Do Sustainability Ratings Induce ESG Window-Dressing by Mutual Funds?*

Kevin C. W. Chen

Hong Kong University of Science and Technology

Wilbur X. Chen

Hong Kong University of Science and Technology

Haifeng You

Hong Kong University of Science and Technology

Abstract

We find that after the initiation of Morningstar's mutual fund ESG ratings, the implied returns from funds' quarter-end holdings exhibit significantly higher ESG β than their actual returns. The increase in the ESG β gap between the two is greater in the period following higher ESG factor returns, and among funds with poor past performance, experiencing outflows, surrounding rating thresholds and bearing explicit ESG labels. We also find that funds with a higher ESG β gap are associated with higher ESG ratings, which tend to attract higher future fund flows. Finally, we show that ESG window dressing also leads to predictable returns. Stocks with the best (worst) ESG ratings earn significantly positive (negative) returns before quarter-ends, which reverse shortly thereafter. This pattern strengthens for stocks with higher ownership by high ESG β-gap funds. Our findings suggest that assessing funds' ESG performance with holdings based ESG scores brings unintended consequences.

Keywords: ESG Investing, Mutual Funds, Rating Agencies, Window-Dressing

JEL Classification Codes: G11, G12, G23, G24, Q56

* Email acchen@ust.hk, acwilbur@ust.hk and achy@ust.hk respectively. We thank seminar participants at HKUST for valuable comments and suggestions. We thank HKUST business school for funding support and Zhisheng Lin for excellent research assistance.

1 Introduction

To cater to increasing investor demand, there has been an explosion of both public and private funds marketed as socially responsible, sustainable, or environmental, social, and governance (ESG) funds. The 2022 *Report on US Sustainable Investing Trends*, published by the SIF Foundation, counts $8.4 trillion of ESG assets under management, representing 12.6 percent of the total assets under management in the United States.

In academic research, the extent to which a fund pursues ESG objectives is usually measured by the average ESG scores of the securities it holds (Dumitrescu et al., 2022; Gibson Brandon et al., 2022; Liang et al., 2022). The same approach is used in investment practice. For example, Morningstar launched sustainability ratings for more than 20,000 mutual funds in March 2016. It bases its ratings on the value-weighted average of Sustainalytics ESG scores of funds' portfolio firms. Hartzmark and Sussman (2019) show that investors reacted to Morningstar's sustainability ratings, as funds with high ratings enjoyed significant net inflows and those with low ratings experienced substantial net outflows.

Yet the use of ESG scores in assessing funds' ESG performance is not without problems. Raghunandan and Rajgopal (2022) show that funds that self-identify as ESG-oriented hold portfolio firms with higher average ESG scores but weaker ESG performance. We extend this literature by studying another unintended consequence of relying on aggregated ESG scores of mutual funds' portfolios to assess their ESG performance—doing so can induce window-dressing by mutual funds. Specifically, we use the initiation of Morningstar's Sustainability Rating for Funds as a shock to the salience of holdings-based ESG assessment and study

whether it induces mutual funds to reshuffle their quarter-end holdings temporarily to achieve better ESG scores and ratings.[1]

To attract more cash inflows and discourage outflows, fund managers have an incentive to hold firms that exhibit better Sustainalytics ESG scores after the launch of the Morningstar sustainability ratings in March 2016. However, stocks with high ESG scores do not always have higher (expected) future returns (Cornell, 2021; Di Giuli and Kostovetsky, 2014; Pástor et al., 2022), and investors may not be willing to sacrifice returns for sustainability considerations (e.g., Larcker and Watts, 2020; Renneboog et al., 2008). To achieve higher overall ESG scores without hurting (expected) return performance, fund managers might temporarily buy (sell) stocks with high (low) ESG scores right before quarter-ends and unwind these positions right afterward, following the initiation of these ratings.[2] If mutual funds pursue this form of ESG window-dressing, then their quarter-end holdings would exhibit higher ESG scores compared to other periods.

However, funds only report holdings at the end of each quarter. To overcome this limitation, we compare mutual funds' exposures to an ESG factor (or ESG β) using both their actual daily returns and implied (hypothetical) daily returns based on their quarter-end portfolio holdings.[3] Consistent with the Morningstar sustainability ratings increasing funds' investment in ESG stocks, we find that both holdings- and return-based ESG βs exhibit a sharp increase after the initiation of these ratings. More importantly, we find that the difference between the two, which

[1] MSCI also launched its own fund ratings in March 2016, based on the aggregated MSCI ESG ratings of a fund's holdings. However, unlike Morningstar, which provides the fund ratings on its website for free, MSCI's rating is provided only to paid subscribers. Thus, its influence on fund flows should be weaker. We follow Hartzmark and Sussman (2019) and focus on the Morningstar ratings.

[2] The temporary reshuffling should only occur in stocks with high (low) ESG scores but low (high) future anticipated returns by fund managers. Otherwise, ESG scores and future returns would be perfectly aligned. Fund managers could then simply buy and hold (avoid) those with high (low) ESG scores. Since fund managers' private assessments of future stock returns are unobservable, we conduct our tests using stocks with extreme ESG scores.

[3] To estimate the ESG factor, we follow the method for computing the value factor (Fama and French, 1993) by performing a double sort on the size and ESG performance dimension, and computing the difference between the value-weighted returns of the size-adjusted top ESG portfolio and the size-adjusted bottom ESG portfolio.

we term as the ESG β gap, also increases significantly after the initiation of the Morningstar sustainability ratings.[4]

Furthermore, we find a greater increase in ESG β gap among mutual funds surrounding the breakpoints of the four-to-five globe ratings, crossing which Hartzmark and Sussman (2019) show to be helpful for attracting (avoiding) fund flows (outflows). In addition, we find that the increase in ESG β gap is more pronounced when the ESG factor is constructed with Sustainalytics ESG scores versus MSCI scores. This comports with the prediction that mutual funds would have a weaker incentive to window-dress for higher MSCI ESG ratings, as they are only available to paid subscribers and therefore less likely to attract retail investor attention and fund flows. Taken together, these results support the hypothesis that the initiation of the Morningstar sustainability ratings induced ESG window-dressing of quarter-end holdings.

As argued earlier, the potential increase in fund flows after achieving a higher Morningstar sustainability rating is one key reason for mutual fund ESG window-dressing. Consistent with this notion, we find first find that funds with weak past performance (as measured by returns) tend to exhibit a greater ESG β gap.[5] Moreover, we also show that the ESG β gap is associated with higher ESG ratings, which are also associated with greater one-to-four-quarter-ahead fund flows.[6] Finally, the ESG β gap tends to be greater during periods following strong ESG performance (as measured by ESG factor returns). Collectively, these empirical patterns suggest that investors' enthusiasm and more capital allocation towards funds with high (and

[4] This approach follows that of Agarwal et al (2014), who measure the window-dressing of good-performing stocks with the difference between the hypothetical return of a fund's reported holdings at the end of the fiscal quarter and its actual return (called backward holding return gap).

[5] In addition to fund performance, we show that ESG labelled funds are more likely to exhibit higher ESG β gap, perhaps due to greater investor demand for high ESG performance. Moreover, we also find that the ESG β gap is more pronounced in time-periods when the ESG factor portfolio exhibits high returns.

[6] Our findings that ESG ratings continue to attract fund flows is somewhat different from Gantchev et al (2023) who show that ESG ratings no longer attract fund flows in the year after the initiation year. One reason could be a difference in sample as their sample ends in 2017, while our ESG ratings sample starts in 2018.

salient) Morningstar sustainability ratings is one of the important incentives for funds to engage in ESG window-dressing.

Quarter-end ESG window dressing by mutual funds also lead to predictable firm return patterns. Specifically, we find significantly positive (negative) abnormal returns over the five trading days for stocks with high (low) ESG scores before calendar quarter-ends and the opposite pattern over the five days afterward.[7] This pattern is more pronounced after March 2016, when Morningstar launched the sustainability fund ratings. Further analyses show that our results are driven primarily by stocks with higher ownership by mutual funds with high ESG β gaps (i.e., those more likely to window-dress).

Finally, we examine whether our hypothesis on the introduction of Morningstar's ESG ratings and the ESG window-dressing extends more broadly to the recent initiation of Morningstar's carbon ratings. Using a measure of the carbon β gap (where the carbon β is based on the returns of a portfolio that shorts fossil fuel firms), we show that the carbon β gap also exhibits a significant increase following the introduction of the Sustainalytics carbon rating. Thus, we find additional evidence that lends support to the claim that sustainability ratings are a key driver of ESG window-dressing.

This study makes several contributions. First, it contributes to the literature on institutional investors' window-dressing of portfolio holdings. The literature finds that these investors tend to buy winners and sell losers at the calendar year-end (e.g., Agarwal et al., 2014; Haugen and Lakonishok, 1988), a practice that results in a persistent annual stock return anomaly known as the January effect. Studies have argued that this phenomenon is due to incentives to mask fund performance or investment strategies (He et al., 2004; Ritter and Chopra, 1989). We

[7] We focus on the calendar quarter-end for our firm analyses, as most mutual funds have fiscal quarters that align with a calendar quarter.

contribute to these studies by showing that reliance on fund-level ESG ratings leads mutual funds to window-dress ESG scores at the quarter-end.[8]

Second, our paper also contributes to the literature on the limitations of ESG ratings for sustainability investing. Some studies find significant disagreement among the ESG ratings of different raters, even after adjusting for the differences in their definitions, suggesting that the ratings may have limited validity and usefulness (e.g. Berg et al., 2022; Chatterji et al., 2016; Serafeim and Yoon, 2022). Our study extends the literature by showing another limitation—these ratings might induce funds to window-dress their quarter-end holdings to temporarily boost investor perception and agency ratings of their ESG performance. Consequently, investors buying funds that window-dressed their ESG performance would not only fail to truly invest in sustainable firms but would also have to bear the additional costs (e.g., transaction costs and forgone profits) incurred by these manipulations.

Finally, our paper also has policy implications. To promote consistent and comparable information for fund investors, the Securities and Exchange Commission (SEC) has proposed mandating more disclosures by ESG funds regarding their strategies (SEC 2022).[9] These enhanced disclosures might direct more investor attention to funds' aggregated ESG scores and performance, which is likely to pressure more funds to window-dress their portfolios. Moreover, the SEC's proposed rule requires certain environmentally focused funds to disclose the greenhouse gas emissions of their portfolio investments. Our study points out a distinct possibility that, if these proposals are implemented, funds could window-dress their quarter-end portfolios to include firms with metrics of greenhouse gas emissions that suit their needs. Given our evidence of prevalent ESG window-dressing by mutual funds, regulators should

[8] Our setting is not confounded with the income tax considerations (e.g., Ng and Wang, 2004; Sias and Starks, 1997), which has been proposed as an alternative reason for the January effect. Thus, our results likely provide unambiguous evidence of ESG window-dressing by mutual funds.

[9] For instance, if a fund considers ESG scores or information from an ESG service provider, it must reveal the name of the provider.

consider possible unintended consequences if they impose portfolio holdings-based sustainability disclosure.

2 Related Literature, Institutional Background, and Hypotheses

Our study relates to several lines of literature, including those on i) window-dressing by institutional investors and ii) sustainable/ESG investing and the role of ESG ratings.

2.1 Window-dressing

Window-dressing by institutional investors (e.g., mutual funds and pension funds) involves purchasing (selling) winning (losing) stocks before reporting dates, in an attempt to present a more attractive image of the fund's performance or investment strategy (e.g., Lakonishok et al., 1991; Ritter and Chopra, 1989). Window-dressing leads to predictable trading and stock return patterns surrounding the (fourth) calendar quarter -ends, which has been used to explain the turn-of-the-year or the January effect (e.g., Haugen and Lakonishok, 1988).[10] Subsequent studies also document window-dressing in other settings. For example, Musto (1999) shows that, to disclose safer portfolios, money market funds hold more government issues relative to private issues around disclosures than at other times. In addition, Chevalier and Ellison (1997) show that growth and growth-and-income funds alter the composition of their portfolios at year-end toward high quality, less risky stocks.

[10] However, the window-dressing view of the January effect has been subject to debate. Several studies argue that the incentive to realize tax losses at year-ends may also induce investors to sell losers and explain the January effect (Poterba and Weisbenner, 2001; Sikes, 2014). Others further argue that the window-dressing incentive is not entirely clear, as this action incurs high portfolio turnover and trading costs, which reduces future performance and could instead drive fund outflows. Consistent with this view, Agarwal et al. (2014) find that, conditional on bad performance, window-dressing may even lead to worse investor outflows.

We contribute to this literature by providing additional evidence of institutional investor window-dressing in a new setting. Specifically, we investigate whether mutual fund managers also window-dress their ESG scores, to present a better ESG performance and attract fund flows. [11]

2.2 Limitations of ESG Scores

Our paper also relates to the recent literature on the role of ESG ratings in sustainable investing. In recent years, ESG investing has surged, despite the controversy on whether the ESG approach helps or hurts investment performance (e.g., Hong and Kacperczyk, 2009; Statman and Glushkov, 2009; Starks, 2023). [12] For example, recent work shows that institutional investors that endorse the United Nations Principles for Responsible Investment (PRI) attract significantly greater investor flows than peers, even if they do not necessarily follow through on their responsible investment commitments (e.g., Gibson Brandon et al., 2022; Kim and Yoon, 2023; Liang et al., 2022).

ESG ratings might facilitate better investment decisions, as the ratings provide investors with a standardized assessment of firms' ESG performance. Recent studies show that ESG

[11] In a contemporaneous study, Parise and Rubin (2023) also study ESG window dressing by mutual funds. Our paper differs from theirs in several ways: First, Parise and Rubin adopt a different approach by showing that the ESG beta in the trading days *immediately before* mandated portfolio disclosure is systematically higher compared to the ESG beta *immediately after* the portfolio disclosure date. This measure relies on the assumption that the ESG betas before and after the portfolio disclosure date are comparable. However, funds might behave differently in the beginning of a calendar quarter. For example, they might rebalance their portfolios more aggressively or take more risk in the beginning of a calendar quarter, which may disturb their return performance and lead to lower ESG beta estimates. In contrast, our measure of the ESG window-dressing compares the ESG beta of the actual portfolio returns and the ESG beta of the implied portfolio returns from the disclosed holdings in the *same* estimation period, which is therefore less likely to be confounded by other economic events. Second, our identification strategy is to use the launch of Morningstar fund ESG rating as a shock to the ESG window dressing incentives, while Parise and Rubin only focus on the post-rating periods without checking whether the effect exist before or not; Finally, we also provide ample evidence on funds' incentives (and consequences) of fund ESG window dressing, showing that achieving higher ESG ratings and attracting fund flows is an important incentive for ESG window-dressing.

[12] According to Bloomberg Intelligence, ESG assets surpassed $35 trillion in 2020, up from $22.8 trillion in 2016, to become a third of the total global AUM. Furthermore, assuming a third of the past five-year growth rate (15%), ESG assets could exceed $50 trillion by 2025. Source: https://www.bloomberg.com/company/press/esg-may-surpass-41-trillion-assets-in-2022-but-not-without-challenges-finds-bloomberg-intelligence/

ratings predict future ESG news (Serafeim and Yoon, 2022) and retail investors rely on them when making trading decisions (Rzeźnik et al., 2022). Hartzmark and Sussman (2019) likewise show that investors also rely on ESG ratings when investing in mutual funds. Specifically, when Morningstar began its sustainability ratings for funds, investors allocated $24 billion and $32 billion more to funds ranked five globes and $12 billion to $15 billion in assets less to funds ranked one globe.[13]

Yet ESG ratings are not a panacea in ESG investing. Simply relying on ESG ratings does not guarantee that investors will reach the goal of promoting sustainability. Notably, the literature shows that there is considerable disagreement among different raters, jeopardizing the ratings' reliability and validity in capturing ESG performance (Berg et al., 2022; Chatterji et al., 2016).[14] Consistent with the inability of ESG scores to capture ESG performance, Raghunandan and Rajgopal (2022) find that, while self-labelled ESG funds hold portfolio firms with higher average ESG scores, these firms tend to have worse track records in labor and environmental compliance and higher carbon emissions per unit of revenue.

2.3 Hypothesis Development

As discussed above, the ESG label and ratings of mutual funds increasingly drive fund flows, which in turn determine the amount of AUM and management fee revenues (Liang et al., 2022). This relationship has likely increased with the introduction of the publicly available Morningstar sustainability ratings. Thus, we would expect that, after the introduction of these ratings, fund managers will have a stronger incentive to temporarily buy (unload) some stocks

[13] The five and one globes correspond to the top and bottom 10% respectively.
[14] European Commission Report (2021) identifies between 10 to 15 major sustainability-related data providers, including Bloomberg, CDP, FTSE Russell, ISS-ESG, MSCI, Refinitiv, RepRisk, RobecoSAM, Sustainalytics, and Vigeo Eiris. Several of these agencies provided ESG ratings and related data before 2016 but only at the firm level.

with high (low) ESG scores (i.e. ESG window dressing).[15] ESG window dressing not only helps improve their quarter-end aggregated ESG scores and potentially their ESG rating but also allows them to (temporarily) hold some hot ESG stocks with high media attention, which Solomon et al. (2014) find to also drive fund flows.

However, ESG window-dressing does entail costs. First, it brings additional transaction costs, which could reduce funds' returns. Second, funds might miss good investment opportunities (and returns) from window-dressing, as they would temporarily unload some securities with poor ESG scores but good expected returns. Third, funds may incur reputational (and regulatory) costs if their managers are caught committing window-dressing. In recent years, greenwashing of investment funds has become a concern to investors as well as regulators, and there has been increasing litigation against greenwashing.[16] Therefore, whether and to what extent mutual funds collectively engage in these activities remain empirical questions.

If mutual funds do pursue ESG window-dressing at quarter-ends, we expect the aggregated ESG score to be abnormally high at the quarter-end compared to other periods. However, mutual funds only report their holdings at quarter-ends, which prevents us from testing this prediction directly. Therefore, we construct a variable called ESG β gap to detect mutual funds' ESG window-dressing. ESG β gap is the difference in the exposure to an ESG factor between the implied daily returns to the portfolios reported at the start of the quarter and the actual daily

[15] Managers would choose to do this temporarily because stocks with high ESG scores do not always have higher (expected) returns (Cornell, 2021; Di Giuli and Kostovetsky, 2014; Pástor et al., 2022). If the ESG score of a stock conflicts with the managers' private assessment of its future expected returns, such temporary strategic manipulation would enable funds to maintain a high ESG perception while minimizing the negative impact on their expected portfolio returns.

[16] In March 2021, the SEC announced the creation of its Climate and ESG Enforcement Task Force, responsible for overseeing ESG-related disclosures, investments, and compliance efforts by securities issuers and advisers. In the March release of its 2022 Examination Priorities, the SEC Division of Examinations announced that it would be specifically focusing on "ESG-related advisory services and investment products (e.g., mutual funds, exchange-traded funds (ETFs), and private fund offerings)." https://www.reuters.com/legal/legalindustry/greenwashing-wave-hits-securities-litigation-2022-09-22/

portfolio returns.[17] A positive ESG β gap indicates that the ESG exposure of the reported holdings is higher than the average ESG exposure of the mutual fund's actual portfolio. Thus, if the initiation of the Morningstar sustainability ratings increases fund managers' incentives to window-dress the ESG profile of their quarter-end holdings, we would expect:

P1(a): The mean ESG β gap of mutual funds will increase after the initiation of the Morningstar sustainability ratings.

As discussed above, MSCI launched similar ratings (but based on firm-level MSCI ESG ratings) at around the same time as Morningstar. However, instead of providing free access to the ratings as Morningstar did, MSCI's fund ratings were available only to its subscribers, mostly institutional investors. Given that most of the mutual fund investors are retail investors, MSCI's fund ratings have lower salience and accessibility for retail investors. Thus, we expect that this would damp fund managers' incentive to window-dress their quarter-end portfolio's aggregated MSCI ESG ratings, which motivates our second prediction:

P1(b): The ESG β gap increase after the initiation of the Morningstar sustainability ratings would be larger for ESG factors based on Sustainalytics scores than MSCI ESG scores.

Mutual funds' ESG window-dressing could also result in predictable returns around mutual fund portfolio reporting dates at the stock level. Specifically, to improve their aggregated ESG scores at quarter-ends without reducing their expected returns, fund managers would purchase (sell) some stocks with high (low) ESG scores and unwind these transactions shortly afterward. If such trades are prevalent and lead to stock price movements, we would expect:

[17] As discussed in Section 3, to compute the ESG β gap, we estimate the difference between (a) the sensitivity of daily returns of a mutual fund's quarter-end portfolio holdings to the ESG factor and (b) the sensitivity of a fund's actual daily returns to the ESG factor over the fiscal quarter. This approach resembles that of Agarwal et al. (2014), who measure the window-dressing of mutual funds as the difference between the hypothetical return of a fund's reported quarter-end holdings and its actual return (called the backward holding return gap).

P2(a): After the initiation of the Morningstar sustainability rating, stocks with high (low) ESG scores would exhibit more positive (negative) abnormal returns right before quarter-ends and more negative (positive) abnormal returns right after quarter-ends.

Finally, not all mutual funds window-dress due to differences in culture, ethics, and cost/benefit trade-offs. Therefore, we would expect the predictable pattern in returns for firms with extreme ESG scores to be more pronounced if these firms are owned by mutual funds more likely to pursue manipulative activities. Using the ESG β gap as a proxy for the likelihood of ESG window-dressing by mutual funds, we make our final prediction as follows:

P2(b): The predictable abnormal returns as in P2(a) will be more pronounced for stocks with higher ownership by mutual funds with high ESG β gaps.

3 Data, Sample Selection and Variable Construction

3.1 Sample Selection

We construct our sample from a collection of databases, namely, CRSP/Compustat, Sustainalytics, and the CRSP mutual fund data. We perform analyses at both the fund and firm level, and so we describe the breakdowns of the sample selection in Table 1.

For the fund analyses, our sample begins with the CRSP mutual fund data from 2011–2022. We restrict the sample to US domestic equity funds, funds with valid fiscal year-end dates, and funds that are not classified as exchange-traded funds or index funds. We further drop funds with less than $5 million in total net assets or less than 10 portfolio firms (following Kacperczyk et al. 2009). Additionally, following Morningstar's cutoff, we drop funds if less than two-thirds of their portfolio firms by net assets are covered by Sustainalytics. Lastly, we drop funds that do not have sufficient daily data at the quarter level for computing the ESG β,

as well as observations with missing entries for controls. After these restrictions, our final panel data of fund-quarters consists of around 3,000 mutual funds and around 81,000 fund-quarter observations over 2011–2022.

We also perform analyses at the firm level, and this sample begins with the universe of CRSP/Compustat firm-quarter observations from 2011–2022. As we rely on Sustainalytics data to compute the ESG scores in our sample, we further restrict our sample to observations that are covered by Sustainalytics. Lastly, we restrict our sample to observations that have non-missing entries for the control variables. In total, our final sample has a total of over 4,000 companies with around 84,000 firm-quarter observations.

We present the key summary statistics for the firm sample in Panel A and the fund sample in Panel B of Table 2 respectively. For the fund data, we winsorize continuous variables at the 5 and 95% levels. In Panel A, we find that the average size of the fund is roughly $1 billion, and these funds tend to earn a return of roughly 2.6% on a monthly basis. For the firms in our sample, we winsorize fundamental variables at the 5% and 95% levels. In Panel B of Table 2, we report the summary statistics of the firms. The average size is roughly $9 billion, and the average book-to-market ratio is 0.58. The firms also have a positive operating profit on average of roughly 5%, and they tend to increase assets at an average growth rate of 1.7%.

3.2 Variable Construction: Sustainalytics ESG Score

Our main independent variable in our firm analysis is the Sustainalytics ESG score. In the months before October 2019, the Sustainalytics ESG score was computed as an absolute score of ESG ratings, where higher ESG ratings indicated better ESG performance. To increase comparability of the ESG score across industries, we follow Morningstar's methodology in computing an industry-adjusted normalized ratings score:

$$ESG\ Score_{i,t} = 50 + 10 \times \frac{ESG\ Rating_{i.t} - \mu_{ESG\ Rating,\ j,\ t}}{\sigma_{ESG\ Rating,\ j,\ t}}, \quad (1)$$

where the normalized ESG score is the ESG rating, relative to the average ESG rating at the industry-month level, scaled by the standard deviation in ESG ratings at the industry-month level.

After October 2019, Morningstar used the ESG risk ratings for its fund ratings, and so we use these ratings to compute the ESG score after this month. The ESG risk rating is a score of ESG risk, where *lower* values indicate better ESG performance relative to industry peers. The methodology for computing the scores evolves, and so we standardize the scores by computing a percentile rank in each quarter. After October 2019, the ordering of the ranks is reversed (as higher scores indicates weaker ESG performance), and so we multiply the ESG risk rating by minus one, before ranking the firms by their ESG scores.

3.3 Variable Construction: ESG β gap

For the fund analyses, our main independent variable is the ESG β gap, which proxies for the extent of ESG performance window-dressing. We follow Agarwal et al. (2014) to compute the backward holding return gap, and our methodology to develop the ESG β gap measure involves several steps.

First, we construct a long-short ESG portfolio that holds high ESG firms on the long side and low ESG firms on the short side. To construct this portfolio, we first perform a two-way sort of stocks based on the latest available Sustainalytics ESG score for stocks, and the market capitalization of stocks at the end of the prior month. We then follow the method for constructing the value factor in Fama and French (1992) by creating the breakpoints for ESG performance and size, and using the breakpoints to form 2 x 2 portfolios for ESG performance and size. Firms that are below the 30^{th} or above the 70^{th} percentile of ESG score are classified as the low and high ESG firms respectively. For the size dimension, firms below and above the median market capitalization are classified as the small and big firms respectively. We then compute the factor returns with the following formula:

$$ESG\ Factor_t = \frac{1}{2}\left(ESG_{H,S} + ESG_{H,B}\right) - \frac{1}{2}\left(ESG_{L,S} + ESG_{L,B}\right), \quad (2)$$

where $ESG_{H,S}$ and $ESG_{H,B}$ are the daily value-weighted returns of high ESG firms that are classified as small or big, respectively. $ESG_{L,S}$ and $ESG_{L,B}$ are the daily value-weighted returns of low ESG firms that are classified as small or big, respectively.

Next, we estimate an ESG β for both the actual returns of mutual funds and the implied daily return based on the reported holdings as of the end of the quarter, using the methodology outlined in Figure 1. For the actual and holding returns, we implement the following regression for each fund-fiscal quarter at the daily level:

$$Actual\ Ret_{p,t}\ or\ Hold\ Ret_{p,t} = \alpha_p + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_{AESG}\ or\ \beta_{HESG}\ ESG_t + \varepsilon_{p,t}, \quad (3)$$

where $Actual\ Ret_{p,t}$ is the reported actual daily returns of mutual fund p for day t and $Hold\ Ret_{p,t}$ is the value-weighted daily returns of firms in the disclosed portfolio of mutual fund p for day t. We implement these regressions for each fund-quarter with at least 50 trading days in the fiscal quarter. If actual returns are used as the dependent variable, our main coefficient is β_{AESG}, which measures the mutual fund's actual exposure to the ESG factor defined above. Alternatively, if holding-implied returns are used as dependent variable, our main coefficient of interest is β_{HESG}, which measures the mutual fund's exposure to the ESG portfolio as implied by its disclosed portfolio holdings. In addition, we include as controls in the regression the five Fama-French factors (Fama and French, 2015) for market ($MKTRF_t$), size ($SIZE_t$), value (HML_t), profitability (RMW_t), and investment (CMA_t) risks.

Conceptually, the actual ESG β (β_{AESG}) measures the exposure of the fund to ESG firms, as rated by Sustainalytics. If a fund is purchasing (selling) high (low) ESG score firms, it will exhibit a large and positive actual beta, and vice versa. Similarly, the intuition for the holdings-based ESG β (β_{HESG}) is that it tracks the exposure of the fund's reported holdings to ESG firms.

Thus, if a fund reports holdings that are tilted toward (away from) high (low) ESG funds, it will exhibit a large and positive holding beta, and vice versa.

Lastly, we compute an ESG β gap by taking the difference between the holding ESG β (β_{HESG}) and actual ESG β (β_{AESG}). A positive gap suggests that the ESG β based on a fund's disclosed holdings at the quarter-ends is higher than the ESG β based on the actual returns over the quarter. This would indicate that the reported portfolio at the quarter-ends has a higher aggregated ESG scores than the average aggregated ESG scores of the fund throughout the whole quarter, an empirical pattern that one would expect if a fund engages in ESG window-dressing.

4 Empirical Results

4.1 Sustainability Ratings and ESG Window-dressing in Mutual Funds

We begin by studying whether mutual funds engage in more ESG window dressing following the introduction of the Morningstar sustainability ratings. Specifically, we test our first prediction *P1(a)* that the mean ESG β gap of mutual funds increases after the initiation of the ratings. We first describe the related variables. In Figure 2(a), we plot the rolling 4-quarter average ESG beta estimated from mutual funds' actual daily returns (β_{AESG}) over time. As the figure shows, this beta is below zero before the introduction of the Morningstar sustainability ratings in March 2016. After March 2016, β_{AESG} increases to above zero, suggesting that mutual funds on average invest more in stocks with high ESG scores. ESG beta estimated from the implied daily returns to firms in mutual funds' disclosed portfolio at the quarter-end (β_{HESG}) also exhibits a similar pattern. Figure 2(b) shows that the rolling 4-quarter average β_{HESG} jumps up starting from the initiation of the Morningstar sustainability ratings in March 2016.

In Figure 2(c), we plot the rolling 4-quarter average ESG β gap over time. Consistent with the prediction in *P1(a)* that the ESG β gap should rise following the introduction of the

Morningstar sustainability ratings, we find an increase in the average ESG β gap since March 2016. Thus, the analyses presented in this figure provide some descriptive evidence that the window-dressing of ESG has become more common since the introduction of the Morningstar sustainability ratings.

We also study the changes over the March 2016 event quarter with different proxies of the ESG β gap, while holding fund characteristics fixed. To implement this analysis, we run the following regression:

$$ESG\ Beta_{p,q}\ or\ Beta\ Gap_{p,q} = \alpha_p + \beta 1_{2016} + \varepsilon_{p,q}, \quad (4)$$

where the dependent variable is either the actual holding ESG β ($ESG\ Beta_{p,q}$) or the ESG β gap ($Beta\ Gap_{p,q}$) for fund p in quarter q. To examine the changes in fund window-dressing, we include fund fixed effects (α_p). Our coefficient of interest is the coefficient on the indicator for the post-March 2016 period, and we expect this coefficient to be positive if the ESG βs and the ESG β gap increase following the introduction of the Morningstar sustainability ratings. For all regressions models, we cluster standard errors at the fund level.

Table 3 presents the results of this analysis for the actual ESG β in Panel A, the holding ESG β in Panel B, and the ESG β gap in Panel C. For each of the panels, we present two versions of the ESG βs or ESG β gaps, namely, the ESG βs/β gap computed with the three-factor Fama-French model (Fama and French, 1993) and the primary specification in Equation 3, the five-factor Fama-French model (Fama and French 2015). In addition, we also present an alternative specification that includes lagged control variables for size, quarterly returns, quarterly fund flows, management fee, expense ratio, portfolio concentration and the ESG portfolio returns.

In Panel A of Table 3, we find that the changes in the actual ESG β are robust across both specifications of the actual ESG β. These results suggest that the average holding ESG β in the post-Morningstar sustainability ratings period is 0.51–0.54 higher compared to the period

beforehand. Moreover, in Panel B, we also see an increase in the ESG β computed with holding-implied returns of roughly 0.58-0.60 in the post-Globe period compared to beforehand. Thus, our analyses across both panels and figures suggest that there is an increasing focus on ESG issues, which encourages greater (less) exposure to firms with higher (lower) ESG scores.

Panel C of Table 3 presents the main results of this table, which are the changes in the ESG β gap (holding ESG β minus the actual ESG β) around the year of the introduction of the Morningstar sustainability ratings. Consistent with our prediction *P1(a)*, we find an increase in the ESG β gap across both specifications of the beta gap after the initiation of the Morningstar sustainability ratings. Specifically, we find a 0.008 increase in the ESG β gap (or roughly 15% of the increase in the actual ESG beta), which suggests that ESG window-dressing has increased since the introduction of the Morningstar sustainability ratings.[18]

Notably, the event year 2016 coincides with a growing interest in ESG assessment of mutual funds, as other rating providers, such as MSCI, also started computing fund ESG scores in that year. As discussed in *P1(b)* of Section 2, unlike the Morningstar sustainability ratings, which are publicly available, the aggregated MSCI ESG ratings at the fund level are only available to subscribers. As a result, fund-level ESG ratings from MSCI would likely have a more muted impact on attention and fund flows from (retail) investors. Thus, we would expect mutual funds to have a weaker incentive to inflate their MSCI ESG ratings compared to their Morningstar sustainability ratings. Since the MSCI ESG ratings are based on the aggregated MSCI ESG scores of the portfolio firms, we analyze the MSCI-based ESG β gap with the ESG factor (return) that is constructed from the MSCI ESG ratings.

Table 4 presents the analysis with the MSCI-based ESG β gap. In Panel A, we report the changes in the MSCI-based ESG β gap, and we find that this version of the β gap decreases

[18] In untabulated additional analyses, we show that our results are robust to including the momentum factor in the β estimation model (Equation 3). Our results are also robust to using equal-weighting or score-weighting methods in constructing the ESG factor portfolio returns.

after March 2016.[19] Table 4 Panel B tests the prediction that the increase in ESG β gap is greater for ESG factor constructed from the Morningstar's scores than the one constructed from the MSCI scores. We compute the dependent variable as the difference between the ESG β gap computed using Sustainalytics ESG scores and the gap computed using MSCI ESG scores and regress it against an indicator variable of post-March 2016 together with fund fixed effects. Our analysis shows that the Sustainalytics-based ESG β gap consistently increases after March 2016, as the difference in the ESG β gap increases by 0.01 across the two different specifications of the beta gap with fund-level controls. This result indicates that the event quarter of March 2016 has a more pronounced effect on the Sustainalytics ESG β gap than the MSCI ESG β gap, which is consistent with our prediction *P1(b)*.

Thus overall, our analysis across Tables 3 and 4, suggests that the introduction of the Morningstar sustainability ratings has played a significant role in increasing window-dressing of the ESG performance, as measured by Sustainalytics.

4.2 Incentives of ESG Window-dressing in Mutual Funds

To unpack the incentives that drive ESG window-dressing in mutual funds, we study the key determinants of the ESG β gap. Following prior literature, we predict that four main factors could influence ESG window-dressing behavior in mutual funds.

First, fund performance could play a role in driving ESG window-dressing, as funds with weak performance could use ESG window-dressing as a tool to offset the negative implications of poor performance. Second, the investment objective of the fund could play a role in driving incentives for ESG window-dressing. Some funds advertise themselves as "ESG" funds, and for these funds, investors may pay additional attention to their ESG performance, leading to additional incentives for ESG window-dressing. Third, the profitability of the ESG investment

[19] On the other hand, consistent with rising interest in ESG, in untabulated additional analyses we find that MSCI-based ESG β is also increasing after March 2016.

strategy could also be a reason for ESG window-dressing.[20] Following periods of strong ESG stock performance, investors would likely allocate more capital to mutual funds with high ESG ratings, which should give funds a stronger incentive to engage in ESG window-dressing. Fourth, we predict that funds surrounding the Morningstar globe ratings thresholds will also have a greater incentive to engage in ESG window-dressing. Hartzmark and Sussman (2019) show that investors allocate more money to funds ranked five globes. If attracting fund flows is one of the objectives of ESG window-dressing, we would also expect mutual funds with pre-managed ESG scores right below and right above the thresholds of the four-to-five globes to have a stronger incentive to window-dress. Since we do not observe the pre-managed ESG scores, we instead predict that funds with aggregated ESG scores (based on reported holdings) near these breakpoints to exhibit a greater increase in ESG β gap.

We examine these determinant factors by implementing the following regression model:

$$ESG\ \beta\ Gap_{p,q} = \alpha + \beta_1 \text{Ret}_{p,q-1} + \beta_2 \text{Flow}_{p,q-1} + \beta_3 \text{ESG Fund}_p +$$

$$\beta_4 \text{ESG Return}_{p,q-1} + \beta_5 Near\ Breakpoint_{i,q-m} + \beta_6 \text{Size}_{p,q-1} +$$

$$\beta_7 \text{Management Fee}_{p,q-1} + \beta_8 \text{Expense Ratio}_{p,q-1} + \beta_9 \text{Concentration}_{p,q-1} + \varepsilon_{p,q}, \quad (5)$$

where, $\text{Ret}_{p,q-1}$is the quarterly return, $\text{Flow}_{p,q-1}$is the quarterly net flow and measures fund performance. ESG Fund_p is an indicator for funds labelled as "ESG" and measures the difference in ESG window-dressing incentives for these funds. $\text{ESG Return}_{p,q-1}$ is the quarterly return of the Sustainalytics ESG factor, which proxies for the profitability of the ESG investment strategy over time.

For the fourth potential determinant of ESG window-dressing, we test this determinant by using an indicator for funds that are near the four-and-five ratings threshold. To construct this variable, we first compute the breakpoints of Morningstar's five globe sustainability ratings

[20] Prior work shows that green assets have outperformed non-green assets in recent years (Pastor et al, 2023).

and each fund's distance from these breakpoints. As we do not observe the internal aggregated ESG scores used to compute ESG ratings, we construct a synthetic internal ranking using the aggregated ESG scores based on the methodology outlined in the 2021 version of the Morningstar sustainability ratings guide. Specifically, we first compute the month-weighted aggregated ESG score over rolling 12-month windows and then run a logistic regression at the month level of an indicator for a higher rating on the weighted ESG score for each ratings group (i.e. the one to two, two to three, three to four, and four to five group), to obtain predicted values for the likelihood of a higher rating based on the weighted ESG score, and we use these predicted values to form the synthetic ranks of the ESG score.[21] We then code a variable, $Near\ Breakpoint_{i,q-m}$, an indicator for whether the fund is within 1% of the four-to-five ratings breakpoint using the synthetic rankings, in the month before the end of the quarter where the ESG β gap is estimated. (See Figure 3 for an illustration of the timing of the variable measurement.)

To control for key fund characteristics in our determinant analysis, we also include as controls in Equation 5, several other variables, namely, the logarithm of net asset value ($\text{Size}_{p,q-1}$), management fee ($\text{Management Fee}_{p,q-1}$), expense ratio ($\text{Expense Ratio}_{p,q-1}$), and the HHI index of portfolio holdings ($\text{Concentration}_{p,q-1}$). As we require the Morningstar globe rating data some of the variables in Equation 5, which is only available from 2018, we implement the regression for the subsample of 2018–2022.

In Table 5, we report the regression results for two models with and without fund fixed effects respectively. Consistent with the hypothesis that fund performance drives incentives for ESG window-dressing, we find that negative return performance is associated with a higher

[21] Morningstar uses the transformed ESG score before October 2019 and the raw ESG risk score after October 2019 to compute the rankings. Following the methodology outlined in the Morningstar sustainability ratings guide, if less than 12 months of historical ESG risk scores or ESG score information are available, we compute the weighted scores based on the most recent history of the ESG risk score after October 2019 (Sustainalytics provides data on these scores as of March 2019) and the most recent history of the transformed ESG score before October 2019.

ESG β gap. This suggests that when fund performance is poor, funds tend to engage in more ESG window-dressing. [22] Moreover, at the fund-level, we find that ESG-labelled funds are also associated with greater ESG β gap, which suggests that these types of funds face greater incentives to engage in ESG window-dressing. At the quarter-level, we also examine the relationship between the aggregate ESG factor returns and the ESG β gap. We find a positive association, suggesting that when the ESG investment strategy yields positive returns, funds tend to engage in more ESG window-dressing, perhaps as a way to obfuscate their stock-picking ability.

In addition, as predicted, we show that ESG window-dressing tends to be more prominent near the four to five breakpoints. Specifically, we find that funds located around these breakpoints exhibit an increase in ESG β gap by 0.029 (which is roughly 33% of the sample standard deviation of the ESG β gap). Thus, we find some evidence that funds near the extreme sustainability globe ratings tend to engage in more ESG window-dressing activity.

Furthermore, we shed further light on the role of ESG ratings in driving ESG window-dressing incentives in our analyses on the consequences of the ESG β gap and ESG ratings. Specifically, when funds engage in ESG window-dressing in order to achieve a higher rating, these funds presumably receive some benefits for doing so. Similar to Agarwal et al (2014), we expect that one potential benefit of funds' ESG window-dressing is attracting fund flows through inflated ESG performance (ratings). To test this idea, we examine (1) whether the ESG β gap increases the fund ratings and (2) whether higher fund ratings attracts fund flows.

Table 6 reports these two sets of analyses. As the Morningstar globe rating data is only available from 2018, we also implement these analyses for the subsample of 2018–2022. In Panel A of Table 6, we show that the ESG β gap is associated with a higher ESG rating, after

[22] While fund flows are not statistically related with the ESG β gap, the relationship is negative and when we remove fund returns from the regression model, the coefficient on the fund flows variable is negative and statistically significant.

controlling for lagged returns, the control variables in Equation 5 and quarter fixed effects. Moreover, in Panel B, we find that funds with a higher ESG rating tend to exhibit higher fund flows in the one-to-four-quarter ahead, after controlling for the past returns and other control variables. Specifically, the estimates suggest that moving from one rating to another, increase future fund flows in the one-, two-, three- and four- quarter ahead by roughly 0.2%.

In addition, we examine whether investors are able to see through window-dressing behavior by penalizing funds that have achieved higher ESG ratings through window-dressing. To study this possibility, we add the ESG β gap variable to our analysis in Panel B of Table 6 and evaluate whether the ESG β gap is negatively associated with fund flows, after controlling for ESG ratings, in Panel C of the same table. Across the one-to-four quarter ahead fund flow horizons, we generally find a weakly positive relationship between the ESG β gap and future fund flows. Thus, we find suggestive evidence that investors do not penalize firms for achieving higher ESG ratings through ESG window-dressing.

Taken together, we find several results that shed light on the key drivers of ESG window-dressing behavior in mutual funds. We find that funds with negative performance and are labelled as ESG, tend to engage in more ESG window-dressing. In addition, ESG window-dressing tends to be more pronounced during periods of positive ESG factor return performance, After controlling for these determinant factors, we further show that the fund flow benefits of achieving higher ESG ratings, is an important incentive for ESG window-dressing. In support of this claim, we show that the ESG β gap tends to be more pronounced around the extreme globe ratings breakpoints. Moreover, we show that the ESG β gap helps funds achieve a higher ESG globe rating, and these ratings also attract fund flows over the one-to-four-quarters ahead.

4.3 Quarter-End Return Patterns in High and Low ESG Firms

In this section, we test our prediction *P2(a)* by analyzing firm return patterns around the mutual funds' portfolio reporting dates. As discussed earlier, we would expect that funds would

increase (decrease) their holdings of high (low) ESG firms right before the end of a quarter. After the portfolio reporting dates, we would expect them to unwind these trades. For stocks with high (low) ESG scores, these trades would lead to positive (negative) returns before the end of the quarter and a reversal afterward.

To test this prediction, we first identify a sample of high and low ESG firms by sorting firms based on their ESG scores in each quarter and creating two subsamples of firms in the top and bottom deciles. We then examine the return patterns for these two subsamples in the five-days before and five days after the calendar quarter-end.[23]

Panel A of Table 7 presents the results for the full sample, showing that, in the five days before the quarter ends, stocks with high ESG scores earn an average abnormal return of 6.4 basis points (bps) and those with the lowest ESG scores earn an average negative abnormal return of 13.5 bps. While the positive return for high ESG stocks is statistically insignificant, the negative abnormal return for the low ESG stocks and the difference between the two (19.8 bps) are both statistically significant. As predicted, the pattern reverses after the quarter ends, where high ESG stocks earn a negative return of 16.1 bps and low ESG stocks earn a positive return of 17.2 bps on average. Both numbers as well as the difference between them (-33.3 bps) are statistically significant.

Panel B of Table 7 shows that the results are mainly driven by the post-ratings period (i.e. 2016–2022), when high ESG stocks outperform low ESG stocks by 26.1 bps in the five days before the holding reporting dates but underperform them by 36.6 bps in the subsequent five days. Panel C of Table 7 reports the results for the pre-ratings period (2011–2015), showing no significant difference in returns between high and low ESG stocks either before or after the fiscal period-ends.

[23] As most mutual funds report portfolio holdings at the calendar quarter-end (53% of our sample), we focus on the calendar quarter-end to study the average effects of mutual fund reshuffling at quarter-end.

To further probe the return patterns, we examine the relationship between the percentile rank of ESG scores and the return patterns in the five days before and after the calendar quarter-end. Specifically, we implement the following return regression at the firm-quarter level:

$$BHAR(-4{,}0\ or\ 1{,}5)_{i,q} = \alpha_j + \alpha_q + \beta_1 ESG\%_{i,q} + \sum \gamma X_{s,i,q-1} + \varepsilon_{i,q}, \quad (6)$$

where $BHAR(-4{,}0\ or\ 1{,}5)_{i,q}$ is the buy-and-hold market-adjusted return in the (-4,0) daily trading window before the quarter-end or the (1,5) daily trading window after the quarter-end for firm *I* in quarter q. The main coefficient of interest is β_1, which estimates the slope between the ESG percentile score ($ESG\%_{i,q}$) and abnormal returns. In certain specifications, we also control for industry (α_j) and quarter fixed effects (α_q) as well as lagged size, book-to-market, asset investment rate, and operating profit. We also cluster standard errors at the firm level.

In Table 8, we report the results of the regression model. In Panel A, we present the regression analysis for the full sample from 2011–2022 for the BHAR (-4,0) window before the quarter-end in the first four columns and the BHAR (1,5) window after the quarter-end in the last four columns. The results in the first four columns of Panel A show that stock returns over the five days before the quarter ends exhibit a significantly positive association with the ESG percentile score across all specifications. The results in the last four columns of Panel A show that the pattern largely reverses in the five days after the fiscal quarter ends. We find that the five-day abnormal returns are negatively associated with ESG percentile score. These results are again consistent with ESG window-dressing, that is, funds buying (selling) stocks with high (low) ESG scores right before the quarter-end reporting dates and unwinding these positions right afterward.

Across Panels B and C of Table 7, we conduct the same return analysis in the periods after and before the introduction of the Morningstar sustainability ratings in March 2016. The results show that that the predictable return patterns are concentrated in the period after March 2016. Specifically, the results in Panel B are largely consistent with and generally stronger than those

in Panel A, showing that stock returns are positively associated with the ESG percentile score over the five days before quarter ends but that the association turns negative in the five days after the quarter ends. Panel C presents the results for the period before March 2016, showing no statistically significant relationship between the ESG scores and the abnormal returns in the period before or after the quarter-end. Thus, our analysis across Tables 7 and 8 provides evidence consistent with ESG window-dressing by mutual funds; that is, funds tend to buy (sell) stocks with high (low) Sustainalytics ESG scores right before the quarter-end portfolio reporting dates and reverse the position soon after, and that the introduction of the Morningstar sustainability ratings has increased their incentives to do this.

4.4 ESG Beta Gap and Quarter-End Returns

In this section, we test our prediction *P2(b)* by studying whether the return patterns that we document above are concentrated among firms that are followed by mutual funds that engage in more ESG window-dressing. To construct the extent of following by ESG window-dressing mutual funds, we first compute the ESG β gap at the fund level, according to the method in Section 3.3. We then aggregate the ESG β gap measure to the firm level by taking the holdings-weighted average of the fund-level ESG β gap, and then we split the sample of firms in each quarter by the median value of the aggregated firm-level ESG β gap.

In Table 9, we present these cross-sectional analyses based on the return regression specification in Equation 7. Panel A presents the cross-sectional results for the full sample period from 2011–2022, and we find consistent evidence that firms that are followed by more mutual funds with higher ESG β gap tend to exhibit a stronger return pattern, which suggests ESG window-dressing. Specifically, we find that in the regression of abnormal returns in the period before (after) the quarter-end, the positive (negative) coefficient on the ESG score percentile variable is only statistically significant in the subsample of firms with an above median aggregate ESG β gap. Moreover, we compare the slope of the relationship between the

ESG score percentile and abnormal returns for firms that are below and above median ESG β gap, and we find that the slope coefficient of the regression of before (after) the quarter-end returns is significantly more positive (negative) for firms that are in the upper median of the ESG β gap. Thus, the evidence suggests that the predictable return patterns documented in Table 8 are concentrated in the subsample of firms with higher ownership by mutual funds with higher ESG β gaps.

Panels B and C of Table 9 further show that the results are mainly driven by the subperiods after the introduction of the Morningstar sustainability ratings in March 2016. Panel B reports the results for the post-ratings period, and we find stronger results in terms of economic magnitude compared to the full-sample results presented in Panel A. Furthermore, results for the pre-ratings period presented in Panel C of Table 9 is generally statistically insignificant. First, coefficients on the ESG percentile score are insignificant across all sub-samples, consistent with weaker incentives of ESG window-dressing in the period. Second, there is no significant differences in the coefficient between the two subgroups partitioned on ESG beta gap when we use BHAR(-4,0) or BHAR(1,5) as the dependent variable. Thus, the results in Table 9 show that the predictable return patterns are driven by transactions done for the purpose of ESG window-dressing.

4.5 Carbon Ratings and Carbon Window-Dressing in Mutual Funds

In April 2018, Morningstar also rolled out a low-carbon designation (LCD) for mutual funds, based on their new fund-level carbon risk score. Much like the Sustainability globe ratings, the LCD increased the importance of achieving higher carbon performance, as funds with LCD tend to attract greater fund flows (Ceccarelli et al, 2024). Thus, we also expect that funds would engage in window-dressing activity on carbon performance after the introduction of the LCD rating, similar to the prediction in *P1(a)*.

To test whether funds engage in more carbon window-dressing after the introduction of the carbon ratings, we examine the same set of analyses as Table 3 but with a new measure for the carbon β gap. To measure the carbon β, we measure the exposure to the carbon factor portfolio, which we estimate as the returns to a short-side position on a portfolio of fossil fuel firms.[24] We construct the returns to this portfolio with the following formula:

$$Carbon\ Factor_t = -\frac{1}{2}(FF_S + FF_B), \tag{7}$$

where FF_S and FF_B are the daily value-weighted returns of fossil fuel firms (from fossil fuel industries defined in Appendix B) that are classified as small or big, respectively. We then construct the actual, holding carbon βs and the carbon β gap, using the same methodology for the ESG β gap, outlined in Section 3.3.

We test our prediction that the average carbon β gap of mutual funds increases after the initiation of Morningstar's carbon ratings by analyzing the changes in the carbon β's and β gap in regression models that hold fund characteristics fixed. Specifically, in Table 10, we present the estimates in regressions, which follows the structure of Equation 4 by analyzing the changes in the carbon β's and β gap, around the 2018Q2 event for the sample from 2016-2022.

Panel A of Table 10, shows that the average actual carbon β measured using the 3 factor and 5 factor model, increases by 0.014-0.019 in the period after the introduction of the Morningstar carbon ratings compared to the period beforehand. Moreover, in Panel B, we also see an increase in the carbon β computed with holding-implied returns of roughly 0.017-0.021 in the post-carbon rating period compared to beforehand. Thus, our analyses across both panels suggest that there is an increasing focus on carbon-related issues, which encourages less exposure to fossil fuel firms.

[24] The LCD requires the fund to have a 12-month trailing average carbon risk score of below 10, and a 12-month trailing average exposure to fossil fuels of below 7% of total assets. As we do not have access to the carbon risk scores, we use the companies defined as fossil fuel companies following Sustainalytics' industry definition to construct exposures to the second criteria of the LCD.

In Panel C of Table 10, we present the main results of this table, which are the changes in the carbon β gap (holding carbon β minus the actual carbon β) around the quarter of the introduction of the Morningstar carbon ratings. Like our findings in Table 3, we find an increase in the carbon β gap across both specifications of the beta gap after the initiation of the Morningstar carbon ratings. Specifically, we find a 0.003 increase in the carbon β gap (or roughly a 15% of the increase in the actual carbon beta), which suggests that carbon window-dressing has increased since the introduction of the Morningstar carbon ratings.

5 Conclusion

We examine whether the introduction of Morningstar's fund ESG ratings induces window-dressing of ESG performance among mutual funds. To address this question, we develop several sets of analyses to detect window-dressing. Our results suggest that, after the launch of Morningstar's Sustainalytics fund ratings, ESG window-dressing becomes more prevalent among mutual funds. Specifically, we find that the difference in ESG exposure of reported portfolio holdings *at quarter-ends* and the ESG exposure of average portfolios *over the quarter* (which we define as the ESG β gap) increases following the introduction of the Morningstar sustainability ratings. Moreover, we find that the ESG β gap tends to be more pronounced following periods with strong ESG factor performance and when funds have poor past return performance, experience outflows, bear explicit ESG labels, and are near the breakpoints of the four-to-five Morningstar sustainability globe ratings. Further building on the fund-flow incentive story, we also find that the ESG β gap tends to increase the ESG rating of funds, which in turn, helps attract fund flows.

Consistent with mutual funds window-dressing by temporarily buying (selling) firms with high (low) ESG scores before quarter-ends, we find that firms with high (low) ESG ratings tend to exhibit higher (lower) returns in the period before the quarter-end and that the patterns

reverse sharply in the opposite direction right after the quarter-end. These patterns are particularly strong after the launch of the Morningstar sustainability ratings and are more pronounced among firms with high ownership by mutual funds with a high ESG β gap.

Lastly, we also show that our hypothesis on sustainability fund ratings as a driver of ESG window-dressing extends to more recently introduced fund ratings on carbon performance. Specifically, we find that the difference in carbon performance exposure of reported portfolio holdings *at* quarter-ends and the carbon performance exposure of average portfolios over the quarter (which we define as the carbon β gap) also increases following the introduction of the Morningstar carbon ratings for mutual funds.

Overall, our findings are likely to raise concerns for practitioners and regulators. Specifically, we find evidence that the initiation of one form of ESG disclosure for mutual funds, fund ESG ratings, increases window-dressing in funds. This finding suggests that investors should exercise caution in assessing the ESG performance of funds through ratings, as these ratings can be manipulated via window-dressing. More broadly, our findings could also be of particular concern to regulators, several of whom have proposed the mandatory disclosure of several new forms of fund ESG disclosures. These disclosures may also bring unintended consequences in the form of window-dressing.

References

Agarwal, V., Gay, G.D., Ling, L., 2014. Window dressing in mutual funds. *The Review of Financial Studies 27*, 3133–3170.

Berg, F., Koelbel, J.F., Rigobon, R., 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance 26*, 1315–1344.

Carhart, M. (1997), On Persistence in Mutual Fund Performance, *The Journal of Finance*, 52(1), 57-82.

Ceccarelli, M., Ramelli, S. and A. F. Wagner (2024), Low Carbon Mutual Funds, *Review of Finance*, 45-74.

Chatterji, A.K., Durand, R., Levine, D.I., Touboul, S., 2016. Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal 37*, 1597–1614.

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

Cornell, B., 2021. ESG preferences, risk and return. *European Financial Management 27*, 12–19.

Di Giuli, A., Kostovetsky, L., 2014. Are red or blue companies more likely to go green? Politics and corporate social responsibility. *Journal of financial economics 111*, 158–180.

Dumitrescu, A., Gil-Bazo, J., Zhou, F., 2022. Defining greenwashing. Available at SSRN 4098411.

European Commission Report, 2021. Study on sustainability-related ratings, data and research. Available at: https://op.europa.eu/en/publication-detail/-/publication/d7d85036-509c-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-183474104

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

Fama E. and K. French (2015), A five-factor asset pricing model, *Journal of Financial Economics*, *116*(1), 1-22.

Gantchev, N., Giannetti, M. and R. Li (2023), Sustainability or Performance? Ratings and Fund Managers' Incentives, *ECGI Working Paper.*

Gibson Brandon, R., Glossner, S., Krueger, P., Matos, P., Steffen, T., 2022. Do responsible investors invest responsibly? *Review of Finance 26*, 1389–1432.

Hartzmark, S.M., Sussman, A.B., 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance 74*, 2789–2837.

Haugen, R.A., Lakonishok, J., 1988. The incredible January Effect: the stock market's unsolved mystery. Dow Jones-Irwin, Homewood, Ill.

He, J., Ng, L., Wang, Q., 2004. Quarterly trading patterns of financial institutions. *The Journal of Business 77*, 493–509.

Hong, H., Kacperczyk, M., 2009. The price of sin: The effects of social norms on markets. *Journal of financial economics 93*, 15–36.

Kacperczyk, M., Sialm, C. and L. Zheng (2008), Unobserved Actions of Mutual Funds, The *Review of Financial Studies*, 21(6), 2379–2416.

Kim, S., Yoon, A., 2023. Analyzing active fund managers' commitment to ESG: Evidence from the United Nations Principles for Responsible Investment. *Management Science 69*, 741–758.

Lakonishok, J., Shleifer, A., Thaler, R.H., Vishny, R.W., 1991. Window dressing by pension fund managers. National Bureau of Economic Research Cambridge, Mass., USA.

Lakonishok, J., Smidt, S., 1986. Volume for winners and losers: Taxation and other motives for stock trading. *The Journal of Finance 41*, 951–974.

Larcker, D.F. and E.M., Watts 2020. Where's the greenium? *Journal of Accounting and Economics 69*, 101312.

Liang, H., Sun, L., Teo, M., 2022. Responsible hedge funds. *Review of Finance 26*, 1585–1633.

MSCI, 2023. MSCI ESG Fund Ratings Methodology, MSCI ESG Research LLC, July 2023.

Munoz, F. Ortiz, C. and L. Vicente (2022), Ethical window dressing: SRI funds are as good as their word, *Financial Research Letters*, *49*.

Musto, D.K., 1997. Portfolio disclosures and year-end price shifts. *The Journal of Finance 52*, 1563–1588.

Ng, L., Wang, Q., 2004. Institutional trading and the turn-of-the-year effect. *Journal of Financial Economics 74*, 343–366.

Pástor, Ľ., Stambaugh, R.F., Taylor, L.A., 2022. Dissecting green returns. *Journal of Financial Economics 146*, 403–424.

Parise, G. and M. Rubin (2024), Green Window Dressing, *Proceedings of the EUROFIDAI-ESSEC Paris December Finance Meeting 2023*

Poterba, J.M., Weisbenner, S.J., 2001. Capital gains tax rules, tax-loss trading, and turn-of-the-year returns. *The Journal of Finance 56*, 353–368.

Raghunandan, A., Rajgopal, S., 2022. Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies 27*, 822–863.

Renneboog, L., Ter Horst, J., Zhang, C., 2008. Socially responsible investments: Institutional aspects, performance, and investor behavior. *Journal of Banking & Finance 32*, 1723–1742.

Ritter, J.R., Chopra, N., 1989. Portfolio rebalancing and the turn-of-the-year effect. The *Journal of Finance 44*, 149–166.

Rzeźnik, A., Hanley, K.W., Pelizzon, L., 2022. Investor reliance on ESG ratings and stock Price performance.

Securities and Exchange Commission (SEC), 2022. Enhanced Disclosures by Certain Investment Advisers and Investment Companies About Environmental, Social, and Governance Investment Practices, Federal Register Vol. 87, No. 117, June 17, 2022.

Serafeim, G., Yoon, A., 2022. Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Review of accounting Studies* 1–31.

Sias, R.W., Starks, L.T., 1997. Institutions and individuals at the turn-of-the-year. The *Journal of Finance 52*, 1543–1562.

Sikes, S.A., 2014. The turn-of-the-year effect and tax-loss-selling by institutional investors. *Journal of Accounting and Economics 57*, 22–42.

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

Starks, L.T., 2023. Presidential Address: Sustainable Finance and ESG Issues—Value versus Values. *The Journal of Finance* 78, 1837-1872.

Statman, M., Glushkov, D., 2009. The wages of social responsibility. *Financial Analysts Journal 65*, 33–46.

Tables and Figures

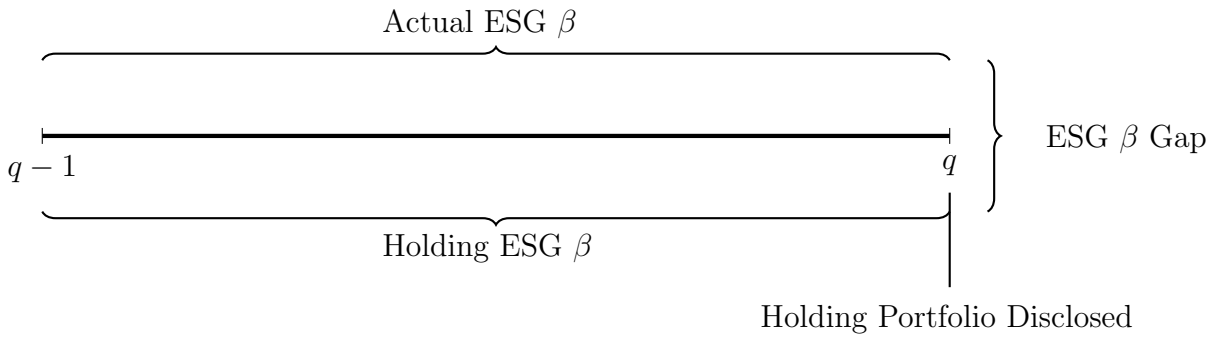
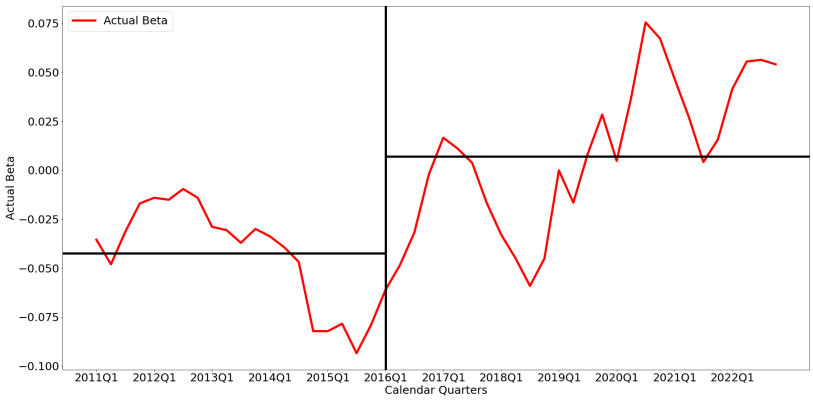
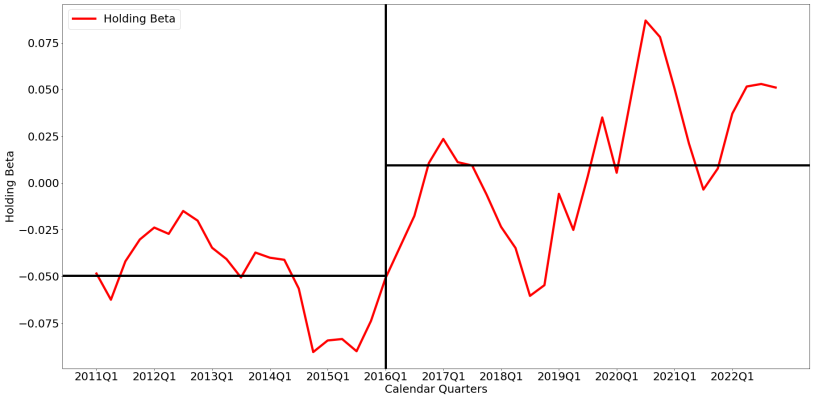


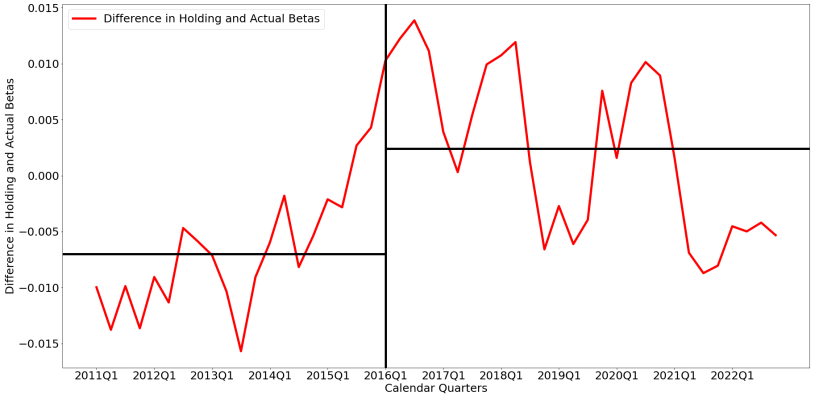
Figure 1: ESG β Gap Measurement Methodology.



(a) Actual Sustainalytics ESG Beta



(b) Holding Sustainalytics ESG Beta



(c) Sustainalytics ESG Beta Gap

Figure 2: Time-series average trends of ESG βs and ESG β Gap. We present the rolling 4-quarter average of ESG β of the actual fund returns in Panel A and the implied mutual fund returns using portfolio holding information in Panel B. We plot the rolling 4-quarter average of ESG β gap (holding minus actual β) in Panel C. The black vertical line denotes the introduction of the Morningstar sustainability ratings in Q1 2016. The horizontal lines denote the average values before and after Q1 2016.

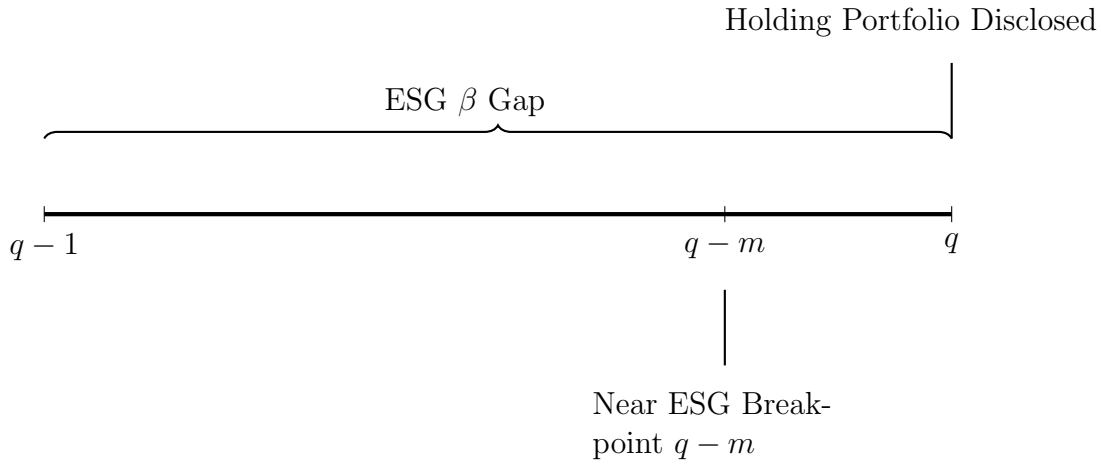


Figure 3: Timeline of ESG β Gap and ESG Ratings.

Table 1: Sample Selection

This table presents the sample formation statistics at the fund-level in Panel A and firm-level in Panel B.

Panel A: Fund-Level

	Number of Funds	Number of Fund-Quarter Observations
CRSP Data from 2011-2022	18,516	554,404
CRSP Data with Fiscal-Year End Dates	13,955	396,312
Domestic and Equity Funds	6,622	179,758
Funds less ETFs and Index Funds	4,656	128,111
Funds with $>$ 5 M. in TNA and $>$ 10 Portfolio Firms	3,477	97,966
Funds with $>$ 2/3 TNA with Sustainalytics Universe Coverage	3,277	94,565
Funds with Sufficient Daily Data for ESG Beta Estimation	3,217	84,094
Funds with Valid Control Observations	3,164	81,273

Panel B: Firm-Level

	Number of Firms	Number of Firm-Quarter Observations
CRSP Data from 2011-2022	14,938	366,776
Data with Sustainalytics Coverage	4,338	91,256
Data with Valid Control Observations	4,173	84,359

Table 2: Summary Statistics

This table reports the summary statistics of the main variables used in our sample. All variables are defined in Appendix A. In Panel A, we present the summary statistics of the fund-level variables, namely, the total net asset value, expense ratio, management fee, fund flows (in percentages), actual returns (in percentages), portfolio concentration, actual ESG beta, holdings-implied ESG beta and ESG beta gap. In Panel B, we present the summary statistics of firm-level variables, namely, market capitalization, book-to-market ratio, investment ratio, operating profit ratio and the sustainalytics ESG score percentile.

Panel A: Fund-Level

	Mean	Standard Deviation	Median	25%	75%	N
Total Net Asset Value (Billion USD)	1.265	2.164	0.319	0.074	1.313	81273
Expense Ratio	0.011	0.003	0.01	0.009	0.013	81273
Management Fee	0.692	0.267	0.713	0.547	0.867	81273
Fund Flows (%)	-0.229	8.111	-1.525	-4.305	1.912	81273
Actual Returns (%)	2.647	8.109	3.122	-0.856	7.021	81273
Portfolio Concentration	0.02	0.018	0.017	0.005	0.029	81273
Actual ESG Beta	-0.014	0.22	-0.006	-0.146	0.12	81273
Holdings-Implied ESG Beta	-0.015	0.231	-0.009	-0.15	0.121	81273
ESG Beta Gap	-0.001	0.086	-0.001	-0.046	0.044	81273

Panel B: Firm-Level

	Mean	Standard Deviation	Median	25%	75%	N
Market Capitalization (Billion USD)	9.458	14.933	3.178	1.01	9.959	84359
Book-to-Market	0.587	0.456	0.471	0.248	0.811	84359
Investment	0.017	0.061	0.009	-0.014	0.037	84359
Operating Profit	0.049	0.077	0.049	0.02	0.081	84359
Sustainalytics ESG Score Percentile	0.501	0.284	0.501	0.252	0.751	84359

Table 3: Introduction of the Morningstar Sustainabiliy Globes and the ESG β Gap

This table reports the ESG β Gap over time. In Panel A, we present the changes in actual ESG β after the introduction of the Morningstar sustainability ratings in 2016. In Panel B, we present the changes in holdings-implied ESG β after the introduction of the Morningstar sustainability ratings in 2016. In Panel C, we present the changes in the ESG β gap, defined as the difference between the holdings-implied and actual ESG β. The holding-implied and actual ESG β are computed as the exposures of the implied or actual fund returns to the ESG factor, which is computed as the difference between the value-weighted high sustainalytics ESG score portfolio and the low sustainalytics ESG score portfolio. We use two risk factor models to compute the ESG β exposures, namely the Fama-French 3-factor model and the Fama-French 5-factor model. All regressions control for fund fixed effects. Additionally, in columns 2 and 4, we control for lagged size, quarterly returns, quarterly fund flows, portfolio concentration, expense ratio, management fees and the quarterly ESG factor portfolio return. Standard errors are clustered at the fund-level, and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: Actual ESG β				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2016Q1}$	0.065***	0.054***	0.060***	0.051***
	(0.003)	(0.003)	(0.003)	(0.004)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	81,273	81,273	81,273	81,273
Adj. R^2	0.2861	0.2926	0.1201	0.1293
Panel B: Holding ESG β				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2016Q1}$	0.070***	0.060***	0.067***	0.058***
	(0.003)	(0.004)	(0.003)	(0.004)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	81,273	81,273	81,273	81,273
Adj. R^2	0.2940	0.3008	0.1266	0.1372
Panel C: ESG β Gap (Holding - Actual)				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2016Q1}$	0.006***	0.008***	0.007***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	81,273	81,273	81,273	81,273
Adj. R^2	0.0772	0.0786	0.0506	0.0518

Table 4: Sustainalytics ESG β Gap Relative to MSCI ESG β Gap

This table reports the ESG β gap computed based on the Sustainalytics ratings relative to the ESG β gap computed based on the MSCI ratings over time. In Panel A, we present the changes in the ESG β gap computed based on the MSCI ratings, after the introduction of the Morningstar sustainability ratings in 2016. In Panel B, we present the changes in ESG β gap computed with the Sustainalytics ratings relative to the same ESG β gap computed with the MSCI ratings, after the introduction of the Morningstar sustainability ratings in 2016. The holding-implied and actual sustainalytics ESG β are computed as the exposures of the implied or actual fund returns to the ESG factor, which is computed as the difference between the value-weighted high sustainalytics ESG score portfolio and the low sustainalytics ESG score portfolio. The holding-implied and actual MSCI ESG β are computed as the exposures of the implied or actual fund returns to the ESG factor, which is computed as the difference between the value-weighted high MSCI ESG score portfolio and the low MSCI ESG score portfolio. We use two risk factor models to compute the ESG β exposures, namely the Fama-French 3-factor model and the Fama-French 5-factor model. All regressions control for fund fixed effects. Additionally, in columns 2 and 4, we control for lagged size, quarterly returns, quarterly fund flows, portfolio concentration, expense ratio, management fees and the quarterly ESG factor portfolio return. Standard errors are clustered at the fund-level, and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: MSCI ESG β Gap (Holding - Actual)				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2016Q1}$	-0.005***	-0.003**	-0.004***	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	81,273	81,273	81,273	81,273
Adj. R^2	0.0432	0.0475	0.0378	0.0415
Panel B: Sustainalytics ESG β Gap Relative to MSCI ESG β Gap				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2016Q1}$	0.011***	0.010***	0.011***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	81,273	81,273	81,273	81,273
Adj. R^2	0.0523	0.0539	0.0425	0.0456

Table 5: Drivers of ESG β Gap

We report the determinants of the ESG β gap in this table. In our analysis, we regress the ESG β gap on several lagged determinant variables, namely, quarterly returns, quarterly net flow, an indicator for ESG-labelled funds, the value-weighted returns of the ESG factor portfolio, an indicator for a fund that has an ESG score that is within $\pm 1\%$ from the aggregated ESG score-implied threshold between four and five globe ratings, logarithm of net asset value (size), management fee, expense ratio and portfolio concentration, for the sample of funds from 2018-2022 (we restrict the sample to 2018-2022, as the ESG rating data is only available after 2018). Standard errors are clustered at the fund-level and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Dependent Variable	ESG β Gap	ESG β Gap
$\text{Returns}_{i,q-1}$	-0.037***	-0.036***
	(0.003)	(0.003)
$\text{Net Flow}_{i,q-1}$	-0.010	-0.008
	(0.007)	(0.008)
ESG Fund_{i}	0.006***	
	(0.002)	
ESG Return_{q-1}	0.089***	0.086***
	(0.008)	(0.008)
$\text{Near Breakpoint}_{i,q-m}$	0.031**	0.029**
	(0.015)	(0.014)
$\text{Log(NAV)}_{i,q_1}$	0.000	-0.004***
	(0.000)	(0.002)
$\text{Management Fee}_{i,q-1}$	-0.002	0.009
	(0.003)	(0.010)
$\text{Expense Ratio}_{i,q-1}$	-0.188	-1.989*
	(0.301)	(1.142)
$\text{Portfolio Concentration}_{i,q-1}$	-0.002	-0.008
	(0.008)	(0.007)
Fund FE	No	Yes
Observations	23,282	23,282
Adj. R^2	0.0100	0.0637

Table 6: ESG β Gap, ESG Ratings and Fund Flows

This table reports the relationship between the ESG β gap and ESG ratings, as well as the relationship between ESG ratings and future fund flows. In Panel A, we examine the relationship between the ESG β gap and the ESG ratings. In Panel B, we examine the relationship between the ESG ratings and one-to-four quarter ahead fund flows. In Panel C, we add the ESG β gap as an independent variable in Panel B. All regressions control for quarter fixed effects, and controls for lagged size, management fee, expense ratio, quarterly returns and portfolio concentration. Standard errors are clustered at the fund-level and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: ESG β Gap and ESG Ratings

Dependent Variable	ESG Rating
ESG Beta Gap$_{i,q}$	0.362**
	(0.158)
Returns$_{i,q-1}$	0.033
	(0.059)
Log(NAV)$_{i,q_1}$	-0.041***
	(0.015)
Management Fee$_{i,q-1}$	0.477***
	(0.130)
Expense Ratio$_{i,q-1}$	-61.024***
	(12.996)
Portfolio Concentration$_{i,q-1}$	0.325
	(0.267)
Quarter FE	Yes
Observations	23,282
Adj. R^2	0.0091

Panel B: ESG Ratings and Fund Flows

Fund Flow Horizon	1 Quarter Ahead	2 Quarter Ahead	3 Quarter Ahead	4 Quarter Ahead
ESG Rating$_{i,q}$	0.002***	0.002***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Returns$_{i,q-1}$	0.067***	0.049***	0.003	0.023***
	(0.005)	(0.005)	(0.006)	(0.005)
Log(NAV)$_{i,q_1}$	-0.005***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Management Fee$_{i,q-1}$	-0.012**	-0.019***	-0.022***	-0.023***
	(0.005)	(0.005)	(0.005)	(0.005)
Expense Ratio$_{i,q-1}$	-1.622***	-1.031**	-0.628	-0.442
	(0.474)	(0.487)	(0.489)	(0.476)
Portfolio Concentration$_{i,q-1}$	-0.008	0.003	0.005	0.007
	(0.007)	(0.008)	(0.008)	(0.009)
Quarter FE	Yes	Yes	Yes	Yes
Observations	22,400	20,661	18,955	17,283
Adj. R^2	0.0474	0.0458	0.0452	0.0484

Panel C: ESG Ratings, ESG β Gap and Fund Flows

Fund Flow Horizon	1 Quarter Ahead	2 Quarter Ahead	3 Quarter Ahead	4 Quarter Ahead
ESG Rating$_{i,q}$	0.002***	0.002***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
ESG Beta Gap$_{i,q}$	0.006	0.015*	0.016*	0.007
	(0.009)	(0.008)	(0.008)	(0.009)
Returns$_{i,q-1}$	0.067***	0.049***	0.003	0.023***
	(0.005)	(0.005)	(0.006)	(0.005)
Log(NAV)$_{i,q_1}$	-0.005***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
Management Fee$_{i,q-1}$	-0.012**	-0.019***	-0.022***	-0.023***
	(0.005)	(0.005)	(0.005)	(0.005)
Expense Ratio$_{i,q-1}$	-1.622***	-1.028**	-0.627	-0.441
	(0.474)	(0.487)	(0.489)	(0.475)
Portfolio Concentration$_{i,q-1}$	-0.008	0.003	0.005	0.007
	(0.007)	(0.008)	(0.008)	(0.009)
Quarter FE	Yes	Yes	Yes	Yes
Observations	22,400	20,661	18,955	17,283
Adj. R^2	0.0474	0.0460	0.0454	0.0484

Table 7: High or Low Sustainalytics ESG Firms and Abnormal Returns

This table reports the 5-day BHARs before and after the quarter end of high and low Sustainalytics ESG rating firms. We define high and low ESG rated firms as the top and bottom decile of ESG scores from Sustainalytics. 5-day BHARs before the quarter-end are computed as the buy-and-hold daily returns over the -4, 0 window relative to the buy-and-hold daily value-weighted market portfolio over the same window. The 5-day BHARs after the quarter-end is computed as the buy-and-hold daily abnormal returns over the 1, 5 window. Panel A, reports the analysis for the full sample. Panels B and C reports the analysis after the introduction of the morningstar sustainability globe ratings and the period before. We drop the first quarter of 2020, due to the market volatility in the first few months of COVID-19. Estimates are reported in percentages and standard errors are clustered at the firm-level, and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: Full Sample (2011-2022)				
Return Inteval	BHAR(-4,0)		BHAR(1,5)	
Sample	High ESG	Low ESG	High ESG	Low ESG
Estimate	0.064	-0.135*	-0.161***	0.172**
	(0.042)	(0.077)	(0.043)	(0.078)
Observations	8,516	8,464	8,516	8,464
Difference High - Low ESG	0.198**		-0.333***	
	(0.087)		(0.089)	
Observations	16980		16980	
Panel B: Post-Globe Ratings (2016-2022)				
Return Inteval	BHAR(-4,0)		BHAR(1,5)	
Sample	High ESG	Low ESG	High ESG	Low ESG
Estimate	0.085*	-0.176*	-0.197***	0.169*
	(0.048)	(0.094)	(0.049)	(0.095)
Observations	6,778	6,746	6,778	6,746
Difference High - Low ESG	0.261**		-0.366***	
	(0.106)		(0.106)	
Observations	13524		13524	
Panel C: Pre-Globe Ratings (2011-2015)				
Return Inteval	BHAR(-4,0)		BHAR(1,5)	
Sample	High ESG	Low ESG	High ESG	Low ESG
Estimate	-0.018	0.028	-0.022	0.184*
	(0.080)	(0.085)	(0.088)	(0.103)
Observations	1,738	1,718	1,738	1,718
Difference High - Low ESG	-0.046		-0.207	
	(0.115)		(0.134)	
Observations	3456		3456	

Table 8: Sustainalytics ESG Scores and Returns Analysis

This table reports the associations between the Sustainalytics ESG score rank percentile and the 5-day BHARs before and after the quarter-end. 5-day BHARs before the quarter-end are computed as the buy-and-hold daily returns over the -4, 0 window relative to the buy-and-hold daily value-weighted market portfolio over the same window. The 5-day BHARs after the quarter-end is computed as the buy-and-hold daily abnormal returns over the 1, 5 window. Panel A, reports the analysis for the full sample. Panel B and C studies the analysis after the introduction of the morningstar sustainability globe ratings and the period before. The set of control variables includes size, book-to-market, investment and operating profit. Estimates are reported in percentages and standard errors are clustered at the firm-level, and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: Full Sample (2011-2022)								
BHAR Window	(-4,0)	(-4,0)	(-4,0)	(-4,0)	(1,5)	(1,5)	(1,5)	(1,5)
ESG $\%_{i,q}$	0.132*	0.170**	0.168**	0.166**	-0.212***	0.064	-0.213***	-0.228***
	(0.071)	(0.067)	(0.076)	(0.076)	(0.070)	(0.064)	(0.078)	(0.078)
$\text{Size}_{i,q-1}$			-0.007	-0.011			0.118***	0.121***
			(0.017)	(0.017)			(0.018)	(0.018)
$\text{BM}_{i,q-1}$			-0.228***	-0.222***			0.142**	0.136**
			(0.056)	(0.056)			(0.065)	(0.065)
$\text{Inv}_{i,q-1}$				0.252				-0.701**
				(0.296)				(0.329)
$\text{OP}_{i,q-1}$				0.225				-0.003
				(0.165)				(0.047)
Quarter FE	Yes	No	No	No	Yes	No	No	No
GICS4-Quarter FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	84,359	84,359	84,359	84,359	84,359	84,359	84,359	84,359
Adj. R^2	0.0178	0.1228	0.1233	0.1234	0.0612	0.1932	0.1938	0.1939
Panel B: Post-Globe Ratings (2016-2022)								
BHAR Window	(-4,0)	(-4,0)	(-4,0)	(-4,0)	(1,5)	(1,5)	(1,5)	(1,5)
ESG $\%_{i,q}$	0.179**	0.249***	0.253**	0.248**	-0.260***	0.109	-0.247**	-0.263***
	(0.085)	(0.084)	(0.098)	(0.099)	(0.083)	(0.080)	(0.099)	(0.099)
$\text{Size}_{i,q-1}$			-0.011	-0.016			0.132***	0.136***
			(0.019)	(0.019)			(0.020)	(0.020)
$\text{BM}_{i,q-1}$			-0.221***	-0.216***			0.135*	0.129*
			(0.060)	(0.060)			(0.070)	(0.070)
$\text{Inv}_{i,q-1}$				0.252				-0.753**
				(0.328)				(0.364)
$\text{OP}_{i,q-1}$				0.283				-0.041
				(0.235)				(0.066)
Quarter FE	Yes	No	No	No	Yes	No	No	No
GICS4-Quarter FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	67,550	67,550	67,550	67,550	67,550	67,550	67,550	67,550
Adj. R^2	0.0184	0.1248	0.1252	0.1253	0.0650	0.1906	0.1913	0.1914
Panel C: Pre-Globe Ratings (2011-2015)								
BHAR Window	(-4,0)	(-4,0)	(-4,0)	(-4,0)	(1,5)	(1,5)	(1,5)	(1,5)
ESG $\%_{i,q}$	-0.060	-0.071	-0.056	-0.049	-0.019	-0.076	-0.100	-0.108
	(0.088)	(0.083)	(0.084)	(0.084)	(0.097)	(0.086)	(0.090)	(0.090)
$\text{Size}_{i,q-1}$			-0.009	-0.011			0.016	0.019
			(0.025)	(0.025)			(0.025)	(0.025)
$\text{BM}_{i,q-1}$			-0.317***	-0.332***			0.253**	0.275***
			(0.087)	(0.087)			(0.103)	(0.105)
$\text{Inv}_{i,q-1}$				0.188				-0.223
				(0.461)				(0.529)
$\text{OP}_{i,q-1}$				-0.146				0.214**
				(0.095)				(0.088)
Quarter FE	Yes	No	No	No	Yes	No	No	No
GICS4-Quarter FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Observations	16,809	16,809	16,809	16,809	16,809	16,809	16,809	16,809
Adj. R^2	0.0108	0.1318	0.1330	0.1330	0.0158	0.2467	0.2471	0.2472

Table 9: Sustainalytics ESG Scores, Returns and Mutual Funds with ESG β Gap

This table reports the associations between the Sustainalytics ESG rank percentile and the 5-day BHARs before and after the quarterend for firms that are followed by mutual funds with low or high ESG β gap. 5-day BHARs before the quarter-end are computed as the buy-and-hold daily returns over the -4, 0 window relative to the buy-and-hold daily value-weighted market portfolio over the same window. The 5-day BHARs after the quarterend is computed as the buy-and-hold daily abnormal returns over the 1, 5 window. We compute the firm-level measure of the ESG β gap by weighting the the ESG β gap of mutual funds by the shares held in the firm at the end of the quarter. We assign firms as low or high ESG β gap if the firm is below or above the quarterly median of the β gap. Panel A, reports the analysis for the full sample. Panel B and C studies the analysis after the introduction of the morningstar sustainability globe ratings and the period before. Estimates are reported in percentages and standard errors are clustered at the firm-level, and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: Full Sample (2011-2022)				
BHAR Window	(-4,0)	(-4,0)	(1,5)	(1,5)
Cross-Section	Below Median ESG β Gap	Above Median ESG β Gap	Below Median ESG β Gap	Above Median ESG β Gap
ESG $\%_{i,q}$	-0.118	0.474***	-0.040	-0.469***
	(0.092)	(0.128)	(0.104)	(0.126)
Controls	Yes	Yes	Yes	Yes
GICS4-Quarter FE	Yes	Yes	Yes	Yes
Observations	45,165	39,194	45,165	39,194
Adj. R^2	0.1400	0.1075	0.1972	0.1934
Difference in β	0.592***		-0.430***	
	(0.155)		(0.164)	
Controls	Yes		Yes	
GICS4-Quarter FE	Yes		Yes	
Observations	84,359		84,359	
Adj. R^2	0.1233		0.1955	
Panel B: Post-Globe Ratings (2016-2022)				
BHAR Window	(-4,0)	(-4,0)	(1,5)	(1,5)
Cross-Section	Below Median ESG β Gap	Above Median ESG β Gap	Below Median ESG β Gap	Above Median ESG β Gap
ESG $\%_{i,q}$	-0.128	0.617***	0.036	-0.591***
	(0.124)	(0.155)	(0.138)	(0.150)
Controls	Yes	Yes	Yes	Yes
GICS4-Quarter FE	Yes	Yes	Yes	Yes
Observations	34,643	32,907	34,643	32,907
Adj. R^2	0.1444	0.1116	0.1966	0.1916
Difference in β	0.745***		-0.627***	
	(0.196)		(0.205)	
Controls	Yes		Yes	
GICS4-Quarter FE	Yes		Yes	
Observations	67,550		67,550	
Adj. R^2	0.1272		0.1943	
Panel C: Pre-Globe Ratings (2011-2015)				
BHAR Window	(-4,0)	(-4,0)	(1,5)	(1,5)
Cross-Section	Below Median ESG β Gap	Above Median ESG β Gap	Below Median ESG β Gap	Above Median ESG β Gap
ESG $\%_{i,q}$	-0.072	-0.101	-0.194	0.008
	(0.107)	(0.146)	(0.120)	(0.166)
Controls	Yes	Yes	Yes	Yes
GICS4-Quarter FE	Yes	Yes	Yes	Yes
Observations	10,522	6,287	10,522	6,287
Adj. R^2	0.1381	0.1478	0.2443	0.2807
Difference in β	-0.029		0.202	
	(0.180)		(0.215)	
Controls	Yes		Yes	
GICS4-Quarter FE	Yes		Yes	
Observations	16,809		16,809	
Adj. R^2	0.1433		0.2611	

Table 10: Introduction of Morningstar Carbon Rating and Carbon β Gap

This table reports the carbon β gap over time. In Panel A, we present the changes in actual carbon β after the introduction of the Morningstar carbon ratings in 2018Q2. In Panel B, we present the changes in holdings-implied carbon β after the introduction of Morningstar carbon ratings in 2018Q2. In Panel C, we present the changes in the carbon β gap, defined as the difference between the holding-implied and actual ESG β. The holding-implied and actual sustainalytics carbon β are computed as the exposures of the implied or actual fund returns to the carbon factor, which is computed as minus one multiplied by the returns to the value-weighted fossil fuel firm portfolio (industry definition presented in Appendix B). We use two risk factor models to compute the ESG β exposures, namely the Fama-French 3-factor model and the Fama-French 5-factor model. All regressions control for fund fixed effects. Additionally, in columns 2 and 4, we control for lagged size, quarterly returns, quarterly fund flows, portfolio concentration, expense ratio, management fees and the quarterly ESG factor portfolio return. Standard errors are clustered at the fund-level, and reported in parentheses. ***, **, * denote significance at the 1%, 5% and 10% level.

Panel A: Actual Carbon β				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2018Q2}$	0.012***	0.014***	0.018***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	46,417	46,417	46,417	46,417
Adj. R^2	0.3552	0.3641	0.2592	0.2655
Panel B: Holding Carbon β				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2018Q2}$	0.015***	0.017***	0.020***	0.021***
	(0.001)	(0.001)	(0.002)	(0.002)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	46,417	46,417	46,417	46,417
Adj. R^2	0.3671	0.3728	0.2781	0.2832
Panel C: Carbon β ***Gap (Holding - Actual)***				
Risk Model	3 Factor Model		5 Factor Model	
$\text{Post}_{q=2018Q2}$	0.002***	0.003***	0.002***	0.003***
	(0.000)	(0.001)	(0.001)	(0.001)
Fund FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Observations	46,417	46,417	46,417	46,417
Adj. R^2	0.1644	0.1680	0.1171	0.1179

Appendix A: Variable Definitions

Variable Name	Variable Description
Firm-Level Variables:	
Sustainalytics ESG Score	Before October 2019, this score is computed as the transformation of the ESG rating: $50 + 10 \times \frac{Score - \mu_{score}}{\sigma_{score}}$, where the averages ($\mu_{score}$) and standard deviations (σ_{score}) are computed at the industry group-month-level. After October 2019, this score is the ESG Risk Rating score multiplied by -1 to align the order of the score with the ESG ratings.
Sustainalytics ESG Score Percentile	The monthly percentile rank of the sustainalytics ESG score described above. Missing values of ESG scores are imputed on a 3-month rolling basis.
Size	Natural log of market capitalization at the fiscal period end.
Book-to-Market (BM)	Book equity ($seqq + txditcq - pstkrq$ in Compustat divided by market capitalization at the fiscal period end.
Operating Profit (OP)	Sales ($saleq$ in Compustat) minus cost of good sold ($cogsq$ in Compustat) minus interest expense ($xintq$ in Compustat) minus SG&A expenditures ($xsgaq$ in Compustat) divided by book equity ($seqq + txditcq - pstkrq$ in Compustat) plus minority interest ($miiq$ in Compustat).
Investment (Inv)	Quarterly change in total assets (atq in Compustat) relative to total assets one quarter prior.
Abnormal BHAR Returns	Buy-and-hold daily returns over either the (-4,0) or (1,5) window around the quarter-end minus the buy-and-hold market portfolio returns over the same return window.
Aggregated Firm-Level ESG β Gap	Holdings-weighted average of the ESG β gap of mutual funds following the firm.
Fund-Level Variables:	
Actual Mutual Fund Returns	Net asset value-weighted average of daily returns across all share classes for each mutual fund. Computed at the fiscal quarterly or daily-level.
Holding Mutual Fund Returns	Value-weighted average daily returns of firms in a mutual fund's disclosed holdings at the end of the fiscal quarter. The returns are computed at the daily-level.

Fund Flows	The net fund flow over the fiscal quarter. The fund flow is computed as $\frac{TNA_q}{TNA_{q-1}} - (1 + R_q)$, where TNA_q is the total net assets at the end of the fiscal quarter, and R_q is the returns over the fiscal quarter.
Actual ESG β	The beta on the ESG portfolio return (β_{ESG}) from a fund-quarter-level regression of daily actual mutual fund returns (defined above) on the Fama-French 5 factors and the ESG portfolio return: $R_{it} = \alpha + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 ESG_t$. The ESG portfolio return is constructed as $ESG_t = \frac{ESG_{H,S}+ESG_{H,B}}{2} - \frac{ESG_{L,S}+ESG_{L,B}}{2}$, where $ESG_{H,S}$ and $ESG_{H,B}$ are the value-weighted returns for the top 30% of firms by ESG score that are in the upper and bottom median of market capitalization, respectively. $ESG_{L,S}$ and $ESG_{L,B}$ are the value-weighted returns for the bottom 30% of firms by ESG score that are in the upper and bottom median of market capitalization, respectively.
Holding ESG β	The beta on the ESG portfolio return (β_{ESG}) from a fund-quarter-level regression of daily holding mutual fund returns (defined above) on the Fama-French 5 factors and the ESG portfolio return: $R_{it} = \alpha + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 ESG_t$.
ESG β Gap	The difference between the holding ESG β and the actual ESG β.
Actual Carbon β	The beta on the ESG portfolio return (β_{ESG}) from a fund-quarter-level regression of daily actual mutual fund returns (defined above) on the Fama-French 5 factors and the ESG portfolio return: $R_{it} = \alpha + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 Carbon_t$. The carbon portfolio return is constructed as $Carbon_t = -\frac{FF_{H,S}+FF_{H,B}}{2}$, where $FF_{L,S}$ and $FF_{L,B}$ are the value-weighted returns for the companies in the industries involved with the extraction and production of fossil fuels (see Appendix B for industry definitions), and that are in the upper and bottom median of market capitalization, respectively.
Holding Carbon β	The beta on the ESG portfolio return (β_{ESG}) from a fund-quarter-level regression of daily holding mutual fund returns (defined above) on the Fama-French 5 factors and the ESG portfolio return: $R_{it} = \alpha + \beta_1 MKTRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 Carbon_t$.
Carbon β Gap	The difference between the holding carbon β and the actual carbon β.

Expense Ratio	The ratio of operating expenses to total investments.
Management Fees	The ratio of management fees to net assets.
Portfolio Concentration	The Herfindahl index of portfolio holdings. Computed as the sum of squared shares of share price $\times$ shareholding relative to total net assets.
Near Breakpoint	An indicator variable that is coded as 1 if the implied morningstar ESG rating is within $\pm 1\%$ of the implied breakpoint between the four-to-five globe ratings. All other observations are coded 0. We compute the breakpoints and the associated probabilities around the breakpoint by running a logistic regression of $Higher\ Rating_{i,q} = \beta ESG\ Score_{i,q}$ for each month and for each rating group (1,2), (2,3), (3,4) and (4,5) and taking the predicted value as the associated probability of achieving higher rating. We select the midpoint probability (i.e. 50%) from the logistic regression as the breakpoint that separates two ratings. The ESG score in the logistic regression is computed as the 12-month rolling weighted average of the holdings-weighted Sustainalytics ESG score defined above, where the weights for the ith month from the current month is $(12 - i)$.
ESG Fund	An indicator variable that is coded as 1 if the fund is in the list of sustainable funds in Morningstar's 2018 *Sustainable Funds U.S. Landscape Report* or in the list of funds in the Sustainable Investment Forum.

Appendix B: Fossil Fuel Industries

Sustainalytics Sub-Industry
Oil & Gas Exploration and Production Coal
Integrated Oil & Gas
Electric Utilities
Oil & Gas Equipment
Oil & Gas Drilling
Multi-Utilities
Oil & Gas Storage and Transportation
Gas Utilities
Independent Power Production and Traders
Oil & Gas Refining and Marketing