
China	5	-	117	2	-	-	9	-	-	-	-	-	-	-	-	-	-	-	-	1	1	135
Denmark	13	-	4	26	51	-	105	3	7	8	-	-	1	3	1	-	-	-	-	17	36	275
Finland	12	-	1	13	61	-	51	5	7	2	-	-	-	11	3	-	-	-	-	11	36	213
France	25	-	3	14	384	135	99	4	19	1	-	-	-	-	-	-	-	-	1	51	2	738
India	-	-	-	-	-	-	-	177	-	-	-	-	-	-	-	-	-	-	-	-	-	177
Ireland	16	-	4	16	30	-	42	2	10	5	-	-	-	-	-	1	-	-	3	27	9	165
Italy	17	-	-	13	41	3	36	-	-	-	-	-	-	-	-	-	-	-	-	17	28	155
Luxembourg	118	2	22	120	390	7	477	23	69	23	-	-	1	15	5	32	5	4	26	171	99	1609
Malaysia	58	-	11	1	1	-	26	-	-	-	116	-	-	-	-	-	-	-	-	-	-	213
Netherlands	10	-	-	6	41	-	52	1	1	-	-	-	-	1	-	-	-	-	-	7	12	131
Norway	2	-	1	6	34	-	56	-	1	-	-	-	58	-	2	-	-	-	-	1	1	162
Sweden	16	-	2	12	52	-	103	2	4	3	-	-	-	6	112	1	-	-	1	11	6	331
Switzerland	19	1	1	14	57	-	106	1	13	2	-	-	-	-	-	105	-	-	4	30	4	357
Taiwan	34	-	6	12	5	-	38	5	5	4	-	-	-	1	-	-	156	1	-	2	3	272
Thailand	11	-	12	4	8	-	22	1	6	-	-	-	-	-	-	-	-	96	-	4	-	164
UK	73	14	12	63	145	-	253	9	54	6	-	-	-	2	2	2	-	-	283	182	18	1118
USA	75	2	27	237	73	4	1139	6	24	1	-	-	-	3	-	-	-	-	7	3346	4	4948
other	70	-	8	31	79	4	142	4	18	-	-	42	-	2	-	10	1	2	6	42	91	552
Total	615	477	242	649	1574	153	3374	244	246	56	116	42	60	47	125	152	162	103	331	4223	387	13378

Table 4.2: Descriptive statistics

Panel A of this table reports the number of fund-month observations, distinct funds and fund management countries as well as investment areas in the final sample. The sample includes all equity mutual funds for which return and turnover data during the 1992−2018 time period is available from MD. A fund's investment area is determined by its Global Category, its investment area according to MD, the Morningstar Category, Morningstar Index, and its MPT index. The fund management country is determined by the management company's address and phone number in MD. Panel B and C report summary statistics for the main variables for funds that are managed in and outside the USA respectively. Turnover is the annual fund turnover, gross return is the fund's return before fee, and excess return is a fund's gross return minus the return of the Morningstar Index which MD assigns as a benchmark to the fund's Morningstar Category. All variables are winsorized at the 1%- and 99%-level.

	# Obs.	Mean	Std.dev.	Min.	25%	Median	75%	Max.
Panel A: Number of funds and countries								
Fund-month observations	1,129,597							
Funds	13,378							
Fund management countries	40							
Investment areas	41							
Panel B: US managed funds								
Fund turnover (in %)	608,883	77.9	74.5	2.0	31.0	59.0	100.0	593.0
Monthly gross return (in %)	608,883	0.81	5.02	−15.13	−1.88	1.16	3.82	14.59
Monthly excess return (in %)	608,883	0.06	2.09	−6.59	−0.95	0.04	1.02	6.99
Total net assets (mn. USD)	601,195	1,167.0	2,393.8	16.1	88.5	295.6	1,001.6	13,420,6
Total expense ratio (in %)	566,220	1.21	0.42	0.18	0.95	1.17	1.44	3.06
Panel C: Non-US managed funds								
Fund turnover (in %)	520,714	91.8	113.1	−38.3	21.6	56.0	120.6	593.0
Monthly gross return (in %)	520,714	0.81	5.63	−15.13	−2.26	1.01	4.12	14.59
Monthly excess return (in %)	520,714	0.02	1.99	−6.59	−0.97	0.01	1.00	6.99
Total net assets (mn. USD)	506,345	352,3	821,5	16,2	45,4	115,1	317,6	13,420.6
Total expense ratio (in %)	290,848	1.64	0.56	0.18	1.34	1.67	1.95	3.06

I collect data on the Pástor and Stambaugh (2003) liquidity factor from WRDS. To determine the business cycle of global economies, I use the composite leading indicator published by the OECD.[20] To estimate the sentiment of international equity markets, I use the consumer confidence index, which is also available from the OECD.[21] Among other, Lemmon and Portniaguina (2006), Schmeling (2009), and Stambaugh et al. (2012) use the consumer confidence index as an estimate for market sentiment. Moreover, the positive correlation (ρ=0.39) of the US consumer confidence index and the Baker and Wurgler (2006) sentiment factor[22] legitimates the use of consumer confidence as an indicator for market sentiment. I measure the return of international equity markets by the return of a country-specific index. For most countries, I chose an index of the S&P BMI series or the S&P IFCI index series, both representing a broad market. Solely for Luxembourg, I chose the Russell Luxembourg index as the S&P indices do not cover the complete 1992−2018 time period. Broad market indices are highly correlated across index providers (ρ>0.95 for all countries and ρ>0.99 in many cases) and the choice of S&P as the favored index provider is therefore arbitrary. I download monthly index returns from MD.

4.3 Fund turnover and mutual fund performance

The main result of PST is the positive time-series regression slope of a within-fund regression of the benchmark-adjusted gross return on lagged turnover. Following the original paper, I test this prediction using the panel data and regressing benchmark-adjusted gross returns during month *t* on a fund's turnover during the last year that ended prior to *t*. The inclusion of fund fixed effects yields the average within-fund regression slope with funds being weighted by their number of observations and the fluctuation in turnover. Including month fixed effects instead of fund fixed effects results in the corresponding estimator for the cross-sectional relationship. Table 4.3 displays the results for three disjunct subsamples: domestic US equity funds as they were used by PST (panel A), non-domestic US equity funds (panel B) and non-US funds (panel C). Thereby, I define US funds and non-US funds by the country the fund

[20] I downloaded data on the composite leading indicators from https://data.oecd.org/leadind/composite-leading-indicator-cli.htm.
[21] Data on the consumer confidence indicator are available form https://data.oecd.org/leadind/consumer-confidence-index-cci.htm.
[22] Data of the Baker and Wurgler (2006) sentiment factor are available from Jeffrey Wurgler's website: http://people.stern.nyu.edu/jwurgler/.

Table 4.3: Turnover-performance relation in the cross-section and time series

This table displays the estimated slope coefficients from panel regressions of funds' monthly excess return on lagged fund turnover. Excess return is a fund's gross of fee return minus the return of the Morningstar Index which MD assigns as a benchmark to the fund's Morningstar Category. Lagged turnover is a fund's turnover during the last calendar year that ended prior to the respective month. Each panel reports the results of four regressions that differ in their inclusion of fund and/or month fixed effects. Panel A reports the results for funds that are managed in the US and invest in the US equity market. Panel B reports the results for non-domestic US managed funds and Panel C for funds that are managed outside the US. All data are from the 1992−2018 time period. T-statistics are reported in parentheses. Standard errors are heteroscedasticity-robust and clustered by Morningstar style category times month as well as by fund unless fund fixed effects are included. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

Panel A: Domestic US equity funds (3,346 funds)		
	Month fixed effects	
Fund fixed effects	No	Yes
Yes	0.00083*** (5.64)	0.00047*** (4.41)
No	0.00032** (2.29)	0.00013 (1.11)
Panel B: Non-Domestic US equity funds (1,602 funds)		
	Month fixed effects	
Fund fixed effects	No	Yes
Yes	0.00036* (1.69)	0.00005 (0.29)
No	−0.00004 (−0.27)	−0.00014 (−1.04)
Panel C: Non-US equity funds (8,430 funds)		
	Month fixed effects	
Fund fixed effects	No	Yes
Yes	0.00024*** (2.94)	0.00013* (1.94)
No	0.00004 (0.40)	0.00000 (0.03)

management is located in. For the US, this definition is almost equivalent to using a fund's domicile as the selection criteria.[23] Domestic US funds are those US managed funds that restrict their investment area to the US. Other funds, including globally invested funds and single funds with an investment area in the USA plus Canada are not considered domestic. The upper left cell of each panel of Table 4.3 contains the regression coefficient for the time-series relationship, the lower right cell the coefficient for the cross-sectional relationship. For completeness I also report the regression slopes including both fund and time fixed effects (upper right) and regression slopes without any fixed effects.[24] The results confirm PST's results of a positive time-series turnover-performance relationship for domestic US equity funds. The regression slope of 0.00083 is roughly 33% below the value calculated by PST, which might be due to a slightly different data sample and particularly due to different time periods.[25] The decline of the results in an out-of-sample test is in line with the findings of McLean and Pontiff (2016)[26]. The positive turnover-return relationship is statistically and economically significant. The t-value of 5.62 is close to the 6.67 reported by PST and with an average within-fund standard deviation of turnover of 34.07%, a one standard deviation increase in turnover translates into an increase in expected annual return of 0.34%. Panel A also reports a positive cross-sectional relationship between turnover and subsequent fund performance. With a regression coefficient of 0.00013 this effect, however, is roughly 67% smaller than in the original paper and is not statistically significant different from 0 at a 10% level.

Using the subsamples of non-domestic US equity funds and non-US funds, both consisting of funds that were not included in the original PST paper, I find evidence for a positive within-fund turnover-performance relationship also for international mutual funds. The slopes of 0.00036 and 0.00024 are much smaller than for domestic US equity funds, but remain statistically significant at least at the 10% level. Economically, a one standard deviation increase

[23] 96 % of all US-domiciled funds are managed in the US and 97% of all US managed funds are domiciled in the US.

[24] By including fund and time fixed effects I calculate the within-fund dependency while controlling for unobserved changes of fund turnover or performance across time, e.g. due to the economic environment. A regression without any fixed effects will capture both within-fund and across-fund dependencies at a time.

[25] Only 65.8% of my observations on domestic US equity funds fall before 2011 and should thus be included in the analysis of PST. Vice versa does my sample not include the 1979−1991 time period that was also investigated by PST.

[26] McLean and Pontiff (2016) show that published research results on market anomalies are usually 26% lower in out-of-sample tests.

in turnover leads to an increase in fund performance of a 0.12% and 0.13% for non-domestic US equity funds and non-US equity funds respectively during the subsequent year. In contrast to the predictions of PST, I do not find any positive cross-sectional relationship between lagged turnover and performance. The regression slope is zero for non-US funds and even negative for non-domestic US funds.

To test the robustness of these results, I split the sample into subsamples of funds, each representing a country or geographic area. As discussed above, funds can be classified by either their investment area, their management country or their domicile. Whereas the latter is mainly driven by regulatory and tax considerations, a fund classification by investment area or management country separates funds by their common underlying equity market or fund market structures respectively. If, for example, equity markets differ in terms of market efficiency, then all funds with the same investment area will share common profit opportunities that emerge from market inefficiencies. If markets for actively managed funds around the globe differ in their competitiveness or investors' willingness to pay for asset management services, then I would expect some fund management locations to attract more skilled managers than others. Either characteristic might potentially lead to differences in the turnover-return relationship. Whereas the impact of skill on the turnover-performance relationship follows directly from PST's model, the degree of market efficiency is not explicitly considered by PST. I assume that some fraction of a fund's trades does not contribute to a better performance but is noise trading and that this fraction is larger among funds that invest in more efficient markets. This assumption would lead to a stronger turnover-performance relationship in less efficient markets.

To determine which dimension, the investment area or the fund management country, is more decisive for the turnover-performance relationship, I proceed as follows. For each fund I regress the monthly benchmark-adjusted return on a constant and the lagged fund turnover to determine fund specific regression coefficients β. I then regress the cross-section of βs on the average β of all funds with the same investment area or the same fund management location excluding the fund itself. Table 4.4 presents the results. Univariate regressions reveal a significant similarity in the turnover-performance relationship among funds with the same investment area, but not among funds with the same management location. A single fund's turnover-performance relationship is on average 0.444 larger if the average turnover-performance beta of funds investing in the same geographical region increases by 1. The dependence is statistically significant after clustering standard errors by the investment area. I additionally calculate the average coefficient among funds of the same investment advisor as this might be an even better category of skill than the fund management location. Yet, a single fund's performance-turnover regression coefficient does not depend on the average coefficient of funds with the same advisory firm. These result remain robust when I use the three average turnover-performance coefficients of all funds with the same investment area, management country and the same advisory firm in one regression (column 4 of Table 4.4). In unreported results, I also test this relationship using the average coefficient of funds with the same legal domicile. The commonality of the turnover-performance relationship is close to zero. A multivariate analysis including the domicile adds no additional insights because domicile and management country are highly correlated.

Table 4.4: Similarity of the turnover-performance relationship across countries

This table reports the results of a two-step regression procedure to express the similarity in the turnover-performance relationship across funds from the same country. In a first step, each fund's monthly fund excess returns over the benchmark are regressed on the fund's lagged turnover. In a second step, the fund-specific first-step regression slopes β^{ts} are regressed on the average regression slopes. The average regression slope is calculated across all funds that share the same investment area ($\overline{\beta^{ts}_{investment}}$), fund management country ($\overline{\beta^{ts}_{management}}$), or investment advisor ($\overline{\beta^{ts}_{advisor}}$). The fund's own regression slope is excluded from the calculation of averages. All data are from the 1992−2018 time period and constructed as described in Section 2. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Dependent variable: β^{ts}			
	(1)	(2)	(3)	(4)
$\overline{\beta^{ts}_{investment}}$	0.444***			0.460***
	(4.84)			(4.87)
$\overline{\beta^{ts}_{management}}$		−0.006		−0.042
		(−0.05)		(−0.33)
$\overline{\beta^{ts}_{advisor}}$			0.011	0.008
			(0.43)	(0.32)

Based on the results of a high commonality of the turnover-performance relationship among funds with the same regional investment area, it appears rational to split the fund sample into subsets based on the investment area and to estimate common time-series and cross-sectional regression coefficients within each subsample. Panel A of Table 4.5 presents these results for the 20 most common investment regions, each being represented by more than 3,000 fund-month observations in my sample. The results are striking: The regression slope of benchmark-adjusted returns on lagged turnover including fund fixed effects and thus measuring the time-series relationship is positive for 18 of the 20 subsamples[27]. It is statistically significant different from 0 at least at the 10% confidence level in 6 of the 20 subsamples and for Canada, Finland and Russia the regression slope is even steeper than reported in the original paper. The only two investment areas with a negative turnover-performance relationship are China and India where the coefficient is less than one standard deviation below 0. When I use month fixed effects instead of fund fixed effects to estimate the cross-sectional relationship, the coefficient is positive for only 12 of the 20 subsamples.[28] It is statistically significant different from zero at least at the 10% level for 2 subsamples with positive and 5 subsamples with negative coefficients. As a result, the time-series relationship is stronger than the cross-sectional relationship in 17 of the 20 subsamples.[29] Altogether, these results confirm the earlier findings of a positive within-fund relationship between turnover and subsequent performance for a clear majority of investment areas. PST's prediction of a positive but less pronounced cross-sectional relationship can only partially be approved. The cross-sectional relationship is significantly smaller than the time-series relationship, but it is not generally above 0. Despite the earlier presented results that suggest a separation of funds by their investment area, I redo the analysis splitting my sample by the country from which a fund is managed. Panel B of Table 4.5 presents the respective results for the 20 most common fund management countries in my sample.[30] The results are less supportive for the main prediction of PST. There is a positive time-series relationship between turnover and performance for only 13 of the 20 subsamples, which is statistically not significantly different from 50%. A closer look into the results reveals that the relationship is rather positive for countries with more observations and the results become diluted by countries with fewer data. It is positive for 8 of the 10 largest subsamples and for 12 or the largest 15 subsamples. Thus, despite the only borderline significant fraction of 13/20 subsamples with a positive regression coefficient, the main prediction of PST is rather supported by the results within the most common management countries. Results on the cross-sectional regression slope and the higher

[27] This is statistically significantly different from 50% at a 1% level.
[28] This is statistically not significantly different from 50% at a 10% level.
[29] This is statistically significantly different from 50% at a 1% level.
[30] There might be countries with more funds, e.g. Japan. The sample, however, is restricted to funds with data on turnover and returns as described in section 4.2.

Table 4.5: Regression results by management country and investment country

This table reports the estimated slope coefficients from panel regressions of funds' monthly excess return on lagged fund turnover. Excess return is a fund's gross of fee return minus the return of the Morningstar Index which MD assigns as a benchmark to the fund's Morningstar Category. Lagged turnover is a fund's turnover during the last calendar year that ended prior to the respective month. Either fund fixed effects or year-month fixed effects are used to measure either a time-series or a cross-sectional relationship. Funds are sorted into subsamples by either their management country (Panel A) or their investment area (Panel B). Within each panel, results are displayed for the 20 largest subsamples. All data are from the 1992−2018 time period. T-statistics are reported in parentheses. Standard errors are heteroscedasticity-robust and clustered by Morningstar style category times month as well as by fund unless fund fixed effects are included. The mean slope coefficients and t-statistics as well as the proportion of positive and negative coefficients are reported at the bottom of each panel. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***. The significance of the inequality of the time-series and cross-sectional regression slopes is determined using a bootstrapping method.

Panel A: Regression results by investment area

Country	Time-series regression		Cross-sectional regression	Country	Time-series regression		Cross-sectional regression
Asia Pacific	0.00013 (0.56)	>*	−0.0003* (−1.87)	Latin America	0.00004 (0.07)	>	−0.00009 (−0.24)
Canada	0.00170*** (4.23)	>***	0.00052** (1.98)	Malaysia	0.00079 (1.30)	>	0.00043 (1.13)
China	−0.00028 (−0.27)	<	−0.00015 (−0.34)	Norway	0.00044 (0.37)	<	0.00079 (1.66)
Emerging Markets	0.00034 (1.57)	>**	−0.00025* (−1.77)	Russia	0.00207** (2.23)	>*	0.00039 (0.99)
Europe	0.00017 (1.31)	>	0.00009 (1.15)	Sweden	0.00028 (0.51)	>	−0.00041* (−1.89)
Finland	0.00188* (1.68)	>	0.00038 (1.48)	Switzer-land	0.00070** (2.33)	>**	−0.00012 (−0.56)
France	0.00064 (1.02)	>	0.00031 (0.68)	Taiwan	0.00047 (0.92)	>	0.00021 (0.93)
Global	0.00016 (1.34)	>***	−0.00018** (−2.24)	Thailand	0.00034 (1.12)	>	0 (−0.01)
India	−0.00022 (−0.93)	<	0.00007 (0.53)	UK	0.00050* (1.66)	>	0.00044** (2.31)
Japan	0.00007 (0.19)	>	−0.00039** (−2.06)	USA	0.00074*** (5.21)	>***	0.00007 (0.75)
Mean	0.00055 (1.32)		0.00009 (0.13)				
Positive/ Negative	18/2	>: 17 <: 3	12/8				

(continued)

(continued)

Panel B: Regression results by management country							
Country	Time-series regression		Cross-sectional regression	Country	Time-series regression		Cross-sectional regression
Austria	0.00032 (0.79)	>	0.00000 (0.00)	Malaysia	0.00017 (0.34)	<	0.00018 (0.56)
Canada	0.00073*** (3.63)	>**	0.00023 (1.35)	Nether-lands	−0.00004 (−0.09)	>	−0.00035 (−1.22)
China	−0.00094 (−0.62)	<	0.00054 (1.4)	Norway	0.00135* (1.8)	>	0.00051 (1.41)
Denmark	−0.00096** (−2.10)	<*	−0.0001 (−0.31)	Singapore	−0.00009 (−0.2)	>**	−0.00107 (−2.81)
Finland	0.00087* (1.91)	>	0.00031 (1.19)	Sweden	0.00002 (0.07)	>	−0.00033 (−2.01)
France	−0.00011 (−0.55)	<	−0.00002 (−0.14)	Switzer-land	0.00029* (1.83)	>***	−0.00025 (−1.79)
India	−0.00029 (−1.21)	<	−0.00005 (−0.4)	Taiwan	0.0005 (1.10)	>	0.0004 (1.83)
Ireland	0.00087** (2.35)	>**	0.00006 (0.35)	Thailand	0.00026 (0.92)	>	0.00009 (0.44)
Italy	−0.00012 (−0.5)	<	0.00023 (1.63)	UK	0.00067*** (3.8)	>***	0.0001 (1)
Luxem-bourg	0.00019* (1.75)	>**	−0.0001 (−1.51)	USA	0.00073*** (5.31)	>***	0.00008 (0.85)
Mean	0.00022 (1.02)	>: 14 <: 6	0.00002 (0.09)				
Positive/ Negative	13/7		12/8				

time-series than cross-sectional regression coefficient remain qualitatively unchanged when splitting subsamples by the fund management country instead of the investment area.

To summarize, the presented results support the main prediction of PST—the positive time-series relationship between fund turnover and subsequent fund return—even in an international fund universe. The variation of the regression slope across subsamples of funds with different investment areas and management countries calls for further investigations on characteristics that might explain this variation. Section 4.6 presents potential sources. I find the time-series relationship to be stronger than the cross-sectional relationship. In contrast to

PST, however, I do not find the latter to be consistently positive across investment area and management area subsamples.

4.4 Fund characteristic differences across funds

The model of PST further predicts that the time-series turnover-performance relationship is stronger for funds that encounter higher trading costs or whose fund managers possess more skill. To test this prediction, PST break down their sample into subsets of funds by either the market capitalization of the stocks a fund invests in, by their book-to-market-ratio, by funds' size or the funds' expense ratios. I follow this approach and sort each fund into one of three stock size categories (large cap, mid cap, small cap) and one of three book-to-market categories (growth, blend, value) as indicated by the Morningstar equity style box.[31] Additionally, I sort funds into size and fee terciles. For all four dimensions I repeat the time-series-regression within each category or tercile. Table 4.6 reports the results. The first and third line of each panel display the results sorting on one dimension without controlling for the other dimensions. The second and fourth line of each panel controls for the other dimensions by including an interaction of fund turnover and category dummies (excluding the mid cap/blend and middle tercile to avoid multicollinearity) in the regression. Trading costs are assumed to be higher among funds trading small cap stocks as those stocks are usually less liquid. Panel A of Table 4.6 presents the respective results. Using the full sample of international funds, I find the turnover-performance relationship to be larger for funds trading small cap stocks than for funds investing into mid-cap stocks and it is larger for funds trading mid-cap stocks than for funds that select large caps. This result is in line with the PST model's prediction. The difference between small cap and large cap funds is 0.00043 and thus more than 50% of the total effect reported in Table 4.3. The result, however, is largely driven by US-funds for which the difference is even statistically significant at the 5% level (unreported results). If I exclude US-funds from the sample, the result disappears, and the difference even turns negative. The pattern remains unchanged if I control for the book-to-market, fund size and fee dimensions.

[31] The Morningstar style box classifies funds in a 3x3 matrix combining a stock size dimensions (large cap, mid cap, small cap) and a book-to-market dimension (growth, blend, value). The style box is usually updated on a quarterly basis. For fund-months-observations with missing style box data, I use the last available style box information unless this is older than 12 months.

Smaller funds trade smaller dollar values and are thus better suited to invest in illiquid stocks. Therefore, it is plausible to assume that the turnover-performance relationship is more pronounced among funds with less assets under management. Panel C of Table 4.6 tests this prediction and in line with PST I find the within-fund relationship to be much larger for smaller funds, both when I include and when I exclude US funds. The relationship is even negative for the largest 33% funds in my sample.

Besides trading costs, management skill should influence the turnover-performance relationship as the model predicts. Skilled fund managers deliver a better performance and are thus able to charge higher fees. Following PST, I therefore test the impact of skill by comparing high-fee funds to low-fee funds. As panel D of Table 4.6 reveals, I do find a (statistically not significantly) larger turnover-performance relationship among more expensive funds and thus the more skilled fund managers, but this effect disappears and turns negative when I exclude US funds. Unreported results show that among US funds the turnover-performance relationship is much larger for fund with high fees and this difference is significant at the 1% level.

For completeness I also compare growth to value stocks (Table 4.6, Panel B). PST report no difference in their turnover-performance relationship and point to the result of Edelen et al. (2013) who report comparable trading costs for both fund types. Surprisingly, I find a strong difference with growth funds having a positive and value funds having a negative turnover-performance relationship. This difference persists after the exclusion of US-funds and when controlling for stock size, fund size, and fund fee categories.

According to the PST model, besides trading costs and skill, differences in the average autocorrelation of fund turnover ρ could also explain the differences in the turnover-performance relationship. PST find respective patterns of a lower autocorrelation for small-cap as opposed to large-cap funds as well as for small funds (i.e. funds with low total net assets) as opposed to large funds. Table 4.7 presents descriptive statistics for fund subsamples constructed by the Morningstar style box, the fund size tercile and the total expense ratio tercile. Other than reported by PST, small-cap funds do not have a lower but a larger turnover autocorrelation and the difference is mainly driven by US-funds. Thus, a difference in ρ cannot explain the differences in the turnover-performance relationship and its disappearance when excluding US funds. The empirical evidence does, however, match the model predictions for the subsamples constructed on fund size, fund fees and the book-to-market ratio of funds' portfolio holdings. Smaller funds have a lower autocorrelation of turnover in both samples, including

Table 4.6: Turnover-performance relation by fund characteristic categories

The table documents the heterogeneity of estimated slope coefficients from panel regressions of funds' monthly excess return over the respective Morningstar Index on lagged turnover and fund fixed effects. Lagged turnover is measured during the previous calendar year. Funds are sorted into subsets by the size (Panel A) or book-to-market ratio (Panel B) of their portfolio holdings (as indicated by the Morningstar equity style box), into fund size terciles (Panel C) or expense ratio terciles (Panel D). Within each panel, results are shown for all funds as well as excluding US managed funds. For all subsamples, I calculate univariate regressions as well as regressions with additional control variables. These controls are an interaction of fund turnover with the following variables: dummy variables for small- and large-cap funds (except in Panel A), dummy variables for growth and value funds (except in Panel B), dummies for fund in the bottom and top size tercile (except in Panel C), and dummies for fund in the bottom and top fee tercile (except in Panel D). Therefore, the regression slope reported in Panel A including control variable should be interpreted as the turnover-performance relationship of a blend fund with medium fund size and a medium expense ratio. All data are from the 1992−2018 time period. T-statistics are reported in parentheses. Standard errors are heteroscedasticity-robust clustered by Morningstar style category x month. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

Panel A: Stock Size Categories

Funds	Small Cap	Mid Cap	Large Cap	Small-Large	Controls
All	0.00057***	0.00021	0.00014*	0.00043	No
	(2.65)	(1.25)	(1.92)	(1.11)	
All	0.00011	−0.00015	−0.00018	0.00029	Yes
	(0.41)	(−0.66)	(−1.28)	(0.66)	
Non-US	−0.00007	−0.00006	0.00007	−0.00014	No
	(−0.21)	(−0.25)	(0.77))	(−0.78)	
Non-US	−0.00033	−0.00034	−0.00028	−0.00005	Yes
	(−0.90)	(−1.04)	(−1.40)	(−0.61)	

Panel B: Stock Value-Growth-Categories

Funds	Growth	Blend	Value	Growth-Value	Controls
All	0.00061***	0.00006***	−0.00027**	0.00088**	No
	(4.70)	(0.55)	(−1.87)	(2.41)	
All	0.00041*	−0.00015	−0.00051**	0.00092***	Yes
	(1.72)	(−0.66)	(−2.03)	(2.99)	
Non-US	0.00057***	−0.00011	−0.00042**	0.00099**	No
	(2.98)	(−0.80)	(−2.21)	(2.18)	
Non-US	0.00034	−0.00034	−0.00066*	0.00100**	Yes
	(0.95)	(−1.04)	(−1.86)	(2.40)	

Panel C: Fund Size Categories

Funds	Small	Medium	Large	Small-Large	Controls
All	0.00073***	0.00027**	−0.00036***	0.00109***	No
	(6.89)	(2.51)	(3.05)	(5.07)	
All	0.00069***	0.00024	−0.00039*	0.00108***	Yes
	(3.41)	(1.15)	(−1.83)	(4.96)	
Non-US	0.00047***	0.00008	−0.00023	0.00070*	No
	(3.17)	(0.50)	(−1.512)	(1.91)	
Non-US	0.00056*	0.00014	−0.00017	0.00073**	Yes
	(1.93)	(0.47)	(−0.57)	(2.06)	

(continued)

(continued)

Panel D: Fund Expense Ratio Categories					
Funds	High	Medium	Low	High-Low	Controls
All	0.00038***	0.00006	0.00008	0.00030	No
	(2.87)	(0.62)	(0.69)	(0.55)	
All	0.00014	−0.00004	0.00007	0.00007	Yes
	(0.82)	(−0.32)	(0.52)	(0.14)	
Non-US	0.00003	0.00008	0.00019	−0.00016	No
	(0.14)	(0.50)	(1.41)	(−1.02)	
Non-US	−0.00015	−0.00027	0.00013	−0.00028	Yes
	(−0.64)	(−1.64)	(0.82)	(−0.99)	

and excluding US funds. This is in line with the stronger turnover-performance relationship of smaller funds. Funds with a higher fee have on average a lower turnover autocorrelation, but this difference is mainly driven by US funds and disappears when I exclude US-funds. That is in line with the previously reported stronger time-series turnover-performance relationship among high-fee funds in a global sample but no such difference in a sample excluding US funds. Growth funds have a higher ρ than value funds, which according to the PST model is consistent with the stronger turnover-performance slope coefficient of growth funds reported above. Finally, it is worth noting that the model's predictions about the cross-sectional turnover-performance relationship do not persistently hold in my analysis outside the US. As predicted by the model, growth funds (as opposed to value funds) and funds charging higher fees (as opposed to funds charging lower fees) have higher turnover values and higher benchmark-adjusted returns. In contrast to that, small cap funds (as opposed to large cap funds) and large funds (as opposed to small funds) that are managed outside the US have lower turnover ratios but higher benchmark-adjusted returns.

Altogether, the analyses on the differences across fund subsamples only partially support the predictions of the PST model outside the US. Higher trading costs should imply a stronger turnover-performance relationship. I find this effect comparing small and large funds but do not observe such a pattern when comparing non-US small cap and large cap funds. The relationship should be stronger among more skilled managers, but I do not find a larger regression slope among more expensive funds. The impact of the turnover autocorrelation on the turnover-performance relationship is negative, as predicted by the PST model, for three of the four subsample analyses. The model's predictions of a positive cross-sectional relationship are not supported by my results outside the US.

Table 4.7: Fund turnover, performance, fees, and size by fund characteristic categories

This table reports the average fund turnover, its standard deviation and autocorrelation, the average benchmark-adjusted return, expense ratio, and median fund size for subsets of funds sorted by the stock size (Panel A) or stock book-to-market ratio (Panel B) of the funds' portfolio holdings (both determined by the Morningstar equity style box), by fund size terciles (Panel C) or expense ratio terciles (Panel D). All subsets are constructed either using the full fund sample or excluding US managed funds, where the management country is determined by the address and phone number of the fund's management company as described in Section 4.2. The benchmark-adjusted return is a fund's gross of fee return minus the return of the Morningstar Index which MD assigns as a benchmark to the fund's Morningstar Category. All variables are downloaded from MD.

Included Funds		Annual Turnover			Monthly excess return	Expense Ratio	Size
		Mean (in %)	Standard Deviation	Autocorre-lation	Mean (in %)	Mean (in %)	Median (mn. USD)
Panel A: Descriptive Statistics by Stock Size							
All	Large	79.83	90.39	0.28	0.02	1.34	198.3
All	Mid	95.89	105.55	0.33	0.07	1.42	166.3
All	Small	81.34	79.41	0.33	0.11	1.37	166.4
Non-US	Large	86.66	105.49	0.21	−0.02	1.64	127.6
Non-US	Mid	106.24	127.59	0.23	0.12	1.73	96.8
Non-US	Small	80.44	104.57	0.21	0.17	1.73	104.1
Panel B: Descriptive Statistics by Stock Book-to-Market Ratio							
All	Growth	92.31	95.78	0.33	0.10	1.36	210.2
All	Blend	80.09	91.34	0.27	0.01	1.38	169.3
All	Value	74.40	88.46	0.29	0.00	1.34	182.6
Non-US	Growth	95.10	114.55	0.22	0.10	1.69	123.2
Non-US	Blend	88.05	107.58	0.20	−0.01	1.65	120.0
Non-US	Value	86.69	108.90	0.21	−0.03	1.65	113.1
Panel C: Descriptive Statistics by Fund Size							
All	Small	100.29	114.88	0.23	0.03	1.59	41.0
All	Mid	84.10	89.85	0.28	0.04	1.39	181.1
All	Large	67.02	68.15	0.38	0.05	1.15	1033.0
Non-US	Small	105.63	126.71	0.19	0.01	1.79	39.1
Non-US	Mid	85.65	103.52	0.21	0.03	1.61	168.0
Non-US	Large	71.13	90.70	0.26	0.03	1.51	741.8
Panel D: Descriptive Statistics by Fund Expense Ratio							
All	Low	69.46	68.12	0.34	0.04	0.84	442.8
All	Mid	81.46	82.94	0.34	0.03	1.32	221.1
All	High	108.01	118.99	0.28	0.07	1.94	95.7
Non-US	Low	86.28	102.17	0.19	0.00	0.77	155.4
Non-US	Mid	90.24	108.59	0.23	0.01	1.41	149.4
Non-US	High	111.25	130.43	0.20	0.04	2.03	82.7

4.5 International comovement in turnover and the predictive power of turnover

PST report a strong comovement of turnover across funds and higher turnover ratios in times of higher market sentiment, larger return dispersion and less liquid markets. I expand the respective analyses to my international fund sample and investigate whether the comovement of mutual fund turnover is primarily within investment areas and fund management countries or also across countries. Following PST, I calculate average monthly turnover values for each fund as the mean of all other funds' turnover ratios (excluding the fund itself) during the same month. In the same manner I calculate the average turnover of all funds with the same investment area, the same fund management origin, and the same fund advisory firm. I also calculate the average turnover of funds of the same stock capitalization style and stock book-to-market style as indicated by the Morningstar 3x3 style box as well as the average fund turnover for funds in the same size and expense ratio tercile. Finally, for each fund-month observation I calculate the average turnover of similar funds, which are funds with the same style box and within the same size and expense ratio tercile. I then regress the pooled turnover data on the average turnover values and fund fixed effects. Table 4.8 presents the results. Regressing turnover on the average turnover across all funds (column 1) indicates significant commonalities in the international sample. The regression coefficient is 0.790 and thus even larger than PST's result of 0.651. The within-R^2 of 2.75 percent also indicates a higher comovement than in the original paper (PST report an R^2 of 1.28 percent). The explanatory power of average turnover increases to an R^2 of 3.69 and 4.43 percent when regressing turnover on the average value of funds with the same investment area and fund management country (columns 2 and 3). This result indicates that the comovement of fund turnover is slightly stronger within-fund management countries than within investment areas. Unreported results show that adding the average turnover across all funds to either the investment area or fund management country-specific average turnover increases R^2 only marginally by 0.21 and 0.12 percentage points. Therefore, the comovement is mainly within a fund management country (and to some extend within an investment area) and not across regions by either definition. One channel of these results might be common investment advisors among funds that are managed in the same country. Column 4 regresses turnover on the average turnover of funds that share the same advisory firm. With an R^2 of almost 10% the within-fund advisory average explains a significant fraction of the turnover variation. Notably, this result is not driven by the small number of funds per fund advisor and thus by construction since average

Table 4.8: Commonality in fund turnover

This table displays commonalities in fund turnover. It reports the coefficients of a panel regression of single funds' turnover on the average turnover across funds during the same year. All regressions include fund fixed effects. The average turnover is calculated across all funds ($AvgTurn_{all}$), funds with the same management country ($AvgTurn_{man}$), the same investment area ($AvgTurn_{inv}$), the same investment advisor ($AvgTurn_{adv}$), funds from the same fund size tercile ($AvgTurn_{fund\ size}$) or expense ratio tercile ($AvgTurn_{expenses}$), funds that invest in the same stock size category ($AvgTurn_{stock\ size}$) or stock book-to-market category ($AvgTurn_{book/market}$) as determined by the MD equity style box or across funds from the same size and fee tercile that also belong to the same style box ($AvgTurn_{similar}$). The fund itself is excluded from the calculation of average turnover values. Standard errors are clustered on a fund and month level. The within fund R^2 is reported at the bottom of the table.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AvgTurn_{all}$	0.790 (19.13)				−0.012 (−0.26)			0.565 (10.93)		0.256 (5.81)
$AvgTurn_{man}$		0.705 (18.86)			0.270 (5.71)					
$AvgTurn_{inv}$			0.670 (18.05)		0.170 (4.87)					
$AvgTurn_{adv}$				0.534 (24.34)	0.457 (24.52)				0.442 (21.44)	0.419 (18.95)
$AvgTurn_{fund\ size}$						0.300 (7.71)				
$AvgTurn_{expenses}$						0.163 (4.87)				
$AvgTurn_{stock\ size}$						0.181 (3.64)				
$AvgTurn_{book/market}$						0.226 (4.44)				
$AvgTurn_{similar}$							0.378 (13.08)	0.255 (8.82)	0.254 (12.15)	0.204 (8.95)
Within R^2 (in %)	2.75	4.43	3.69	9.85	11.12	4.13	3.12	4.24	9.26	9.48
Observations	1,129,597	1,129,016	1,128,534	1,055,833	1,054,703	822,610	855,935	855,935	800,470	800,470

turnover values are calculated for each fund-month observation and do not include the turnover of the fund itself. This result is noteworthy because Section 4.3 has shown that the strong comovement in fund turnover across funds that share the same fund advisor does not translate into a strong similarity of the turnover-performance relationship across those funds (Table 4.4). Including the total average turnover as well as investment area, management country and fund advisor specific averages in one regression (column 5) confirms the previously discussed results: the comovement is strongest among funds that share the same advisor as indicated by the highest regression coefficient. Management country and investment area specific averages are also linked to fund turnover. The overall average turnover does not add additional explanatory power to the model, but its coefficient is close to zero. Columns 6 and 7 support the findings of PST. When turnover is regressed on the average turnover of funds in the same stock size, stock book-to-market, fund size or fund expense ratio category (column 6) or the average turnover of funds that share all four categories (column 7), the regression coefficients are all significantly positive. The coefficients indicate that turnover particularly comoves across funds with similar characteristics. This result remains valid when controlling for the average turnover across all funds (column 8), the average turnover of funds with the same advisory firm (column 9) or both (column 10).[32] All results of Table 4.8 remain qualitatively unchanged if US managed funds are excluded from the sample (unreported results).

A central assumption of the PST model is that trading activity and thus turnover arises in response to profit opportunities, i.e. mispricing in the market. Consequently, the overall level of turnover should be higher during times when mispricing is more likely. PST argue that a high market sentiment, large cross-sectional standard deviations of stock returns, that is dispersion, and low market liquidity indicate market mispricing. Sentiment-driven investors will engage more actively in stock trading when sentiment indicators are high, and their trading will cause mispricing (e.g., Stambaugh et al. (2012)). A high dispersion indicates a heterogeneity in fund returns and thus investment opportunities. Finally, reduced market liquidity will allow mispricing to persist and thus indicates profit opportunities. To test whether the presumed relationship between turnover and the abovementioned mispricing

[32] The R^2 of regressions (10) which includes the advisor specific average turnover as an explanatory variable, is below the R^2 of regression (4) with only the advisor specific average turnover as an explanatory variable. This result is due to the reduced date sample in regression (10) where funds without style box information are dropped. Applying regression (4) to the reduced sample would return an R^2 of 8.00.

indicators holds internationally, I regress the pooled sample of fund turnovers on market sentiment, dispersion and liquidity indicators. Each month, I use country-specific consumer confidence indices from the OECD as an indicator of sentiment[33] and calculate country-specific cross-sectional standard deviations of the return of domestic stocks within each country. I use the Pástor and Stambaugh (2003) liquidity factor for all international stock markets. Among other, Chordia et al. (2000), Hasbrouck and Seppi (2001), and Huberman and Halka (2001) and Brockman et al. (2009) have shown commonalities in market liquidity across countries and it thus seems rational to use a well-established US liquidity measure as an estimator of international liquidity. Following PST, I also control for the current business cycle as measure by the country-specific composite leading indicator and the contemporaneous market return. I include fund fixed effects and a time trend since turnover tends to decrease over time. Table 4.9 presents the results. Surprisingly, I find a negative relationship between sentiment and fund turnover for both domestic US funds and non-US funds. This result is contrary to the finding of PST and can only partially be explained by the use of the consumer confidence index instead of the Baker and Wurgler (2006) sentiment factor. Using the latter yields a relationship between sentiment and turnover that is close to zero (unreported results). In line with PST, turnover is positively related to return dispersion and negatively related to liquidity for both domestic US funds and non-US funds. Notably, I can also confirm the finding of higher fund turnover during recessions and in times of low market returns. The overall low explanatory power of the model is in line with the results of PST. Altogether, my analysis supports the idea that funds trade more when market opportunities are prevalent, that is in times of return uncertainty (high dispersion) and low liquid markets. I do, however, not find a positive relationship between consumer sentiment and fund turnover.

The positive time-series relationship between fund turnover and subsequent performance derived by the PST model is based on the assumptions that fund managers mainly trade on mispricing. If fund managers did not trade in response to mispricing but executed trades randomly, I would not expect any turnover-performance relationship or, because turnover autocorrelation is positive and trading is associated with costs, even a negative relationship. I therefore split the sample into funds with a positive and funds with a negative turnover-

[33] Among others, Lemmon and Portniaguina (2006), Schmeling (2009), and Stambaugh et al. (2012) use the consumer confidence as an estimator of market sentiment. As I match the country-specific consumer sentiment indices to funds' investment areas, I lose observations with multinational investment regions, e.g. Asia Pacific, Europe, Global.

performance relationship. I determine this relationship by regressing benchmark-adjusted returns on lagged turnover values separately for each fund. The sign of the regression coefficient determines the relation between turnover and performance for the respective fund. For each of the two subsamples I then regress turnover on the abovementioned mispricing indicators. As reported in Table 4.9, I find that low liquidity and high dispersion are closer connected to turnover for funds with a positive turnover-performance relationship than for funds with a negative turnover-performance relationship. The surprisingly negative relationship between sentiment and turnover is more pronounced for funds that do not share the positive turnover-performance relationship either. This observation gives reason to the idea that funds that do not primarily trade on mispricing do not have the positive turnover-performance relationship either. I will further investigate this explanation for the heterogeneity of my main results in Section 4.6.

Another result of PST is that not only a fund's own turnover but even the turnover of other funds, and even more of funds with similar investment patterns and characteristics, predict fund performance. To test this result for my international sample, I regress fund performance, measured as before by the benchmark-adjusted return, on different versions of average turnover. Average turnover is calculated across all funds in my sample, all funds with the same investment area, funds with the same management country, the same fund advisor firm, across funds with similar fund characteristics[34] as well as only across US managed and non-US managed funds. A fund's own turnover is excluded from the calculation of the average turnover values the fund's performance is regressed on. Table 4.10 reports the results of multiple regressions where panel A results are based on the full sample and panel B results are based on a subsample where US managed funds are excluded. Regressions (1) – (5) indicate that any single average turnover value positively predicts future fund performance. A comparison of the within-fund R^2 values reveals that the average turnover of all funds and the average turnover of funds with similar characteristics can best explain a fund's performance variation. Adding the fund's own lagged turnover to the model (columns 6 and 7) and using all three explanatory variables in one regression (column 8) further improves the model's explanatory power, although the within-R^2 remains very low. These results hold within the complete sample as well as the subsample that excludes US managed funds.

---

[34] That is funds with the same Morningstar style box, from the same fund size and fund expense ratio tercile.

Table 4.9: Fund turnover and indicators of market mispricing

The table reports the relationship between fund turnover and measures of market mispricing. It reports the coefficients of multivariate panel regressions of fund turnover on the consumer confidence index of the country a fund invests in ($Sentiment_{i,t}$), the cross-sectional return volatility ($Dispersion_{i,t}$) in this equity market, the Pástor and Stambaugh (2003) liquidity factor ($Liquidity_t$), the OECD composite leading indicator ($Business\ Cycle_{i,t}$), and the return of a leading market index ($Market\ Return_{i,t}$) in the fund's investment area. Additionally, a linear time trend and fund fixed effects are added to the regression. The sample is split into US managed and non-US managed funds, where the management country is determined by the address and phone number of the fund's management company as described in Section 4.2, and funds with a positive and negative turnover-performance relationship. The within fund R^2 as well as its difference to a within fund R^2 from a regression of turnover on only the trend component are reported. Standard errors are clustered on a fund and year level. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	US- versus Non-US funds		Single funds' turnover-performance relationship	
	US funds	non-US funds	positive	negative
$Sentiment_{i,t}$	−0.154** (−2.38)	−0.371*** (−4.72)	−0.127 (−1.70)	−0.343*** (−3.98)
$Dispersion_{i,t}$	71.738*** (2.87)	48.781*** (3.14)	66.172*** (3.71)	44.364** (2.50)
$Liquidity_t$	−0.020 (−0.01)	−4.337 (−0.49)	−2.651 (−0.54)	−1.487 (−0.30)
$Business\ Cycle_{i,t}$	−1.266** (−2.06)	−0.017 (−0.03)	−1.547*** (−3.06)	−0.005 (−0.01)
$Market\ Return_{i,t}$	−5.635 (−1.37)	−14.022 (−0.98)	−14.262*** (−2.87)	−4.068 (−0.73)
Time trend	−0.150*** (−14.05)	−0.181*** (−6.36)	−0.186*** (−14.51)	−0.109*** (−7.09)
Within R^2 (in %)	4.23	1.61	4.72	1.41
R^2-R^2 (trend only)	0.72	0.62	0.80	0.55
Observations	439,191	211,992	380,431	277,694

The average turnover of US managed funds has a particularly strong predictive power for fund performance as columns 9 and 10 reveal. Regressing benchmark-adjusted returns on the average turnover of US managed and non-US managed funds yields a positive coefficient for the former but a coefficient close to 0 for the latter. Surprisingly, this result holds even when excluding US managed funds from the sample (panel B). It remains puzzling why the turnover of US managed funds predict the performance not only of US managed funds but also of funds managed outside the US and why this predictive power is even stronger than the average turnover of non-US managed funds.

Table 4.10: Relation between fund performance and average turnover

This table reports the estimated slope coefficients from panel regressions of funds' monthly excess return on measures of average lagged turnover. Excess return is a fund's gross of fee return minus the return of the Morningstar Index which MD assigns as a benchmark to the fund's Morningstar Category. Lagged turnover is the turnover of a fund during the last calendar year that ended prior to the respective month. The average turnover is calculated across all funds ($AvgTurn_{all}$), funds with the same management country ($AvgTurn_{man}$), the same investment area ($AvgTurn_{inv}$), the same investment advisor ($AvgTurn_{adv}$), or across funds from the same size and fee tercile that are also assigned to the same equity style box ($AvgTurn_{similar}$). Furthermore, it is calculated across all US managed ($AvgTurn_{US}$) and non-US managed funds ($AvgTurn_{Non\text{-}US}$). Panel A reports the results for the entire sample, Panel B for funds that are managed outside the US. All data are from the 1992−2018 time period. T-statistics are reported in parentheses. Standard errors are heteroscedasticity-robust and clustered by Morningstar style category times month. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

Panel A: All funds

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AvgTurn_{all}$	0.00467					0.00446		0.00295		
	(5.59)					(5.35)		(2.83)		
$AvgTurn_{man}$		0.00263								
		(4.00)								
$AvgTurn_{inv}$			0.00249							
			(3.95)							
$AvgTurn_{adv}$				0.00660						
				(4.05)						
$AvgTurn_{similar}$					0.00191		0.00177	0.00115		
					(5.45)		(5.24)	(3.39)		
$AvgTurn_{US}$									0.00638	0.00616
									(3.86)	(3.73)
$AvgTurn_{Non\text{-}US}$									−0.00016	−0.00015
									(−0.25)	(−0.24)
Fund Turnover						0.00027	0.00037	0.00031		0.00023
						(4.23)	(5.23)	(4.44)		(3.67)
Within R^2 (in %)	0.07	0.04	0.04	0.01	0.05	0.07	0.06	0.08	0.10	0.10
Observations	1,129,597	1,129,016	1,128,534	1,055,833	855,935	1,129,597	855,935	855,935	1,129,597	1,129,597

(continued)

(continued)

Panel B: Non US-funds										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AvgTurn_{all}$	0.00348 (3.75)					0.00334 (3.62)		0.00034 (0.27)		
$AvgTurn_{man}$		0.00084 (1.25)								
$AvgTurn_{inv}$			0.00118 (1.54)							
$AvgTurn_{adv}$				0.00350 (1.83)						
$AvgTurn_{similar}$					0.00086 (1.43)		0.00077 (1.31)	0.00071 (1.08)		
$AvgTurn_{USA}$									0.00353 (1.67)	0.00344 (1.62)
$AvgTurn_{Non\text{-}USA}$									0.00061 (0.41)	0.00055 (0.38)
FundTurn						0.00016 (2.04)	0.00021 (2.33)	0.00021 (2.28)		0.00015 (2.01)
Within R^2 (in %)	0.04	0.01	0.01	0.00	0.01	0.04	0.02	0.05	0.04	0.04
Observations	520,714	520,133	519,966	489,410	290,717	520,714	290,717	290,717	520,714	520,714

Altogether, the average fund turnover predicts future fund performance. The average turnover of funds that share the same investment area or fund management country does not do a better job in predicting fund performance than the average turnover across all funds. The average turnover of similar funds in terms of fund characteristics, however, adds additional performance predictability. Surprisingly and puzzling, the turnover of US managed funds predicts future fund performance around the globe better than country-specific turnover measures.

4.6 Explaining international differences

The Section 4.3 analyses have shown that there exists a positive turnover-performance relationship even outside the US and within most investment areas and fund management countries. Nevertheless, the results presented in Table 4.5 indicate significant differences in this relationship between investment areas and between fund management countries. Whereas the regression slope for funds investing in the US is 0.00074, it is −0.00028 for funds investing in China. Funds that are managed in Norway have a regression coefficient of 0.00135, whereas funds managed in Denmark have an economically and statistically negative relationship of −0.00096. How can these international differences be explained? According to the model of PST, only the autocorrelation of turnover, trading costs and management skill should have an impact on the time-series turnover-performance relationship. Whereas I do find such an impact of skill, I cannot confirm the impact of trading costs and turnover autocorrelation. Instead, I establish a relationship between the extent to which fund managers trade on mispricing and the turnover-performance relationship and I find market momentum to be negatively correlated to the estimated time-series regression slopes.

4.6.1 Impact of turnover autocorrelation

PST argue that the time-series turnover-performance relationship is weaker among funds with a higher turnover autocorrelation. The rationale behind it is as follows. Turnover is associated with contemporaneous trading costs but future profits. Thus, a high lagged turnover will predict a high future performance. If, however, turnover is strongly correlated over time, a high lagged turnover also predicts a high present turnover and therefore high trading costs. These trading costs reduce the profit that results from past turnover. As a result, the positive relationship between lagged turnover and subsequent performance gets diluted. To test this prediction of PST, I calculate the average turnover autocorrelation within each country. Table 4.11 displays, among others, the country-specific autocorrelation of fund turnover, either clustered by the funds' investment area (panel A) or their management country (panel B). A first impression does not support the prediction of PST. Average turnover autocorrelation is 0.37 for funds that invest in the US and only 0.28 for funds investing in China. Norwegian funds have an average turnover autocorrelation of 0.31 whereas it is only 0.18 for funds managed in Denmark. These differences are going into the opposite direction of what the PST model predicts. In a more structured approach, I calculate the correlation between the country-specific time-series turnover-performance relationship and the average turnover autocorrelation within the respective country. Table 4.12 reports the Pearson correlation coefficient and the Spearman rank correlation for all investment areas (panel A) or management countries (panel B) with at least 1000 fund-month observations. It also reports the results of a regression of the country-specific turnover-performance slopes on the average turnover autocorrelation. Correlation coefficients and regressions are calculated in an unweighted version and a weighted version where weights are given by the natural logarithm of the number of observations. In contrast to the prediction of the PST model, the relationship between turnover autocorrelation and the turnover-performance regression slope is positive but statistically not significantly different from 0 in all specifications but for the weighted Pearson correlation when funds are sorted by their investment area. I thus conclude that the autocorrelation of fund turnover cannot explain the international differences in the turnover-performance relationship.

4.6.2 Impact of trading costs

According to the PST model, the turnover-performance relationship should be stronger among funds with higher trading costs. Other than profits, trading costs occur contemporaneously and should thus not impact the subsequent fund performance. If trading costs are high, however, a fund manager will only trade on profit opportunities if those are large enough to cover the trading costs. Therefore, the relationship between turnover and future performance should be more pronounced in markets with higher trading costs. Among others, Domowitz et al. (2001) and Bollerslev et al. (2018) analyze international trading costs. The former report trading costs to be particularly high in emerging markets such as India, Malaysia and Thailand, the latter report low trading costs in the US, UK, Switzerland and France. The correlation between the trading costs reported by Domowitz et al. (2001) or Bollerslev et al. (2018)

Table 4.11: Fund turnover, performance, fees and size by countries and regions

This table reports the average fund turnover, its standard deviation and autocorrelation, the average benchmark-adjusted return, expense ratio, and median fund size for subsets of funds sorted by the funds' investment areas (Panel A) or the country the funds are managed from (Panel B). A fund's investment area is determined by its Global Category, its investment area according to MD, the Morningstar Category, Morningstar Index, and its MPT index. The fund management country is determined by the management company's address and phone number in MD. Within each panel, data on the 20 largest subsets are displayed. All variables are downloaded from MD.

	Annual turnover			Monthly excess return	Expense ratio	Size
	Mean (in %)	Standard deviation	Auto-correlation	Mean (in %)	Mean (in %)	Median (mn. USD)
Panel A: Descriptive Statistics by Investment Area						
Asia Pacific	91.88	95.64	0.23	0.05	1.74	90.2
Canada	55.49	65.71	0.27	0.00	1.29	211.7
China	150.47	152.17	0.28	−0.07	1.83	103.4
Emerging Markets	75.06	81.10	0.23	−0.01	1.58	212.8
Europe	96.53	117.08	0.18	0.05	1.71	120.0
Finland	94.35	119.96	0.27	0.27	1.74	123.6
France	70.89	99.11	−0.11	0.10	2.22	95.9
Global	73.47	81.60	0.24	−0.01	1.39	172.8
India	126.59	157.56	0.29	0.25	2.02	113.3
Japan	92.27	99.81	0.19	0.14	1.58	101.6
Latin America	73.01	98.07	0.07	0.04	1.84	105.0
Malaysia	90.66	71.59	0.35	0.06	1.67	45.7
Norway	57.27	55.57	0.30	0.16	1.45	118.5
Russia	58.93	78.37	0.15	0.16	2.35	121.6
Sweden	78.72	89.18	0.21	0.21	1.33	254.5
Switzerland	52.45	94.19	0.19	0.13	1.33	109.3
Taiwan	258.70	163.08	0.39	0.12	1.73	41.3
Thailand	201.58	175.59	0.41	0.04	1.83	56.8
UK	91.49	93.19	0.06	0.09	1.42	180.8
USA	81.26	80.93	0.37	0.05	1.19	263.3

(continued)

(continued)

	Annual turnover			Monthly excess return	Expense ratio	Size
	Mean (in %)	Standard deviation	Auto-correlation	Mean (in %)	Mean (in %)	Median (mn. USD)
Panel B: Descriptive Statistics by management country						
Austria	133.24	134.35	0.02	−0.11	1.98	49.5
Canada	58.66	65.36	0.22	−0.01	1.61	144.0
China	269.73	170.80	0.27	−0.27	1.90	136.8
Denmark	49.17	41.96	0.18	0.04	1.52	84.4
Finland	72.09	96.45	0.18	0.10	1.85	89.0
France	87.63	122.08	−0.07	−0.01	1.98	100.7
India	152.98	169.12	0.35	0.29	2.49	99.6
Ireland	108.92	118.87	0.24	0.03	1.47	152.4
Italy	147.87	159.03	0.33	−0.03	2.39	106.8
Luxembourg	93.50	115.04	0.16	0.02	1.72	154.8
Malaysia	90.16	68.57	0.31	−0.03	1.62	43.7
Netherlands	81.22	85.20	0.15	0.01	1.53	127.7
Norway	63.68	62.55	0.31	0.12	1.55	115.9
Singapore	66.02	58.31	0.14	−0.02	1.73	56.7
Sweden	84.46	89.30	0.17	0.07	1.47	205.9
Switzerland	95.59	124.75	0.19	−0.01	1.41	87.5
Taiwan	211.64	166.46	0.36	0.05	1.87	43.1
Thailand	193.30	162.93	0.38	0.03	1.75	47.5
UK	92.11	107.31	0.23	0.06	1.42	177.8
USA	77.90	74.53	0.36	0.06	1.21	295.6

and the turnover-performance relationship in my sample is −0.10 and −0.29. This result must be interpreted with caution due to the low number of countries in my sample for which Domowitz et al. (2001) or Bollerslev et al. (2018) report trading costs (13 respectively 8 countries). Nevertheless, the negative correlations do not support the prediction of the PST model that higher trading costs imply a stronger turnover-performance relationship.

4.6.3 Impact of fund management skill

According to the PST model, the third determinant of the turnover-performance relationship is management skill. Skilled managers will buy and sell stocks whenever there are profit opportunities. Unskilled managers are not able to select the right stocks and thus their trading activity does not impose future profits. As a result, the turnover-performance relationship should be stronger in countries with highly skilled fund managers. Skill is a fund manager characteristic and should therefore rather be determined by the country a fund is managed from than by the fund's investment area. It appears more likely that highly-developed fund management markets and asset management hubs such as the US or Switzerland can attract more skilled managers than that the most skilled fund managers cluster in funds with common investment areas. I would therefore rather expect a positive correlation between skill and the turnover-performance relationship when comparing funds from different fund management countries. I use five different indicators of management skill to test this hypothesis: the funds' total expense ratio, average fund size[35], funds' average benchmark-adjusted return, their average Berk and van Binsbergen (2015) skill measure and the IMF Financial Development Index introduced by Svirydzenka (2016). Whereas the former measures might be endogenous because a stronger turnover-performance relationship might lead to a better performance and therefore might attract flows and allow fund managers to charge higher fees, the Financial Development Index measures how developed financial institutions and financial markets of a country are in terms of their depth, access, and efficiency, independent from fund performance characteristics. Table 4.12 presents the Pearson and Spearman rank correlation coefficients between the turnover-performance relationship, measured by the country-specific regression slope, and the different skill measures. Funds are clustered by their investment areas (panel A) or their management country (panel B) and only countries with at least 1,000 fund-month observations are included. The table also shows the results of a regression of country-specific turnover-performance slopes on the average skill measures. I calculate unweighted and

[35] Berk and Green (2004) argue that skilled fund managers earn higher fees and attract more money, and as a result manage on average larger funds.

Table 4.12: Country-specific turnover-performance relationships and national characteristics

The table reports the relationship between country-specific turnover-performance regression slopes and characteristics of the respective country's funds and the domestic equity market. Funds are sorted into subsamples by either their investment area (Panel A) or their fund management country (Panel B). For all subsamples with more than 1,000 fund-month observations I calculate the time-series turnover-performance regression slopes from a panel regressions of funds' monthly excess return on lagged fund turnover including fund fixed effects as describes in Section 4.3. I then relate those slopes to country-specific characteristics, which are the average turnover autocorrelation of all funds from that country, the average fee, size, excess return and Berk and van Binsbergen (2015) skill measure of those funds as well as the average Financial Development Index (FDI) of this country and the average momentum effect in the country's domestic equity market. I regress the cross-section of monthly equity returns on the prior 12-month equity returns to determine a country's return momentum. The country-specific factors are averaged across time and I thereby weight monthly data by the number of monthly fund observations for the respective country. I calculate Pearson's correlation coefficient and Spearman's rank correlation coefficient as well as a linear regression slope to relate the turnover-performance regression slopes to any one country-specific characteristic. For each method I calculate an unweighted and weighted version, where a country's weight is the natural logarithm of fund-month observations for the respective country. All data are from the 1992−2018 time period. T-statistics are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

Panel A: By Investment Area

		Correlations				Regression	
		Pearson	Pearson weighted	Spearman	Spearman weighted	Univariate	Univariate weighted
Turnover Autocorrelation		0.17	−0.17	0.07	0.09	0.0014 (0.80)	0.0014 (0.79)
Skill Measures	Fee	−0.10	−0.12	−0.20	−0.23	−0.00032 (−0.52)	−0.00037 (−0.62)
	Size	0.26	0.25	0.22	0.19	1.15e−12 (1.43)	9.00e−13 (1.32)
	Excess Return	0.01	−0.01	−0.01	−0.01	0.00007 (0.03)	−0.00006 (−0.02)
	Skill measure	0.21	0.24	−0.01	0.02	2.94e−12 (1.07)	3.19e−12 (1.31)
	FDI	0.20	0.22	0.41**	0.45**	0.0015 (0.98)	0.0015 (1.59)
Momentum		−0.53***	−0.54***	−0.19	−0.21	−0.0639*** (−2.85)	−0.0634*** (−2.93)

(continued)

(continued)

Panel B: By Management Country

		Correlations				Regression	
		Pearson	Pearson Weighted	Spearman	Spearman weighted	Univariate	Univariate weighted
Turnover Autocorrelation		0.22	0.24	0.17	0.19	0.0011 (1.12)	0.0012 (1.22)
Skill Measures	Fee	−0.19	−0.23	−0.17	−0.22	−0.0004 (−0.99)	−0.0005 (−1.20)
	Size	0.14	0.20	0.00	0.06	2.98e−13 (0.72)	3.84e−13 (1.02)
	Excess Return	0.23	0.24	0.20	0.22	0.0014 (1.18)	0.0015 (1.25)
	Skill measure	0.41**	0.44**	0.32	0.36*	3.35e−12** (2.25)	3.48e−12** (2.45)
	Financial Development	0.45**	0.48**	0.40**	0.43**	0.0028** (2.44)	0.0029** (2.63)
Momentum		−0.05	−0.09	0.00	−0.04	−0.0065 (−0.23)	−0.0112 (−0.41)

weighted versions of correlations and regression slopes with weights being given by the logarithm of fund-month observations. As expected, the results are weak when clustering funds by their investment area. The only statistically significant correlation is measured for the Financial Development Index when using the Spearman rank correlation. The results become more striking when comparing funds from different fund management countries. The correlations between the turnover-performance relationship and fund size as well as between the turnover-performance relationship and the average benchmark-adjusted return is positive, which supports the predicted impact of skill. Other than expected, funds managed in countries with higher fees have a lower turnover-performance relationship than funds managed in countries with lower fees.[36] None of these relationships, however, is statistically significant. Comparing the two most advanced measures of management skill, the Berk and van Binsbergen (2015) skill measure and the IMF Financial Development Index, across fund management countries, I find a strong and statistically significant positive connection between skill and the

[36] Because the descriptive statistics in Table 4.11 indicate that fees are higher in financially less developed countries, fees might not be a good indicator of skill.

turnover-performance relationship that holds in (almost) all specifications.[37] The correlations between the Berk and van Binsbergen (2015) skill measure respectively the IMF Financial Development Index and the turnover-performance relationship are 0.41 and 0.45. These results provide strong evidence that skilled fund managers are better at spotting mispricing and thus funds managed in financially highly developed countries have a stronger turnover-performance relationship.

4.6.4 Impact of market momentum

One basic assumption behind the PST model is that an optimal investment strategy requires turnover to adjust the portfolio to realize return opportunities. In different stock markets, an optimal portfolio strategy might require different levels of turnover. In a market with a strong cross-sectional momentum effect, past winners will tend to also outperform in the future. Therefore, an investor who held an optimal portfolio in the past does not have to trade many of her positions to also hold an optimal portfolio in the future. If she yet trades excessively, trading is most likely not connected to future performance. If, however the momentum effect is low or even dominated by a reversal, then an investor would have to adjust larger parts of her portfolio even if she held an optimal portfolio before. As a result, turnover would be the key to future overperformance. Therefore, I would expect a stronger turnover-performance relationship in markets with a low momentum than in markets with a high momentum.

To measure the momentum effect within an equity market during a certain month, I regress the cross-section of stock returns within the respective market during that month on the cross-section of lagged returns over the previous 12 months.[38] The regression coefficient indicates whether the 12-month past returns impact the one-month future returns in the cross-section. For each country, I calculate the average momentum effect as the weighted average of monthly momentum measures where the weights are determined by the monthly number of fund observations for the respective market. As before, I regress the cross-section of within-country turnover-performance slopes on the variable of interest, which is the average momentum effect. As expected, panel A of Table 4.12 reports a much stronger turnover-performance

[37] Results become statistically borderline insignificant for the Berk and van Binsbergen (2015) skill measure when using the unweighted version of the Spearman rank correlation.
[38] I use the 12-month momentum because the turnover-performance relationship also considers the turnover during the previous year.

relationship among funds that invest in a country with a lower return momentum. The correlation is −0.53 and statistically different from 0 at a 1%-level. The result remains qualitatively unchanged when using a weighted correlation and the rank correlation is negative, but statistically not significantly different from zero. The regression of the turnover-performance slope on the momentum factor yields a negative and statistically significant coefficient. Altogether, these results confirm the prediction of a stronger turnover-performance relationship in markets with a low return momentum. The results of panel B serve as a placebo test. There is hardly any correlation between the turnover-performance relationship and the momentum effect in the country the fund management is located.

4.6.5 Impact of trading behavior

Fund managers are expected to trade more when mispricing is more likely, that is in times of a high return uncertainty (high dispersion), a high market sentiment and low market liquidity. Section 4.5 results confirm this expectation for dispersion and liquidity but not for the market sentiment. As reported above, the comparing of funds with a positive and negative turnover-performance relationship reveals that funds with a positive turnover-performance relationship are more likely to trade when dispersion is high and liquidity low (Table 4.9). In the same vein, the unexpected results of negative correlation of sentiment and turnover is mainly driven by funds with a negative turnover-performance relationship. I therefore test whether different motivations for trading across funds might lead to a heterogeneity in the turnover-performance relationship. Other than skill or market momentum, trading behavior is unlikely to be common among funds with the same investment area or management country. I therefore split the fund sample as follows. For each fund, I regress the annual turnover on measures of return dispersion, market sentiment and liquidity as well as the business cycle and market return as described in Section 4.2 and keep the regression coefficients. For each of the determinants dispersion, sentiment and liquidity, I spilt the sample into funds with a positive and funds with a negative coefficient. I then redo the fund fixed effects turnover-performance analysis for each subsample. Table 4.13 reports the results as well as the differences of the regression slopes. Funds that trade more in times of a high market dispersion have a within-fund turnover-performance regression slope of 0.00087 and thus more than three times as much as funds that trade more in times of low dispersion. The turnover-performance relationship is 0.00098 for funds that trade more when sentiment is high and only 0.00020 for funds that trade more when sentiment is low. Comparing funds that trade

Table 4.13: Turnover determinants and the turnover-performance relationships

This table reports the time-series turnover-performance relationship within fund subsamples. In a first step, for each fund, turnover is regressed on the consumer confidence index of the country a fund invests in (*Sentiment*), the cross-sectional return volatility (*Dispersion*) in this equity market, the Pástor and Stambaugh (2003) liquidity factor (*Liquidity*), the OECD composite leading indicator and the return of a leading market index in the fund's investment area. Additionally, a linear time trend is added to this multivariate regressions. Funds are then sorted into subsamples by one of the regression coefficients with respect to the *Sentiment*, *Dispersion* and *Liquidity* variables: For each dimension funds with a positive coefficient are separated from funds with a negative coefficient. Within each of the 3x2 subsamples I then estimate the slope coefficients from a panel regression of the funds' monthly excess returns on their lagged fund turnover including fund fixed effects. Excess return is a fund's gross of fee return minus the return of the Morningstar Index which MD assigns as a benchmark to the fund's Morningstar Category. Lagged turnover is a fund's turnover during the last calendar year that ended prior to the respective month. Slope coefficients for all subsamples as well as the difference in slope coefficients between funds with a positive and a negative first step regression coefficient are reported. T-statistics are reported in parentheses. Standard errors are heteroscedasticity-robust and clustered by Morningstar style category times month. Statistical significance at the 10%, 5%, and 1% level is denoted by *, **, and ***.

	Sentiment	Dispersion	Liquidity
Positive	0.00098***	0.00087***	0.00036**
	(4.55)	(4.53)	(2.44)
Negative	0.00020	0.00024*	0.00077***
	(1.48)	(1.71)	(4.95)
Positive-Negative	0.00078***	0.00063***	−0.00041**
	(2.80)	(2.48)	(2.01)

more when market liquidity is low to funds that trade more when markets are liquid shows that the former have a more than double as strong turnover-performance relationship as the latter (0.00077 versus 0.00036). For all three comparisons the differences are statistically significant at least at a 5% level.

Together with the results of Section 4.5 these findings should be interpreted as follows. Funds are expected to trade more when mispricing is more likely, that is in times of low market liquidity, high return dispersion and high sentiment. International empirical evidence only partially supports this prediction. Funds that trade more during times of higher mispricing, however, have a stronger turnover-performance relationship. This is fully in line with the economic idea of the PST model. Among funds that trade more when mispricing is unlikely, a large part of the trading should not be connected to future performance and thus the turnover-performance relationship should be less prevalent among those funds.

To summarize, this section presents several channels that can explain the heterogeneity of the turnover-performance relationship across funds and countries. Skill, market momentum and the degree to which fund managers trade on mispricing positively impact the turnover-performance relation. Contrary to the prediction of the PST model, I do not find an impact of trading costs and the turnover autocorrelation on the turnover-performance relationship.

4.7 Conclusion

This paper provides international evidence on the model of PST. The model assumes that fund managers trade on market mispricing and that the purchase of underpriced stocks leads to future profits. From that, it establishes a positive within-fund time-series relationship between fund turnover and the subsequent performance. Besides this main results, PST derive a battery of predictions about a fund's turnover and its relationship to fund performance. The international evidence in this paper is based on more than 13,000 funds from 40 countries and serves as an out-of-sample test of the PST predictions. Most importantly, I confirm the main result of PST, that is the positive within-fund time-series relationship between fund turnover and subsequent fund performance, even outside the US. In contrast to PST but in line with prior literature, I do not find evidence for a cross-sectional turnover-performance relationship outside the US, neither in a pooled sample nor within a vast majority of single country fund subsamples. PST predict that the turnover-performance relationship is stronger for funds that encounter higher trading costs or charge higher fees. Comparing different fund styles and fund characteristics, I can only partially approve this result outside the US. I find a stronger relationship among small funds, which are assumed to encounter higher trading costs, but do not find a larger within-fund relationship for small cap funds (which usually also face higher trading costs) or funds with higher fees outside the US. Furthermore, I can approve PST's result of a higher fund turnover when mispricing is more prevalent, that is in times of return uncertainty (high dispersion) and low liquid markets. My results, however, do not indicate a positive relationship between consumer sentiment and fund turnover. In line with PST, I find strong comovements in turnover across funds and I find the average fund turnover to predict future fund performance. This predictive power gets stronger when using the average turnover of similar funds. Surprisingly, the turnover of US managed funds is a better predictor of future fund performance around the globe than country-specific turnover averages.

Besides a pure out-of-sample test of the PST model and PST's results, the international structure of both the fund management country and the funds' investment area allows to investigate how international differences relate to differences in the turnover-performance relationship. In line with the prediction of PST, I find the turnover-performance relationship to be stronger in countries where funds have a higher Berk and van Binsbergen (2015) skill measure and countries with a higher Financial Development Index, which are thus more likely to attract skilled fund managers. I furthermore find the degree to which fund managers trade on mispricing to be positively correlated to the turnover-performance relationship and discover market momentum to be negatively related to this relationship. Both results fit the economic rationale behind the PST model. These results do not only support the validity of the PST model but they help to understand why fund managers trade and under which circumstances this trading leads to a future fund outperformance. Notably, my analyses do not approve all the predictions of PST and I particularly do not find a correlation between the turnover-performance relationship and turnover autocorrelation or trading costs in my international sample. Altogether, these results strongly support the general idea of the PST model but they raise doubt about some of the extensive predictions of the model, in particular the role of trading costs and return autocorrelation. The model might require further refinements to fully capture the drivers and return effects of fund turnover.

References

Amihud, Y., Goyenko, R., 2013. Mutual fund R2 as predictor of performance. Review of Financial Studies, 26, 667–694.

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

Baker, M., Litov, L., Wachter, J. A., Wurgler, J., 2010. Can mutual fund managers pick stocks? Evidence from their trades prior to earnings announcements, Journal of Financial and Quantitative Analysis, 45, 1111–1131.

Berk, J. B., Green, R. C., 2004. Mutual fund flows and performance in rational markets. Journal of Political Economy, 112, 1269−1295.

Berk, J. B., Van Binsbergen, J. H., 2015. Measuring skill in the mutual fund industry. Journal of Financial Economics, 118, 1−20.

Bollen, N. P., Busse, J. A., 2001. On the timing ability of mutual fund managers. The Journal of Finance, 56, 1075–1094.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L. H., 2018. Risk everywhere: Modeling and managing volatility. The Review of Financial Studies, 31, 2729−2773.

Brockman, P., Chung, D. Y., Pérignon, C., 2009. Commonality in liquidity: A global perspective. Journal of Financial and Quantitative Analysis, 44, 851−882.

Brown, N. C., Wei, K. D., Wermers, R., 2013. Analyst recommendations, mutual fund herding, and overreaction in stock prices. Management Science, 60, 1−20.

Busse, J. A., Goyal, A., Wahal, S., 2013. Investing in a global world. Review of Finance, 18, 561−590.

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

Chen, J., Hong, H., Huang, M., Kubik, J., 2004. Does fund size erode mutual fund performance? American Economic Review, 94, 1276–1302.

Chen, H.-L., Jegadeesh, N., Wermers, R., 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, 343–368.

Chordia, T., Roll, R., Subrahmanyam, A., 2000. Commonality in liquidity. Journal of Financial Economics, 56, 3−28.

Cohen, R., Coval, J., Pástor, Ľ., 2005. Judging fund managers by the company they keep. Journal of Finance, 60, 1057–1096.

Cremers, M., Ferreira, M. A., Matos, P., Starks, L., 2016. Indexing and active fund management: International evidence. Journal of Financial Economics, 120, 539−560.

Cremers, M., Pareek, A., 2016, Patient capital outperformance: The investment skill of high active share managers who trade infrequently. Journal of Financial Economics, 122, 288−306.

Cremers, M., Petajisto, A., 2009. How active is your fund manager? A new measure that predicts performance. The Review of Financial Studies, 22, 3329−3365.

Cumby, R. E., Glen, J. D., 1990. Evaluating the performance of international mutual funds. The Journal of finance, 45, 497−521.

Dahlquist, M., Engström, S., Söderlind, P., 2000. Performance and characteristics of Swedish mutual funds. Journal of Financial and Quantitative Analysis, 35, 409−423.

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

Dasgupta, A., Prat, A., Verardo, M., 2011. The price impact of institutional herding. The Review of Financial Studies, 24, 892−925.

Domowitz, I., Glen, J., Madhavan, A., 2001. Liquidity, volatility and equity trading costs across countries and over time. International Finance, 4, 221−255.

Edelen, R. M., Evans, R., Kadlec, G. B., 2007. Scale effects in mutual fund performance: The role of trading costs. Working paper.

Edelen, R. M., Evans, R., Kadlec, G. B., 2013. Shedding light on “invisible” costs: Trading costs and mutual fund performance. Financial Analysts Journal, 69, 33–44.

Elton, E. J., Gruber, M. J., Blake, C. R., 1996. Survivorship bias and mutual fund performance. Review of Financial Studies, 9, 1097−1120.

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

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

Ferreira, M. A., Keswani, A., Miguel, A. F., Ramos, S. B., 2012. The flow-performance relationship around the world. Journal of Banking and Finance, 36, 1759−1780.

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

Griffin, J. M., Kelly, P. J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. The Review of Financial Studies, 23, 3225−3277.

Grinblatt, M., Titman, S., 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. Journal of Business, 62, 393–416.

Grinblatt, M., Titman, S., 1993. Performance measurement without benchmarks: An examination of mutual fund returns. Journal of Business, 66, 47–68.

Grinblatt, M., Titman, S., Wermers, R., 1995. Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. American Economic Review, 85, 1088−1105.

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

Hasbrouck, J., Seppi, D. J., 2001. Common factors in prices, order flows, and liquidity. Journal of Financial Economics, 59, 383−411.

Huang, J., Sialm, C., Zhang, H., 2011. Risk shifting and mutual fund performance. The Review of Financial Studies, 24, 2575−2616.

Huberman, G., Halka, D., 2001. Systematic liquidity. Journal of Financial Research, 24, 161−178.

Investment Company Institute, 2018. Investment Company Fact Book.

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

Jiang, G. J., Yao, T., Yu, T., 2007. Do mutual funds time the market? Evidence from portfolio holdings. Journal of Financial Economics, 86, 724–758.

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

Kacperczyk, M., Van Nieuwerburgh, S., Veldkamp, L., 2014. Time-varying fund manager skill. Journal of Finance, 69, 1455–1484.

Kosowski, R., 2011. Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions. The Quarterly Journal of Finance, 1, 607−664.

Lemmon, M., Portniaguina, E., 2006. Consumer confidence and asset prices: Some empirical evidence. The Review of Financial Studies, 19, 1499−1529.

Malkiel B. G., 1995. Returns from investing in equity mutual funds 1971 to 1991. Journal of Finance, 50, 77–91.

Mamaysky, H., Spiegel, M., Zhang, H., 2008. Estimating the dynamics of mutual fund alphas and betas. Review of Financial Studies, 21, 233–264.

McLean, R. D., Pontiff, J., 2016. Does academic research destroy stock return predictability? The Journal of Finance, 71, 5−32.

Nofsinger, J. R., Sias, R. W., 1999. Herding and feedback trading by institutional and individual investors. The Journal of Finance, 54, 2263−2295.

Pástor, Ľ., Stambaugh, R. F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy, 111, 642−685.

Pástor, Ľ., Stambaugh, R. F., 2012. On the size of the active management industry. Journal of Political Economy, 120, 740−781.

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

Pástor, Ľ, Stambaugh, R. F., Taylor, L. A., 2017. Do funds make more when they trade more? The Journal of Finance, 72, 1483−1528.

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

Schmeling, M., 2009. Investor sentiment and stock returns: Some international evidence. Journal of Empirical Finance, 16, 394−408.

Sias, R. W., 2004. Institutional herding. The Review of Financial Studies, 17, 165−206.

Sias, R. W., Starks, L. T., Titman, S., 2001. The price impact of institutional trading, Working paper.

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics, 104, 288−302.

Svirydzenka, K., 2016. Introducing a new broad-based index of financial development. International Monetary Fund.

Walter, A., Weber, M. F., 2006. Herding in the German mutual fund industry. European Financial Management, 12, 375−406.

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

Wermers, R., 2000. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. The Journal of Finance, 55, 1655–95.

Yan, X., 2008. Liquidity, investment style, and the relation between fund size and fund performance. Journal of Financial and Quantitative Analysis, 43, 741–768.

Curriculum Vitae

Sebastian Fischer, born on June 28, 1987, in Stuttgart (Germany)

Education

05/2015 – 01/2019: Ph.D. in Finance, University of St.Gallen (Switzerland)

10/2010 – 11/2012: University of Ulm (Germany)
Business and Economics (M.Sc.), Major in Finance

07/2011 – 12/2011: National University of Singapore
Exchange Semester

10/2008 – 12/2010: University of Ulm (Germany)
Business and Economics (B.Sc.), Major in Accounting & Audit

09/1987 – 06/2007: Herzog-Christoph-Gymnasium, Beilstein (Germany)

Professional Experience

Since 02/2019: Zurich Insurance Group (Switzerland)

05/2015 – 01/2019: University of St.Gallen (Switzerland)

07/2012 – 04/2015: Kampmann, Berg & Partner (Germany)

04/2010 – 05/2010: Ernst & Young GmbH WPG (Germany)