School of Finance

University of St.Gallen

THE IMPACT OF THE MORNINGSTAR SUSTAINABILITY RATING ON MUTUAL FUND FLOWS

MANUEL AMMANN
CHRISTOPHER BAUER
SEBASTIAN FISCHER
PHILIPP MÜLLER

The Impact of the Morningstar Sustainability Rating on Mutual Fund Flows

Manuel Ammann[a], Christopher Bauer[b], Sebastian Fischer[c],
Philipp Müller[d]

University of St.Gallen

First Version: May 5, 2017

This Version: November 17, 2017

ABSTRACT

We examine the effect of the introduction of Morningstar's Sustainability Rating in March 2016 on U.S. equity mutual fund flows. Exploiting this shock to the availability of sustainability information we establish a causal relation between sustainability and fund flows. Using panel regressions, propensity score matching, and an event study methodology we find strong and robust evidence that retail investors shift money away from low-rated and into high-rated funds. An average high-rated retail fund receives between $4.1m and $10.1m higher net flows and an average low-rated retail fund suffers from $1.0m to $5.0m lower net flows than an average-rated fund during the first year after the publication of the Rating. The effect is caused by the publication of the Morningstar Sustainability Rating and not by a general attractiveness of sustainable funds. Institutional investors react much more weakly to the publication of the Rating.

JEL Classification: G11, G14, G23
Keywords: Mutual Fund, Sustainability, Investment Decisions, Information

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 attention has been paid to the increasing attention of investors to sustainable investments and its impact on fund flows. As assets under professional management utilizing sustainable investment criteria grew by 33 percent from 2014 to 2016, more than one out of 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 percent of investors have expressed their interest in sustainable investments.[5] Thus, a new variable is added to the decision-making process of an investor.

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. Bollen (2007) argues that investors have a multi-attribute utility function and therefore profit from owning socially responsible investments. His findings, especially a lower

---

[1]E.g. 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 to 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 higher marketing efforts.

[4]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 United States. Thereof $8.72 trillion have been invested according to sustainable investment strategies.

[5]The study can be found in the Morgan Stanley's 2017 edition of the Sustainable Signals series, 'New Data from the Individual Investor'.

[6]Filbeck et al. (2009) provide an overview over 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).

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 the 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. As a result, academic literature has documented the substantial impact of information intermediaries who provide free access to clearly displayed information.[8] Due to the prevalent dissent about the definition of sustainability and the fact that information on sustainability has been available on 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 judgement in order to align their investments to their preferences. Prior to March 2016 there was no such freely accessible and reliable information.

With the publication of its Sustainability Rating, Morningstar, one of the leading information providers in the mutual fund industry, has transferred sustainability from a difficult-to-grasp characteristic to 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 percent are assigned a Sustainability Rating of 5, whereas the bottom 10 percent of funds receive a rating of 1. An analysis of the relative google search interest, displayed in Figure 1, reveals great attention to the Morningstar

---

[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 for 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.

Sustainability Rating, not only upon its public launch but also during 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.

[Figure 1 about here.]

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 U.S. 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 a vast majority of mutual funds on February 29 (available for Morningstar Direct users only) and March 17, 2016 (publicly available without costs). 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 yet derive a sound estimate of the effect of the Morningstar Sustainability Rating on fund flows, we employ three 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

---

[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 over existing studies.

significant for retail share classes and does not appear prior to the launch of the Rating. A high-rated retail fund receives a 0.73 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 non-linear relationships between fund characteristics and fund flows (as argued in the case of performance by e.g. Chevalier and Ellison (1997), Sirri and Tufano (1998), Ivković and Weisbenner (2009)) we apply a propensity score matching. We match retail funds with a high as well as 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 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 to that, 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 percent during the first 6 months after the ranking was published. Low-rated funds suffer from an abnormal flow of 1.01 percent 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 $4.1m and $10.1m higher inflows p.a. than it would have expected in case of an average rating. A low-rated retail fund suffers from a $1.0m to $5.0m lower net flow p.a. compared to an average-rated fund. We find that the impact of Morningstar's

Sustainability Rating is much weaker for institutional share classes.

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 three 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 into 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. 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. 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 reminder is structured as follows: Section 2 introduces the Morningstar Sustainability Rating in greater detail and describes our data set. Section 3 contains our results for panel regressions (Subsection 3.1), the propensity score matching (Subsection 3.2) and the event study setting (Subsection 3.3). Section 4 concludes.

2. Data and Summary Statistics

2.1. Background on Morningstar's Sustainability Rating

The Morningstar Sustainability Rating indicates to what extent a fund holds securities whose issuers are successfully managing environmental, social and governance (ESG) 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 percent 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 percent highest/lowest portfolio sustainability scores receive a Sustainability Rating of high/low. The next top and bottom 22.5 percent are rated above and below average and the middle 35 percent 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.[10]

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.

---

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

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.[11] We extrapolate the Sustainability Ratings from November 2015 to October and September 2015, so that our sample comprises full six months prior to the Sustainability Rating publication.[12] 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 percent. Thus, extending our sample of pre-publication months by 50 percent outweighs the minor approximation error caused by extrapolation.

Our sample focuses on actively managed domestic U.S. 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.[13] We exclude share classes closed to investors and share classes with total net assets of below $1m. 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 performance, we are not double-counting observations using individual share classes. Table 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 Sustainabil-

---

[11]We thank Morningstar for providing a unique dataset on the historical Morningstar Sustainability Rating from November 2015 to March 2017.

[12]Our results remain unchanged repeating the analysis without extrapolating unpublished data.

[13]CRSP variables RETAIL_FUND and INST_FUND

ity 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 percent over the whole period. In the following, all analyses are based on share classes and we refer to them as funds.

[Table 1 about here.]

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.

2.3. Summary Statistics and Variable Definitions

By means of 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:

$$\text{FLOW}_{i,t} = \frac{\text{TNA}_{i,t} - \text{TNA}_{i,t-1} * (1 + \text{R}_{i,t})}{\text{TNA}_{i,t-1}}, \tag{1}$$

where $\text{TNA}_{i,t}$ are total net assets of fund i at the end of month t, and $\text{R}_{i,t}$ 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 percent and 99 percent levels. Additionally, we report data on various performance measures, 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

8

fund flows (see, for example Ippolito (1992), Gruber (1996), Chevalier and Ellison (1997), Goetzmann and Peles (1997), and Sirri and Tufano (1998)), but that the rating itself 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 (further referred to as "Performance Rating") as a widely respected medium to long term performance measure, supplemented by monthly raw returns and 12-months 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.[14]

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. 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 account for the size of a fund measured 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 2 provides summary statistics for these characteristics for all funds in our sample from September 2015 to February 2016, i.e. prior to the launch of the Sustainability Rating. In the six-months period, funds on average experienced outflows (-0.24 percent) and had negative average monthly returns (-0.84 percent), respectively annualized 12-months four-factor alphas (-1.51 percent). 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 percent), but insignificant. The least outflows can be observed for funds with a below average Sustainability Rating (-0.03 percent) and the highest outflows for funds

---

obtaining the highest Sustainability Rating (−0.61 percent). Regarding the control variables, funds in the highest rating category tend to have significantly superior performance measures, twice as much 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.

[Table 2 about here.]

Panels B and C of Table 2 describe retail and institutional funds separately from September 2015 to February 2016, i.e. prior to the publication of the Sustainability Rating. Institutional funds on average suffer from stronger outflows, have a higher Performance Rating and less negative 12-months four-factor alphas, are about 30 percent smaller, 5 years younger and have lower expense ratios than retail funds. Comparing relative net flows between the highest and the 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 least net outflows (−0.03 percent) compared to the highest rating category which observes the strongest net outflows (−0.68 percent).

Remarkably, after the launch of the Sustainability Rating the flow-sustainability relationship changed substantially, as shown by Panel D of Table 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 percent. Further, in line with our conjecture that different investor clienteles appreciate the new sustainability measure unequally, Panels E and F of Table 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 percent) is much more pronounced than for institutional funds (0.49 percent). Moreover, there is a sharp increase in flows (0.56 percent) moving from retail funds rated as above average to funds rated with the highest Sustainability Rating. Similarly, a markedly drop in flows comparing below-average-rated to low-rated retail funds (−0.40 percent) 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 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 12-months alpha than funds with the lowest Sustainability Rating. A similar result holds for the turnover and the expense ratio, both decreasing in the Sustainability Rating, revealing significant differences between funds of the top and bottom rating category. 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.

3. Results

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 non-published 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 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, i.e. we use fund characteristics measured at the end of the previous month. We additionally include month-style-fixed effects to account for a 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 due to the publication of the Rating. We therefore split our sample in two parts, one covering the months September 2015

to February 2016, i.e. the time prior to the public launch, and one covering March 2016 to March 2017, where 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 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 0.20 percentage points per month lower net flow than average-rated funds and high-rated funds even a 0.27 percentage points higher net flow than average-rated funds. Those differences are statistically significant and flows are monotonously increasing in the Sustainability Rating.

[Table 3 about here.]

Investors are expected to react only to the availability of the Sustainability Rating 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 U.S. investors have a much lower interest in sustainable investments.[15] 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 OLS-regressions for both. As reported in columns 3 and 4 of Table 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 0.31 percentage points per month (0.11 percentage points per month) lower net flow than an average-rated fund. A high-rated (above-average-rated) fund receives a 0.42 percentage points per month (0.10 percentage points per month) higher net flow. Given those significant results for retail funds only we will focus all further analysis on retail funds. Our results strongly suggest that retail investors gain information

---

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

from the publication of Morningstar's Sustainability Rating and adapt their investments to that information. Given the median size of a retail share class of $89.4m, an average fund receives 4.50m more inflows over one year if it obtains a high instead of an average Sustainability Rating. On the other hand it receives $3.34m 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 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 into higher-rated funds.

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 from 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, 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 end of February 2016 fund characteristics 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 vs. Average"), funds with a low versus a matched control sample of funds with an average Rating ("Low vs. Average") and funds with a high Rating versus a matched control sample of funds with a low Rating ("High vs. 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 vs. 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 vs. Average" and "High vs. Low" setting.

In a third step, we match each fund of the treatment group to the one 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 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 a high quality of our matching. In the "High vs. Average" setting we observe a significant difference only in the median fund age. For the "Low vs. 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 vs. 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 to our results. Since turnover ratios are not found to influence fund flows significantly, neither difference harms the quality of our matching.

[Table 4 about here.]

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 5 displays the results. During the three and the six-months period prior to the launch of the Morningstar Sustainability Rating, i.e. during December 2015 to February 2016 and September 2015 to February 2016, there is no significant difference in fund flows between our treatment and the matched control group. Mutual funds with a high Sustainability Rating get neither significantly higher nor significantly lower net flows than average 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 publications. During the months March to May 2016, high-rated funds receive a 3.01 percentage points higher net flow than average-rated funds and even 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.[16] The effect of higher net flows into high-rated funds and out of low-rated funds continues during the subsequent months and adds to a total of 9.41 percentage points during the first 12 months for high-rated versus average-rated funds, and even 12.04 percentage points for high 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 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.

[Table 5 about here.]

[Figure 2 about here.]

Since the median size of a retail share class in our matching with a high (low) Sustainability Rating is \$108m (\$65m), the effect of a 9.41 percentage points (6.80 percentage points) flow difference should be interpreted as follows: A high-rated fund according to Morningstar's Sustainability Rating receives on average \$10.1m higher net inflows during March 2016 to February 2017 than it would have expected if it had received an average Sustainability Rating. In contrast, an average low-rated fund receives about \$2.1m 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.

---

[16]We should not expect the effect of "High vs. Low" to equal the sum of the "High vs. Average" and "Low vs. Average" effects since the fund characteristics of the treatment group in the "Low vs. Average" setting differs significantly from the treatment groups of the other settings and since we are measuring an average effect on the treated.

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 change our results qualitatively. 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 vs. High"). Results confirm our finds from "High vs. 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.

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. The following section describes the details of how the event study approach is implemented and the diagnostic tests employed to assess the reliability of the benchmark model.

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[17], we compute the expected flows by estimating the subsequent time-series benchmark regression for each individual fund i:

$$F_t^i = \gamma^i + \beta_1^i SPCF_t^i + \beta_2^i F_{t-1}^i + \beta_3^i R_{t-1}^i + \beta_4^i \Delta Alpha1Y_{t-1}^i + \beta_5^i (\Delta Alpha1Y_{t-1}^i)^2 + e_t^i, \quad (2)$$

where F_t^i is the relative net flow to fund i at time t, $SPCF_t^i$ is the average relative net flow at

---

[17] 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 B1 in the appendix shows the model's capability of replicating the convex flow-performance relation. Table A1 in the appendix 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 a fund's initial rating grouped by the funds' respective Performance Rating.

time t to all funds in the same style and same lagged performance category as fund i and F_{t-1}^i is the lagged relative net flow to fund i at time $t-1$. R_{t-1}^i is the monthly net return of fund i observed in the preceding month. $\Delta \text{Alpha1Y}_{t-1}^i$ 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.

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 to 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[18], and were 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:

$$AF_t^i = F_t^i - \gamma^i - \beta_1^i SPCF_t^i - \beta_2^i F_{t-1}^i - \beta_3^i R_{t-1}^i - \beta_4^i \Delta Alpha1Y_{t-1}^i - \beta_5^i (\Delta Alpha1Y_{t-1}^i)^2. \quad (3)$$

Thereby, 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

---

[18]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.

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 i at time t within the event window are standardized by the square root of their estimated forecast variance (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 N funds for each Sustainability Rating group, the average standardized relative abnormal fund flow $\overline{SAF_t}$ 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).

3.3.2. Results

Table 6 and Table 7 contain the average standardized relative abnormal flow $\overline{SAF_t}$ and the average cumulative standardized relative abnormal flow $\overline{CSAF_t}$, respectively, 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 6 about here.]

Table 6 reports the results for the average standardized relative abnormal fund flow $\overline{SAF_t}$. For the highest sustainability category the $\overline{SAF_t}$ 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 of the event months. Standardized abnormal flows to the lowest sustainability category are insignificantly 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 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 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 differences-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 the $\overline{SAF_t}$ and $\overline{CSAF_t}$ provide strong evidence for a significant difference in the flow responses and therefore provide further indication for investors' reaction to the Sustainability Rating.

22

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 is also existent for the sustainability-flow relation. 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.

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 six months, i.e., the event window then spans the period of six months prior to the initially defined event month: $t = [-6; 0]$ and the estimation window captures 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 $\overline{SAF_t}$, which are reported in Table 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 insignificant 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 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.

23

Finally, to allow for an economic interpretation, we repeat the event study, but waive the standardization procedure. By doing so, we find that the positive abnormal flow response identified for funds in the highest rating category adds 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 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 Sustainability Rating[19]. Since the median size of a retail share class with a high (low) Sustainability Rating is \$111m (\$48m), the flow response con be interpreted as follows: A fund that is assigned the highest Sustainability Rating experiences on average \$4.05m p.a. higher net inflows than expected without the introduction of the Rating. In contrast, the launch of the Rating leads on average to a \$0.98m p.a. 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.

---

[19] 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 six months respectively 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 N funds for each Sustainability Rating category.

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 into 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 0.42 percentage points per month higher net flow than average-rated and a 0.73 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 high-rated retail retail funds to receive an abnormal inflow of 1.83 percent during the first 6 months after the initial rating publication whereas low-rated retail funds receive an abnormal flow of -1.01 percent during the same half-year period. For an average fund, these findings translate into \$4.1m to \$10.1m higher inflows for high-rated funds and \$1.0m to \$5.0m 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 smaller for institutional investors.

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, Alberto, and Guido W. Imbens, 2016, Matching on the Estimated Propensity Score, *Econometrica* 84, 781–807.

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

Benson, Karen L., and Jacquelyn E. Humphrey, 2008, Socially responsible investment funds: Investor reaction to current and past returns, *Journal of Banking & Finance* 32, 1850–1859.

Boehmer, Ekkehart, Jim Masumeci, and Annette B. Poulsen, 1991, Event-study methodology under conditions of event-induced variance, *Journal of Financial Economics* 30, 253 – 272.

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

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *Event-Study Analysis*, 149–180 (Princeton University Press).

Carhart, Mark M., 1997, On Persistence in Mutual Fund Performance, *The Journal of Finance* 52, 57–82.

Chevalier, Judith, and Glenn Ellison, 1997, Risk Taking by Mutual Funds as a Response to Incentives, *Journal of Political Economy* 105, 1167–1200.

Del Guercio, Diane, and Paula A. Tkac, 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, Diane, and Paula A. Tkac, 2008, Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow, *Journal of Financial and Quantitative Analysis* 43, 907–936.

Dodd, Peter, and Jerold B. Warner, 1983, On corporate governance: A study of proxy contests, *Journal of Financial Economics* 11, 401 – 438.

El Ghoul, Sadok, and Aymen Karoui, 2017, Does corporate social responsibility affect mutual fund performance and flows?, *Journal of Banking & Finance* 77, 53–63.

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

Filbeck, Greg, Raymond Gorman, and Xin Zhao, 2009, The "Best Corporate Citizens": Are They Good for Their Shareholders?, *Financial Review* 44, 239–262.

Goetzmann, William N., and Nadav Peles, 1997, Cognitive Dissonance and Mutual Fund Investors, *Journal of Financial Research* 20, 145–158.

Gruber, Martin J., 1996, Another Puzzle: The Growth in Actively Managed Mutual Funds, *The Journal of Finance* 51, 783–810.

Huang, Jennifer, Kelsey D. Wei, and Hong Yan, 2007, Participation Costs and the Sensitivity of Fund Flows to Past Performance, *The Journal of Finance* 62, 1273–1311.

Ippolito, Richard A., 1992, Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry, *The Journal of Law and Economics* 35, 45–70.

Ivković, Zoran, and Scott Weisbenner, 2009, Individual investor mutual fund flows, *Journal of Financial Economics* 92, 223 – 237.

Massa, Massimo, 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.

Patell, James M., 1976, Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Test, *Journal of Accounting Research* 14, 246–276.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2008, The price of ethics and stakeholder governance: The performance of socially responsible mutual funds, *Journal of Corporate Finance* 14, 302–322.

Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2011, Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds, *Journal of Financial Intermediation* 20, 562–588.

Sirri, Erik R., and Peter Tufano, 1998, Costly Search and Mutual Fund Flows, *The Journal of Finance* 53, 1589–1622.

Statman, Meir, 2008, Quiet conversations: The expressive nature of socially responsible investors., *Journal of Financial Planning* 21.
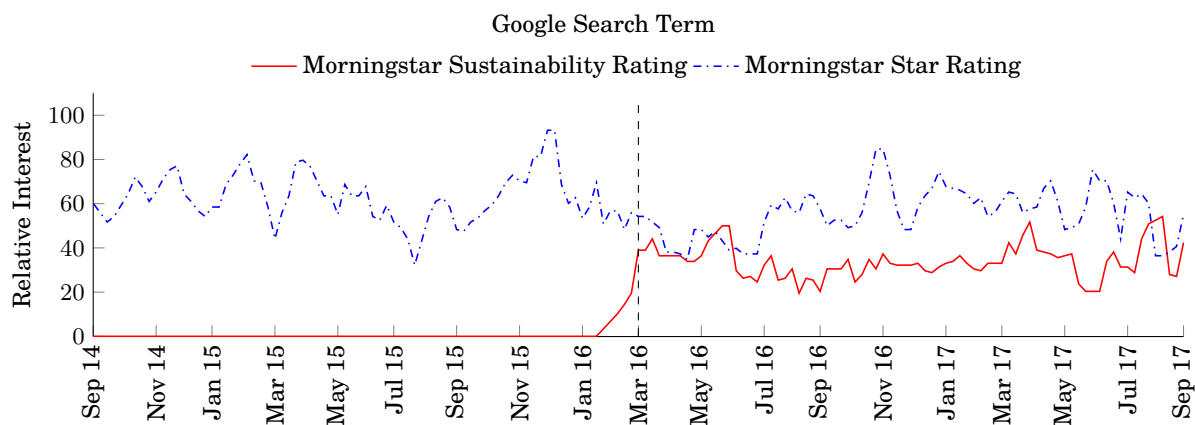
Figure 1. Google Search Interest in Morningstar Ratings.
This graph shows the four weeks moving average of the relative google search interest in the search terms "Morningstar Sustainability Rating" and "Morningstar Star Rating". The dashed vertical line indicates the initial publication date of the Sustainability Rating.
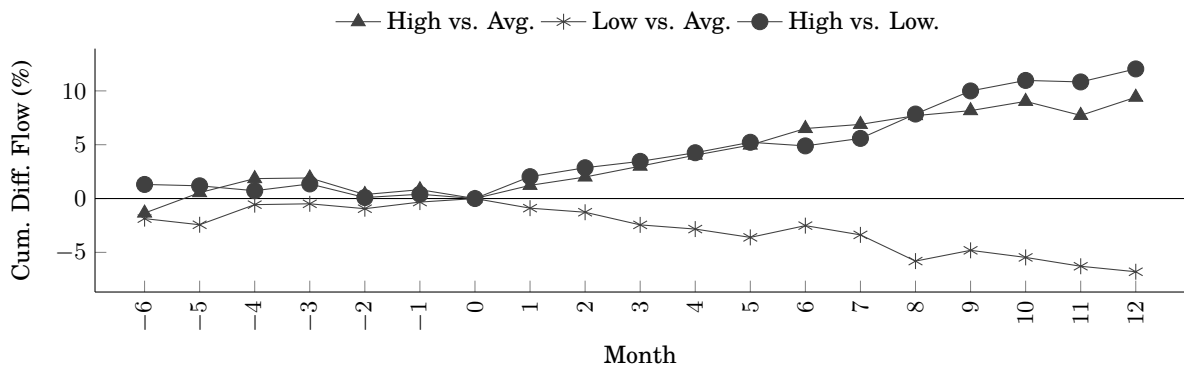
Figure 2. Propensity Score Matching - Cumulative Differences in Flows.
This figure visualizes the average treatment effect on the treated from a nearest neighbor propensity score matching. It shows the cumulative flow differences between the treatment and the control group for the 6 months prior to and the 12 months after the launch of the Morningstar Sustainability Rating. Funds are matched based on their characteristics (12-months alpha, Performance Rating, size, fund age, expense ratio, and turnover ratio) as of 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 vs. Average" and "Low vs. Average" the treatment group consists of all high or low-rated funds according to the Morningstar Sustainability Rating in March 2016; the control group is selected from all average-rated funds. For "High vs. 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 $[\text{TNA}_t - (1 + \text{R}_t) * \text{TNA}_{t-1}]/\text{TNA}_{t-1}$. 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.
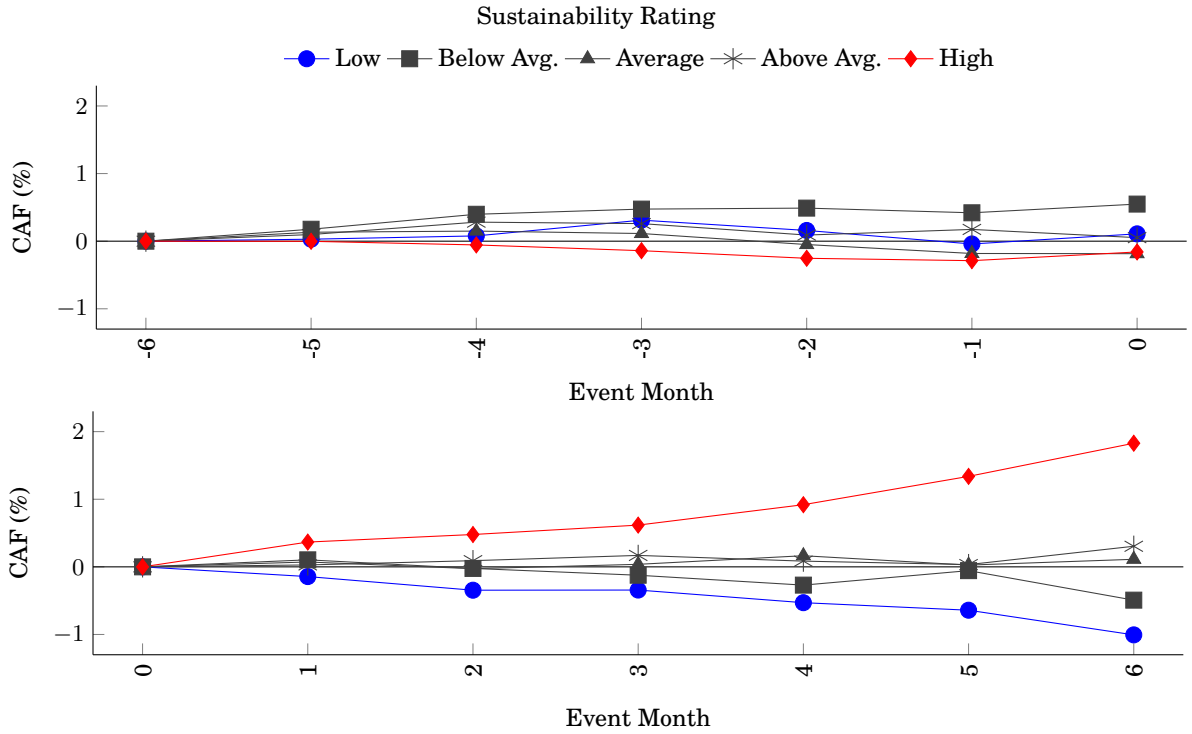
Figure 3. Cumulative Relative Abnormal Flow Before and After Launch of the Sustainability Rating.

The figure visualizes the average cumulative relative abnormal fund flows $\overline{CAF_t}$ of each Sustainability Rating category for the 6-months period before and after a fund initially received a Sustainability Rating. To ensure the availability of the Sustainability Rating for the full sample of six 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 N 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 waive 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 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 parameter for each fund individually over an estimation window of 24 months $t = [-24; 0]$, while we shift the estimation window by 6 months, i.e. $t = [-30; -6]$, to calculate the $\overline{CAF_t}$ prior to the introduction.

Table 1
Sample Size by Month

This table lists the monthly number of distinct U.S. 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	

Table 2

Mean Fund Characteristics Sorted by the Sustainability Rating After the Initial Rating Publication

This table presents the mean values of fund characteristics of our sample. All values are reported on 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. Panel B and Panel C display mean values for the subsample of retail and institutional share classes, respectively. Panel D to 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. Panel E and F split the sample into retail, respectively institutional share classes. 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 $[\text{TNA}_t - (1 + \text{R}_t) * \text{TNA}_{t-1}]/\text{TNA}_{t-1}$. 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 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 obtained respectively, from the MTNA, TURN_RATIO and EXP_RATIO variables in the CRSP mutual fund database. Column $(5) - (1)$ presents difference in means tests for mean characteristics of high and low-rated share classes. *, **, and *** indicate significance at the 10 %, 5 %, and 1 % levels, respectively.

	Sustainability Rating						
	Low	(2)	(3)	(4)	High	$(5) - (1)$	Total
Panel A: Mean fund characteristics for retail and institutional funds (09/2015 − 02/2016)							
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-months alpha (%)	−2.17	−1.03	−1.49	−1.59	−1.81	0.32**	−1.51
Total net assets ($m)	428.93	770.15	944.21	704.18	991.19	484.67***	794.51
Fund age in years	14.33	14.12	15.02	14.54	14.31	0.04	14.57
Turnover ratio (%)	56.33	62.51	57.95	57.67	42.00	−14.45***	57.44
Expense ratio (%)	1.27	1.17	1.15	1.19	1.11	−0.14***	1.17
Retail share classes (%)	58.10	49.36	50.19	53.13	57.61	−1.49	52.18
Number of observations	1568	3264	5099	3813	1189		14933
	(10.5%)	(21.86%)	(34.15%)	(25.53%)	(7.96%)		(100.00%)
Panel B: Mean fund characteristics for retail funds (09/2015 − 02/2016)							
Relative net flow (%)	−0.60	0.20	−0.30	0.01	−0.56	0.05	−0.17
Monthly return (%)	−1.28	−0.92	−0.94	−0.65	−0.50	0.71***	−0.86
Performance rating	2.59	2.89	2.81	2.86	3.01	0.42***	2.83
12-months alpha (%)	−2.61	−1.24	−1.89	−1.93	−1.85	0.64***	−1.84
Total net assets ($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%)		(100.00%)
Panel C: Mean fund characteristics for institutional funds (09/2015 − 02/2016)							
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-months alpha (%)	−1.55	−0.82	−1.08	−1.21	−1.74	−0.08	−1.14
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%)		(100.00%)

(continued)

Table 2 – *Continued*

	Low	(2)	(3)	(4)	High	(5) – (1)	Total
				Sustainability Rating			

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

	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-months alpha (%)	−4.56	−3.47	−3.20	−2.63	−1.89	2.68***	−3.13
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%)		(100.00%)

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

	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-months alpha (%)	−4.82	−3.73	−3.49	−2.99	−2.16	2.65***	−3.43
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%)		(100.00%)

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

	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-months alpha (%)	−4.24	−3.22	−2.90	−2.23	−1.57	2.71***	−2.82
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%)		(100.00%)

Table 3
Relative Net Flow – Sustainability Regression

This table reports the results from OLS 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 $[\text{TNA}_t - (1 + \text{R}_t) * \text{TNA}_{t-1}]/\text{TNA}_{t-1}$. 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) and the Carhart four-factor alpha is 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 2 with single observations for each share class and month. We split the sample in 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 are 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	−0.312	−0.200*	−0.311**	−0.085	0.677
	(−1.46)	(−1.64)	(−2.28)	(−0.42)	(1.54)
Below average	−0.043	−0.125*	−0.113	−0.133	0.993*
	(−0.33)	(−1.72)	(−1.25)	(−1.12)	(1.94)
Above average	0.134	0.021	0.100	−0.056	0.440
	(0.92)	(0.25)	(0.95)	(−0.35)	(1.23)
High	−0.049	0.267**	0.419***	0.116	−0.033
	(−0.21)	(2.19)	(3.10)	(0.64)	(−0.04)
1-month lagged Performance Rating					
Low	−0.444**	−0.334	−0.100	−1.200***	−1.283*
	(−2.26)	(−1.34)	(−0.50)	(−3.11)	(−1.73)
Below average	−0.548***	−0.595***	−0.454***	−0.913***	−1.155***
	(−4.06)	(−4.44)	(−5.18)	(−3.67)	(−3.39)
Above average	1.080***	1.030***	1.019***	1.065***	1.781***
	(6.94)	(10.32)	(7.05)	(8.15)	(4.26)
High	3.302***	3.199***	3.340***	3.153***	3.078***
	(6.81)	(11.19)	(7.60)	(9.31)	(4.66)
1-month lagged 12-months alpha (%)	0.234***	0.147***	0.144***	0.148***	0.156***
	(6.20)	(5.94)	(5.63)	(5.12)	(4.35)
1-month lagged monthly return (%)	−0.031	0.088	0.157***	−0.010	−0.144**
	(−0.49)	(1.59)	(3.27)	(−0.14)	(−2.77)
1-month lagged turnover ratio (%)	−0.000	−0.001	−0.000	−0.002	0.008
	(−0.14)	(−0.99)	(−0.21)	(−1.18)	(1.35)
1-month lagged expense ratio (%)	0.582*	−0.390**	−0.182	−1.122***	−0.139
	(1.91)	(−2.28)	(−1.39)	(−3.71)	(−0.22)
Log 1-month lagged total net assets ($m)	−0.202***	−0.249***	−0.238***	−0.302***	−0.204*
	(−3.00)	(−8.62)	(−6.06)	(−6.76)	(−1.78)
Fund age in years	−0.002	−0.004	0.005	−0.036***	−0.023
	(−0.40)	(−1.12)	(1.32)	(−3.24)	(−0.62)
Month–style fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	14945	38202	19676	18526	2418
Adjusted R^2	0.08	0.07	0.09	0.07	0.08

Table 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 the 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 vs. 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 vs. 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 vs. 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 are 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-months 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 ($m)	4.67	4.68	4.78	4.34	0.71	0.68
Fund age in years	16.63	13.01	17.97	15.01	0.46	0.04**
Expense ratio (%)	1.30	1.23	1.26	1.22	0.47	0.89
Turnover ratio (%)	48.29	35.00	43.21	40.00	0.28	0.68
Panel B: Matched sample comparison – Model: Low vs. Average						
12-months alpha (%)	−3.52	−2.93	−3.46	−3.24	0.89	0.29
Performance rating	2.47	3.00	2.49	2.00	0.83	0.41
Log total net assets ($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-months 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 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-months alpha, Performance Rating, size, fund age, expense ratio, and turnover ratio) as of 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 vs. Average" and "Low vs. Average" the treatment group consists of all high or low-rated funds according to the Morningstar Sustainability Rating in March 2016, the control group is selected from all average-rated funds. For "High vs. 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: six months prior (-6 to 0), and three 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 $[\text{TNA}_t - (1 + \text{R}_t) * \text{TNA}_{t-1}]/\text{TNA}_{t-1}$. 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 are denoted by *, **, and ***, respectively.

| | N | Months | | | | | |
		-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)

Table 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 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 $\overline{SAF_t}$ 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.

Panel A: Average standardized relative abnormal flow by Sustainability Rating and event month (coefficients)

Event month	Low $\overline{SAF_t}$	Low % > 0	Below Avg. $\overline{SAF_t}$	Below Avg. % > 0	Average $\overline{SAF_t}$	Average % > 0	Above Avg. $\overline{SAF_t}$	Above Avg. % > 0	High $\overline{SAF_t}$	High % > 0	High–Low $\overline{SAF_t}$
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***	37.5***	−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)

Event month	Low $\overline{SAF_t}$	Low % > 0	Below Avg. $\overline{SAF_t}$	Below Avg. % > 0	Average $\overline{SAF_t}$	Average % > 0	Above Avg. $\overline{SAF_t}$	Above Avg. % > 0	High $\overline{SAF_t}$	High % > 0	High–Low $\overline{SAF_t}$
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 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 N 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 % levels, respectively.

Panel A: Average cumulative standardized relative abnormal flow by Sustainability Rating and event month (coefficients)

Event month	Low $\overline{CSAF_t}$	Low $\% > 0$	Below Avg. $\overline{CSAF_t}$	Below Avg. $\% > 0$	Average $\overline{CSAF_t}$	Average $\% > 0$	Above Avg. $\overline{CSAF_t}$	Above Avg. $\% > 0$	High $\overline{CSAF_t}$	High $\% > 0$	High–Low $\overline{CSAF_t}$
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)

Event month	Low $\overline{CSAF_t}$	Low $\% > 0$	Below Avg. $\overline{CSAF_t}$	Below Avg. $\% > 0$	Average $\overline{CSAF_t}$	Average $\% > 0$	Above Avg. $\overline{CSAF_t}$	Above Avg. $\% > 0$	High $\overline{CSAF_t}$	High $\% > 0$	High–Low $\overline{CSAF_t}$
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***

Table 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 six 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 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 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	53.2	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 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

Appendix A.   Event Study - Benchmark Diagnostics

Table 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 the 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 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
	$\overline{\text{SAF}_t}$	$\% > 0$	$\overline{\text{SAF}_t}$	$\% > 0$	$\overline{\text{SAF}_t}$	$\% > 0$	$\overline{\text{SAF}_t}$	$\% > 0$	$\overline{\text{SAF}_t}$	$\% > 0$	$\overline{\text{SAF}_t}$
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

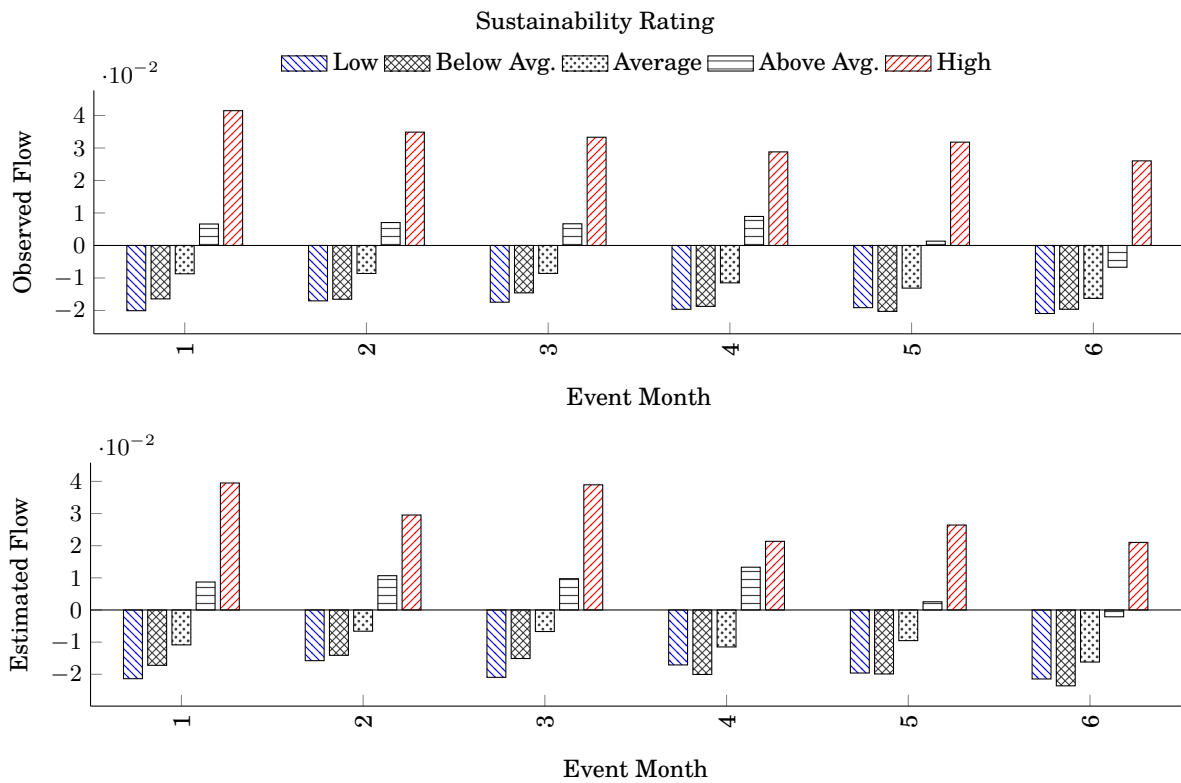# Appendix B.   Event Study - Observed vs. Estimated Flows

Figure B1. Observed vs. 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, i.e. 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 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.