Green Window Dressing[†]

Gianpaolo Parise
Tilburg University & CEPR
g.parise@tilburguniversity.edu

Mirco Rubin
EDHEC
mirco.rubin@edhec.edu

This version: 8th July 2024; First version: 28th February 2023

Abstract

This paper establishes that mutual funds strategically time the trades of ESG stocks around disclosure to inflate their sustainability ratings. This claim is supported by four analyses. First, we show that funds' ESG betas increase shortly before disclosure and decrease shortly afterwards. Second, we establish that funds outperform the portfolios they disclose. Third, we document an increase in ESG buys (sells) before (after) disclosure based on imputed fund trades. Fourth, we provide evidence that ESG stock prices temporarily rise before disclosure and decline afterwards. Overall, we document that green window dressing positively impacts fund sustainability ratings, performance, and flows. (*JEL Codes: G11, G23, Q56*)

[†]We thank for valuable comments Martijn Boons, Hamid Boustanifar, Laurent Calvet, Huaizhi Chen, Mathijs van Dijk, Elroy Dimson, Frederic Ducoulombier, Jonas Eriksen, Gianfranco Gianfrate, Marcin Kacperczyk, Oguzhan Karakas, Philipp Krueger, Pedro Matos, Daniel Metzger, Steffen Meyer, Lasse Pedersen, Kim Peijnenburg, Andrea Polo, Stefano Ramelli, Raghu Rau, Luke Taylor, Raman Uppal, Olivier Scaillet, Chia-Yi Yen, David Zerbib, and participants to the 2023 Cornell University ESG investing research conference, the 2023 Eurofidai Paris conference, the 2023 CEPR Caffe Research Seminar Series, and to seminars at Aarhus BSS, EDHEC Business School, ESSEC Paris Business School, Luiss University Rome, Luxemburg University, Tilburg University, and the University of Cologne. Part of this project was conducted while Gianpaolo Parise was a visitor at Copenhagen Business School.

Green Window Dressing

This version: 8th July 2024

Abstract

This paper establishes that mutual funds strategically time the trades of ESG stocks around disclosure to inflate their sustainability ratings. This claim is supported by four analyses. First, we show that funds' ESG betas increase shortly before disclosure and decrease shortly afterwards. Second, we establish that funds outperform the portfolios they disclose. Third, we document an increase in ESG buys (sells) before (after) disclosure based on imputed fund trades. Fourth, we provide evidence that ESG stock prices temporarily rise before disclosure and decline afterwards. Overall, we document that green window dressing positively impacts fund sustainability ratings, performance, and flows. (*JEL Codes: G11, G23, Q56*)

— The goal [of institutional investors] is not to pick environmentally conscious companies; the goal is to track the S&P 500 while having "ESG" in the name. — Matt Levine.[1]

The proliferation of investment vehicles marketed as responsible is among the most prominent trends in asset management. Investing responsibly presents however an inherent tension, as fund managers need to deliver financial returns while adhering to a responsibility mandate. From a theoretical standpoint, imposing an Environmental, Social, and Governance (ESG) constraint into standard portfolio optimization implies a trade-off between ESG and Sharpe ratios (Pedersen, Fitzgibbons, and Pomorski (2021)).

This paper investigates whether fund managers mitigate the trade-off between sustainability and performance by strategically timing the sale and purchase of ESG assets around disclosure dates. Specifically, we seek to understand whether ESG funds hold sustainable stocks when portfolio positions are publicly observable, but tilt allocations toward less sustainable assets when holdings are undisclosed. To that end, we address two questions. First, do managers increase their sustainable asset holdings just before disclosure, only to divest them shortly afterward? Second, does this trading strategy relax ESG constraints post-disclosure, leading to higher sustainability ratings *and* better performance?

The economic incentive for this behavior, which we label *green window dressing*, stems from asymmetric monitoring frequency by investors.[2] Although funds report their performance daily, portfolio holdings need to be publicly disclosed only every three months. As a result, investors cannot monitor whether funds comply with their responsibility mandate

[1]From "ESG Stocks Are Graded on a Curve" Bloomberg Opinion, 11 November 2019.

[2]In the text, we use the term "green" in a broad sense to encompass ESG-focused investments.

between disclosure dates, implying that the constraint to invest in ESG stocks is binding only four times a year. In short, fund managers arguably have a stronger incentive to perform well at all times than to be responsible all the time.

Detecting green window dressing presents however a key empirical challenge. Similar to investors, we do not observe fund portfolios within a quarter. While some funds do disclose portfolio holdings at higher frequencies, voluntary disclosure is equally susceptible to window dressing and afflicted by positive selection bias (i.e., only virtuous funds may have an incentive to disclose). In our paper, we design four different empirical tests to detect green window dressing when daily portfolio holdings cannot be directly observed.

First, we compare individual funds' ESG exposures shortly before and shortly after mandated portfolio disclosure to ESG exposures in periods when fund managers lack the incentive to inflate sustainability ratings. Our findings indicate that ESG mutual funds increase their average ESG beta by 0.12 in the 10 days leading to disclosure, only to revert back to previous levels in the 10 days after disclosure. This increase is accompanied by contemporaneous decreases in exposure to the market and to a long/short pollution factor. Our results are robust to employing different ESG indexes, risk models, and estimation windows. By contrast, we find no evidence of green window dressing for ESG ETFs or when assigning placebo filing dates.

Second, we construct counterfactual fund returns based on disclosed portfolio positions and compare them to actual returns reported by funds on a daily basis, building on the return gap approach proposed by Kacperczyk, Sialm, and Zheng (2008). Our findings reveal that, after disclosure, ESG funds exhibit substantially higher returns (but lower ESG exposure) than portfolios disclosed only days before. This observed outperformance, which we estimate at 25 basis points over a month, aligns with the hypothesis that funds substitute ESG stocks with higher-yielding assets when they are able to conceal it. However, we document that this additional performance is predominantly attributable to an increase in market exposure. Overall, we find no evidence that fund managers can generate alpha through green window dressing alone.

Third, we rely on the trade imputation method developed by Bongaerts, van Brakel, van Dijk, and Huij (2024) to estimate the most probable path through which fund holdings evolve between two disclosure dates. This approach reverse-engineers fund trades by utilizing observable inputs that are available at a higher frequency than portfolio holdings, such as daily fund and stock returns, and monthly TNAs. Based on these imputed trades, we document that, on average, funds purchase ESG stocks amounting to 1% of their portfolio value in the 10 days before disclosure, with a sharp increase on the disclosure date itself. However, these trades are subsequently reversed in the 10 days following disclosure.

Fourth, we analyze daily stock returns around end-of-quarter filings. Our results indicate that ESG stocks generate positive cumulative abnormal returns of 0.20% in the three days before fund portfolio disclosure, but these returns revert shortly after disclosure. This finding suggests that the pre-disclosure increase in stock returns is driven by price pressure from a surge in demand rather than fundamental news. Overall, while each of our empirical tests presents specific strengths and weaknesses, they all consistently indicate that funds engage in green window dressing.

In the second part of the paper, we examine the economic incentive for engaging in green window dressing. Our analysis establishes that a one-standard-deviation increase in pre-disclosure ESG exposure results in a 2.1 percentage point increase in the likelihood of receiving the highest Morningstar sustainability rating of five globes, and a 1.0 percentage point decrease in the probability of receiving the lowest rating of one globe. Furthermore, we find that expensive funds, funds at both extremes of the return distribution, and funds affiliated with signatories to the United Nations Principles for Responsible Investment are more likely to inflate ESG holdings prior to disclosure.

A natural question is whether green window dressing is a desirable strategy for fund investors. From a purely return-maximizing perspective, there is little evidence that responsible funds consistently outperform (see, e.g., Renneboog, Ter Horst, and Zhang (2008a)). Nevertheless, for investors committed to ESG (because of their mandates or as part of a marketing strategy), opting for green window dressers is advantageous, as it helps closing the performance gap with ESG-unconstrained funds while investing in funds that are awarded higher sustainability ratings. In line with this argument, our findings indicate that green window dressing is prevalent among funds that target institutional clients. This implies that institutional investors might be relying on "delegated window dressing" as a way to comply with their ESG mandate while earning returns similar to the market portfolio.

Our paper adds to a growing body of literature on sustainable investing. Recent research establishes that there is a trade-off between being sustainable and delivering performance (Gibson, Glossner, Krueger, Matos, Steffen, et al. (2020); Gantchev, Giannetti, and Li (2024); Orlov, Ramelli, and Wagner (2023); and Ceccarelli, Ramelli, and Wagner (2024)). This cost in terms of performance may explain why U.S. signatories to the United Nations Principles for Responsible Investment do not hold more sustainable assets than the others (Gibson, Glossner, Krueger, Matos, and Steffen (2022) and Kim and Yoon (2022)). Furthermore, several papers explore how sustainable funds perform with respect to non-sustainable ones with mixed findings.[3] Our paper sheds light on a potential explanation for why sustainable funds do not substantially under-perform, as we document that green window dressing helps ESG funds to close the performance gap with funds that do not have an ESG mandate. More broadly, much of what we know about ESG funds is based on regulatory portfolio disclosure. We contribute to the literature by showing that such disclosure is biased and may not reflect the portfolio positions of mutual funds, except on dates when those positions are observable to investors.

A number of papers address similar research questions to ours with mixed findings. Muñoz, Ortiz, and Vicente (2022) compare end-of-month to end-of-quarter fund portfolio holdings. As funds have an obligation to publicly disclose only the latter, and the authors find little difference in ESG holdings between the two, they conclude that window dressing does not occur. However, this empirical approach relies on two implicit assumptions: first, that voluntarily disclosed end-of-month portfolio holdings are not window dressed, and second, that the decision to voluntarily disclose is exogenous to window dressing. We show that neither of these assumptions holds true in our data.[4] Furthermore, Kempf and Osthoff (2008) find no evidence of green window dressing on a period prior to the one we analyze.

[3]E.g., Renneboog, Ter Horst, and Zhang (2008a,b); Gil-Bazo, Ruiz-Verdú, and Santos (2010); El Ghoul and Karoui (2017); López-Arceiz, Bellostas-Pérezgrueso, and Moneva (2018); Dolvin, Fulkerson, and Krukover (2019); and Liang, Sun, and Teo (2022).

[4]Furthermore, Muñoz, Ortiz, and Vicente (2022) base their analysis on Refinitiv ESG ratings, which exhibit relatively low correlation with Morningstar/Sustainalytics ratings (Berg, Koelbel, and Rigobon (2022)) and might be revised retroactively by the rating provider (Berg, Fabisik, and Sautner (2024)). To the extent that fund managers window dress to improve their Morningstar rating, such behavior might not be fully captured by Refinitiv scores.

Consistent with their results, we show that ESG funds did not engage in window dressing before the introduction of Morningstar sustainability ratings. By contrast, two subsequent papers confirm our finding that funds green window dress their portfolios by conducting a subset of our analyses. Specifically, Chen, Chen, and You (2024) analyze the difference between fund returns and the returns on their disclosed portfolios, whereas An, Huang, Lou, Wen, and Xu (2024) use a different imputation method. Notably, our paper is the only one to estimate the impact of green window dressing on fund ratings, performance, and flows, thereby providing direct evidence of fund managers' incentives.

Finally, our findings expand an extensive literature on window dressing by asset managers. Previous research indicates that, ahead of revealing portfolio holdings, fund managers tend to purchase assets that have shown past success while divesting from past losers, particularly in the final quarter of the year (Haugen and Lakonishok (1987), Lakonishok, Shleifer, Thaler, and Vishny (1991), Sias and Starks (1997), Musto (1997), Musto (1999), and Ng and Wang (2004)). Similar to traditional window dressing, green window dressing aims at providing a biased representation of portfolio holdings to attract greater capital inflows. However, our paper highlights two key differences. First, by substituting high-ESG with high-return assets after disclosure, green window dressers earn higher returns than funds that maintain the same asset allocation. By contrast, traditional window dressing is commonly associated with lower realized returns post disclosure (see, e.g., Agarwal, Gay, and Ling (2014)). Second, the economic mechanisms underlying the two behaviors are different. We find that one of the main incentive for green window dressing stems from fund managers' desire to achieve better sustainability ratings, whereas traditional window dressing does not have a beneficial impact on ratings.

I. Empirical design

A. Data

We obtain mutual fund data from CRSP mutual funds (returns, characteristics, portfolio holdings) and Morningstar (ESG fund identifiers, sustainability ratings). We only consider U.S. domestic equity mutual funds for comparability reasons, as disclosure requirements and ESG benchmarks vary between countries. We exclude ETFs and index funds from our main analysis. We obtain CO2 emissions from Trucost (as in Bolton and Kacperczyk (2021)) and stock ESG ratings from Morningstar/Sustainalytics and MSCI.

We define ESG mutual funds those classified as Sustainable Investment Funds by Morningstar based on the information included in their prospectus (*Sustainable Investment - Overall*), similar to Albuquerque, Koskinen, and Santioni (2023). Specifically, Morningstar identifies as Sustainable Investment Funds those that explicitly indicate any kind of sustainability, impact, or ESG strategy in their prospectus or offering documents (Morningstar (2020)). We then augment this set to include funds that have a clear reference to ESG in their name. Specifically, we classify as ESG funds those with names that include one of the following words: "responsible", "green", "environment", "climate", "ESG", "sustainable", "sustainability", "clean", "social", or "SRI" (similar to, e.g., Di Giuli, Garel, Michaely, and Petit-Romec (2024) and He, Kahraman, and Lowry (2023)). In Section IV, we present results for U.S. active equity mutual funds that we do *not* classify as having an ESG mandate according to the criteria outlined above. It is important to note that we do not classify all signatories to the United Nations Principles for Responsible Investment (PRI) as ESG funds, as these include large fund complexes like Fidelity and Blackrock that mostly offer funds without an explicit ESG mandate.

We conduct our analyses on the period March 2016-December 2022, as we conjecture that the incentive to inflate ESG profiles stems from the introduction of Morningstar sustainability ratings in March 2016. We test for the presence of green window dressing on an earlier sample in Section II.A.

We conduct all analyses at the fund level. We aggregate returns on different share classes based on their relative weight at the beginning of each year. Our final sample includes 223 ESG funds and 5,793 non-ESG funds (analyzed in Online Appendix Section C). Table I reports the summary statistics for our sample of ESG funds.

B. Methodology

Our first empirical test seeks to identify how fund ESG exposure changes shortly before and shortly after mandated portfolio disclosure with respect to earlier periods within the fiscal quarter. Mutual funds are required to disclose to the Security and Exchange Commission (SEC) portfolio holdings as of the end of the last day of each month by filing form N-PORT (see https://www.sec.gov/rules/final/2016/33-10231.pdf). The SEC make publicly available only the information reported the third month of each fiscal quarter. However, some funds make portfolio holdings information publicly available at other month-ends as well by sending it to data vendors such as CRSP and Morningstar (see Schwarz and Potter

(2016) for a detailed analysis of fund disclosure).

While funds report their performance daily, market participants only observe snapshots of portfolio positions a minimum of four times per year at fiscal quarter-ends, assuming funds do not voluntarily disclose additional portfolios, and up to 12 times per year if funds voluntarily make portfolio holdings publicly available at every month-end. These disclosed portfolio holdings serve as the primary input for determining sustainability ratings. Morningstar Sustainability Ratings, for example, are based on the combination of portfolio holdings and the ESG profile of each corporate issuer. Notably, while funds need to report their holdings as of the end of the fiscal quarter, this information becomes publicly available to market participants with a delay of up to 60 days (see, e.g., Agarwal, Gay, and Ling (2014)).

To detect green window dressing, we rely on the fact that the incentive to hold ESG assets increases exogenously approaching the date of disclosure. In our first empirical analysis, we seek a structural break in the betas between fund daily returns and the returns on the Morningstar US Sustainability Index, which includes exclusively the U.S. companies with the lowest ESG risk based on the scores from Sustainalytics/Morningstar.[5] We focus on this index as it is based on the same ESG scores employed by Morningstar to assign fund sustainability ratings. We consider eight alternative ESG indexes in Section C of the Online Appendix.[6]

[5] The Morningstar US Sustainability Index is long only and market capitalization-weighted. It includes 50% of the Morningstar US Large-Mid Index by float market capitalization and selects securities with the lowest ESG risk. Ineligible firms include those that are not rated, receive a ESG risk rating of 4 ("high") or 5 ("severe"), are involved in activities related to firearms or controversial weapons, or derive more than 50% of their revenues from Tobacco.

[6] A potential complication for our analysis could arise if ESG rating providers retroactively change scores. However, Berg, Fabisik, and Sautner (2024) only find evidence of systematic retroactive changes to Refinitiv ESG ratings, which we do not utilize in our analysis.

Note that, in principle, green window dressing involves two trades. A trade to greenwash the portfolio before disclosure, and a subsequent trade to reverse the greenwashing after disclosure. In our empirical analysis, we seek to identify these two trades by estimating separately the change in ESG exposure in a period shortly before (respectively, shortly after) mandated disclosure compared to a control period when funds are less likely to strategically alter their ESG exposure. Our methodology requires us to run separate regressions in two symmetric event windows before and after the disclosure date t_e:

$$R_{i,t} = a_{-,i,e} + \beta_{-,i,e}^{MKT} MKT_t + \beta_{-,i,e}^{ESG} ESG_t + \varepsilon_{i,t}, \qquad t \in [t_e - n, t_e - 1], \tag{1}$$

$$R_{i,t} = a_{+,i,e} + \beta_{+,i,e}^{MKT} MKT_t + \beta_{+,i,e}^{ESG} ESG_t + \varepsilon_{i,t}, \qquad t \in [t_e + 2, t_e + n + 1], \tag{2}$$

where $R_{i,t}$ is the daily return of fund i on day t, MKT_t is the daily return on the market, and ESG_t is the daily return on the Morningstar US Sustainability Index. It is important to include MKT_t in our specifications, as ESG indexes have a high correlation with market returns, while we want to measure the part of fund returns that is orthogonal to the market and explained by ESG exposure. In Online Appendix Section C, we present similar results obtained using less parsimonious risk models that add to the ESG index the Fama and French (1993) 3 factors or the Carhart (1997) 4 factors.

We estimate ESG and market betas over windows of n trading days. The first regression is estimated on the interval $t \in [t_e - n, t_e - 1]$, whereas the second regression is estimated on the interval $t \in [t_e + 2, t_e + n + 1]$, where t_e is the disclosure date. We exclude the disclosure date itself, t_e, and the trading day immediately after $(t_e + 1)$ to avoid conflating the effect of green window dressing with last-minute trades to inflate quarter-end portfolio prices, as documented by Carhart, Kaniel, Musto, and Reed (2002).[7] Notably, we face a trade-off in the choice of n, the length of the event window for the estimation. On the one hand, a shorter window aligns with the incentive for fund managers to inflate ESG exposure as close to disclosure as possible, thereby holding the financially sub-optimal "green portfolio" for a shorter duration. On the other, a longer window increases the statistical power of our tests but assumes that funds start window dressing several days before disclosure. As a compromise, we present results for $n = 10$ trading days as our baseline and for $n = 5$ and $n = 15$ days as robustness.

We estimate different (pre- and post-disclosure) ESG and market betas for each fund i and each quarter event e, thereby recognizing that funds have different exposures and the same fund can change exposure over time. We then compare them to the exposures estimated in a pre-event window:

$$R_{i,t} \;=\; a_{0,i,e} \,+\, \beta_{0,i,e}^{MKT}\, MKT_t \,+\, \beta_{0,i,e}^{ESG}\, ESG_t \,+\, \varepsilon_{i,t}, \qquad t \in \mathcal{T}_{0,e}, \tag{3}$$

where $\mathcal{T}_{0,e}$ is the entire second month of the fiscal quarter except the first and last trading day. We consider the entire month rather than a shorter window to increase the precision and stability of our estimates. We find similar results when we consider alternative control periods (we present robustness analyses for the estimation windows in Online Appendix Section C). Our null hypothesis is that funds do not green window dress their portfolios before disclosure, implying that ESG betas remain stable from the pre-event to the event

[7] Our results remain similar if we include in our estimation windows those dates as well.

window or, equivalently, that ESG beta is the same in Eq. (1) and (3) above:

$$H_0: \; \Delta\beta^{ESG}_{-,i,e} \;\equiv\; \beta^{ESG}_{-,i,e} \;-\; \beta^{ESG}_{0,i,e} \;=\; 0, \qquad \text{for all} \;\; i = 1, ..., N \;\; \text{and} \;\; e = 1, ..., E, \tag{4}$$

where N is the total number of funds, and E is the total number of disclosure dates. In essence, our objective is to assess whether funds exhibit statistically equivalent ESG exposure shortly before disclosure as in the month *before* disclosure. Symmetrically, we explore how ESG exposure changes in the days shortly *after* disclosure with respect to the month before disclosure by comparing the ESG betas from Eq. (2) and (3).

Our main test statistic is the average of the differences of the estimated ESG exposures before each disclosure date and that in the month before, computed across all N funds and E event dates, that is:

$$\overline{\Delta\beta}^{ESG}_{-} \;\equiv\; \frac{1}{N \cdot E} \sum_{i=1}^{N} \sum_{e=1}^{E} \Delta\hat{\beta}^{ESG}_{-,i,e}, \tag{5}$$

where $\Delta\hat{\beta}^{ESG}_{-,i,e} \;\equiv\; \hat{\beta}^{ESG}_{-,i,e} \;-\; \hat{\beta}^{ESG}_{0,i,e}$, and $\hat{\beta}^{ESG}_{-,i,e}$ (resp. $\hat{\beta}^{ESG}_{0,i,e}$) is the OLS estimator of $\beta^{ESG}_{-,i,e}$ (resp. $\beta^{ESG}_{0,i,e}$) obtained by estimating regression model (1) (resp. (3)) for each fund i. To improve readability, in the following we use β^{ESG}_{+} and β^{ESG}_{-}, rather than the more accurate notation $\beta^{ESG}_{+,i,e}$ and $\beta^{ESG}_{-,i,e}$, to indicate fund i's exposures to ESG. Notably, our approach yields unbiased estimates even in the presence of omitted variables, as long as their effect is relatively "slow moving."[8] Symmetrically, we define $\overline{\Delta\beta}^{ESG}_{+}$ as the average change in ESG

[8]Note that, in the presence of omitted variables, the conditional expected values of our estimators are $E[\hat{\beta}^{ESG}_{-,i,e}|X] \;=\; \beta^{ESG}_{-,i,e} \;+\; \upsilon_{-,i,e}\,\varphi_{-,i,e}$ and $E[\hat{\beta}^{ESG}_{0,i,e}|X] \;=\; \beta^{ESG}_{0,i,e} \;+\; \upsilon_{0,i,e}\,\varphi_{0,i,e}$, where φ is the impact of the omitted factors on fund returns and υ is the impact of our covariates X (the returns on the market and on the ESG index) on the omitted variable. As a result, our approach yields unbiased estimates $E[\Delta\hat{\beta}^{ESG}_{i,e}|X] \;=\; \beta^{ESG}_{-,i,e} \;-\; \beta^{ESG}_{0,i,e}$ under the assumption that the effects above are invariant for a short enough period, as the omitted variable biases simplify: $\upsilon_{-,i,e}\,\varphi_{-,i,e} \;-\; \upsilon_{0,i,e}\,\varphi_{0,i,e} \;=\; 0$.

exposure in the days *after* disclosure relative to the month before disclosure.

The problem under analysis presents a further empirical challenge, as our approach requires us to compare betas obtained from short estimation windows. Our setting may invalidate inference based on large sample asymptotics for $\overline{\Delta\beta}_{-}^{ESG}$ and $\overline{\Delta\beta}_{+}^{ESG}$. For this reason, we conduct bootstrap inference for our statistic as described in Online Appendix Section A. In essence, we assess the value of $\overline{\Delta\beta}^{ESG}$ in our sample by contrasting it with the counterfactual distribution generated by imposing the null hypothesis that $\beta_{-}^{ESG} = \beta_{0}^{ESG}$ (respectively, that $\beta_{+}^{ESG} = \beta_{0}^{ESG}$) on the bootstrapped data generating process for $R_{i,t}$. In simpler terms, we aim to determine the likelihood, within our empirical context, of estimating a value for $\overline{\Delta\beta}_{-}^{ESG}$ (respectively, $\overline{\Delta\beta}_{+}^{ESG}$) as large as the one empirically observed in the data when its true value is 0.

Different from us, previous papers detect *traditional* window dressing using a variety of other methods including examining fund abnormal selling in the last quarter of the year (Lakonishok, Shleifer, Thaler, and Vishny (1991)) and calculating the spread between fund returns and the proportions of winner and loser stocks disclosed (Agarwal, Gay, and Ling (2014)). However, those empirical approaches have no clear implications for ESG exposures and therefore cannot be directly implemented to address our research question.

II. Green window dressing: the empirical evidence

A. Evidence from factor loadings

Table II relies on the empirical design outlined in Section I to test for the presence of green window dressing. Panel A reports the average change in exposures during the days leading up to the disclosure dates relative to the month before disclosure. Columns 1 to 3 document an increase in ESG betas, ranging from 0.05 in the 15 days before disclosure to 0.15 in the 5 days before disclosure. Notably, the coefficients' magnitude increases monotonically as we restrict the estimation window to the days closer to the disclosure date. This pattern indicates that funds maintain a comparatively lower ESG exposure during the month and then progressively increase it as the disclosure date approaches. Our preferred specification, which relies on an estimation window of 10 days, yields a coefficient of 0.12. This increase almost doubles funds' baseline ESG exposure from the previous month, estimated at 0.16 using Eq. (3). Column 4 of Panel A reports a 0.11 decrease in market loading in the 10 days leading up to disclosure, indicating that funds substitute market exposure with ESG exposure. This reduction corresponds to a 15% decrease compared to the average market exposure of 0.72 estimated in the month before mandated disclosure. Standard errors based on the standard deviation of the average delta betas lead us to consistently reject the null hypothesis that exposures remain stable in the days leading up to disclosure. Tests based on the wild bootstrap, presented in Online Appendix Section A, also comfortably reject the null hypothesis for our main specification based on the 10-day window, thereby mitigating the potential concern that the documented effects are due to estimation noise.

Panel B of Table II documents how fund exposures change shortly *after* disclosure. Columns 1 to 3 report that there is no statistically significant difference in ESG betas between the 5, 10, and 15 trading days following the day after the quarter-end and the month before disclosure. Column 4 also shows that there is no significant difference in market beta. Overall, we find that post-disclosure exposure levels closely align with those estimated on non-mandated disclosure months. Similar to what we find for the period before disclosure, the magnitude of the ESG beta coefficients decreases the more we expand our estimation window to days away from disclosure, suggesting the funds hold progressively less ESG stocks. When considered alongside the results in Panel A, these findings indicate that funds tilt their allocation towards ESG assets before disclosure, when it matters for their ratings, before progressively reverting back to a higher market exposure after disclosure. The ex-post incentive for engaging in this trading behavior is illustrated in Online Appendix Figure A.2, which shows that the market portfolio outperformed portfolios of high-ESG and low-CO2-emission stocks during our sample period. These results indicate that ESG fund managers move in and out of ESG stocks around disclosure to benefit from market exposure while showcasing stock holdings that better align with their ESG mandate. Overall, our tests support the green window dressing hypothesis.

To interpret the beta changes estimated above in relation to actual portfolio changes, Table III reports the share of disclosed holdings invested in high-ESG stocks and the pre-disclosure ESG betas for the funds in our sample, obtained using Eq. (1).[9] We sort funds

[9] *High-ESG share* is defined as the share of funds' disclosed portfolios invested in stocks with Sustainalytics/Morningstar scores corresponding to ESG risk "negligible" or "low."

based on the Morningstar sustainability ratings they receive two months after disclosure, as portfolios are disclosed with a delay up to 60 days and Morningstar may need time to revise its scores. To attenuate the impact of outliers due to the short estimation windows, we focus on median estimates (Panel A) and report average estimates as well for completeness (Panel B). A number of facts emerge from the table. First, 5-globe rated funds hold 28 percentage points more of their portfolios in high-ESG stocks compared to 1-globe rated funds. However, the difference in median pre-disclosure ESG betas is much larger, amounting to 94 percentage points ($= 0.56 + 0.38$).[10] Second, median pre-disclosure ESG betas increase monotonically with the number of globes, indicating that the ESG exposure we estimate in the 10 days before disclosure is highly informative of the sustainability rating funds receive two months later, thereby validating our approach.[11] Third, the gap in ESG exposures for funds with a difference in rating of one globe tends to be larger than our estimate for the average increase in pre-disclosure ESG exposure (which is 0.12, see Table II). This finding implies that the average change in ESG beta moves up funds' ESG rating by one globe or less.

In Online Appendix Section B, we link delta ESG betas and portfolio information to quantify the amount of trading implied by the estimates above and the ensuing trading costs. We calculate that a pre-disclosure change in ESG beta of 0.12 corresponds to a change in portfolio allocation ranging between 2.1 and 3.6 percentage points over a 10-day period. Considering that the annual portfolio turnover for the funds in our sample is 0.53 (see Table I), this change aligns with their average 10-day turnover assuming that trading

[10] Importantly, the average and median factor loadings for funds rated 1 or 2 globes are less precisely estimated, as relatively few ESG funds in our sample receive low sustainability ratings.

[11] Note that, as there is substantial overlap between ESG and market portfolios, a negative ESG beta does not imply that funds are shorting ESG stocks, but that they are under-weighting them relative to the composition of the market portfolio.

is uniformly distributed over time, as $(0.53/252) * 10 = 0.021$. In terms of trading costs, we estimate that green window dressers incur an additional cost that reduces their performance by 3.1 basis points (bps) each quarter. This loss amounts to 91% of the average daily return for the funds in our sample, which is 3.4 bps. In other words, green window dressing erodes slightly less than one day of performance. Overall, we conclude that window dressers increase moderately their trading activity around disclosure, which result in a modest increase in trading costs.

We further validate our empirical approach by conducting three placebo tests:

- *Placebo disclosure dates.* In Column 1 of Online Appendix Table A.1 we assign 5,000 random dates to the funds in our sample and compare ESG exposures in the 10 days before placebo disclosure to the ESG exposures in the month before. With this analysis, we seek to understand whether the documented increase in ESG betas may be generated by trends in the data rather than by funds engaging in green window dressing. For instance, because funds are buying increasingly more ESG assets (Curtis, Fisch, and Robertson (2021)). We find that ESG exposure does not increase before placebo disclosure dates, thereby validating our approach.

- *Before Morningstar ratings.* We replicate the analysis for the period preceding the introduction of Morningstar sustainability ratings. Specifically, we conduct the analysis on ESG funds for the period from January 2010 to February 2016 (Morningstar sustainability ratings were introduced in March 2016). Column 2 of Table A.1 in the Online Appendix documents that, before the introduction of Morningstar sustainability ratings, ESG funds did *not* increase ESG exposure before portfolio disclosure. This result is consistent with the findings of Hartzmark and Sussman (2019), indicating that fund flows only respond to easily understandable, salient ESG scores that did not exist before March 2016. Moreover, this finding allows us to reconcile our results with those of Kempf and Osthoff (2008), who find no evidence of green window dressing in an earlier sample period.

- *ESG ETFs and passive funds.* We replicate our analysis on passive mutual funds and index trackers with an ESG mandate. We do not find increases in ESG exposure before disclosure for this set of funds (see Column 3 of Online Appendix A.1). This finding indicates that the empirical pattern documented above requires active trading by fund managers and, therefore, helps us to dismiss alternative explanations based on trends or mechanical effects.

In Online Appendix Section C, we perform a number of additional robustness and falsification tests. Specifically, we show that our results are robust to controlling for additional risk factors (Online Appendix Table A.2), relying on alternative ESG indexes (Online Appendix Table A.3), considering alternative estimation windows (Online Appendix Table A.4), considering alternative pre-event windows (Online Appendix Table A.5), controlling for expected flows (Online Appendix Table A.6), relying on an approach analogous to Fama and MacBeth (1973) to account for cross-sectional correlation in the residuals as an alternative to the bootstrap (Online Appendix Table A.7), and excluding the last calendar quarter of the year (Online Appendix Table A.8). Furthermore, we find that, shortly before fiscal quarter-ends, funds decrease exposure to a pollution factor based on US-listed firms' tons of CO2 emissions, further supporting the argument that funds temporarily reduce their loadings on profitable but "forbidden" factors before disclosure (Online Appendix Table A.9).

B. Evidence from disclosed portfolios

In this section, we compare the returns realized by funds with the counterfactual returns on their disclosed portfolios. For our analysis, we build on Kacperczyk, Sialm, and Zheng (2008) and define the daily return gap for fund i on day t as:

$$GAP_{i,t} = R_{i,t} - \underbrace{(R^H_{i,t} - EXP_{i,t})}_{R^C_{i,t}}, \tag{6}$$

where $R_{i,t}$ is the (net-of-fees) return realized by fund i on day t, and $R^C_{i,t}$ is the counterfactual return obtained as the difference between the return on a hypothetical buy-and-hold portfolio that invests in the stock positions disclosed at fiscal quarter-ends, $R^H_{i,t}$, and fund fees, $EXP_{i,t}$.[12] Intuitively, a fund that holds its publicly disclosed positions fixed should exhibit a return gap of zero.[13]

Table IV presents the $GAP_{i,t}$ in the days immediately before and after mandated portfolio disclosure. We find an average daily $GAP_{i,t}$ of -1.3 basis points (bps) in the 10 days leading up to disclosure. This result implies that funds underperform the portfolio they are about to disclose. This finding aligns with the hypothesis that funds engage in costly trading in the days leading to disclosure, potentially to window dress their positions. Consequently, they

[12] To obtain daily fees, we assume that annual expenses are uniformly charged throughout the year.

[13] This claim also assumes that fund inflows/outflows do not give rise to a different relative asset composition and that cash and other non-equity assets do not yield different rates over the period considered.

compress realized returns compared to counterfactual returns (as hypothetical portfolios do not incur transaction costs). However, after disclosure, funds outperform the portfolio they have just reported, as $GAP_{i,t}$ becomes increasingly positive (see Figure I). We find an average daily $GAP_{i,t}$ of 1.2 bps in the 10 days after disclosure (Column 2 of Table IV). Over a month, this implies that funds earn a 25 bps higher return than they would have by keeping the same stocks they disclosed (assuming there are 21 trading days in a month), which could help explain why ESG funds do not substantially under-perform funds without an ESG mandate (see, e.g., Renneboog, Ter Horst, and Zhang (2008b)). This finding confirms that fund managers tilt their portfolios towards higher-paying stocks after disclosure.

The result in Column 2 of Table IV indicates that funds outperform their disclosed assets, but do they achieve this by decreasing ESG exposure? To address this question, we separately estimate the ESG betas on realized returns, $R_{i,t}$, and counterfactual returns $R^C_{i,t}$, respectively. While our first empirical test in Section II.A evaluates the difference in ESG betas for the same fund over time (i.e., shortly before disclosure versus the month before), this alternative approach examines the difference in ESG betas in the cross-section. Specifically, it compares the ESG exposure for the same fund over the same days based on realized and counterfactual returns. Consistent with our initial empirical approach, we rely on a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index as a proxy for the ESG portfolio. Column 3 of Table IV shows that, in the 10 days after disclosure, funds exhibit an ESG exposure of 0.12 based on actual returns. However, we estimate an ESG beta of 0.17 on the counterfactual returns over the same period (Column 4). Consequently, funds exhibit a 0.05 *lower* ESG beta than that of the portfolio they have just disclosed. Overall, our results indicate that funds trade-off ESG exposure for performance.[14]

In Figure II, we sort funds into 20 quantiles based on delta ESG betas and plot their average cumulative returns over the following three months. We define delta ESG betas as the difference between the ESG betas estimated on realized and counterfactual returns, with a low value indicating green window dressing. The graphs show that while window dressers outperform in terms of excess returns (Figure II, top panel), this additional performance is primarily achieved by increasing exposure to the market portfolio. In fact, the relationship between delta ESG betas and CAPM alphas is nearly flat (Figure II, bottom panel). In other words, although fund managers can boost their returns and narrow the performance gap with the market portfolio through green window dressing, they cannot generate (CAPM) alpha. We come back to this claim in Section IV.B below.

C. Evidence from imputed trades

In this section, we investigate daily holding changes by ESG funds around disclosure dates using imputed trades. Imputation methods seek to reconstruct how stock holdings evolve between two disclosure dates based on observable quantities such as fund and stock daily returns, end-of-month TNAs, and monthly holdings (when available). Since there are theoretically infinite individual stock holding trajectories compatible with the above inputs,

[14]One potential concern with the return gap analysis is the uncertainty around when funds charge fees to their investors. In our main analysis, we compare realized returns reported by the fund (net of fees) to counterfactual returns for which we assume that fees are uniformly charged across the year. This approach could result in a positive return gap after disclosure not due to window dressing, but because the assumption above might lead us to overestimate fees relative to those charged in reality. To address this concern, we replicate our analysis without subtracting any fees from counterfactual returns. This approach biases our analysis against finding a positive return gap, as fees are subtracted from realized returns but not from counterfactual returns. Nonetheless, all results remain consistent, mitigating the concern that our findings might be driven by how we account for fees (see Online Appendix Table A.11).

different approaches have been proposed to determine the most probable path. These methods typically impose various ex-ante restrictions to how stock holdings can evolve and employ different statistical and computational techniques to choose among the solutions that mimic the observed fund returns while matching the other observables.

In our analysis, we leverage the methodology developed by Bongaerts, van Brakel, van Dijk, and Huij (2024), which uses hierarchical preferences and an iterative method that applies random and adaptive constraints.[15] This method builds on the approach first developed by Farrell (2018) but, rather than using a genetic algorithm, uses lexicographic preferences to rank alternative solutions based on their plausibility, using an Occam's razor approach. With respect to alternative imputation methods proposed in the literature, that of Bongaerts, van Brakel, van Dijk, and Huij (2024) provides the greatest ex-post estimation precision when benchmarked against real trading data while making the fewest assumptions ex-ante. We refer the reader to the original paper for detailed methodological information.[16]

Compared to our first empirical test that relies on changes in fund betas, analyzing imputed trades has both advantages and limitations. Among the advantages, imputed trades allow us to estimate which stocks are traded, the trade dates, and the quantities exchanged. Furthermore, since Bongaerts, van Brakel, van Dijk, and Huij (2024)'s methodology was not specifically developed to detect trading in ESG stocks, it provides a neutral testing ground

[15] We are very grateful to the authors for sharing their data with us.

[16] Alternative methods for trade imputation include those by Farrell (2018) and An, Huang, Lou, Wen, and Xu (2024). Farrell (2018)'s approach uses a greedy genetic algorithm and imposes several restrictions to limit the number of possible solutions, including no round-trip trades, constant cash positions throughout the quarter, and quarterly net position changes divided into four identically sized trades. An, Huang, Lou, Wen, and Xu (2024) use a parametric approach assuming linear changes in holdings within subperiods and evaluate the accuracy of the imputation based on weekly rather than daily stock holding changes. The authors also limit the imputation set to 500 stocks, which can artificially inflate trading volumes when trades involve stocks outside of the selected set, and assume constant cash holdings throughout the quarter.

to bring the green window dressing hypothesis to the data. However, in our research setting, there are some important limitations. First, to keep the problem manageable, imputation methods need to limit the number of possible solutions by either making strong assumptions about how funds trade or by using rules of thumb to choose between alternative solutions.[17] Second, to reduce the number of possible solutions, the imputation assumes that funds only trade intra-quarter the stocks they disclose either at the beginning or at the end of that quarter. This assumption makes this approach unsuitable to detect trades in low-ESG stocks that funds may want to hide from investors at all disclosure dates. More broadly, imputed holding trajectories are sensitive to the constraints imposed ex-ante and to the methodology chosen to solve the problem. As a result, different approaches estimate different trading patterns. Importantly, none of these limitations apply to the analysis based on changes in ESG betas.

To provide further evidence in support of the green window dressing hypothesis, we explore how funds trade ESG stocks in the days around fiscal quarter-ends. Specifically, we define the net buying of ESG stocks by fund i on day t, as

$$Net\ buys_{i,t}^{ESG} = \frac{\sum_j \left(Buys_{i,j,t}^{ESG} - Sells_{i,j,t}^{ESG}\right)}{Ptfvalue_{i,t-1}}, \tag{7}$$

where $Buys_{i,j,t}^{ESG}$ ($Sells_{i,j,t}^{ESG}$) is the dollar amount purchased (sold) by fund i of ESG stock j on day t, and $Ptfvalue_{i,t-1}$ is the total dollar amount of the fund portfolio at the end of the

[17] Bongaerts, van Brakel, van Dijk, and Huij (2024)'s algorithm selects the solution that minimizes the number of intra-quarter round-trip trades and the amount of cash held by the fund. Furthermore, it shrinks very small but non-zero estimated trade sizes to zero. Notably, choosing the solution that involves the lowest number of intra-quarter round-trip trades could bias our analysis against finding evidence of green window dressing. This is because a way to engage in green window dressing is to sell ESG assets at the beginning of the quarter and buy them back before the quarter ends.

previous day. To be consistent with our previous analyses, we define as ESG stocks those eligible for inclusion in the Morningstar US Sustainability Index.[18] We then estimate the following dynamic event-study specification:

$$Net\ buys^{ESG}_{i,t} \ = \ \gamma_{i \times m} \ + \sum_{\tau=-10}^{+10} \beta_\tau \ \mathbf{1_{i,t}} \ + \ \epsilon_{i,t}, \tag{8}$$

where $\mathbf{1_{i,t}}$ is a vector of indicator functions taking a value of 1 when day t corresponds to event date $\tau \in [-10, +10]$ for fund i, and the term $\gamma_{i\times m}$ represents a battery of fund-by-month fixed effects that account for funds' average net purchases of ESG stocks during the month. We retain in the sample dates from 20 days before to 20 days after mandated disclosure dates. Note that, as we do not include indicator variables for the event dates from -20 to -11 and from $+11$ to $+20$, those dates serve as baseline. The coefficient β_τ therefore measures the additional net purchase of ESG stocks on event date τ relative to dates further away from disclosure.

Figure III illustrates fund net buying of ESG stocks around disclosure dates. We find that purchases of ESG stocks begin to increase approximately eight days before the quarter ends, ranging from 0.06% to 0.13% of fund portfolio values in the five days before the disclosure date. We also document a sharp increase in net buying on the disclosure date itself.[19] Overall, if we sum the pre-disclosure coefficients, the total increase in ESG holdings amounts to 1.0% of the fund portfolio value. This rise in ESG stocks trading is consistent with the evidence in

[18]That is, stocks of firms with "negligible," "low," or "average" ESG risk that do not sell firearms and do not earn most of their revenues from tobacco products. We present similar results for the stocks in the top tercile of MSCI ESG ratings in Online Appendix Figure A.4.

[19]If we define ESG stocks based on MSCI ESG ratings, funds begin purchasing ESG stocks already 10 days before the disclosure date (see Online Appendix Figure A.4).

Albuquerque, Koskinen, and Santioni (2023) that ESG-prospectus funds increased trading in ES stocks during the Covid crisis, potentially to greenwash their portfolios. However, trading directions reverse shortly after the disclosure date, with fund managers progressively divesting ESG stocks. The net daily buying after disclosure fluctuates around -0.08%, leading to total net sales of ESG stocks amounting to 0.60% of the portfolio value over the first 10 days of the quarter. Considering that the average annual turnover in our sample is 0.53 (see Table I), the quantities traded for window dressing in the 10 days before disclosure are about half the average 10-day fund turnover (1.0% *vs.* $(0.53/252) * 10 = 2.1\%$). Notably, this estimate is slightly below the one implied by the change in ESG betas, consistent with the fact that this imputation method minimizes the number of intra-quarter round-trip trades (which could be important to green window dressing, as funds may sell and buy back the same ESG assets within the same quarter). In Online Appendix Section B, we further extend the analysis on imputed trades to provide an alternative estimation of the trading costs associated with green window dressing. Overall, the results from this section corroborate those from our previous tests.

D. Evidence from stock returns

To investigate the asset pricing implications of window dressing, we examine the returns on ESG stocks around portfolio disclosure dates. If money managers trade ESG assets in large quantities to inflate their sustainability ratings, we should observe systematic patterns in stock returns induced by price pressure. In the following, we estimate abnormal returns around portfolio disclosure for ESG stocks (defined as in the previous section, i.e., stocks eligible for inclusion in the Morningstar US Sustainability Index). We conduct our event study around four dates each year: March 31st, June 30th, September 30th, and December 31st. This is because, for a majority of funds in our sample, fiscal quarters coincide with calendar quarters. In cases where a disclosure date falls on a holiday, we use the last trading day of the same month. We only consider stocks disclosed by the funds in our sample rather than all US stocks. To calculate abnormal stock returns, we estimate a market model in a 100-day window that ends 50 days before each disclosure event. Following the literature, we consider a short $[-3, +3]$-day window around portfolio filings to minimize the impact of unrelated events.

Figure IV shows the abnormal returns an investor would earn by buying an equally weighted portfolio of ESG stocks on event date $t = -3$ and holding it until event date $t = +3$. We document large positive abnormal returns at event dates $t = -1$ and $t = 0$. This pattern is consistent with the hypothesis that, in the aggregate, funds bid up ESG stock prices just before portfolio disclosure. Furthermore, we find a negative effect immediately after disclosure, as the cumulative abnormal return progressively reverts to zero. This finding suggests that the positive abnormal returns before disclosure are driven by price pressure that reverts when funds stop buying, due to arbitrage forces. Another possibility is that funds start to immediately liquidate ESG stocks after sending portfolio information to the regulator, thereby exerting negative price pressure. Overall, these price patterns are consistent with the hypothesis that institutional investors engage in green window dressing.

Importantly, the price patterns from this event study do not need to match our findings at the fund level, as what we document here is the result of aggregate market dynamics, whereas in the rest of the paper we focus on ESG funds only. In fact, the end of the calendar quarters correspond to the filing date for all institutional investment managers (Form 13F). Other investors such as non-ESG mutual funds, hedge funds, and pension funds may have a similar incentive to engage in green window dressing.

III. Economic rationale for green window dressing

In this section, we explore the rationale behind green window dressing. Importantly, window dressing stock holdings may prompt legal actions if it results in fund ESG holdings falling below the minimum threshold required by law for an extended period of time or if the fund's allocation is inconsistent with what is specified in the fund prospectus.[20] To evaluate asset managers' incentives to engage in green window dressing and risk legal or reputational repercussions, we investigate which funds are more likely to window dress (Section A), the effect on fund sustainability ratings (Section B), and the impact on fund flows (Section C).

A. Heterogeneous green window dressing

To understand which funds have stronger incentive to window dress, we explore which fund characteristics correlate with large future pre-disclosure increases in ESG exposure. Specifically, we define a fund as a green window dresser ($Window\ dresser_{t+1}$) if its $\Delta\hat{\beta}_{-}^{ESG}$ in quarter

[20]Specifically, the September 2023 amendments to rule 35d-1 under the Investment Company Act (informally known as the "Names Rule") require funds with a name suggesting a particular investment strategy to have at least 80% of the fund value invested in assets in line with that strategy (see www.sec.gov/files/rules/final/2023/33-11238.pdf and Fisch and Robertson (2023)). For example, an ESG fund should hold 80% of its portfolio in ESG assets. However, the amended rule does not specify which assets qualify as "ESG."

$t+1$, estimated as described in Section I, ranks in the top 10% of the sample distribution.

Table V reports our findings. Column 1 shows that more expensive funds are more likely to inflate their ESG profile. Specifically, a one-percentage-point increase in the fund expense ratio corresponds to a 9.1 percentage point higher probability of window dressing. This finding is consistent with the notion that expensive funds are more likely to engage in opportunistic behavior, as documented by, e.g., Gaspar, Massa, and Matos (2006) and Eisele, Nefedova, Parise, and Peijnenburg (2020).

Column 2 shows that both the best and the worst performing funds are more likely to window dress: star funds by 4.2 percentage points and laggard funds by twice as much (9.0 percentage points). As these correlations are not causal, we cannot determine whether it is window dressing that causes a fund to rank high (or low) against its peers or it is the achieved performance that leads the manager to window dress. In light of previous findings in the literature, it is plausible that laggard funds window dress in an attempt to improve their ranking (Brown, Harlow, and Starks (1996); Kempf and Ruenzi (2008); Cutura, Parise, and Schrimpf (2023)), whereas some funds reach the status of "stars" also thanks to opportunistic behaviors (Nanda, Wang, and Zheng (2004)).

We examine the effect of fund and family size in Column 3. We find a negative coefficient for fund size, which is however not statistically significant. This (lack of) result may appear surprising when considering that larger funds incur greater costs to trade. However, there are two opposing forces to consider. First, larger funds tend to hold more ESG stocks, which we find are cheaper to trade (see Online Appendix Section B and Online Appendix Figure

A.6).[21] Second, the relationship between ESG and size is arguably endogenous, as window dressers attract greater investor flows, thereby growing more rapidly. When considering fund family size, we find that funds affiliated with large fund families are 2.6 percentage points less likely to window dress. This finding is consistent with the hypothesis that larger fund families have greater reputational capital at stake. Although green window dressing does not necessarily violate funds' mandates, inflating fund ESG profiles ahead of disclosure may raise ethical concerns and have negative repercussions on the asset management's reputation. Our results indicate that larger asset management firms may take this into consideration and refrain from window dressing to any large extent.

Column 4 shows that funds that do not voluntarily disclose stock holdings are 2.0 percentage points more likely to window dress, indicating that the decision of voluntarily disclosing portfolio information is endogenous with respect to the choice of window dressing. This finding also suggests that funds that choose to report infrequently, may maintain an overall lower ESG exposure for most of the time, thereby window dressing their portfolios to a larger extent around mandated disclosure.

In Column 5, we regress $Window\ dresser$ on a dummy variable that takes a value of 1 if the fund management company is among the signatories to the United Nations Principles for Responsible Investment (PRI). We find that signatories are more likely to window dress, consistent with the argument that members have stronger incentive to "look green" and in line with previous research that documents that PRI funds engage in greenwashing (Kim and Yoon (2022)).

[21] This result is in line with the finding by Busse, Chordia, Jiang, and Tang (2021) that larger funds realize lower percentage transaction costs than smaller funds.

Finally, Columns 6 and 7 show a negative correlation between *Window dresser* and *Retail investors*, a dummy variable that takes a value of one if 50% or more of a fund's beginning-of-year assets are managed on behalf of retail investors. At a first look, this result is perhaps surprising since retail investors are generally perceived as less sophisticated and easier to deceive. Yet, it is possible that institutional investors actively choose to delegate their money to green window dressers, as that allows them to elude investment restrictions while still claiming to be environmentally conscious. This, in turn, may lead mutual funds to window dress to cater to their institutional clients' preferences. We come back to this hypothesis in Section C below.

B. Green window dressing and Morningstar globes

In this section, we document that window dressing has a positive impact on Morningstar sustainability ratings. Starting from March 2016, Morningstar began assigning sustainability ratings to mutual funds. While there are other providers of such ratings, Morningstar stands out as the most widely followed, with its influence on investor decisions well-documented in the literature (Hartzmark and Sussman (2019)) and verified in our data in Online Appendix Section C (see Online Appendix Table A.10). The most salient among Morningstar's ratings is a discrete score, ranging from one globe (indicating the lowest sustainability) to five globes (representing the highest sustainability).[22]

Disclosed portfolio holdings are relevant for the ratings for two reasons. First, ratings are

[22]Note that Gantchev, Giannetti, and Li (2024) document that most investors rely less on globes as they learn that these ratings do not predict performance. However, as our focus is on ESG funds, globes remain highly valuable because they provide a third-party assessment of whether a fund complies with its ESG mandate.

based on the sustainability profile of the assets in the fund's portfolio, meaning that a fund is deemed sustainable if it invests in sustainable companies. Second, Morningstar retrieves asset holdings mainly from regulatory filings and portfolio holdings voluntarily provided by the asset managers.

In our analysis, we regress future sustainability ratings on a fund's increase in ESG exposure before disclosure ($\Delta\hat{\beta}^{ESG}_{-}$), which we standardize to ease the interpretation of the results. We follow existing literature and focus on ratings issued two months after the end of the fiscal quarter (see, e.g., Agarwal, Gay, and Ling (2014)) because portfolio holdings are publicly disclosed with up to a 60-day delay, and Morningstar may need time to update its scores. In our regressions, we estimate both between- and within-fund effects by excluding (including) fund fixed effects. This empirical choice allows us to understand whether window dressers earn higher ratings compared to peer funds, and whether the same fund is awarded comparatively higher ratings than average when it increases its ESG exposure.

Table VI establishes that green window dressing positively impacts future fund sustainability ratings. In the analysis, we focus on globe ratings of five and one, as Hartzmark and Sussman (2019) find that those scores lead to large investor inflows and outflows, respectively. Column 1 reports that a one-standard-deviation increase in $\Delta\hat{\beta}^{ESG}_{-}$ leads to a significant 2.1 percentage point increase in the probability of receiving a 5-globe rating (1.5 percentage points when we include fund fixed effects, see Column 2). In Columns 3 and 4 of Table VI, we examine the impact of green window dressing on the likelihood of receiving a rating of one globe (the worst possible rating in Morningstar). Our results indicate that a one-standard-deviation increase in $\Delta\hat{\beta}^{ESG}_{-}$ reduces the probability of receiving the lowest sustainability rating of one globe by one percentage point (Column 3). By contrast, we find no statistically significant effect when we include fund fixed effects (Column 4). This lack of statistical significance may be attributed to the small number of ESG funds receiving the worst sustainability rating in our sample, which implies that our test may lack statistical power. Importantly, while this section establishes Morningstar fund ratings as a plausible channel through which fund managers are incentivized to engage in green window dressing, it does not preclude the existence of other channels. For instance, fund managers may window dress because their portfolios have to appear compliant with their investment mandate or because institutional clients evaluate the ESG profile of each individual stock in their portfolio.

C. Delegated window dressing

We examine how investors respond to green window dressing by estimating the effect of the change in pre-disclosure ESG exposure on cumulative flows for the three months after disclosure ($t+1$ to $t+3$). Column 1 of Table VII shows that investor flows chase window dressers. A one-standard-deviation increase in $\Delta\hat{\beta}_{-}^{ESG}$ increases future 3-month flows by 0.65 percentage points. Considered together with the results in the previous sections, our findings indicate that green window dressing improves sustainability ratings and increases returns. These effects, in turn, prompt investors to allocate more money to these funds.

An interesting question is why investors chase window dressers. One possibility is that they genuinely believe that these fund managers deliver superior performance while adhering to their responsibility mandates. Another possibility is that they select these funds exactly because they window dress. If investors that need to be compliant with a responsibility mandate seek money managers that hold CAPM-optimal portfolios *and* receive high sustainability ratings, delegating to ESG manipulators may be optimal. We cannot disentangle these two hypotheses directly, as we do not observe why investors make their choices. However, it is reasonable to assume that the selection process of retail investors is more behavioral and prone to deception, whereas that of institutional investors is driven by rational (but possibly opportunistic) considerations. While it is entirely possible that institutional investors select green window dressers because they are themselves "fooled," they are arguably less likely to be deceived than retail investors.

Columns 2 and 3 report the effect on flows for funds that target institutional and retail investors, respectively. We define a fund as retail if it manages 50% or more of its assets on behalf of retail investors based on fund share class information. We find an effect stronger than the baseline for institutional funds (0.71 percentage points) and a positive but statistically insignificant effect for retail funds. The fact that we do not find an effect for funds whose shares are mostly offered to retail investors can be explained by previous findings in the literature that individuals select responsible funds because of social preferences or for social signaling and are willing to forgo financial performance (Riedl and Smeets (2017)). In turn, if retail investors do not select ESG funds mainly on performance, the economic incentive for fund managers to window dress is comparatively weaker as they can simply hold lower-earning, high-ESG stocks all the time. This argument is consistent with our previous finding that funds are less likely to window dress if they target retail investors (see Table V).

Overall, our results indicate that institutional investor money chases green window dressers. Our evidence is consistent with two possible explanations. Either institutional clients actively seek window dressers to earn higher performance while complying with their responsibility mandate or window dressing leads to distortions in their desired portfolio allocation.

IV. Extensions

A. Analysis at monthly frequency

Do funds only window dress portfolios at mandated disclosure dates or every time they disclose? Although the SEC only publicly discloses fund holdings at fiscal quarter-ends, several funds report stock holdings to data vendors as of the end of each month. ESG rating providers consider these voluntarily disclosed holdings as well when determining fund ratings, thereby creating a similar incentive for fund managers to engage in green window dressing. In the following, we estimate the average increase in ESG exposure in the 10 days leading up to the end of the month relative to the 10 days prior. We use this reference window, rather than the ESG exposure in the month before as in our baseline specification, because the previous month can be a mandated disclosure month, potentially biasing the analysis. In Column 1 of Table VIII we report the average change in ESG exposure before month-ends. We find an increase in ESG exposure of 0.09, statistically significant at the 1% level. To better understand why funds engage in green window dressing at month-ends, we replicate the analysis separating month-ends that coincide with mandated disclosure dates (Column 2) from all other month-ends (Column 3). Note that the former coincides with the sample we consider in our main analysis and, therefore, gives us similar results.[23] Our results indicate that the magnitude of the effect is 38% higher at fiscal quarter-ends than at other month-ends (0.11 versus 0.08). We further separate non-mandated disclosure months into months in which funds (voluntarily) disclose portfolio holdings, based on whether we can retrieve their stock holdings through data providers (Column 4), and months when funds do not make their portfolio positions publicly available in any form (Column 5). We find that the increase in ESG exposure when funds voluntarily disclose their positions is almost identical to when they are forced to disclose. By contrast, we find no increase in ESG exposure before the end of months when funds do not disclose. Overall, these findings suggest that fund managers tend to either disclose window-dressed portfolio positions or not disclose. This last result is consistent with the argument that, if funds do not disclose, ESG rating providers cannot update their sustainability ratings, which, in turn, eliminates the incentive for managers to inflate their ESG holdings in those months. Our choice of focusing primarily on fiscal quarter-ends is driven by the fact that considering voluntary disclosure dates as well would introduce endogenous selection into the analysis, as funds can decide whether to disclose or not.

[23]The only difference is the reference estimation window which, in our baseline specification, is the entire previous month except the first and last trading day.

B. Portfolio sorts

Green window dressing allows funds to hold more profitable (although less sustainable) assets following disclosure.[24] We rely on a portfolio approach to complement the evidence on the effect of window dressing on performance that we presented in the context of the return gap analysis (see Section II.B). To reduce the noise in the estimation of delta betas, we take the annual average of the four end-of-quarter pre-disclosure delta beta estimates, $\Delta\hat{\beta}_{-}^{ESG}$. We then sort funds into three tercile portfolios and calculate the buy-and-hold monthly return on these annually rebalanced portfolios.

Online Appendix Table A.12 reports calendar-month regressions of tercile portfolios' excess returns on the market (Panel A) and on a 5-factor model that includes as explanatory variables the Fama and French (1993) mimicking portfolios, the Carhart (1997) momentum factor, and the Pástor and Stambaugh (2003) liquidity factor. There are two main takeaways from this analysis. First, although funds that increase pre-disclosure ESG exposure deliver 8 basis-point higher risk-adjusted returns, this difference is not statistically significant. Second, green window dressers exhibit a statistically and economically higher loading on the market factor. This result is in line with our findings from the return gap analysis: Overall, when they can conceal it, window dressers tilt asset allocations away from the ESG portfolio and towards the CAPM-optimal market portfolio. However, this strategy does not allow them

[24] The evidence from the literature on the performance of sustainable stocks is mixed. On the one hand, a number of papers find that ESG characteristics are associated with superior performance because market participants fail to correctly price intangibles (Edmans (2011)), because of increasing environmental concerns (Pástor, Stambaugh, and Taylor (2022)), or because money managers and wealthy investors bid up ESG asset prices (van der Beck (2023); Bansal, Wu, and Yaron (2022)). On the other, a stream of papers documents that polluting and sin stocks pay higher returns (Hong and Kacperczyk (2009); Bolton and Kacperczyk (2021); Bolton and Kacperczyk (2023); and Hsu, Li, and Tsou (2023)) and that expected returns on green stocks are lower (Pástor, Stambaugh, and Taylor (2022)).

to generate (CAPM) alpha.

C. Non-ESG funds

We conduct our main analyses on ESG funds, as these funds have a clear incentive to look green. However, there are arguments that support the hypothesis that non-ESG funds may behave in a similar way. Hartzmark and Sussman (2019) find that investors in non-ESG funds reward (penalize) with extra inflows (outflows) funds that are awarded high (low) Morningstar sustainability ratings. This empirical fact suggests that non-ESG fund managers may have an analogous incentive to deceive investors into believing that they are more sustainable than they really are. In this section, we test whether non-ESG mutual funds increase ESG exposure before mandated portfolio disclosure. Specifically, we replicate our main analysis on all U.S. active equity mutual funds that do *not* have an ESG mandate.

Results reported in Online Appendix Table A.13 indicate that non-ESG mutual funds also display an increase in ESG exposure before disclosure. However, the magnitude of the effect is less than half. Overall, these results suggest that some non-ESG funds also engage in green window dressing, albeit to a lesser extent than their ESG counterparts.

V. Conclusions

ESG fund managers are assigned two conflicting objectives: to deliver performance *and* to invest responsibly. While investors monitor how fund managers fare along the first dimension daily and from unbiased performance metrics, they tend to evaluate funds' responsibility through sustainability ratings. These ratings are based on granular portfolio holdings that must be publicly disclosed four times a year. However, portfolio disclosure is only informative as long as managers disclose portfolio holdings that are representative. If managers move into and out of responsible portfolios to time regulatory filings, sustainability ratings might be uninformative.

In this paper, we establish that money managers engage in "green window dressing." We document that funds move in and out of ESG stocks around disclosure to inflate sustainability ratings. We support this claim with four separate sets of analyses. First, we analyze how daily fund returns load on ESG indexes around portfolio disclosure. Although exposure to ESG is constant when placebo disclosure dates are allocated randomly, we find a sharp increase shortly before funds report their holdings, and a decrease shortly afterwards. Second, we compare realized fund returns with returns on disclosed portfolios, establishing that the former are higher—but have a lower loading on ESG—than the latter. Third, we examine imputed fund trades and document that funds purchase ESG stocks in the days leading up to fiscal quarter-ends, only to sell them at the beginning of the next quarter. Fourth, we document that ESG stocks outperform in the days before disclosure. This pattern is however completely reversed after disclosure, validating an explanation based on price pressure from investors' timing of the regulatory filings. Overall, each of our empirical tests has distinct strengths and weakness but all indicate that funds engage in green window dressing.

In the second part of the paper, we explore the economic rationale for timing ESG trades. We find that expensive funds, as well as star and laggard funds are more likely to engage in green window dressing. Additionally, we document that increases in pre-disclosure ESG exposure predict higher sustainability ratings two months later. In turn, green window dressers end up attracting substantially higher capital flows. This last result holds only for institutional clients, which is consistent with the argument that institutional investors delegate green window dressing to ESG mutual funds.

References

Abdi, Farshid, and Angelo Ranaldo, 2017, A simple estimation of bid-ask spreads from daily close, high, and low prices, *The Review of Financial Studies* 30, 4437–4480.

Agarwal, Vikas, Gerald D Gay, and Leng Ling, 2014, Window dressing in mutual funds, *The Review of Financial Studies* 27, 3133–3170.

Albuquerque, Rui A, Yrjo Koskinen, and Raffaele Santioni, 2023, Mutual fund trading and ESG clientele during the COVID-19 stock market crash, *Unpublished Working Paper*.

An, Li, Shiyang Huang, Dong Lou, Xudong Wen, and Mingxin Xu, 2024, Inferring mutual fund intra-quarter trading An application to ESG window dressing, *Unpublished Working Paper*.

Bansal, Ravi, Di Wu, and Amir Yaron, 2022, Socially responsible investing in good and bad times, *The Review of Financial Studies* 35, 2067–2099.

Barras, L., P. Gagliardini, and O. Scaillet, 2022, Skill, scale, and value creation in the mutual fund industry, *The Journal of Finance* 77, 601–638.

Berg, Florian, Kornelia Fabisik, and Zacharias Sautner, 2024, Is history repeating itself? The (un) predictable past of ESG ratings, *Unpublished Working Paper* pp. 1–59.

Berg, Florian, Julian F Koelbel, and Roberto Rigobon, 2022, Aggregate confusion: The divergence of ESG ratings, *Review of Finance* 26, 1315–1344.

Bolton, Patrick, and Marcin Kacperczyk, 2021, Do investors care about carbon risk?, *Journal of Financial Economics* 142, 517–549.

———, 2023, Global pricing of carbon-transition risk, *The Journal of Finance* 78, 3677–3754.

Bongaerts, Dion, Jean-Paul van Brakel, Mathijs A van Dijk, and Joop Huij, 2024, Imputing mutual fund trades, *Unpublished Working Paper*.

Brown, Keith C, W Van Harlow, and Laura T Starks, 1996, Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry, *The Journal of Finance* 51, 85–110.

Busse, Jeffrey A, Tarun Chordia, Lei Jiang, and Yuehua Tang, 2021, Transaction costs, portfolio characteristics, and mutual fund performance, *Management Science* 67, 1227–1248.

Calonico, Sebastian, Matias D Cattaneo, and Max H Farrell, 2020, Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs, *The Econometrics Journal* 23, 192–210.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, 2014, Robust nonparametric confidence intervals for regression-discontinuity designs, *Econometrica* 82, 2295–2326.

——— , 2015, Optimal data-driven regression discontinuity plots, *Journal of the American Statistical Association* 110, 1753–1769.

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

——— , Ron Kaniel, David K Musto, and Adam V Reed, 2002, Leaning for the tape: Evidence of gaming behavior in equity mutual funds, *The Journal of Finance* 57, 661–693.

Ceccarelli, Marco, Stefano Ramelli, and Alexander F Wagner, 2024, Low carbon mutual funds, *Review of Finance* 28, 45–74.

Chen, Kevin CW, Wilbur X Chen, and Haifeng You, 2024, Do sustainability ratings induce ESG window-dressing by mutual funds?, *Unpublished Working Paper.*

Curtis, Quinn, Jill Fisch, and Adriana Z Robertson, 2021, Do ESG mutual funds deliver on their promises?, *Michigan Law Review* 120, 393.

Cutura, Jannic, Gianpaolo Parise, and Andreas Schrimpf, 2023, Debt de-risking, *Management Science, Forthcoming.*

Davidson, R., and J. G. MacKinnon, 1999, The size distortion of bootstrap tests, *Econometric Theory* 15, 361–376.

——— , 2004, *Econometric Theory and Methods* (Oxford UniversityPress).

Di Giuli, Alberta, Alexandre Garel, Roni Michaely, and Arthur Petit-Romec, 2024, Climate change and mutual fund voting on environmental proposals, *European Corporate Governance Institute–Finance Working Paper*.

Dolvin, Steven, Jon Fulkerson, and Anna Krukover, 2019, Do "good guys" finish last? The relationship between morningstar sustainability ratings and mutual fund performance, *The Journal of Investing* 28, 77–91.

Edmans, Alex, 2011, Does the stock market fully value intangibles? Employee satisfaction and equity prices, *Journal of Financial Economics* 101, 621–640.

Eisele, Alexander, Tamara Nefedova, Gianpaolo Parise, and Kim Peijnenburg, 2020, Trading out of sight: An analysis of cross-trading in mutual fund families, *Journal of Financial Economics* 135, 359–378.

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

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

———, 2010, Luck versus skill in the cross-section of mutual fund returns, *The Journal of Finance* 65, 1915–1947.

Fama, Eugene F, and James D MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Farrell, Michael, 2018, Read between the filings: Daily mutual fund holdings and liquidity provision, *Unpublished Working Paper*.

Fisch, Jill R., and Adriana Z. Robertson, 2023, What's in a name? ESG mutual funds and the SEC's Names Rule, *Law Working Paper 697/2023*.

Frino, Alex, Andrew Lepone, and Brad Wong, 2009, Derivative use, fund flows and investment manager performance, *Journal of Banking & Finance* 33, 925–933.

Gantchev, Nickolay, Mariassunta Giannetti, and Rachel Li, 2024, Sustainability or performance? Ratings and fund managers' incentives, *Journal of Financial Economics* 155, 103831.

Gaspar, José-Miguel, Massimo Massa, and Pedro Matos, 2005, Shareholder investment horizons and the market for corporate control, *Journal of Financial Economics* 76, 135–165.

Gaspar, Jose-Miguel, Massimo Massa, and Pedro Matos, 2006, Favoritism in mutual fund families? Evidence on strategic cross-fund subsidization, *The Journal of Finance* 61, 73–104.

Gibson, Rajna, Simon Glossner, Philipp Krueger, Pedro Matos, and Tom Steffen, 2022, Do responsible investors invest responsibly?, *Review of Finance* 26, 1389–1432.

——— , et al., 2020, Responsible institutional investing around the world, Discussion paper Swiss Finance Institute.

Gil-Bazo, Javier, Pablo Ruiz-Verdú, and André AP Santos, 2010, The performance of socially responsible mutual funds: The role of fees and management companies, *Journal of Business Ethics* 94, 243–263.

Hahn, Jinyong, and Whitney Newey, 2004, Jackknife and analytical bias reduction for non-linear panel models, *Econometrica* 72, 1295–1319.

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

Harvey, C. R., and Y. Liu, 2017, Luck vs. skill and factor selection, *The Fama Portfolio: Selected Papers of Eugene F. Fama* pp. 250–260.

——— , 2022, Luck versus skill in the cross section of mutual fund returns: Reexamining the evidence, *The Journal of Finance* 77, 1921–1966.

Haugen, Robert A, and Josef Lakonishok, 1987, *The incredible January effect: The stock market's unsolved mystery* (Irwin Professional Pub).

He, Yazhou Ellen, Bige Kahraman, and Michelle Lowry, 2023, ES risks and shareholder voice, *The Review of Financial Studies* 36, 4824–4863.

Hong, Harrison, and Marcin Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hounyo, U., and J. Lin, 2023, Can mutual fund "stars" really pick stocks? New evidence from a wild bootstrap analysis, *Working Paper*.

Hsu, Po-hsuan, Kai Li, and Chi-yang Tsou, 2023, The pollution premium, *The Journal of Finance* 78, 1343–1392.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved actions of mutual funds, *The Review of Financial Studies* 21, 2379–2416.

Kempf, Alexander, and Peer Osthoff, 2008, SRI funds: Nomen est omen, *Journal of Business Finance & Accounting* 35, 1276–1294.

Kempf, Alexander, and Stefan Ruenzi, 2008, Tournaments in mutual-fund families, *The Review of Financial Studies* 21, 1013–1036.

Kim, Soohun, and Aaron Yoon, 2022, Analyzing active fund managers' commitment to ESG: Evidence from the United Nations Principles for Responsible Investment, *Management Science*.

Lakonishok, Josef, Andrei Shleifer, Richard H Thaler, and Robert W Vishny, 1991, Window dressing by pension fund managers, *The American Economic Review* 81, 227–231.

Liang, Hao, Lin Sun, and Melvyn Teo, 2022, Responsible hedge funds, *Review of Finance* 26, 1585–1633.

López-Arceiz, Francisco José, Ana José Bellostas-Pérezgrueso, and José Mariano Moneva, 2018, Evaluation of the cultural environment's impact on the performance of the socially responsible investment funds, *Journal of Business Ethics* 150, 259–278.

Morningstar, 2020, Leveraging morningstar sustainability data, *Morningstar document.*

———, 2021, Morningstar sustainability rating methodology, *Morningstar document.*

Muñoz, Fernando, Cristina Ortiz, and Luis Vicente, 2022, Ethical window dressing: SRI funds are as good as their word, *Finance Research Letters* 49, 103109.

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

——— , 1999, Investment decisions depend on portfolio disclosures, *The Journal of Finance* 54, 935–952.

Nanda, Vikram, Z Jay Wang, and Lu Zheng, 2004, Family values and the star phenomenon: Strategies of mutual fund families, *The Review of Financial Studies* 17, 667–698.

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

Orlov, Vitaly, Stefano Ramelli, and Alexander F Wagner, 2023, Revealed beliefs about responsible investing: Evidence from mutual fund managers, *Swiss Finance Institute Research Paper*.

Pástor, L'uboš, and Robert F Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

——— , and Lucian A Taylor, 2022, Dissecting green returns, *Journal of Financial Economics* 146, 403–424.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski, 2021, Responsible investing: The ESG-efficient frontier, *Journal of Financial Economics* 142, 572–597.

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

———, 2008b, Socially responsible investments: Institutional aspects, performance, and investor behavior, *Journal of Banking & Finance* 32, 1723–1742.

Riedl, Arno, and Paul Smeets, 2017, Why do investors hold socially responsible mutual funds?, *The Journal of Finance* 72, 2505–2550.

Rohleder, Martin, Dominik Schulte, and Marco Wilkens, 2017, Management of flow risk in mutual funds, *Review of Quantitative Finance and Accounting* 48, 31–56.

Schwarz, Christopher G, and Mark E Potter, 2016, Revisiting mutual fund portfolio disclosure, *The Review of Financial Studies* 29, 3519–3544.

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

van der Beck, Philippe, 2023, Flow-driven ESG returns, *Swiss Finance Institute Research Paper*.

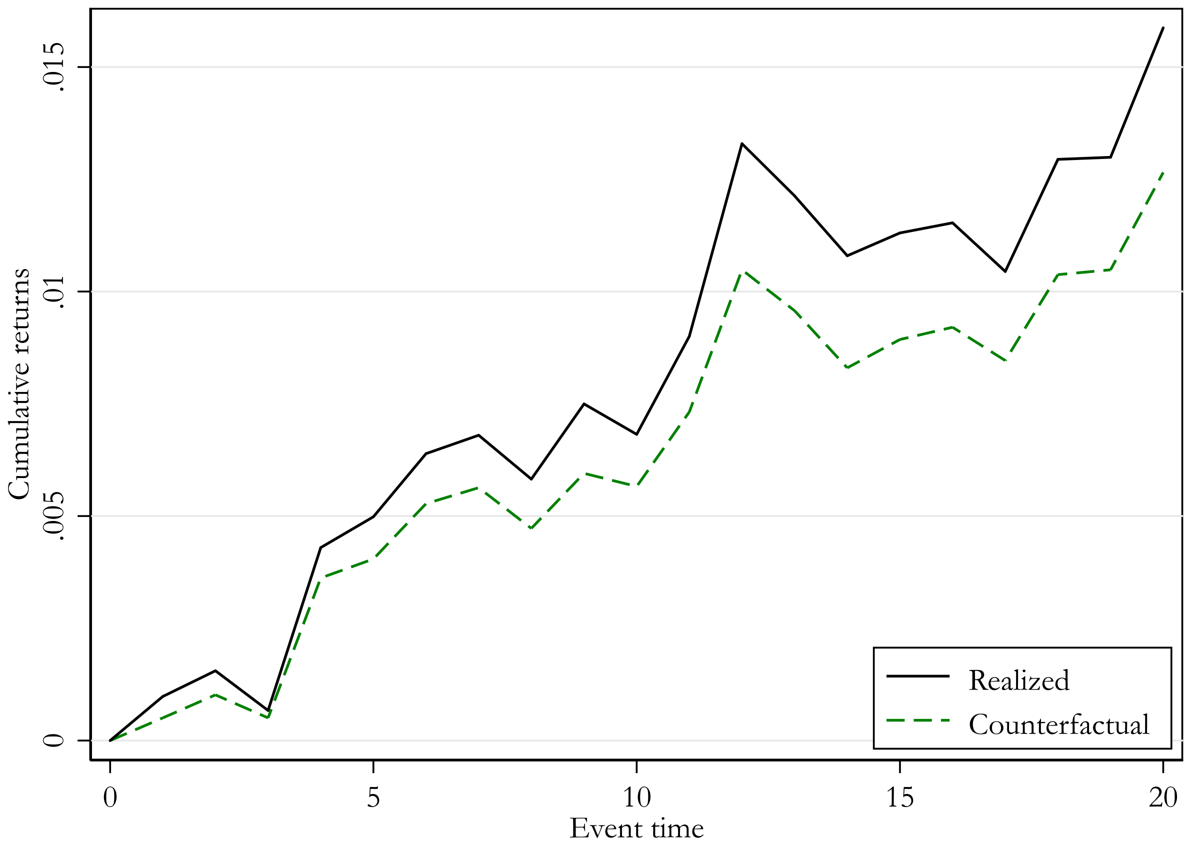
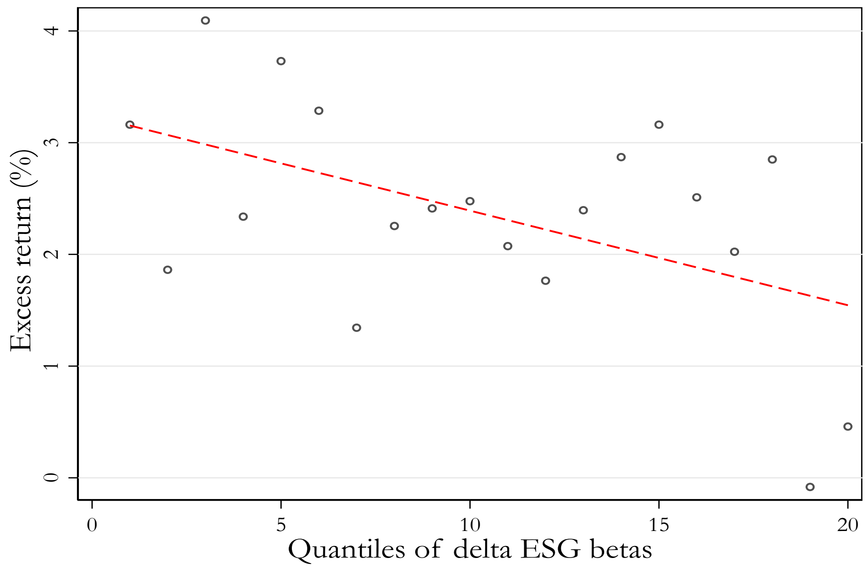


Figure I: Realized versus counterfactual fund returns after mandated disclosure

This figure shows realized and counterfactual cumulative returns for the 20 days after mandated portfolio disclosure. Realized returns are daily returns as reported by funds. Counterfactual returns are daily returns on disclosed portfolios. We construct counterfactual returns by combining the stock holdings disclosed by the funds at fiscal quarter-ends (event time 0) and the subsequent daily stock returns.



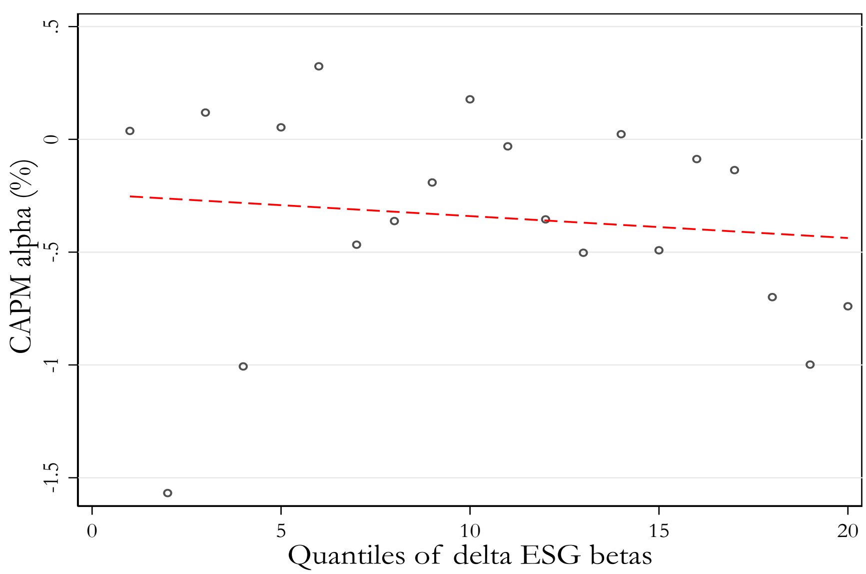


Figure II: Window dressing and fund performance

This figure shows the relationship between delta ESG betas and the subsequent 3-month fund excess returns (top panel) and CAPM alphas (bottom panel). CAPM alphas are obtained by subtracting from the fund excess return its beta times the excess return on the market. We define delta ESG betas as $\beta_R^{ESG} - \beta_C^{ESG}$, i.e., the spread between the ESG exposure estimated on realized and counterfactual returns, respectively. ESG exposures are estimated on 10-day windows starting the day after the first trading day of the fiscal quarter using a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. Counterfactual returns are based on disclosed portfolio holdings. We sort observations into 20 quantiles from the lowest to the highest delta ESG betas; lower delta ESG betas correspond to greater levels of green window dressing.

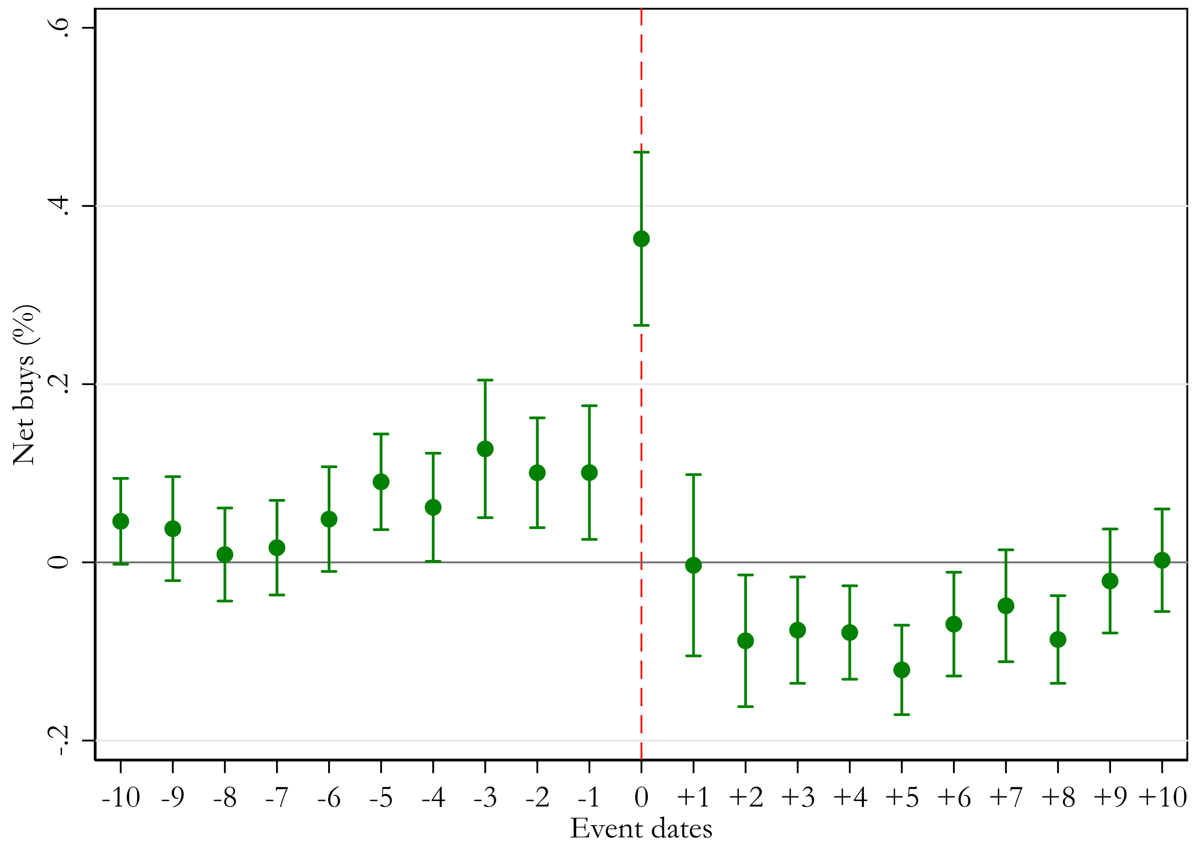


Figure III: Fund trading of ESG stocks around disclosure dates

This figure reports event-study estimates for fund net purchases of ESG stocks in the days around disclosure dates. Coefficients report the effect from 10 days before (event date $\tau = -10$) to 10 days after (event date $\tau = +10$) the disclosure date. Trades are imputed using the methodology by Bongaerts, van Brakel, van Dijk, and Huij (2024). The regressions include fund-by-month fixed effects, and standard errors are clustered at the fund level.

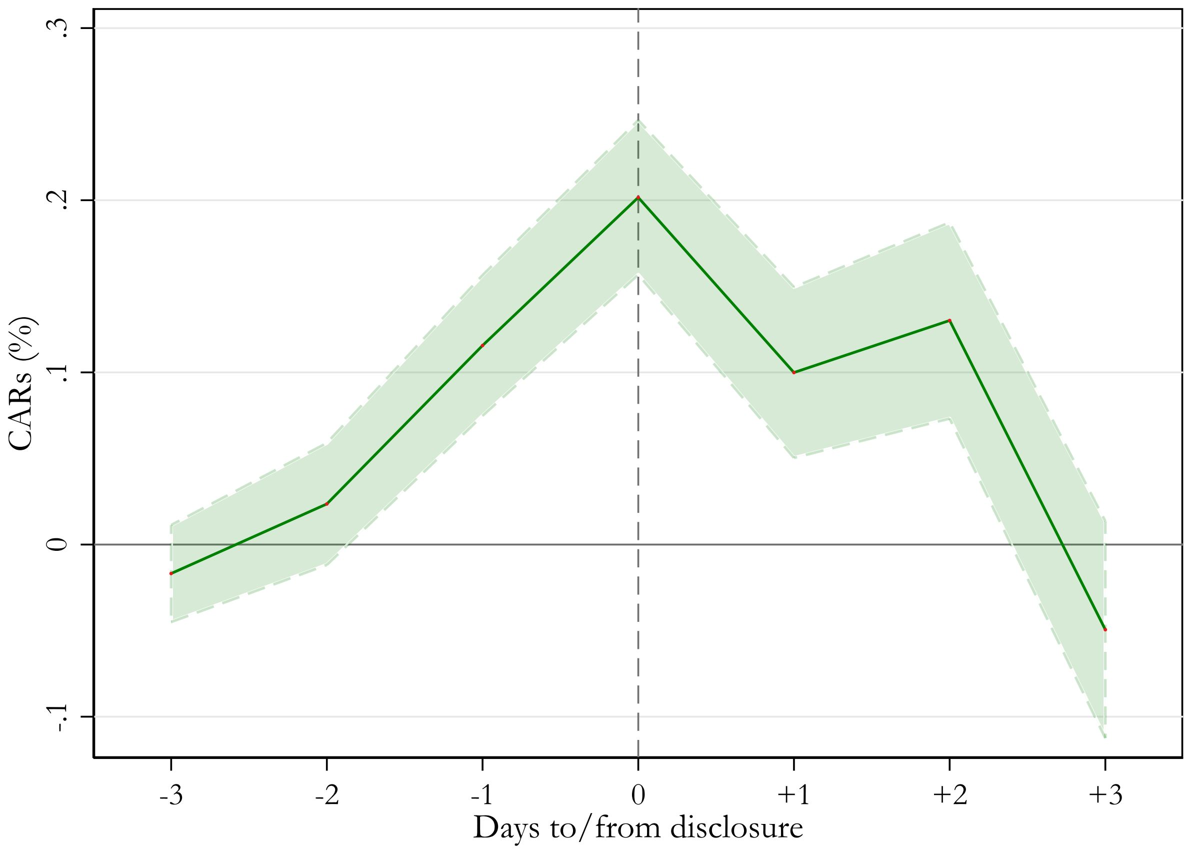


Figure IV: ESG stocks' CARs around disclosure dates

This figure shows average daily cumulative abnormal returns (CARs) in a [−3,+3]-day window around the last trading day (event time = 0) of each calendar quarter for ESG stocks. Abnormal stock returns are from a market model in which the market loadings are estimated on a 100-day window ending 50 days before each event date.

Table I: Summary statistics

This table reports summary statistics for the sample of U.S.-domiciled ESG active equity mutual funds over the period 2016-2022. *TNA* is the fund size in dollar millions, *Fees* is the fund annual expense ratio, *Turnover* is the fund annual turnover from CRSP, *Excess return* is the fund 3-month return in excess of the risk-free rate, *CAPM alpha* is the fund 3-month risk-adjusted return estimated using the Capital Asset Pricing Model, *FF alpha* is the fund 3-month risk-adjusted return estimated using the Fama and French (1993) 3-factor model, *Net flows* are the 3-month cumulative fund flows. *PRI* is a dummy variable that takes a value of 1 if the fund management company is among the signatories to the United Nations Principles for Responsible Investment. *Retail investors* is a dummy that takes a value of 1 if 50% or more of the fund's assets at the beginning of the year are managed on behalf of retail investors. *Morningstar globes* is the Morningstar sustainability rating from 1 (lowest) to 5 (highest).

	Mean (1)	SD (2)	1st (3)	25th (4)	50th (5)	75th (6)	99th (7)
TNA ($ million)	668.864	1,969.003	1.200	30.300	140.500	519.700	7,215.900
Fees (%)	0.920	0.363	0.180	0.751	0.952	1.112	1.823
Turnover	0.532	0.777	0.040	0.200	0.370	0.650	2.210
Excess return (%)	2.121	8.792	-21.420	-2.158	2.900	7.117	23.269
CAPM alpha (%)	-0.426	3.949	-11.641	-2.198	-0.265	1.340	11.727
FF alpha (%)	-0.492	3.859	-11.761	-2.153	-0.330	1.119	12.719
PRI	0.492	0.500	0.000	0.000	0.000	1.000	1.000
Net flows (%)	2.505	14.068	-33.371	-2.898	-0.166	4.496	79.524
Retail investors	0.374	0.484	0.000	0.000	0.000	1.000	1.000
Morningstar globes	4.034	1.074	1.000	3.000	4.000	5.000	5.000

Table II: Do ESG funds engage in green window dressing?

This table presents estimates for the average change in fund ESG and market exposures around mandated portfolio disclosure dates. Panel A reports the average change in exposure in the 15, 10, and 5 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. Panel B reports the average change in exposure in the 5, 10, and 15 days after the day following disclosure relative to the exposure in the entire month before disclosure, excluding the first and last trading day. Exposures are derived from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index and is estimated separately for each fund and disclosure event. Standard errors are reported in round brackets. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively. P-values reported in square brackets are constructed based on the methodology presented in Online Appendix A and represent the fraction of bootstrapped delta betas generated under the null of no window dressing that exceed the critical threshold estimated from the data.

Panel A: Before disclosure

	ESG			MKT
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{MKT}$
Event window (days):	15	10	5	10
	(1)	(2)	(3)	(4)
	0.054***	0.123***	0.146***	-0.110***
	(0.020)	(0.023)	(0.039)	(0.022)
	[0.135]	[0.015]	[0.049]	[0.017]
Obs	4,063	4,063	4,063	4,063

Panel B: After disclosure

	ESG			MKT
	$\overline{\Delta\beta}_{+}^{ESG}$	$\overline{\Delta\beta}_{+}^{ESG}$	$\overline{\Delta\beta}_{+}^{ESG}$	$\overline{\Delta\beta}_{+}^{MKT}$
Event window (days):	5	10	15	10
	(1)	(2)	(3)	(4)
	0.029	0.021	0.004	-0.017
	(0.047)	(0.023)	(0.020)	(0.023)
	[0.271]	[0.332]	[0.427]	[0.367]
Obs	3,952	3,952	3,952	3,952

Table III: ESG holdings, loadings, and globes

This table reports for each fund sustainability rating the share of fund holdings invested in high-ESG stocks, and the median and average pre-disclosure ESG and market betas. Fund sustainability ratings are from 1 (lowest sustainability) to 5 globes (highest sustainability), as assigned two months after mandated disclosure. *High-ESG share* is the share of funds' portfolios invested in stocks with the highest Sustainalytics/Morningstar ESG scores corresponding to "low" or "negligible" ESG risk. ESG ($\hat{\beta}_{-}^{ESG}$) and market betas ($\hat{\beta}_{-}^{MKT}$) are estimated for each fund-disclosure event on the 10 days leading up to the disclosure date using Equation (1) in the paper.

	Panel A: Median values				
Globes:	1	2	3	4	5
High-ESG share	0.098	0.172	0.180	0.249	0.379
$\hat{\beta}_{-}^{ESG}$	-0.381	-0.092	0.027	0.392	0.560
$\hat{\beta}_{-}^{MKT}$	1.232	0.967	0.860	0.486	0.334
	Panel B: Average values				
Globes:	1	2	3	4	5
High-ESG share	0.106	0.167	0.176	0.238	0.369
$\hat{\beta}_{-}^{ESG}$	-0.931	-0.090	-0.126	0.316	0.598
$\hat{\beta}_{-}^{MKT}$	1.833	1.013	0.990	0.541	0.313
Obs	78	194	439	794	1,151

Table IV: Actual versus disclosed portfolios

This table compares fund realized returns and returns on disclosed portfolios. Columns 1 and 2 report the average daily return gap for the 10 days before and after the end of the fiscal quarter, respectively. Return gaps are calculated as the difference between a fund's realized return and the counterfactual return based on the stock positions disclosed at the nearest fiscal quarter-end (event time 0), and are expressed in percentages. Columns 3 to 5 report the average ESG beta estimated on post-disclosure realized (Column 3) and counterfactual returns (Column 4), and their difference (Column 5). ESG betas are estimated using a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Return gap (%)		ESG betas		
	Before	After	β_R^{ESG}	β_C^{ESG}	Diff.
Event days:	[-10, -1]	[2,11]	[2,11]	[2,11]	[2,11]
	(1)	(2)	(3)	(4)	(5)
	-0.013***	0.012***	0.123***	0.173***	-0.050***
	(0.004)	(0.003)	(0.029)	(0.022)	(0.018)
Obs	23,310	23,310	2,331	2,331	2,331

Table V: Which funds window dress?

This table reports coefficients from regressions of $Window\ dresser$ on fund characteristics. $Window\ dresser$ is a dummy variable that takes a value of 1 if $\Delta\hat{\beta}^{ESG}_{-,i,t+1}$ is in the top 10% of the sample distribution. $\Delta\hat{\beta}^{ESG}_{-,i,t+1}$ is the change in exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. *Fees* is the fund's annual expense ratio in percentage. *Star fund* (*Laggard fund*) is a dummy that takes a value of 1 if a fund's Fama and French 3-factor risk-adjusted quarterly return is in the top (bottom) 20% of the quarter. *Large fund* (*Large fund family*) is a dummy that takes a value of 1 if a fund (fund family)'s assets under management are in the top 20% of the quarter. *No voluntary disclosure* is a dummy variable that takes a value of 1 if the fund only publicly discloses stock holdings at the end of the fiscal quarter. *PRI* is a dummy variable that takes a value of 1 if the fund management company is among the signatories to the United Nations Principles for Responsible Investment. *Retail investors* is a dummy that takes a value of 1 if 50% or more of the fund's assets are managed on behalf of retail investors. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable:	$Window\ dresser_{t+1}$						
	Fees	Performance	Size	Disclosure	PRI	Clientele	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Fees_t$	0.091***						0.089***
	(0.014)						(0.015)
$Star\ fund_t$		0.042***					0.028**
		(0.013)					(0.013)
$Laggard\ fund_t$		0.090***					0.072***
		(0.013)					(0.013)
$Large\ fund_t$			-0.017				-0.013
			(0.012)				(0.012)
$Large\ family_t$			-0.026**				-0.026**
			(0.012)				(0.013)
$No\ voluntary\ disclosure_t$				0.020*			0.018*
				(0.010)			(0.010)
PRI_t					0.024**		0.030***
					(0.010)		(0.010)
$Retail\ investors_t$						-0.017*	-0.044***
						(0.010)	(0.011)
Time Fixed Effects	Y	Y	Y	Y	Y	Y	Y
Obs	3,525	3,525	3,525	3,525	3,525	3,525	3,525

Table VI: Does green window dressing improve fund ESG ratings?

This table reports coefficients for the regressions of Morningstar sustainability ratings on funds' increase in ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. $I(Five\ globes)_{i,t+2}$ ($I(One\ globe)_{i,t+2}$) is a dummy variable that takes a value of one if a fund is awarded five globes (one globe) two months after disclosure. $\Delta\hat{\beta}_{-}^{ESG}$ is standardized to have an average of 0 and a standard deviation of 1. Standard errors are clustered at the fund level and reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable:	$I(Five\ globes)_{i,t+2}$		$I(One\ globe)_{i,t+2}$	
	(1)	(2)	(3)	(4)
$\Delta\hat{\beta}_{-,i,t}^{ESG}$	0.021**	0.015***	-0.010**	-0.001
	(0.010)	(0.006)	(0.005)	(0.002)
Fund Fixed Effects	N	Y	N	Y
Time Fixed Effects	Y	Y	Y	Y
Obs	2,656	2,656	2,656	2,656

Table VII: Do green window dressers receive more flows?

This table reports coefficients from OLS regressions of fund 3-month flows on $\Delta\hat{\beta}_{-}^{ESG}$. $\Delta\hat{\beta}_{-}^{ESG}$ is the estimated difference between a fund's ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day, standardized to have an average of 0 and a standard deviation of 1. Column 1 includes all funds, Column 2 includes funds for which less than 50% of the assets are managed on behalf of retail investors, and Column 3 includes funds for which 50% or more of the assets are managed on behalf of retail investors. Standard errors are clustered at the fund level and reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent variable:	Net flows (in %)		
	All	Institutional	Retail
	(1)	(2)	(3)
$\Delta\hat{\beta}_{-,t}^{ESG}$	0.651***	0.710**	0.525
	(0.210)	(0.294)	(0.369)
Fund size$_t$	-1.096***	-0.990***	-1.206***
	(0.181)	(0.220)	(0.378)
Family size$_t$	-0.234	-0.338*	-0.004
	(0.162)	(0.188)	(0.322)
Expense ratio$_t$ *(%)*	-0.039***	-0.025	-0.047***
	(0.011)	(0.018)	(0.018)
Time Fixed Effects	Y	Y	Y
Obs	4,040	2,529	1,511

Table VIII: Do funds window dress portfolios at month-ends?

This table reports estimates for the average change in fund ESG exposure in the 10 days before the end of the month relative to the exposure in the 10 days before that. Exposures are from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. Column 1 reports results for all month-ends, Column 2 includes only month-ends that coincide with fiscal quarter-ends (as in our baseline specification), Column 3 includes only month-ends that do *not* coincide with fiscal quarter-ends, Column 4 only includes month-ends when a fund is voluntarily disclosing portfolio positions, Column 5 only includes month-ends when a fund is not disclosing portfolio positions in any form. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Events:	All month-ends	Only mandated disclosure	No mandated disclosure	Only voluntary disclosure	No disclosure month-ends
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
	(1)	(2)	(3)	(4)	(5)
	0.087***	0.107***	0.078***	0.115***	0.025
	(0.020)	(0.035)	(0.025)	(0.030)	(0.042)
Obs	12,077	4,063	8,014	4,704	3,310

Online Appendix

Green Window Dressing

This Online Appendix includes additional analyses and results and is not for publication with the main paper. Section A describes the wild bootstrap procedure we use to test for the presence of green window dressing. Section B quantifies the trading costs associated with green window dressing. Section C includes additional results and analyses. Specifically, this Online Appendix includes the following items:

- Figure A.1: Graphical representation of the empirical test
- Figure A.2: Cumulative returns on market and ESG portfolios
- Figure A.3: Effective spreads of high-ESG vs. non-high-ESG stocks
- Figure A.4: Trading of stocks with high MSCI ESG ratings around disclosure
- Figure A.5: Trading costs of imputed trades
- Figure A.6: Relationship between fund size and ESG holdings
- Table A.1: Placebo tests (placebo disclosures, before Morningstar, ESG ETFs)
- Table A.2: Alternative risk models
- Table A.3: Alternative ESG indexes
- Table A.4: Alternative estimation windows
- Table A.5: Alternative pre-event windows
- Table A.6: Expected fund flows
- Table A.7: Standard errors à la Fama-Macbeth

- Table A.8: Excluding the last quarter of the year
- Table A.9: Pollution factors
- Table A.10: Fund flows around the 4/5 globe cutoff
- Table A.11: Return gap analysis gross of fees
- Table A.12: Portfolio sorts
- Table A.13: Non-ESG funds

A. Testing for changes in ESG exposure

In order to keep this appendix self-contained, Section A.1 introduces the model, the null, the alternative hypotheses of interest, and the estimators. Section A.2 describes our bootstrap testing procedure. In the following, $\mathbb{1}\{condition\}$ denotes the indicator function, which is equal to 1 if the *condition* inside the brackets is satisfied and 0 otherwise.

A.1 Model, hypothesis, and estimators

The model for the return $R_{i,t}$ on the generic fund i at date t, with $i = 1, ..., N$, is:

$$R_{i,t} = \begin{cases} a_{-,i,e} + \beta^{MKT}_{-,i,e} MKT_t + (\beta^{ESG}_{0,i,e} + \Delta\beta^{ESG}_{-,i,e}) ESG_t + \varepsilon_{i,t}, & t \ = \ t_e - n, ..., t_e - 1 \\ a_{0,i,e} + \beta^{MKT}_{0,i,e} MKT_t + \beta^{ESG}_{0,i,e} ESG_t + \varepsilon_{i,t}, & t \ = \ t_e - 40, ..., t_e - 22, \end{cases} \tag{A.1}$$

where t_e, with $e = 1, ..., E$, is the generic disclosure date at the end of quarter e (i.e., the last trading day of each fiscal quarter), n is a finite number of days such that $0 < n < 21$, so that the n observations around a generic disclosure date do not overlap with the observations in the month before disclosure.[25] We consider a generic finite number of $E \geq 1$ disclosure dates $t_1, t_2, ..., t_E$. Parameter $\Delta\beta^{ESG}_{-,i,e}$ corresponds to the change in ESG exposure of the i-th fund in the last n days of the disclosure quarter e from the previous month, i.e. the second month of the same disclosure quarter. The structure of the panel of changes in ESG exposures $\Delta\beta^{ESG}_{-,i,e}$ is outlined in Online Appendix Figure A.1.

Our goal is to test whether the average ESG exposure $\beta^{ESG}_{-,i,e} = \beta^{ESG}_{0,i,e} + \Delta\beta^{ESG}_{-,i,e}$ in the n days before the disclosure date t_e increases (alternative hypothesis), or if it does not (null hypothesis), with respect to the exposure $\beta^{ESG}_{0,i,e}$ in the previous month. In other words, we want to test weather the average of all the $\Delta\beta^{ESG}_{-,i,e}$s is positive.

In the baseline analysis, we consider $N = 223$ ESG funds, $E = 28$ quarters (from 2016.Q1 to 2022.Q4), $n = 10$ days (with the smallest value of $n = 5$, i.e. one working week, as robustness). Therefore, an appropriate asymptotic scheme is the one with $N \to \infty$, while E and n are fixed (small), which is the typical asymptotic scheme used in "short" panels, i.e., panels with "large" cross sectional dimension N, and "short" time dimension (which in our case is n, the effective sample size over which we estimate the coefficients $\beta^{ESG}_{-,i,e}$). Using the notation of Hahn and Newey (2004), which investigates the distribution of averages of individual fixed effects in short panels, we define $\bar{E}[\Delta\beta^{ESG}_{-,i,e}] := \lim_{N\to\infty} \frac{1}{NE}\sum_{i=1}^{N}\sum_{e=1}^{E}\Delta\beta^{ESG}_{-,i,e}$. The null ($H_0^*$) and alternative ($H_1^*$) hypotheses that we can test in this framework are:

$$H_0^* : \bar{E}[\Delta\beta^{ESG}_{-,i,e}] = 0, \qquad\qquad H_1^* : \bar{E}[\Delta\beta^{ESG}_{-,i,e}] > 0. \tag{A.2}$$

Importantly, the null H_0^* in Eq. (A.2) is implied by the more general null hypothesis H_0 in (4), and a rejection of H_0^*, necessarily implies a rejection of H_0.

[25] To simplify the exposition and notation, we assume that the panel of observations of the funds' returns $R_{i,t}$ is balanced, that is we assume that for each of the N funds we observe returns $R_{i,t}$ at all dates $t = 1, .., T$. We also assume, only to simplify the notation, that there are 21 trading days in each month. Our estimation and testing procedures can be adapted to the more general case of an unbalanced panel and a generic number of working days in every month, at the cost of heavier notation. In all empirical analyses we consider unbalanced panel of returns and the actual number of trading days within each month of the sample.

For each fund i and disclosure date t_e, the two matrices of explanatory variables before the disclosure date ($t_e - n \leq t \leq t_e - 1$), and in the second month of the quarter ($t_e - 40 \leq t \leq t_e - 22$) are:

$$\underset{(n\times 3)}{X_{-,e}} \;:=\; [\; x_{t_e-n},\; ....,x_{t_e-1}\;]', \qquad \underset{(19\times 3)}{X_{0,e}} \;:=\; [\; x_{t_e-40},\; ....,x_{t_e-22}\;]', \tag{A.3}$$

where $x_t = [1,\; MKT_t,\; ESG_t]'$. The vectors of dependent variables for the same dates are:

$$\underset{(n\times 1)}{y_{-,i,e}} \;:=\; [\; R_{i,t_e-n},\; ....,R_{i,t_e-1}\;]', \qquad \underset{(19\times 1)}{y_{0,i,e}} \;:=\; [\; R_{i,t_e-40},\; ....,R_{i,t_e-22}\;]'.$$

Then, the OLS estimators of the vectors of coefficients $\beta_{-,i,e} := [a_{-,i,e}, \beta^{MKT}_{-,i,e}, \beta^{ESG}_{-,i,e}]'$ and $\beta_{0,i,e} := [a_{0,i,e}, \beta^{MKT}_{0,i,e}, \beta^{ESG}_{0,i,e}]'$ for each fund i, and event date e are:

$$\hat{\beta}_{-,i,e} \;:=\; (X'_{-,e}X_{-,e})^{-1}X'_{-,e}y_{-,i,e}\;, \qquad \hat{\beta}_{0,i,e} \;:=\; (X'_{0,e}X_{0,e})^{-1}X'_{0,e}y_{0,i,e}\;, \tag{A.4}$$

where $\hat{\beta}_{-,i,e} := [\hat{a}_{-,i,e}, \hat{\beta}^{MKT}_{-,i,e}, \hat{\beta}^{ESG}_{-,i,e}]'$ and $\hat{\beta}_{0,i,e} := [\hat{a}_{0,i,e}, \hat{\beta}^{MKT}_{0,i,e}, \hat{\beta}^{ESG}_{0,i,e}]'$. By denoting the estimated change in ESG exposure as $\Delta\hat{\beta}^{ESG}_{-,i,e} := \hat{\beta}^{ESG}_{-,i,e} - \hat{\beta}^{ESG}_{0,i,e}$, a natural statistic for the one-sided test in (A.2) is the one in equation (5) in the main paper, that is:

$$\overline{\Delta\beta}^{ESG}_{-} \;=\; \frac{1}{N\cdot E}\sum_{i=1}^{N}\sum_{e=1}^{E}\Delta\hat{\beta}^{ESG}_{-,i,e}\;.$$

Assumptions on the errors $\varepsilon_{i,t}$ are critical to make inference on $\bar{E}[\Delta\beta^{ESG}_{-,i,e}]$. The simplest set of assumptions for which it is relatively easy to derive the asymptotic distribution of $\overline{\Delta\beta}^{ESG}_{-}$, and to prove the validity of the related bootstrap testing procedure described in the next Section A.2, is the following one:

ASSUMPTION 1 *(i)* $E[\varepsilon_{i,t}] = 0$ *for all* i *and* t*; (ii)* $V(\varepsilon_{i,t}) = \sigma^2_{-,i,e} < \infty$ *for all dates* $t = t_e - n + 1, ..., t_e$, $V(\varepsilon_{i,t}) = \sigma^2_{0,i,e} < \infty$ *for all dates* $t = t_e - 40, ..., t_e - 22$*; (iii) the errors* $\varepsilon_{i,t}$ *are independent of* MKT_s *and* ESG_s *for all funds* $i = 1, ..., N$*, and dates* $t, s = 1, ..., T$*.*

Assumption 1 allows the errors to be heteroskedastic across funds and different months, and requires them to be uncorrelated across different days but, importantly, not across funds. As noted by, e.g., Fama and French (2010), Harvey and Liu (2022), and Hounyo and Lin (2023), it might be important to take into account the cross-sectional dependence in the residuals of the different time-series regressions when implementing bootstrap inference to estimate fund alphas.[26] The same issue might arise when making inference on the average of the $\Delta\hat{\beta}^{ESG}_{-,i,e}$, the main object of interest in our paper. In the next Subsection A.2 we propose a wild bootstrap which maintains the contemporaneous cross-sectional dependence of the regression residuals, while imposing the null hypothesis of interest in the bootstrap DGP, namely $\Delta\hat{\beta}^{ESG}_{-,i,e} = 0$. Therefore, our proposed bootstrap procedure is able to generate a cross-sectional distribution of potentially correlated $\hat{\beta}^{ESG}_{-,i,e}$s, which allows us to make correct bootstrap inference on $\overline{\Delta\beta}^{ESG}_{-}$.

Under the above Assumption 1, and by noting that our panel data model is linear in the parameters of interest, $\Delta\hat{\beta}^{ESG}_{-,i,e}$, and that our statistic $\overline{\Delta\beta}^{ESG}_{-}$ is also a linear function of all the $\Delta\hat{\beta}^{ESG}_{-,i,e}$s, it can be shown that our statistic is unbiased.[27] The latter result is also compatible with the observation made by Barras, Gagliardini, and Scaillet (2022) in their Internet Appendix II.A. Namely, that the cross-sectional average of the estimated coefficients "alpha" or "betas" obtained by time-series regressions of mutual funds returns on benchmark returns is an (asymptotically) unbiased estimator of the expected value of the cross-sectional distribution of the coefficients. By rewriting $\overline{\Delta\beta}^{ESG}_{-}$ as $\overline{\Delta\beta}^{ESG}_{-} = \frac{1}{E}\sum_{e=1}^{E}\left[\frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{ESG}_{-,i,e} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{ESG}_{0,i,e}\right]$, we see that $\overline{\Delta\beta}^{ESG}_{-}$ is a linear combination of cross-sectional averages of parameters estimated by time-series regressions for each fund i, i.e., $\frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{ESG}_{-,i,e}$ and $\frac{1}{N}\sum_{i=1}^{N}\hat{\beta}^{ESG}_{0,i,e}$. The latter are the natural unbiased estimators considered

[26]For instance, Harvey and Liu (2017) note that if the benchmark factor model used to compute alphas does not capture entirely the common variation in fund returns, alpha estimates will be cross-sectionally correlated and this correlation must be taken into account when making inference on moments or quantiles of the cross-sectional distribution of alphas.

[27]Therefore, different from the case of the estimation of averages of individual fixed effects in non-linear panel data models considered in, e.g., the seminal work of Hahn and Newey (2004), our statistic does not require any (asymptotic) bias adjustment.

also by Barras, Gagliardini, and Scaillet (2022) for the expected value of the cross-sectional distribution of the coefficients, i.e., $\bar{E}[\hat{\beta}^{ESG}_{-,i,e}]$ and $\bar{E}[\hat{\beta}^{ESG}_{0,i,e}]$, using the notation of Hahn and Newey (2004). As a linear combination of unbiased estimators is also an unbiased estimator, it immediately follows that $\overline{\Delta\beta}^{ESG}_{-}$ is an (asymptotically) unbiased estimator of $\bar{E}[\Delta\hat{\beta}^{ESG}_{-,i,e}] = \bar{E}[\hat{\beta}^{ESG}_{-,i,e} - \hat{\beta}^{ESG}_{0,i,e}]$.

A.2 Bootstrap test accounting for cross-sectional dependence

We now explain the residual (non-parametric) wild bootstrap (WB) for our test which accounts also for a generic form of unmodeled cross-sectional dependence in the regression residuals, which can be implemented by performing the following four steps.

WB.1 *Estimate the vectors of coefficients $\beta_{-,i,e}$ and $\beta_{0,i,e}$ by applying the OLS estimators in (A.4) on the original sample of data $\{R_{i,t},\ MKT_t,\ ESG_t\}_{t=t_e-40,\ldots,t_e-22,t_e-n,\ldots,t_e-1}$, for all funds $i = 1,...,N$ and disclosure dates $e = 1,...,E$. Store the estimated coefficients $\hat{a}_{-,i,e}, \hat{\beta}^{MKT}_{-,i,e}, \hat{\beta}^{ESG}_{-,i,e}$, and $\hat{a}_{0,i,e}, \hat{\beta}^{MKT}_{0,i,e}, \hat{\beta}^{ESG}_{0,i,e}$, and the panel of residuals $\hat{\varepsilon}_{i,t}$ defined as:*

$$
\begin{aligned}
\underset{(n\times 1)}{\hat{\varepsilon}_{-,i,e}} &= [\ \hat{\varepsilon}_{i,t_e-n},\ \ldots.,\hat{\varepsilon}_{i,t_e-1}\ ]' := y_{-,i,e} - X'_{-,e}\hat{\beta}_{-,i,e}\ , \\
\underset{(19\times 1)}{\hat{\varepsilon}_{0,i,e}} &= [\ \hat{\varepsilon}_{i,t_e-40},\ \ldots.,\hat{\varepsilon}_{i,t_e-22}\ ]' := y_{0,i,e} - X'_{0,e}\hat{\beta}_{0,i,e}
\end{aligned}
$$

for all funds $i = 1,...,N$ and $e = 1,...,E$ disclosure dates.

WB.2 *Compute $\Delta\hat{\beta}^{ESG}_{-,i,e} = \hat{\beta}^{ESG}_{-,i,e} - \hat{\beta}^{ESG}_{0,i,e}$ for all i and e, and the statistic*

$$
\overline{\Delta\beta}^{ESG}_{-} = \frac{1}{NE}\sum_{i=1}^{N}\sum_{e=1}^{E}\Delta\hat{\beta}^{ESG}_{-,i,e}.
$$

WB.3 *Let $B = 999$ be the total number of bootstrap iterations. For each bootstrap iteration b, with $b = 1,...,B$, repeat the following steps:*

(a) Generate a panel of bootstrapped errors $\varepsilon_{i,t}^{(b)}$ *as:*

$$\varepsilon_{i,t}^{(b)} = \hat{\varepsilon}_{i,t} \cdot \eta_t, \qquad i = 1, ..., N, \qquad t = t_e - 40, ..., t_e - 20, t_e - n, ..., t_e - 1, \quad e = 1, ..., E,$$

where η_t *is an "auxiliary" random variable which is i.i.d. across dates (i.e.,* η_t *is re-simulated every day) with zero mean and unitary variance. The assumption that* η_t *is the same for all funds at date* t *implies that the* N *bootstrapped fund errors* $\varepsilon_{i,t}^{(b)}$ *maintain the same contemporaneous cross-sectional correlation of the original data.*

(b) Generate the bootstrapped panel of dependent variables $R_{i,t}^{(b)}$ *for all dates and funds as:*

$$R_{i,t}^{(b)} := \begin{cases} \hat{a}_{-,i,e} + \hat{\beta}_{-,i,e}^{MKT} MKT_t + \hat{\beta}_{0,i,e}^{ESG} ESG_t + \varepsilon_{i,t}^{(b)}, & t = t_e - n, ..., t_e - 1 \\ \hat{a}_{0,i,e} + \hat{\beta}_{0,i,e}^{MKT} MKT_t + \hat{\beta}_{0,i,e}^{ESG} ESG_t + \varepsilon_{i,t}^{(b)}, & t = t_e - 40, ..., t_e - 22, \end{cases} \tag{A.5}$$

Importantly, for each couple of values (i, e)*, the ESG exposure is the same before, and after the disclosure date, and corresponds to* $\hat{\beta}_{0,i,e}^{ESG}$*, that is, the value estimated on the original sample in the month before the disclosure date.*

(c) By using the bootstrapped values of $R_{i,t}^{(b)}$ *from the previous step (b), define the vectors of bootstrapped fund returns:*

$$y_{-,i,e}^{(b)} := [\ R_{i,t_e-n}^{(b)},\ ...., R_{i,t_e-1}^{(b)}\ ]', \qquad y_{0,i,e}^{(b)} := [\ R_{i,t_e-40}^{(b)},\ ...., R_{i,t_e-22}^{(b)}\ ]'.$$

Estimate the factor exposures in the last n *days of all disclosure quarters* $e = 1, ..., E$*, all the days of the previous month excluding the first and last ones and funds* $i = 1, ..., N$*, by regressing the bootstrapped returns* $R_{i,t}^{(b)}$ *on the values of* MKT_t *and* ESG_t *in the original sample, which are collected in the vectors* $X_{-,e}$

and $X_{0,e}$ defined in (A.3):

$$\hat{\beta}^{(b)}_{-,i,e} = [\hat{a}^{(b)}_{-,i,e}, \hat{\beta}^{MKT(b)}_{-,i,e}, \hat{\beta}^{ESG(b)}_{-,i,e}]' := (X'_{-,e}X_{-,e})^{-1}X'_{-,e}y^{(b)}_{-,i,e} \, ,$$
$$\hat{\beta}^{(b)}_{0,i,e} = [\hat{a}^{(b)}_{0,i,e}, \hat{\beta}^{MKT(b)}_{0,i,e}, \hat{\beta}^{ESG(b)}_{0,i,e}]' := (X'_{0,e}X_{0,e})^{-1}X'_{0,e}y^{(b)}_{0,i,e} \, .$$

(d) Estimate the bootstrapped changes in ESG exposures as

$$\Delta\hat{\beta}^{ESG(b)}_{-,i,e} := \hat{\beta}^{ESG(b)}_{-,i,e} - \hat{\beta}^{ESG(b)}_{0,i,e}$$

for all i and e, and compute the bootstrapped statistic

$$\overline{\Delta\beta}^{ESG(b)}_{-} = \frac{1}{NE}\sum_{i=1}^{N}\sum_{e=1}^{E}\Delta\hat{\beta}^{ESG(b)}_{-,i,e}.$$

WB.4 *Compute the bootstrap p-value for the one-sided test of $H_0^* : \bar{E}[\Delta\beta^{ESG}_{-,i,e}] = 0$ vs. $H_1^* : \bar{E}[\Delta\beta^{ESG}_{-,i,e}] > 0$ as:*

$$\hat{p} := \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\left\{\overline{\Delta\beta}^{ESG(b)}_{-} > \overline{\Delta\beta}^{ESG}_{-}\right\}.$$

In other words, the bootstrap p-value is the fraction of bootstrap replications B in which the bootstrapped value of the statistic $\overline{\Delta\beta}^{ESG(b)}_{-}$ is strictly larger than the observed statistic $\overline{\Delta\beta}^{ESG}_{-}$ computed on the original data.

WB.3 *(b)* is the critical and innovative step in our wild residual bootstrap (compared to the existing literature on bootstrap for detection of alphas). It allows us to impose the null hypothesis H_0^* by letting H_0 hold, i.e. : $\Delta\beta^{ESG(b)}_{-,i,e} = \beta^{ESG(b)}_{-,i,e} - \beta^{ESG(b)}_{0,i,e} = \hat{\beta}^{ESG}_{0,i,e} - \hat{\beta}^{ESG}_{0,i,e} = 0$, for all funds i in the bootstrap data generating process (DGP), i.e. our model (A.5).[28] At the same time, our bootstrap DGP allows the other two coefficients $\hat{a}_{0,i,e}$ and $\hat{\beta}^{MKT}_{0,i,e}$ to

[28]The fact that the bootstrap DGP satisfies the null hypothesis H_0^*, which implies the null H_0, is generally desiderable for any bootstrap test, as suggested by, e.g. Davidson and MacKinnon (1999), and Davidson and MacKinnon (2004). This property of the bootstrap DGP is commonly referred to as the "golden rule" of bootstrap.

change in the n days before the disclosure date. Note also that, as customary with residual bootstrap, the explanatory variables MKT_t and ESG_t are not re-sampled. Instead, they maintain the same values, and time order, as in the original sample. This implies that the regressors maintain the same time-series dependence in the bootstrap DGP as in the original data. Importantly, this choice allows us to include in the bootstrap DGP the high contemporaneous correlation between the market and the ESG index. Therefore, the way we implement our bootstrapped test also addresses potential finite sample issues arising from the high correlation between the regressors.

Note that, by multiplying in Step **WB.3** *(a)* each residual $\hat{\varepsilon}_{it}$ by the same "auxiliary" random variable η_t independent of the $\hat{\varepsilon}_{it}$, we impose that the bootstrapped residuals, $\hat{\varepsilon}_{it}^{(b)}$, feature the same contemporaneous cross-sectional correlation of the actual residuals. This approach could be considered the wild bootstrap equivalent of the approach put forward by Fama and French (2010) and extended by Harvey and Liu (2022), who instead re-sample with replacement the entire n-dimensional vector of residuals to generate bootstrap residuals. This approach was developed in a contemporaneous work by Hounyo and Lin (2023), who show by using bootstrap and simulation evidence that the cross-sectional dependent wild bootstrap has superior size and power than the classical Fama and French (2010) bootstrap in the standard problem of detecting fund managers' skill (as opposed to luck) from fund alphas. In unreported Monte Carlo experiments calibrated on the properties of our ESG funds and on the time-series properties of the regressors MKT_t and ESG_t (including their high correlation), we have verified that our cross-sectional dependent wild bootstrap has good size and power properties for the null and alternative hypothesis tested in our paper. The procedure to test for changes in exposures *after disclosure* is analogous and, therefore, we do not discuss it in this Online Appendix. We report the results from the bootstrap tests in Table II in the main paper.

B. Costs of green window dressing

To quantify the cost of green window dressing, we estimate trading costs at the stock level based on Abdi and Ranaldo (2017)'s "CHL" two-day-corrected effective spread, defined as:

$$\hat{\kappa}_{two-day,j,t} \quad = \quad \frac{1}{D_t}\sum_{d=1}^{D_t}\hat{\kappa}_{j,d}, \qquad \hat{\kappa}_{j,d} = \sqrt{\max\left\{4(\text{cls}_{j,d}-\text{mid}_{j,d})(\text{cls}_{j,d}-\text{mid}_{j,d+1}),0\right\}}, \tag{A.6}$$

where D_t is the number of trading days in month t, $\text{cls}_{j,d}$ is the closing log-price of stock j on day d, and $\text{mid}_{j,d} = (\text{high}_{j,d} + \text{low}_{j,d})/2$ is the mid-range log-price on day d computed using the daily high ($\text{high}_{j,d}$) and low ($\text{low}_{j,d}$) log-prices. This measure quantifies the percentage cost of a round-trip trade in stock j and accounts for price fluctuations in each day d resulting from large trades, fire-sales, or additional trading around quarter-ends, as it is based on daily realized *close*, *high*, and *low* prices (see Abdi and Ranaldo (2017) for details).

Based on this measure, we find that stocks awarded the two highest ESG scores by Sustainalytics (corresponding to ESG risk "negligible" or "low") are less costly to trade. Restricting our sample to the stocks held by the funds in our sample, we find average effective spreads of 0.85% for high-ESG stocks and 1.05% for non-high-ESG stocks, respectively. These values are similar to the average of 1.39% and median of 1.02% reported by Abdi and Ranaldo (2017). In terms of market capitalization, we find that high-ESG stocks are twice as large than non-high-ESG stocks on average (this result is untabulated). Online Appendix Figure A.3 illustrates the evolution of trading costs for high- and non-high-ESG stocks over our sample period. Overall, the spread in trading costs remains fairly stable over time, with high-ESG stocks consistently being about 19% cheaper to trade.

We then employ the following back-of-the-envelope procedure to evaluate how much portfolios change and the ensuing trading costs. Drawing from the data in Table III, we find that the median 1-globe fund would need to increase the proportion of its portfolio allocated to high-ESG stocks by 28 percentage points $(0.38 - 0.10)$ to achieve a 5-globe rating.[29] This adjustment corresponds to an increase in the fund's ESG beta by 0.94 $(= 0.56 + 0.38)$. Given that the average pre-disclosure change in ESG beta in our sample is 0.12, this translates to an increase in the proportion allocated to high-ESG stocks of approximately 3.6 percentage points $(\frac{0.12}{0.94} \times 0.28)$. When relying on average rather than median beta estimates, this number decreases to 2.1 percentage points $(\frac{0.12}{1.53} \times 0.27)$. The average cost of buying high-ESG stocks before disclosure amounts to half the bid-ask spread estimated above which is 0.87%/2, as Abdi and Ranaldo (2017)'s CHL measures the relative cost of a round-trip trade. When we aggregate these costs at the fund level, window dressing portfolios before disclosure and undoing it after disclosure corresponds to approximately $2 * (0.87\%/2 \times 0.036) = 3.1$ basis points (bps) lower fund performance.[30] This cost, that window dressing funds face once per quarter, represents 91% of the average daily return for the funds in our sample (3.4 bps). In light of these back-of-the envelope calculations, we conclude that engaging in green window dressing results in a moderate increase in trading and trading costs.

As an alternative approach, we build on the analysis in Section II.C and quantify the trading costs associated with green window dressing from trades imputed using the methodology developed by Bongaerts, van Brakel, van Dijk, and Huij (2024). Specifically, we conduct an event-study analysis to determine how fund trading costs vary around disclosure. To that end, we define the percentage cost of trading ESG stocks for fund i on day t as

$$\phi_{i,t} \;=\; \sum_{j} \frac{\kappa_{j,t}}{2} \;\times\; |\Delta w^{ESG}_{i,j,t}|, \tag{A.7}$$

[29] Note that this procedure is a simplification, as funds' ESG Morningstar ratings do not depend only on the fraction of the portfolio allocated to stocks with the highest ESG ratings but, for instance, also on how a fund ranks in terms of ESG holdings with respect to its peers.

[30] Note that this amount measures only the cost of trading high-ESG stocks. If funds entirely replace high-ESG with low-ESG stocks this cost more than doubles (as we find that the cost of trading non-high-ESG stocks is higher).

where $\kappa_{j,t}/2$ is half the Abdi and Ranaldo (2017)'s two-day corrected spread for stock j on date t, and $\Delta w_{i,j,t}^{ESG}$ is the relative change in the portfolio weight of ESG stock j by fund i on date t defined as in Section II.C. in the paper.

Online Appendix Figure A.5 reports the daily additional cost of trading ESG stocks from 10 days before to 10 days after the disclosure date. We find that trading costs increase by 0.005% on the day preceding disclosure, and than by 0.003% and 0.002% on the disclosure date and the following day, respectively. These estimates are about half in magnitude relative to those from our back-of-the-envelope calculation based on changes in ESG betas. This finding is in line with the fact that we find a smaller change in portfolio weights using imputed trades, as the imputation minimizes by construction the number of round-trip trades (which might be necessary if fund managers window dress their portfolios by buying and selling the same ESG stocks during a quarter). Overall, this analysis confirms that the increase in trading costs associated with green window dressing is modest.

C. Further results and robustness tests

Alternative risk models. Table A.2 augments our baseline two-factor model that includes the market and the Morningstar US Sustainability Index by adding 1) the small-minus-big (SMB) factor, 2) the small-minus-big (SMB) and the high-minus-low (HML) factors, and 3) the small-minus-big (SMB), the high-minus-low (HML), and the momentum (MOM) factors, respectively. Columns 4-6 expand the estimation windows to 15 days to increase the estimation's precision. Results remain similar to those from our preferred specification.

Alternative ESG indexes. We consider a battery of alternative ESG indexes in Table A.3. Specifically, we replicate our main analysis replacing the Morningstar US Sustainability Index with the following traded and non-traded ESG indexes (one at the time): i) the MSCI

USA Leaders Index, ii) the MSCI USA Select Index, iii) the KLD 400 Index, iv) the Morningstar Low Carbon Risk Index, v) the Dow Jones USA Sustainability Index, vi) the iShares ESG Leaders ETF, vii) the iShares ESG Select ETF, viii) the MSCI Gender diversity ETF. Note that the time-series for the indexes based on ETFs are shorter, as most of them were introduced *after* the beginning of our sample (March 2016). Our results are robust to using any of these alternative ESG indexes with the exception of the MSCI Gender diversity ETF.

Alternative estimation and control windows. In Table A.4, we present results obtained using alternative estimation windows. In Column 1, we define the event window as the 10 days before the disclosure date and the control pre-event window as the 10 days before that. In Column 2, we define the event window as the 5 days before the disclosure date and we use the same control window as in Column 1. In Column 3, we sort funds into 4 groups based on the average fund churn ratios in the quarter, constructed based on disclosed holdings following Gaspar, Massa, and Matos (2005). We define the event window as the 20 days before the disclosure date for the funds with the lowest churn ratio, as the 15 days before for funds in the second group, as the 10 days before for funds in the third group, and as the 5 days before for the funds with the highest churn ratio. In all cases, results remain similar.

Expected flows and delta betas. A concern with our results is that end-of-month disclosure may be followed by predictable cash inflows (such as, for example, those related to 401(k) contributions). To the extent that funds manage these cash inflows with derivatives based on the S&P 500 (see, e.g., Frino, Lepone, and Wong (2009) and Rohleder, Schulte, and Wilkens (2017)), that could lead to a mechanical decrease (increase) of correlation with the ESG index (the market) after month-ends even when funds are not window dressing. We mitigate this concern by proxying expected flows with the average monthly net flows received by the fund in the previous three months. We then regress the estimated delta ESG betas and delta market betas on expected flows. We find that there is no statistically significant correlation between changes in ESG betas and expected flows, thereby mitigating concerns that the changes in factor loadings are driven by flows (see Online Appendix Table A.6). Importantly, a number of findings we document in the paper are also not consistent with this alternative explanation. Namely, fund managers' response to predictable flows could not explain i) why we do not find any evidence of green window dressing before the introduction of Morningstar sustainability ratings (Online Appendix Table A.1), ii) why the effect is double as strong for ESG than for non-ESG funds (Online Appendix Table A.13), iii) why variations in ESG betas predict sustainability ratings (Table VI), and iv) the ESG stock price dynamics we find around disclosure dates (Figure IV).

Standard errors à la Fama-Macbeth. In our first empirical test, we account for the presence of potential cross-sectional correlation across funds using the wild bootstrap approach described in Online Appendix Section A. In Online Appendix Table A.7, we alternatively account for potential cross-sectional correlation computing the standard errors from the time-series of average delta ESG betas (similar to Fama and MacBeth (1973)). We mitigate the effect of outliers arising from the short estimation window by winsorizing extreme observations. Overall, we reject the null of no change in ESG betas in all specifications with confidence levels ranging from 1% to 10%.

End-of-year effects. Both traditional window dressing and tax motivated trading are prevalent at the end of the year (Lakonishok, Shleifer, Thaler, and Vishny (1991), Sias and Starks (1997)). To make sure our results are not capturing traditional window dressing, we replicate our main analysis after excluding the last calendar quarter of the year. Online Appendix Table A.8 documents that our results are stronger when we focus on the first three calendar quarters of the year. This finding suggests that ESG ratings might not be the primary concern when funds approach the end of the year. For instance, because the main objective is minimizing the tax burden.

CO2 emission factor. The finding that fund managers increase ESG exposure before disclosure implies that they might temporarily decrease exposure to more remunerative factors that look inconsistent with their ESG mandate. In Table A.9 we rerun our two-factor model replacing the Morningstar US Sustainability Index with a long/short pollution factor based on CO2 emissions (see, e.g., Bolton and Kacperczyk (2021, 2023); Hsu, Li, and Tsou (2023)). Specifically, we construct this pollution factor as the value-weighted return on the top 33% minus the bottom 33% emitters among US listed firms based on tons of CO2 emissions.[31]. Our findings show an overall decrease in exposure to high-CO2 emitting companies shortly before disclosure. This analysis is important also because of the high correlation between the market factor and the Morningstar US Sustainability Index, which may raise concerns of multicollinearity. The fact that we find consistent results when we use a long-short pollution factor (that has low correlation with the market by construction) rather than the Morningstar US Sustainability Index reassures on the validity of our findings.

Fund flows around Morningstar globes cutoffs. In this section, we confirm that Morningstar ESG ratings matter for investor flows, which, in turn, incentivizes funds to window dress to achieve a higher ESG rating. Following Hartzmark and Sussman (2019), we estimate the effect of receiving a rating of five globes on flows by conducting a regression discontinuity test around the 4-5 globe rank cutoff. This approach allows us to compare funds that just meet the requirement to receive the highest sustainability rating to funds that just fall short of the rank requirement. Morningstar ranks funds within each global investment category based on how sustainable their disclosed portfolio holdings are. According to Morningstar documentation (Morningstar (2021)), a fund receives five globes if it is in the top 10% of sustainability in its category, four globes if it is ranked between 10% and 32.5%, three globes

[31]Emissions are Scope 1 + Scope 2 + Scope 3 tons of CO2 emissions obtained from Trucost. See, e.g., Bolton and Kacperczyk (2023) for a description of the emission data. The long/short pollution factor is constructed by value-weighting every month stocks in the upper and lower tercile of emissions.

if it is ranked between 32.5% and 67.5%, two globes if it is ranked between 67.5% and 90%, and one globe if it is in the bottom 10% of its category.[32]

Online Appendix Table A.10 tests for a discontinuity on future monthly flows around the 4-5 globe cutoff. We find that funds that just fall short of the 5-globe rank requirement receive around 0.8 percentage point lower flows in the next month (Columns 1 and 2). In untabulated results, we find weaker or no effect around the other globe cutoffs.

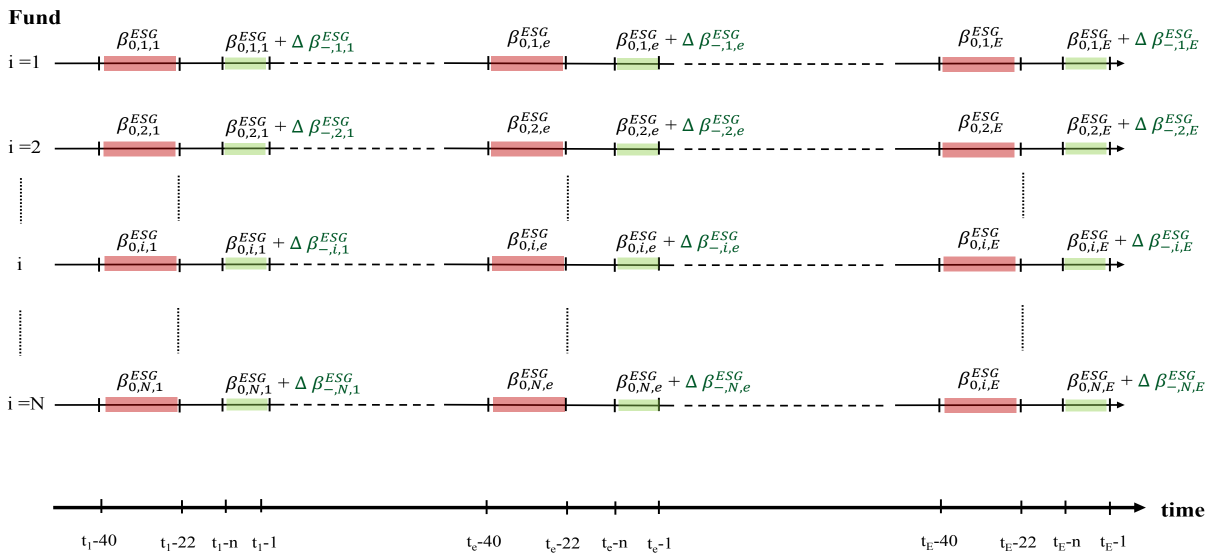


Figure A.1: Multiple changes in ESG exposures across funds and disclosure dates

This figure illustrates the coefficients measuring ESG exposures for different funds and around different disclosure dates in our model.

[32]Exceptions, however, may arise in situations where scores are not normally distributed. This can occur, for instance, when the majority of funds in a category receive similar scores. Another exception to the rule above is made when a fund ranks among the top performers in its category, yet its overall sustainability is low (see Morningstar (2021)).

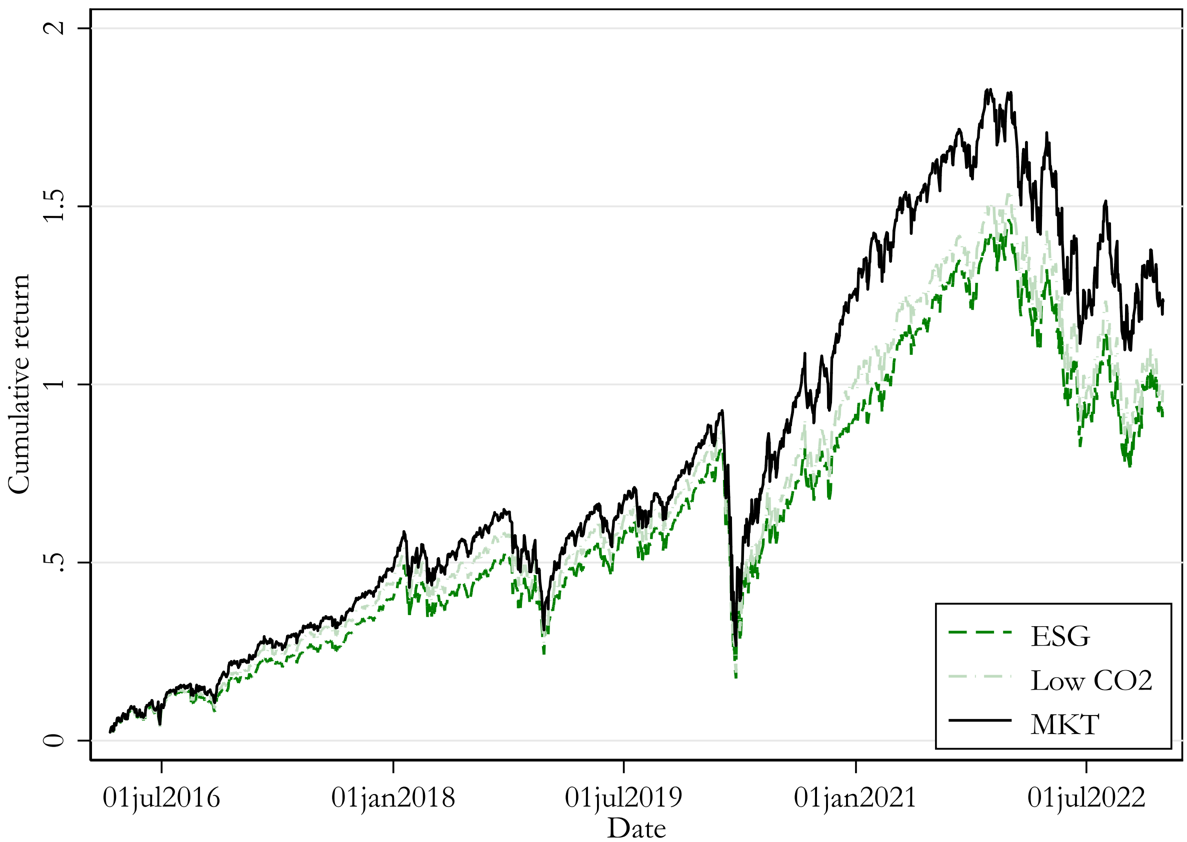


Figure A.2: Cumulative returns on ESG, low carbon, and market portfolios

This figure shows cumulative returns on the Morningstar Sustainability Index (ESG), the Morningstar US Low Carbon Risk Index (Low CO2), and the market portfolio (MKT) from March 2016 to December 2022.

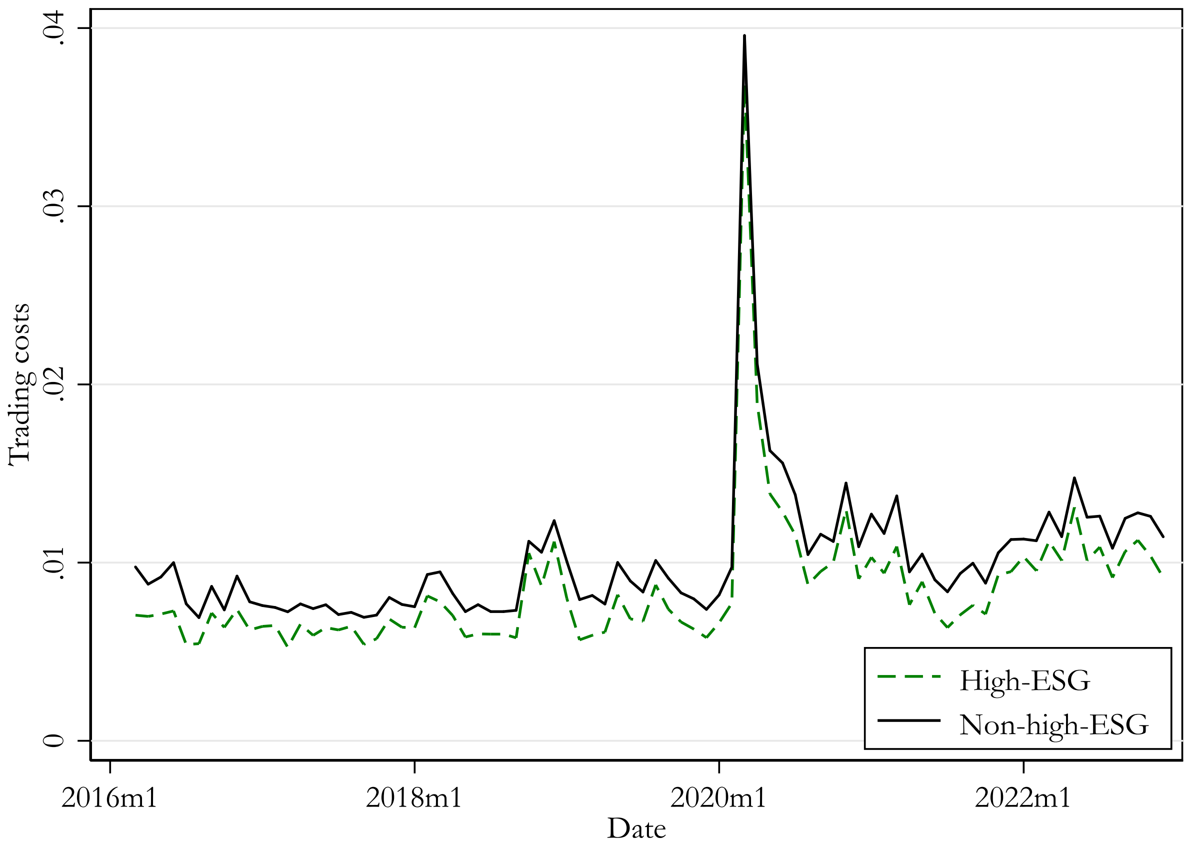


Figure A.3: Costs of trading high-ESG vs. non-high-ESG stocks

This figure reports the average relative cost of a round-trip trade in high-ESG and non-high-ESG stocks over time. Trading costs are calculated based on Abdi and Ranaldo (2017)'s two-day corrected estimated effective spread. The analysis includes exclusively stocks disclosed by the ESG funds in our sample.

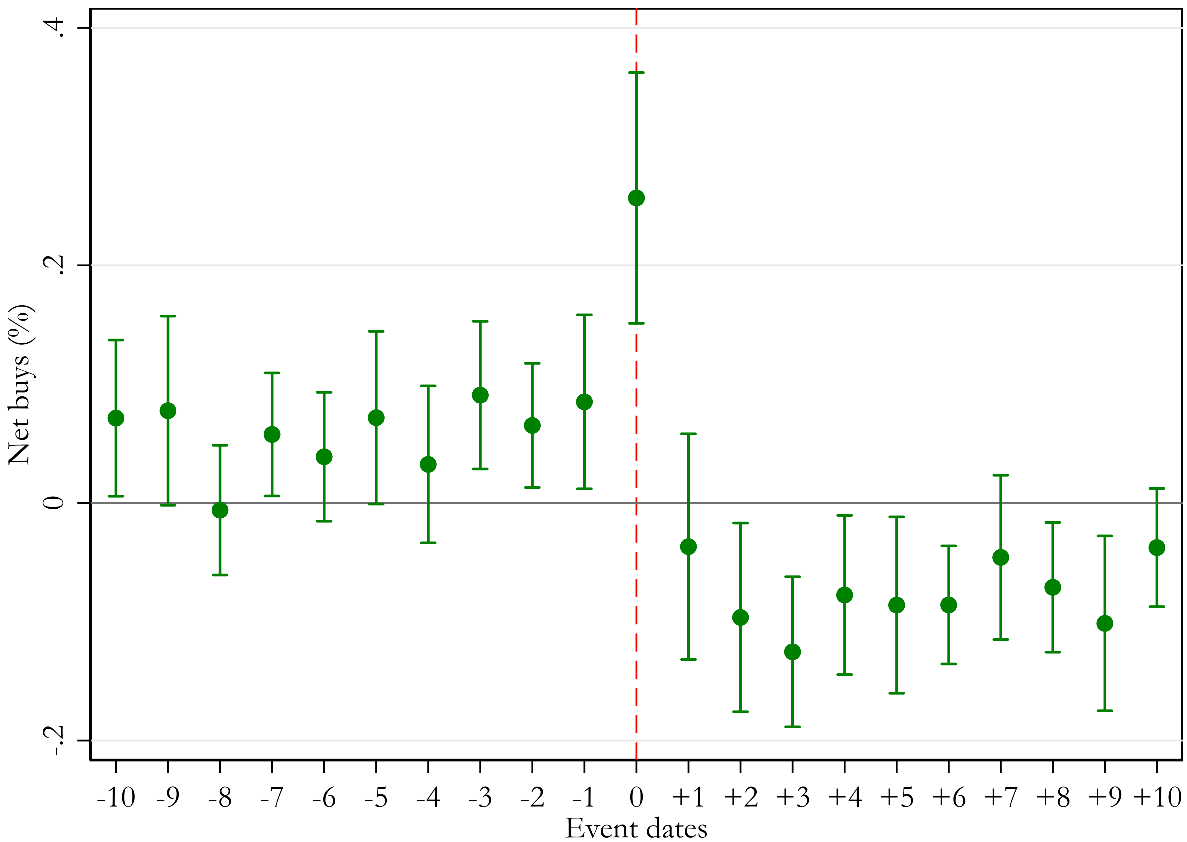


Figure A.4: Trading of MSCI ESG stocks around disclosure dates

This figure reports event-study estimates for fund net purchases of stocks in the top tercile of MSCI ESG ratings in the days around disclosure dates. Coefficients report the effect from 10 days before (event date $\tau = -10$) to 10 days after (event date $\tau = +10$) the disclosure date. Trades are imputed using the methodology by Bongaerts, van Brakel, van Dijk, and Huij (2024). The regressions include fund-by-month fixed effects, and standard errors are clustered at the fund level.

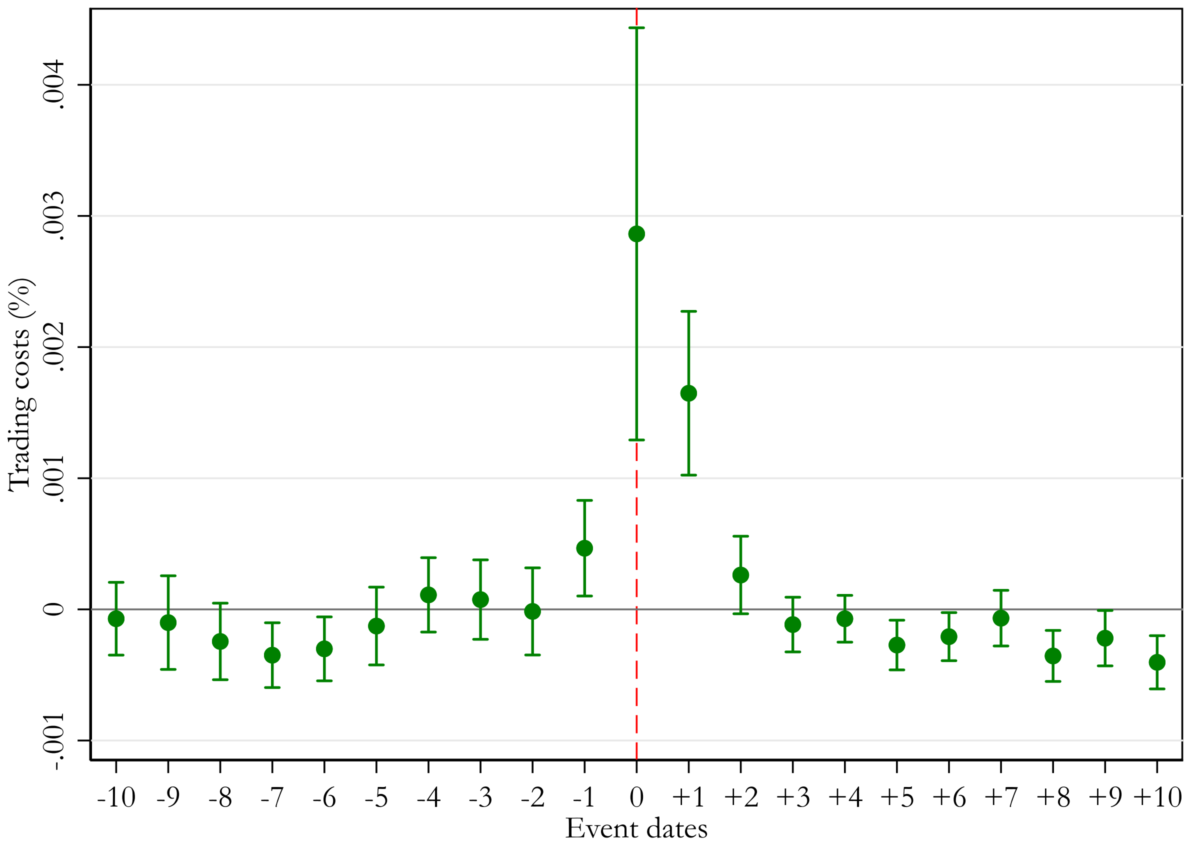


Figure A.5: Costs of green window dressing

This figure reports event-study estimates for the negative impact on fund performance from trading ESG stocks around portfolio disclosure dates. Trading costs at the stock level are estimated based on Abdi and Ranaldo (2017)'s two-day corrected effective spreads and are then aggregated at the fund level. We report coefficients from 10 days before (event date $\tau = -10$) to 10 days after the disclosure date (event date $\tau = +10$). Trades are imputed using the methodology by Bongaerts, van Brakel, van Dijk, and Huij (2024). The regressions include fund-by-month fixed effects, and standard errors are clustered at the fund level.

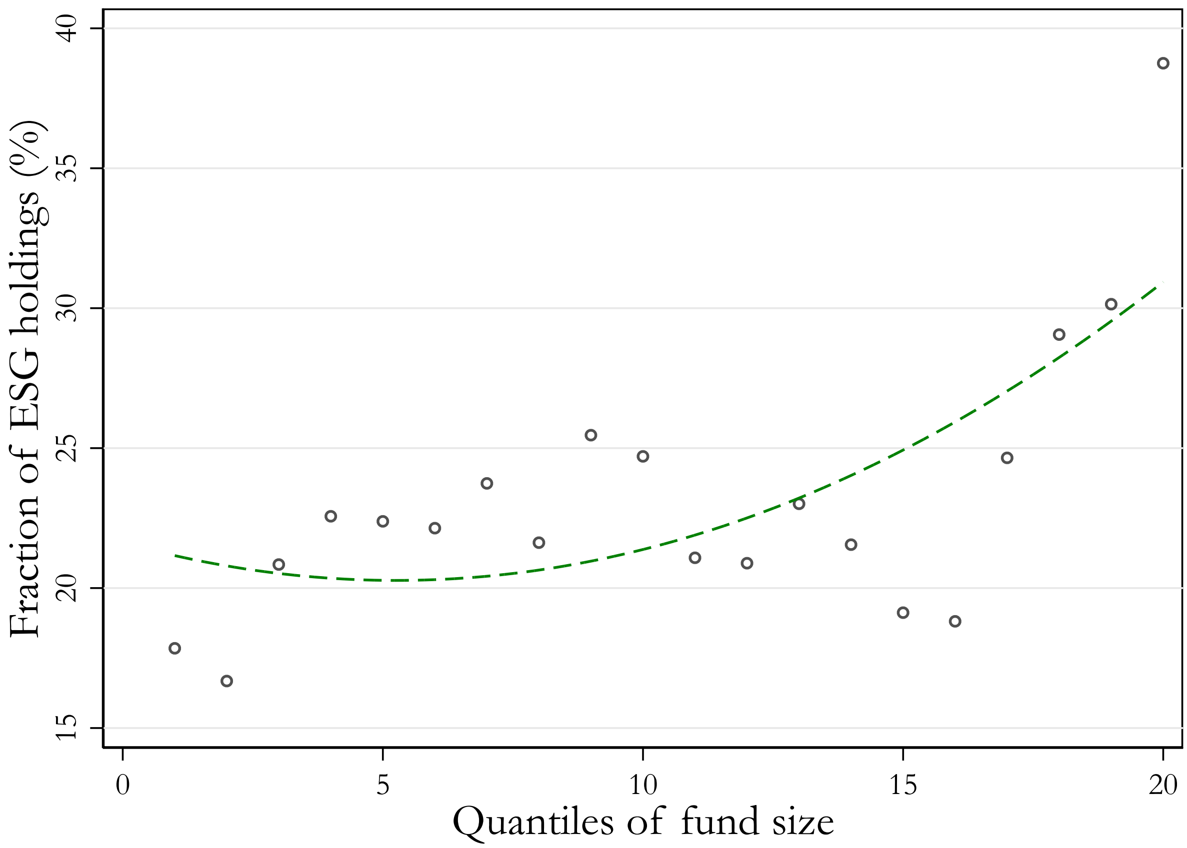


Figure A.6: ESG holdings and fund size

This figure reports the relationship between fund size and the fraction of fund portfolio invested in high-ESG stocks based on Sustainalytics/Morningstar scores. Funds are sorted into 20 quantiles from the smallest (1st quantile) to the largest (20th quantile).

Table A.1: Placebos

This table reports results from three placebo tests. For all tests, we report the average change in ESG exposure in the 10 days before the disclosure date with respect to the average exposure in the previous month (excluding the first and last trading day). ESG exposures are from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. Column 1 conducts the analysis on 5,000 randomly assigned placebo disclosure dates on the same ESG funds and the same period as in our main analysis. Column 2 conducts the analysis on ESG funds on the period from January 2010 to February 2016 included. Column 3 conducts the analysis on ESG passive funds and ETFs for the period from March 2016 to December 2022. Standard errors are reported in parentheses.

	Placebo disclosure	Before Morningstar	ESG ETFs
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
	(1)	(2)	(3)
	-0.007	0.005	0.057
	(0.021)	(0.039)	(0.044)
Obs	5,000	2,271	1,865

Table A.2: Alternative risk models

This table reports the average difference in funds' ESG exposure in in the 10 or 15 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. ESG exposures are from multi-factor models that include the market portfolio, the Morningstar US Sustainability Index, and the additional factors indicated at the top of each column. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Add. risk factors:	SMB	SMB+HML	SMB+HML+UMD	SMB	SMB+HML	SMB+HML+UMD
Event window (days):	10	10	10	15	15	15
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
	(1)	(2)	(3)	(4)	(5)	(6)
	0.117***	0.065***	0.080***	0.076***	0.088***	0.064***
	(0.025)	(0.025)	(0.030)	(0.022)	(0.020)	(0.022)
Obs	4,063	4,063	4,063	4,063	4,063	4,063

Table A.3: Alternative ESG indexes

This table reports the average difference in funds' ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. ESG exposures are from two-factor models that include the market portfolio and a different ESG index. We consider the following ESG indexes: the MSCI USA ESG Leaders, the MSCI USA ESG Select, the KLD 400, the Morningstar US Low Carbon Risk, the Dow Jones US Sustainability, the iShares ESG MSCI USA Leaders ETF, the iShares ESG MSCI USA Select ETF, the MSCI USA Gender Diversity ETF. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

ESG index:	MSCI Leaders	MSCI Select	KLD 400	Morningstar Low Carbon	Dow Jones Sust. Index	MSCI Leaders ETF	MSCI Select ETF	MSCI Gender Div ETF
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	0.096***	0.117***	0.070***	0.087**	0.099***	0.134***	0.105***	0.009
	(0.023)	(0.019)	(0.023)	(0.044)	(0.018)	(0.021)	(0.018)	(0.012)
Obs	4,063	4,063	4,063	4,063	4,063	2,461	4,063	3,982

Table A.4: Alternative estimation windows

This table reports the average difference between funds' ESG exposure shortly before disclosure and the one in a previous period estimated on different windows. ESG exposures are from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. Column 1 defines the pre-disclosure window as the 10 days before the disclosure date and the baseline window as the 10 days before that. Column 2 defines the pre-disclosure window as the 5 days before the disclosure date and the same baseline window as in Column 1. Column 3 sorts funds into 4 groups based on the fund churn ratio constructed as in Gaspar, Massa, and Matos (2005). We define the pre-disclosure window as the 20 days before the disclosure date for funds with the lowest churn ratio, as 15 days for funds in the second group, as 10 days for funds in the third group, and as 5 days for funds with the highest churn ratio. We define as the baseline window the ESG exposure in the month before disclosure excluding the first and last day of the month. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

β_{-}^{ESG}	[-10,-1]	[-5, -1]	CR
β_{0}^{ESG}	[-20,-11]	[-20,-11]	t_{m-1}
	$\overline{\Delta\beta_{-}^{ESG}}$	$\overline{\Delta\beta_{-}^{ESG}}$	$\overline{\Delta\beta_{-}^{ESG}}$
	(1)	(2)	(3)
	0.107***	0.130***	0.089***
	(0.035)	(0.047)	(0.031)
Obs	4,063	4,063	3,336

Table A.5: Alternative pre-event windows

This table reports the average difference between funds' ESG exposure in the 10 days leading up to disclosure and the ESG exposure in the previous month excluding x days. Column 1 excludes from the pre-disclosure (control) month the first and last trading day as in our baseline specification ($x = t_1, t_n$ where n is the number of days in the month). Column 2 excludes the first two and the last two trading days ($x = t_1, t_2, t_{n-1}, t_n$), and Column 3 the first three and the last three trading days ($x = t_1, t_2, t_3, t_{n-2}, t_{n-1}, t_n$). ESG exposures are from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Excluding:	t_1, t_n	t_1, t_2, t_{n-1}, t_n	t_1, t_2, t_3, t_{n-2}, t_{n-1}, t_n
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
	(1)	(2)	(3)
	0.123***	0.131***	0.123***
	(0.023)	(0.024)	(0.024)
Obs	4,063	4,063	4,063

Table A.6: Delta betas and expected flows

This table reports coefficients from OLS regressions of estimated changes in ESG (market) exposures on expected monthly flows. Changes in exposures are funds' ESG (market) exposure in the 10 days leading up to disclosure and that in the entire month before disclosure, excluding the first and last trading day. Exposures are from a two-factor models that include the market portfolio and the Morningstar US Sustainability Index. *Expected flows* is the average monthly net fund flow in the previous three months. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	$\Delta\hat{\beta}_{-}^{ESG}$ (1)	$\Delta\hat{\beta}_{-}^{MKT}$ (2)
Expected flows	-0.408	0.279
	(0.525)	(0.495)
Constant	0.127***	-0.112***
	(0.024)	(0.023)
Obs	3,950	3,950

Table A.7: Standard errors à la Fama-Macbeth

This table reports the average difference between funds' ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. ESG exposures are from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. In Column 2 (3) delta ESG betas are winsorized at the 5% (10%) level. Standard errors are constructed by first computing the average delta ESG beta for each quarter. Then, the standard deviation of these quarterly average delta ESG betas is calculated and divided by the square root of the number of quarters, similar to Fama and MacBeth (1973).

Winsorized	No $\overline{\Delta\beta}_{-}^{ESG}$ (1)	5% $\overline{\Delta\beta}_{-}^{ESG}$ (2)	10% $\overline{\Delta\beta}_{-}^{ESG}$ (3)
	0.123*	0.124**	0.122***
	(0.062)	(0.049)	(0.040)
Obs	4,063	4,063	4,063

Table A.8: Excluding the end of the year

This table reports the average difference in fund ESG exposure estimated in the 5, 10, or 15 days before disclosure and the ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure except the first and last trading day, excluding the last calendar quarter of each year. ESG exposures are from a two-factor models that include the market portfolio and the Morningstar US Sustainability Index. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Event window (days):	5	10	15
	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
	(1)	(2)	(3)
	0.288***	0.190***	0.136***
	(0.046)	(0.027)	(0.023)
Obs	3,024	3,024	3,024

Table A.9: Do funds decrease the exposure to CO2 shortly before disclosure?

This table presents estimates for the average difference between a fund's exposure to a CO2 emissions-based pollution factor in the days leading up to disclosure with respect to the exposure in the previous month excluding the first and last trading day. Exposures are from a two-factor model that includes the market portfolio and the pollution factor. This long/short pollution factor invests in (shorts) a value-weighted portfolio of the top (bottom) 33% of U.S. firms according to tons of CO2 emissions from Trucost. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

		Brown minus Green	
	$\overline{\Delta\beta}_{-}^{CO2}$	$\overline{\Delta\beta}_{-}^{CO2}$	$\overline{\Delta\beta}_{-}^{CO2}$
Event window (days):	15	10	5
	(1)	(2)	(3)
	-0.007	-0.016***	-0.041***
	(0.005)	(0.006)	(0.012)
Obs	4,063	4,063	4,063

Table A.10: Fund flows around the 4/5 globe cutoff

This table reports regression discontinuity tests of one-month-ahead ESG fund flows around the Morningstar 4/5 globe cutoff. The coefficients measure the effect of falling just below the rank threshold to assign 5 globes. MSE-optimal bandwidths are selected using the methods developed in Calonico, Cattaneo, and Titiunik (2014), Calonico, Cattaneo, and Titiunik (2015), and Calonico, Cattaneo, and Farrell (2020). Observations are weighted using a triangular kernel. Controls in Column 2 are fund size, expenses, monthly returns, and a linear time trend variable. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Net flows (%)	
	(1)	(2)
	-0.778**	-0.726**
	(0.347)	(0.333)
Bandwidth	8.568	9.113
Controls	N	Y
Effective obs	4,475	5,018
Obs	8,759	8,759

Table A.11: Actual versus disclosed portfolios gross of fees

This table compares the returns achieved by funds with the returns on their disclosed portfolios. Columns 1 and 2 report the average fund daily return gap for the 10 days before and after the end of the fiscal quarter. Return gaps are calculated as the difference between a fund's realized return and the counterfactual return gross of fees based on the stock positions disclosed at the nearest fiscal quarter end (event time 0) and are expressed in percentages. Columns 3 to 5 reports the average ESG beta estimated on post-disclosure realized (Column 3) and counterfactual returns (Column 4), and their difference (Column 5). ESG betas are estimated using a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	Return gap (%)		ESG betas		
	Before	After	β_R^{ESG}	β_C^{ESG}	Diff.
Event days:	[-10, -1]	[2,11]	[2,11]	[2,11]	[2,11]
	(1)	(2)	(3)	(4)	(5)
	-0.016***	0.009***	0.123***	0.173***	-0.050***
	(0.004)	(0.003)	(0.029)	(0.022)	(0.018)
Obs	23,310	23,310	2,331	2,331	2,331

Table A.12: Portfolio sorts

This table reports calendar-time portfolio abnormal returns and factor loadings of portfolios of mutual funds. The explanatory variables are the monthly returns on the market portfolio (Panel A) and on the Fama and French (1993) mimicking portfolios, the Carhart (1997) momentum factor, and the Pástor and Stambaugh (2003) liquidity factor (Panel B). At the beginning of every calendar year, funds are ranked in ascending order on the basis of the average of the four end-of-fiscal-quarter $\Delta\hat{\beta}^{ESG}_{-}$ estimated in the previous year. $\Delta\hat{\beta}^{ESG}_{-}$s are the estimated difference between a fund's ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day. Funds are assigned to one of three tercile portfolios. Alphas are the intercept of regressions on monthly excess returns on the reported factors, in monthly percent. Standard errors are in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: CAPM

	$\Delta\hat{\beta}^{ESG}_{-}$ terciles			
	1(Low)	2	3(High)	3-1
$\alpha(\%)$	-0.275**	-0.059	-0.195	0.079
	(0.107)	(0.086)	(0.121)	(0.083)
MKT	0.904***	0.901***	0.938***	0.034**
	(0.021)	(0.017)	(0.024)	(0.016)
Obs	84	84	84	84

Panel B: 5-factor model

	$\Delta\hat{\beta}^{ESG}_{-}$ terciles			
	1(Low)	2	3(High)	3-1
$\alpha(\%)$	-0.229**	-0.029	-0.147	0.082
	(0.106)	(0.084)	(0.119)	(0.086)
MKT	0.877***	0.887***	0.910***	0.033*
	(0.024)	(0.019)	(0.027)	(0.019)
SMB	0.058	-0.011	0.085*	0.027
	(0.043)	(0.034)	(0.048)	(0.035)
HML	0.020	0.050**	0.014	-0.006
	(0.029)	(0.023)	(0.033)	(0.024)
MOM	-0.038	-0.037	-0.041	-0.003
	(0.030)	(0.024)	(0.034)	(0.025)
ILL	0.004	0.006	-0.020	-0.025
	(0.030)	(0.024)	(0.033)	(0.024)
Obs	84	84	84	84

Table A.13: Do non-ESG funds engage in green window dressing?

This table reports the average difference between funds' ESG exposure in the 10 days before the disclosure date relative to the exposure in the entire month before disclosure, excluding the first and last trading day, considering only funds that do not have an ESG mandate. ESG exposures are from a two-factor model that includes the market portfolio and the Morningstar US Sustainability Index. Standard errors are reported in parentheses. ***, **, * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$	$\overline{\Delta\beta}_{-}^{ESG}$
Window (days):	5	10	15
	(1)	(2)	(3)
	0.064***	0.058***	0.037***
	(0.010)	(0.006)	(0.006)
Obs	105,467	105,467	105,467