Discretionary Information in ESG Investing: A Text Analysis of Mutual Fund Prospectuses*

Angie Andrikogiannopoulou†
King's College London

Philipp Krueger‡
University of Geneva & SFI

Shema Mitali§
SKEMA Business School

Filippos Papakonstantinou¶
King's College London

Abstract

We construct novel measures of mutual funds' environmental, social, and governance (ESG) commitment by analyzing the discretionary investment-strategy descriptions of their prospectuses. We find that fund flows respond strongly to text-based ESG measures. Using discrepancies between text- and fundamentals-based ESG measures, we identify greenwashing funds. We find that greenwashing is more prevalent since 2016 and among funds with lower past flows and higher expense ratios, and isn't associated with superior subsequent performance. Furthermore, greenwashers attract similar flows to genuinely green funds, suggesting that investors cannot distinguish them. Our results could help regulators' efforts to combat ESG-related misconduct.

Keywords: ESG, Prospectus, Greenwashing, Text Analysis
Mutual Funds, Fund Flows, Fund Performance

JEL Classification: G11, G23

---

*For helpful comments, we thank Simona Abis, Matt Eggerton, Andrew Karolyi, Phil MacInnis, Dimitris Papadimitriou, Eduardo Repetto, and Scott Yonker, and participants at the 2022 Cornell ESG Investing Research Conference, the 2022 Conference of the Geneva Institute for Wealth Management, the 2022 Swiss Finance Institute Research Days, the 2023 IESE Machine Learning for Textual and Unstructured Data Seminar, and seminar participants at Bayes Business School, EPFL, the Financial Conduct Authority, Nottingham University Business School, Surrey Business School, the University of Cyprus, and the University of Mannheim. We are also grateful to the Geneva Institute for Wealth Management for their research award. Finally, we thank Hao Song and Zhaoyi Wang for excellent research assistance.

†King's Business School, Department of Banking & Finance, Bush House, 30 Aldwych, London WC2B 4BG, UK, e-mail: aandriko@kcl.ac.uk.

‡University of Geneva, Geneva School of Economics and Management, 40 Bd du Pont-d'Arve, CH 1211 Geneva 4, Switzerland, e-mail: philipp.krueger@unige.ch.

§SKEMA Business School, 5 Quai Marcel Dassaulty, 92150 Suresnes, France, e-mail: shema.mitali@skema.edu.

¶Corresponding author. King's Business School, Department of Banking & Finance, Bush House, 30 Aldwych, London WC2B 4BG, UK, e-mail: fpapakon@kcl.ac.uk.

1 Introduction

In recent years, interest in environmental, social, and governance (ESG) investing has grown exponentially. Google searches for the term "ESG" have grown at an average rate of fifty percent per year, more than quintupling in the period 2018–2023. The rise in ESG interest has also translated into a rapid growth in capital flows to sustainable mutual funds, with global assets under management reaching $3 trillion in 2021 according to Morningstar. The growing significance of ESG considerations in fund management heightens the importance of the information that is available about funds' ESG investments, as this information should in principle facilitate the matching between like-minded investors and fund managers. But, while there is an extant literature that studies the *objective* information available to investors—e.g., information gleaned from portfolio holdings and returns—much less is known about the *discretionary* information that funds themselves release.[1]

In this paper, we study the discretionary ESG information released through fund prospectuses, which enables us to make several contributions. First, we use text analysis to identify mutual funds that discuss ESG investing in the narrative strategy descriptions of their prospectuses. Second, we compare the discretionary with the objective ESG-related information to identify potential *greenwashers*, i.e., funds that overstate in their legally binding prospectus their actual commitment to ESG investing. Finally, we study how investors react to funds' discretionary ESG disclosures. We find that there is often a discrepancy between funds' words (i.e., the discretionary ESG information released) and their actions (i.e., their ESG fundamentals). We also find that, while investors pay attention to both, they can be deceived, which has implications for investors' welfare, asset prices, corporate policy, and society as a whole. For instance, we calculate that, in 2020, $18 billion was misallocated to greenwashing equity mutual funds in the U.S. alone.

Our study is motivated by a growing literature showing that mutual fund investors have limited attention and information processing capacity so they attend more to fund information that is more salient, i.e., easier to access and process (Barber, Odean and Zheng, 2005; Hartzmark and Sussman,

[1]We note that, even though a fund may manipulate holdings disclosures, e.g., to window dress its performance, these disclosures are still a true snapshot of its holdings. On the other hand, what we call discretionary information may not be objective.

2019; Kostovetsky and Warner, 2020). Fund prospectuses are easy-to-access (e.g., through fund websites) and easy-to-read as they need to comply with the plain language requirement of the U.S.'s Securities and Exchange Commission (SEC). As a result, they constitute a more salient source of information about funds' ESG activities than information on fundamentals, such as holdings and returns.[2] As such, when choosing between funds, investors may react strongly to the information disclosed in prospectuses.

We thus start our analysis by constructing a set of novel measures of each mutual fund's engagement with ESG activities that is based on the number and choice of ESG-related words used to describe the fund's Principal Investment Strategy (PIS) in its prospectus. In our baseline analysis we construct our text-based ESG measures using a keyword-based approach, and as a robustness check we create additional measures using a machine learning algorithm (the random forest). Each approach has its advantages: we find that the keyword approach performs better at capturing the *extensive* margin, i.e., classifying PISs as ESG or non-ESG, while the machine learning approach allows us to better capture the *intensive* margin and to conduct more nuanced analysis separately at the E, S, and G levels. This is because, rather than focusing on the most common keywords, the machine learning approach considers the presence, frequency, and interactions of all useful words as determined by the algorithm itself. Throughout the 2011–2020 period that we study, the number of funds that discuss ESG in their PIS has grown significantly, mirroring the growth in the public's interest in ESG investments and rising to about 450 funds (about 10% of the total number of funds in our sample) in 2020. Among funds that discuss ESG, the ESG-related portion of the PIS is, on average, about a fifth of the entire PIS and it appears about halfway through the PIS, though in 13% of cases it starts at the very first sentence.

In addition to our text-based measures of ESG commitment, we also construct a second set of ESG measures that are based on objective or fundamental information. In line with prior research

[2]Mutual fund prospectuses are mandated by SEC regulations as a means of promoting effective communication between funds and their prospective investors. These regulations specifically require that prospectuses contain key information, updated at least annually, about a fund's investment objectives and strategies, risks, expenses, and performance. Furthermore, this information must be communicated in a clear and concise manner that can be understood by an average or typical investor who may not be sophisticated in legal or financial matters.

and industry practice, we use fund holdings data to calculate a weighted average of the ESG scores of the stocks in each fund's disclosed equity portfolio. To account for a fund's entire portfolio rather than just the part for which we have stock-level ESG scores, we also construct a returns-based ESG measure. Specifically, we perform a style analysis of fund returns and calculate the weighted average of each fund's style weights on stock portfolios sorted by their ESG scores. We refer to these holdings- and returns-based measures as "fundamentals-based" measures of funds' ESG intensity.

We find that fund flows respond strongly to text-based ESG measures, and indeed that they respond more strongly to text-based than to fundamentals-based ESG measures. This demonstrates that salience is important in the context of ESG investing, and that fund prospectuses provide an effective means for fund managers to communicate their ESG commitment to potential investors. Consistent with the idea that investors have limited information processing capacity, we also find that a more readable and less ambiguous writing style of the ESG text in the PIS strengthens the effect of text-based ESG intensity on fund flows. We verify that our findings are robust to using alternative fundamentals-based ESG measures. For instance, our findings hold up to calculating our holdings-based ESG measures using data from a variety of ESG data providers that employ different methodologies to assess corporations' ESG practices, as well as to using primary data on specific ESG dimensions that investors may care about such as greenhouse gas emissions (see Bolton and Kacperczyk, 2021). Likewise, our findings hold up if we replace our measures with commercial fundamentals-based ESG fund ratings such as the Morningstar globe ratings. We also verify that the documented relations hold both for institutional and for retail investors. Finally, they are robust to including fund fixed effects, confirming that within-fund changes in the ESG narrative of the prospectus trigger an increase in investor flows.

In the second part of the paper, we study whether funds engage in greenwashing, i.e., if they overstate in their prospectus their ESG commitment, and whether and how investors respond to greenwashing. Our analysis is topical in light of a series of recent greenwashing allegations and penalties against asset managers, and yields insights for investors and regulators who seek to identify cases of ESG-related misconduct.[3] To conduct our analysis, we contrast our text-based and fundamentals-based ESG measures and construct a time-varying measure of greenwashing at the fund-level. The rationale of our greenwashing measure is that our text-based measures are based on discretionary information that a fund chooses to reveal in its prospectus, while fundamentals-based measures are based on objective information on a fund's holdings and returns, the latter of which also accounts for the possibility of ESG-related window dressing. Thus, our greenwashing measure identifies funds that—relative to their actual ESG fundamentals—are likely overstating their ESG commitments in their prospectuses. Specifically, our baseline greenwashing measure is an indicator variable that identifies funds that include ESG-related words in their PIS—placing themselves in the top 5% of funds in terms of text-based ESG—but exhibit a holdings-based ESG measure below the median across all funds in the same investment category in the same month. In setting a high discrepancy threshold when classifying funds as greenwashers—top 5% versus bottom 50%—we essentially focus on relatively egregious greenwashing.

To corroborate whether our measures capture greenwashing behavior, we first explore which fund characteristics are associated with greenwashing. We find that greenwashing behavior is more likely to be observed in the second half of our sample period (i.e., after 2016, coinciding with the signing of the Paris Agreement and the release of Morningstar's globe ratings), and is associated with funds that have lower past flows and higher expense ratios. These findings are consistent with the hypothesis that funds are more likely to greenwash if they have (i) a potential benefit from doing so (as is the case after 2016, when ESG interest is particularly high), (ii) a strong incentive to do so (as is the case when past flows are low), and/or (iii) the opportunity to go undetected (as is the case for funds with weak governance, which often have high expense ratios). Furthermore, we find that greenwashing funds do not have better investment performance, while non-greenwashing funds—i.e., those that truthfully reveal and actually engage in ESG investing—exhibit better subsequent performance. These results indicate that funds we identify as greenwashers are, indeed, observationally different from non-greenwashers in meaningful and predictable ways.

[3]For example, in May 2022, German police raided the offices of DWS and Deutsche Bank over greenwashing allegations and BNY Mellon agreed to pay to the SEC a $1.5 million fine over misstated ESG investment policies by some of its funds.

Subsequently, we study how investors react to greenwashing, i.e., do they recognize the discrepancy between words and actions, or do they respond by rewarding greenwashing funds with additional capital? We find that investors direct their flows to funds that talk about ESG in their prospectus regardless of whether these funds back their words with actions or not. Using our most conservative coefficient estimates, we calculate that the average fund can get an additional inflow of $90 million per year through greenwashing. This means that, in 2020, $18 billion or about 40% of the wealth flowing into actively managed U.S. equity mutual funds that discuss ESG considerations in their PIS is misallocated. The cost of this misallocation is certainly significant, and it would become truly staggering if this level of greenwashing and deception is extrapolated to the $3 trillion global market of sustainable funds or, even worse, to the $35 trillion invested globally in ESG-labeled assets (Walther and Schildbach, 2022). First, investors' welfare is reduced as they miss out on 40% of the societal benefits they would expect to receive from their investment to funds claiming an ESG focus. Second, society is negatively impacted; at a time when the United Nations Intergovernmental Panel on Climate Policy warns that global climate finance needs to increase by a factor of 3 to 6 to avoid the worst scenarios relating to climate change, greenwashing and the resultant misallocation of wealth constitutes a significant impediment to reaching these financing goals or even knowing where we truly stand relative to these goals. Third, greenwashing could potentially reduce by a half—according to our estimates—the effect of responsible investing on firms' cost of capital, hence on corporate policy. We believe that our findings could be used to motivate and design regulation that tackles greenwashing, and our methodology could subsequently be used to help enforce this regulation as well as to assess its efficacy.

We conduct several additional analyses and robustness checks. First, we construct alternative measures of greenwashing; specifically, we set the discrepancy threshold even higher (top 5% versus bottom 25%) or we require that funds also have a low returns-based ESG measure to be classified as greenwashers. In further robustness checks, we repeat the construction of our greenwashing measures using our text-based ESG measures derived from the machine learning algorithm rather than the keyword approach. We also construct measures that capture greenwashing in the

E, S, and G dimensions separately. This analysis enables us to identify greenwashing behavior at a more granular level, and to address the concern that it may be difficult to measure greenwashing using multi-dimensional ESG scores. For example, a fund may care about and discuss E (but not S or G) issues in its PIS, and so it may have a high E but low overall ESG score, meaning that it could be erroneously classified as a greenwasher using the baseline measure but not using the more granular one. We also find that our results are robust to whether a greenwasher is a United Nations Principles for Responsible Investment (PRI) signatory (e.g., Gibson Brandon et al., 2022). Finally, we exclude activist funds to address the concern that our measures erroneously classify as greenwashers funds that discuss ESG issues in their prospectus but invest in firms with low contemporaneous ESG scores with the intention of subsequently improving them via activism.

This paper is related to several strands of the literature. First, it is related to the literature that applies text analysis to information shared with investors (see Loughran and McDonald, 2016, for an early review). Most studies focus on corporate regulatory filings such as 10-K reports,[4] while a few recent studies apply text analysis to information shared by mutual funds in their prospectuses. For example, Krakow and Schäfer (2020) and Sheng, Xu and Zheng (2022) study the information content of fund risk disclosures. Kostovetsky and Warner (2020) analyze investment strategy descriptions and find that product differentiation—measured by the uniqueness of a fund's strategy description—has a short-term positive effect on fund flows. Abis (2022) applies the random forest algorithm to strategy descriptions of fund prospectuses to classify them as quantitative or discretionary, while Abis et al. (2022) analyze these strategy descriptions to determine funds' optimal disclosure policies. Furthermore, Abis and Lines (2022) use the same data to allocate funds to peer groups and find that funds' portfolios do not deviate much from the average portfolio in their peer group, as such deviations are penalized by the market through capital outflows. Our paper contributes to this literature by studying the ESG content of fund prospectuses and, most importantly, the effect of greenwashing on fund flows. Contrary to Abis and Lines (2022), we

[4]For example, Hoberg and Phillips (2016) construct a firm's set of competitors by calculating product similarities from 10-K report product descriptions, and Cohen, Malloy and Nguyen (2020) show that substantial active changes in firms' 10-K reports predict important changes in their operations and returns.

find that, in the context of ESG investing, there *are* substantial deviations between what funds talk about and what they invest in, and that investors tend to pay more attention to the former and can be deceived. This is possibly because ESG is a hot investment theme, hence divergence between words and actions is more likely to occur as managers are tempted to attract investors' attention. Furthermore, ESG investing is still not well-understood by investors, so evaluating such divergence may be more complicated than for traditional investment styles.

Second, our paper is related to the literature documenting that funds take cosmetic actions to exploit investor sentiment. For example, Cooper, Gulen and Rau (2005) find that funds change their names to reflect "hot" investment styles without making corresponding changes to their holdings. Using different settings and samples, recent papers have highlighted that some institutional investors sign the United Nations Principles for Responsible Investment (PRI), thereby committing to ESG investing, but do not follow through on their ESG commitments (Liang, Sun and Teo, 2022; Gibson Brandon et al., 2022; Kim and Yoon, 2022). **?** show that ESG mutual funds, as defined by Morningstar, invest in companies with bad track records regarding environmental laws. We contribute to this literature by using regulatory filings to construct a novel measure of greenwashing that is based on text analysis of fund prospectuses, which (i) relative to fund names, change more frequently, (ii) relative to hedge funds, represent an economically larger group of institutional investors, and (iii) relative to PRI sign-ups, are legally binding and tend to be a more nuanced way of signaling to investors the fund's commitment to ESG investing.

Third, our paper is related to research concerned with mutual fund ESG strategies, and particularly their effect on fund flows. For instance, Hong and Kostovetsky (2012) show that the political orientation of a fund manager affects the fund's ESG strategy. Other papers have focused on the effect of commercial fundamentals-based ratings like Morningstar's sustainability fund ratings on investor flows (e.g., Hartzmark and Sussman, 2019; Ammann et al., 2019). Our results confirm that commercial ESG ratings affect flows but suggest that the effect of text-based ESG measures is stronger. Furthermore, while the previously documented effect of Morningstar's ratings appears to be short-term (Gantchev, Giannetti and Li, 2021), the effect of text-based ESG measures we find is longer-term. More broadly, our finding on the effect of text-based ESG measures on fund flows contributes to a large literature that studies the determinants of mutual fund flows, including past performance (e.g., Chevalier and Ellison, 1997), fees (Barber, Odean and Zheng, 2005), marketing (e.g., Cooper, Gulen and Rau, 2005, Jain and Wu, 2000, Kostovetsky and Warner, 2020), and fund manager characteristics (e.g., Kumar, Niessen-Ruenzi and Spalt, 2015), among others.

Finally, our results also speak to an important debate on the potential impact of ESG mutual funds and other institutional investors on stock returns and the cost of capital. Berk and van Binsbergen (2022) argue that the shift of institutional investors' investments away from brown firms are too small to affect the cost of capital. van der Beck (2022) and **?** find that, on the other hand, institutional investors have impacted the value of green stocks, due to increased ESG preferences. Our results add that current estimates might be lower than previously documented, when accounting for the presence of greenwashers among U.S. mutual funds.

The rest of the paper is organized as follows. In Section 2, we discuss our data sources. In Section 3, we describe how we construct our text-based and fundamentals-based ESG measures and our greenwashing measure that captures the discrepancy between words and actions. In Sections 4 and 5 we present the results from our analysis, and in Section 6 we conclude.

2 Data

To construct the data we use, we draw from a variety of data sources that we describe below.

2.1 Fund prospectuses

We obtain data on funds' investment strategies from the "Mutual Fund Prospectus Risk/Return Summary Data Sets" that the SEC makes available on its website.[5] The data contain numeric and text information extracted from the risk/return summary section of mutual fund prospectuses

[5] See https://www.sec.gov/dera/data/mutual-fund-prospectus-risk-return-summary-data-sets.

filed with the SEC in a structured XBRL format.[6] For each prospectus (485BPOS form) filed by a fund, we observe fund identifying information (e.g., fund name, fund family's Central Index Key, and Series ID), the document's filing date, and all narrative text blocks included in the summary section (e.g., a description of the fund's investment strategies, exposed risks, and past performance). In our analysis, we focus on the *Principal Investment Strategy* (PIS) text block in which funds are required to provide information about their principal investment strategies, and specifically to identify the types of securities in which the fund principally invests; to describe any policy, practice, or technique used by the fund to achieve its investment objectives; and to explain in general terms how the fund's adviser decides which securities to buy and sell.[7] Our data start in January 2011 (when funds started submitting XBRL data) and stop in June 2020. Funds are required by the SEC to update their summary information at least once per year; they usually keep to this annual schedule, but in about one out of ten years they also provide a mid-year update. Importantly, we note that, for the period up to 2018, all text blocks in the flattened XBRL data provided by the SEC are truncated at 2,048 characters, which affects about half of all PIS text blocks and would be problematic for our analysis as it relies on analyzing this text. To address this problem, we write our own code that downloads the original prospectuses from the SEC's EDGAR website and extracts the full text under the PIS section. In Appendix A, we present some examples of fund PIS descriptions.

2.2 Fund information

We obtain fund-level information for the period January 2010 to June 2020 from the Center for Research in Security Prices (CRSP) Survivorship-Bias-Free U.S. Mutual Fund Database. We focus on U.S. actively managed funds that invest in domestic and foreign equities, so we exclude fixed

[6]The risk/return summary section appears at the beginning of the statutory prospectus (sometimes also filed as a separate summary prospectus) and contains key information about the fund in the following standardized order: (i) investment objectives; (ii) fees and expenses; (iii) principal investment strategies, risks, and performance; (iv) investment advisers and portfolio managers; (v) purchase and sale of fund shares; (vi) tax information; and (vii) financial intermediary compensation. Starting in 2011, all funds are required to report this information in the eXtensible Business Reporting Language (XBRL) format which organizes the filed data under different tags making it easier to analyze and compare over time and across funds. Funds are also required to make this information available on their own websites.

[7]See a reference copy of the form funds need to file, with instructions, at https://www.sec.gov/about/forms/formn-1a.pdf.

income, money market, balanced, index, exchange-traded, and variable annuity funds. To exclude these, we use the CRSP objective codes and flags as well as keyword searches in fund names.[8] The CRSP data include share class-level characteristics such as net returns at a daily frequency, total net assets (TNA) at a monthly frequency, and expense, portfolio turnover, 12b-1 and load fee ratios at an annual frequency. They also contain comprehensive information on all fund holdings at a quarterly frequency, including long and short positions in U.S. and international equities.[9]

Our analysis is conducted at the fund-month level. We use the CRSP_CIK_MAP file provided by Wharton Research Data Services to link the CRSP share class identifiers (crsp_fundno) with the fund identifiers in the SEC filings (CIK-Series ID pairs).[10] We then aggregate all share class data at the fund level using the fund's Series ID. While all fund share classes have the same prospectus and holdings, they have different TNA, fees, and net returns. We compute a fund's TNA as the sum of its classes' TNAs, and a fund's net return (fees) as the weighted average of its classes' returns (fees), with weights equal to the beginning-of-month TNA value of each class. To improve the accuracy of our return data, we omit any monthly return that directly follows a missing one, as it may compound multiple months' returns.

We calculate monthly fund flows for fund i during month t as

$$Flows_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}\left(1 + r_{i,t}\right)}{TNA_{i,t-1}\left(1 + r_{i,t}\right)}, \tag{1}$$

where $TNA_{i,t}$ is total net assets and $r_{i,t}$ is net return of fund i over month t. To reduce the effect of outliers, we winsorize fund flows at the 1% and 99% levels.

Similar to Roussanov, Ruan and Wei (2021), we adjust a fund's expense ratio for amortized loans and we proxy for marketing expenses by combining a fund's 12b-1 fee and front load.

Finally, we calculate monthly fund alphas based on the 4-factor model of Carhart (1997) using

[8] CRSP objective codes are constructed using objective codes from three sources (Wiesenberger, Strategic Insight, and Lipper). We exclude ETFs and variable annuity funds based on the *et_flag* and *vau_fund* variables in CRSP. We remove index funds using the *index_fund_flag* variable, by searching for the words "INDEX", "S&P", "DOW 30", and "NASDAQ" in fund names, and by removing all funds in specific fund families (Dimensional Fund Advisors, Direxion, Potomac, ProFunds, and Rydex) which follow index-like strategies.

[9] We note that, starting in 2008 (hence for our entire sample period), the CRSP mutual fund holdings database is more complete and reliable than the Thomson Reuters s12 database.

[10] We match each prospectus to the CRSP fund-month observation in the month after the filing, and retain this prospectus for subsequent CRSP fund-months until a new prospectus is filed.

daily fund and factor returns. We use the CRSP NYSE/Amex/NASDAQ value-weighted index as the market factor, the one-month Treasury bill rate as the risk-free rate, and the factor-mimicking portfolios for size, book-to-market, and momentum, all downloaded from Kenneth French's website.

Our final sample contains 4,863 funds, which have filed 37,399 prospectuses between January 2011 and June 2020. In Panel A of Table 1, we present summary statistics for the characteristics of all funds in our sample. The average fund is about 13 years old, has total net assets of 1.6 billion U.S. dollars, fund inflows of 0.39% per month, and net returns (4-factor alpha) of 0.78% (−0.10%) per month, and is equally likely to be primarily targeted to institutional or to retail investors.

2.3 ESG ratings and carbon emissions data

We obtain ESG data from several providers, namely MSCI, Sustainalytics, Refinitiv (formerly Thomson Reuters ASSET4), and Morningstar.

From MSCI, Sustainalytics, and Refinitiv, we obtain annual firm-level ESG scores, which we use to calculate fund-level scores (see Section 3.2 below). MSCI ESG scores capture a company's resilience to ESG risks, measured at the GICS (Global Industry Classification Standard) sub-industry level, on a scale from 0 to 10. Sustainalytics computes a score, from 0 to 100, based on a company's ESG risk relative to its industry peers and how well it manages that risk. Refinitiv measures, on a scale from 0 to 100, a company's ESG performance relative to its TRBC (The Refinitiv Business Classification) industry group peers, using several themes such as carbon emissions, human rights, or green innovation. Most often, data providers construct these ESG scores by combining firm-reported data (e.g., from sustainability reports) on how firms manage issues such as environmental impacts, workplace safety, or human rights concerns, and information reported by third parties (e.g., non-governmental organizations). While methodologies can vary across data providers, all these measures generally capture the quality of a firm's ESG policies and practices. Since the MSCI database has the best cross-sectional coverage and has been found to contain the least amount of noise for U.S. firms (Berg et al., 2021), we use it as our primary source. However, as a robustness check, we also construct a combined score that averages the standardized z-scores from all three data providers. For more details on the construction of ESG scores from different providers and their correlations see, for instance, Gibson Brandon, Krueger and Schmidt (2021) and Berg, Koelbel and Rigobon (2022).

From Morningstar, we obtain monthly fund-level globe ratings, which were first made available to investors in March 2016. The globe rating of a fund is based on the percentile rank of its portfolio sustainability score relative to other funds in the same Morningstar category. This portfolio sustainability score is a value-weighted average based on firm-level ESG scores from Sustainalytics and an adjustment that accounts for a company's involvement in ESG-related controversies.

Finally, we complement the ESG scores with firm-level carbon emissions (CO_2) data from Refinitiv. Specifically, for each firm, we obtain its annual total emissions, defined as the sum of direct and indirect emissions. Direct emissions are the equivalent of Scope 1 emissions from the Greenhouse Gas (GHG) Protocol, that is emissions from sources that are owned or controlled by the firm, from the generation of intermediate energy, from transportation/business travel where the transportation methods are owned by the firm, emissions from the operation of buildings, and emissions associated with the sale of own-generated electricity to another firm. Indirect emissions are the equivalent of Scope 2 GHG Protocol emissions, that is they include emissions arising from the generation of imported (purchased) electricity, heat, or steam consumed by the firm; they exclude Scope 3 GHG Protocol emissions, as these emissions are not controlled or owned by the firm.

3 Constructing fund-level ESG measures

We construct two sets of ESG measures for each fund. The first uses fund prospectuses to capture how and how much funds talk about ESG investments. The second uses fund fundamentals (holdings and returns) to capture funds' actual ESG investments. Subsequently, we use our text-based and fundamentals-based ESG measures to devise a measure of greenwashing for each fund. This greenwashing measure captures the discrepancy between words and actions.

3.1 Text-based ESG measures

To construct text-based ESG scores for funds, we perform text analysis on the PIS text block of fund prospectuses. For our baseline analysis, we construct our measures using a keyword-based approach. As a robustness check, we subsequently create additional measures using a machine learning (ML) algorithm. In our context, we find that the keyword approach performs better at capturing the extensive margin, i.e., classifying a PIS as ESG or non-ESG. On the other hand, the ML approach considers the presence, frequency, and interactions of all useful words as determined by the algorithm itself, and so it allows us to better capture the intensive margin and to conduct more nuanced analysis separately at the E, S, and G levels.

3.1.1 Keyword-based method

Our keyword-based approach involves the following steps. First, we pre-process each fund's PIS text by removing non-alphabet characters and punctuation marks that do not end sentences, and by changing all letters to lower case. Each PIS is then split into sentences. Then, we manually read several hundreds of randomly chosen PIS sections and form a list of ESG keywords/phrases. In constructing this keyword list, we are conservative in that we prefer to limit false positives at the expense of more false negatives, because this reduces bias when we use our ESG (and greenwashing) measures in regression analysis; see Gustafson (2003).[11] Finally, we search for these keywords in all funds' PIS sections, and we calculate several text-based measures of ESG intensity, positioning, readability, tonality, and uniqueness. In Table 2, we show our list of ESG keywords and the frequency with which we encounter them across all PISs in our sample; in Figure 1 we show the evolution over time of the ESG keywords' prevalence in fund prospectuses.

ESG intensity. Our main measure of ESG intensity is the relative frequency of ESG mentions in a fund's PIS, i.e., the number of occurrences of ESG keywords divided by the number of words

[11] For example, consider a dummy indicating that a PIS is ESG; a large majority of its values are zero and a small minority are one. In a regression, false positives greatly affect the estimated mean outcome of the (small) ESG class while false negatives have little effect on the estimated mean outcome of the (large) non-ESG class, and so false positives induce bias in the dummy's estimated coefficient while false negatives do not.

in the PIS. As our list of ESG keywords is not exhaustive, we also construct an alternative measure which captures the relative length of the ESG-related portion of the PIS, i.e., the number of words in sentences containing at least one ESG keyword divided by the total number of words in the PIS.

ESG positioning. This measure is based on the location of ESG keywords in the PIS. Our main measure is the proportion of the text (or, in robustness checks, the number of words) from the beginning of the PIS to the beginning of the first sentence containing an ESG keyword. Alternatively, we construct a dummy indicating if at least one ESG keyword appears in the first sentence of the PIS.

ESG readability. We construct two measures of readability for a PIS's ESG portion (i.e., all sentences containing an ESG keyword): the Flesch Reading Ease and the Gunning Fog Index.[12]

ESG tonality. We measure the vagueness of the ESG strategy description based on the frequency of words indicating uncertainty in the ESG portion of the PIS, e.g., some funds write "we *may* consider ESG factors in our investment strategy."

ESG uniqueness. Similar to Kostovetsky and Warner (2020), after removing from all PISs generic stop words that have no information content (e.g., "and" or "the"),[13] (i) we calculate the pairwise unique-word overlap between the ESG portion of a fund's PIS and that of all other funds in the same year, (ii) we divide this overlap by the pairwise minimum of the unique-word counts to turn it into a ratio in [0, 1], and (iii) we calculate the mean across all pairs and flip its sign to turn it from a *similarity* to a *uniqueness* measure. We repeat this procedure for all funds to obtain a raw uniqueness measure, which we then standardize to obtain a measure with mean zero and variance one. Finally, we regress each text's standardized uniqueness on the number of words it contains and use the regression residuals as a measure that captures uniqueness above the expected uniqueness given a text's length.

To control in our analysis for a fund's overall writing style, we also calculate the above measures

[12]Flesch Reading Ease is defined as $206.835 - 1.015\frac{\#\text{words}}{\#\text{sentences}} - 84.6\frac{\#\text{syllables}}{\#\text{words}}$ (Flesch, 1948). The Gunning Fog Index (Gunning, 1952) estimates the years of education needed to understand a text, e.g., a value of 6 corresponds to sixth grade. It is defined as $0.4 \cdot \left(\frac{\#\text{words}}{\#\text{sentences}} + 100\frac{\#\text{complex words}}{\#\text{words}}\right)$, where complex words are roughly those with three or more syllables. We flip this measure's sign so that larger values correspond to more readable text (as for Flesch Reading Ease).

[13]We identify uncertain words (used for the tonality measure above) using the Loughran-McDonald sentiment word list and stop words using their generic and long generic lists; these lists can be found at their website at https://sraf.nd.edu.

of readability, tonality, and uniqueness for the entire PIS as well as for the non-ESG portion of the PIS. All definitions are similar to the above, except that text uniqueness is calculated relative to funds in the same CRSP objective code (rather than the universe of all funds) in the same year, to account for systematic differences in PIS content across investment styles.[14]

We calculate that the entire PIS has a mean (median) length of 404 (347) words, is difficult but readable by college graduates (median Flesch Reading Ease score of 20), and, on average, 2.3% of its words are vague. Focusing on prospectuses that contain an ESG keyword, the ESG-related portion of the PIS has a mean (median) length of 107 (65) words, so about a fifth of the entire PIS. The first sentence containing an ESG keyword appears on average about halfway through, though in 13% of the cases an ESG keyword appears in the very first sentence. Furthermore, the ESG portion of the PIS has substantially lower readability than the entire PIS (median Flesch Reading Ease score of −7 vs. 20 for the entire PIS). Finally, a lower proportion of words (1.6% vs. 2.3% for the entire PIS) are uncertain. The differences in text readability and tonality between the ESG portion and the entire PIS are both statistically significant at the 1% level. See the Internet Appendix for additional summary statistics on the text characteristics of the PIS.

In Panel A of Table 3, we present summary statistics for our text-based ESG measures. ESG keywords make up 0.04% of the words in the average PIS, and 0.62% of the words in a PIS that contains ESG keywords. The PIS contains ESG keywords in a little over 4% of all fund-months.

Finally, in Panel B of Table 1, we present summary statistics for the characteristics of the funds whose PIS contains at least one ESG keyword. Comparing the mean characteristics of the subsample of funds whose PIS contains at least one ESG keyword (in Panel B) with those of the entire sample of funds (in Panel A), we see that the former are younger (11.5 vs. 13.3 years), smaller (632 vs. 1,604 million U.S. dollars in total net asset value), have a lower turnover ratio (57% vs. 74%), higher fund inflows (0.67% vs. 0.39% per month), and higher net returns (1.11% vs. 0.78% per month), with all differences statistically significant at the 1% level. On the other

[14] We do not calculate ESG text uniqueness separately for funds with different investment styles for two reasons. First, we want to capture ESG text uniqueness even if it is related to style. Second, the number of funds with ESG keywords in their PIS is small, so calculating style-specific ESG text uniqueness would result in noisier measures.

hand, the expense ratio, marketing expenses, and the 4-factor alphas are quite similar, and the proportion of funds targeted to institutional investors is 50% for both samples.

3.1.2 Machine learning method

Next, we construct text-based ESG measures using a machine learning approach, specifically a random forest. Here, we discuss our approach and in the Internet Appendix we provide additional details.

We start by pre-processing each fund's PIS text, that is: (i) we remove non-alphabet characters and stop-words of various types,[15] (ii) we remove uninformative words, in the sense that they are in the top 10% of usage in absolute and in relative terms, and (iii) we replace the various inflected forms of words with a single invariant root form using the Porter2 (Porter and Boulton, 2001) stemmer, e.g., we replace "value" and "valuable" with "valu". In Appendix B, we show an example of how we pre-process a PIS.

Next, we select the subsample of PISs that we manually pre-classify to train and test the ML algorithm. To this end, we randomly select one PIS for each of the 4,863 funds in our sample, and among these we randomly select 1,500 PISs. This renders the pre-classified sample quite representative of the full sample in the cross-sectional and time-series dimensions. PIS sections are pre-classified as E, S, G, and/or ESG by two expert classifiers (post-graduate finance students—one pre-doctoral and one doctoral—at King's College London) subject to the following ground rules. First, a PIS with a generic "environmental, social, and governance" reference is classified as ESG but not as E, S, or G. Second, a PIS that only contains a reference to sound accounting practices or management stock ownership is not classified as G or ESG. Most classification disagreements are due to inadvertent errors that are corrected, and the remaining are discussed and resolved. We consider these classifications to be the "ground truth". We then split the pre-classified sample into training (80%) and testing (20%) subsamples using stratified sampling, so that the proportion of

[15] Specifically, we remove some common generic words (e.g., "is", "the"), financial words (e.g., "Nasdaq", "FTSE"), and geographic names, and dates and numbers. For lists of stop-words we use, see Python's sklearn, nltk, and spacy packages, Matlab, and Bill McDonald's website at https://sraf.nd.edu.

ESG and non-ESG PISs is the same in both. Subsequently, we process the pre-classified PISs using the bag-of-words approach. First, we compile a list of all *features*—single words and two-word combinations—mentioned in any PIS in the training subsample. Then, we transform each PIS into a vector of 0s and 1s indicating whether it contains each of the features. Finally, we transform this vector of feature indicators into a vector of feature weights using the term frequency-inverse document frequency (tf-idf) approach; specifically, a feature's weight in a PIS increases with the feature's frequency in the PIS and decreases with the proportion of PISs that contain the feature.

We train the random forest algorithm on the training sample of the pre-classified PISs, or *cases*, each consisting of a vector of feature weights and a ground-truth classification. To construct a random forest, we construct a number of *decision trees*, each consisting of nodes and two branches emanating from each non-leaf node. At each node, starting from the root of a tree, one of the features is chosen to split the cases: the ones with large weights for this feature are allocated to one child node and the rest to the other child node. A variety of metrics can be used to determine the "best" feature split, largely measuring the class homogeneity across cases in the children nodes; a commonly used metric is information gain. This process of feature splitting is repeated at each node until no more information can be gained. Once a decision tree is built, unseen cases can be classified by following the feature splits along the tree until a leaf (hence classification probability) is reached.[16] A drawback of decision trees is that they are prone to overfitting the training sample. To alleviate this, a random forest algorithm constructs a large number of decision trees, introducing randomization at two levels to enable training on different parts of the information in the training sample. First, a bootstrap subsample of the training sample is drawn to generate each tree.[17] Second, at each node in each tree, a subset of all features is randomly selected, among which the optimal feature split is then chosen. Once a random forest is built, unseen cases are allocated to the class selected by the majority of trees.

To further avoid overfitting our random forest to the training sample, we do two more things. First, we use cross-validation to optimally choose the algorithm's hyper-parameters, specifically the

[16]For illustration purposes, Figure 3 shows a fictitious example of a decision tree.

[17]Given that most PISs are non-ESG, we avoid training a classifier that is biased toward this majority class by doing *weighted* bootstrap sampling, i.e., we create bootstrap samples that contain an equal number of cases from each class.

number of decision trees in the forest, the size of the bootstrap subsample used to generate each tree, the maximum depth of each tree, and the number of features considered at each node split.[18] Second, we use Shapley values (Lundberg and Lee, 2017) to identify features that are important in the classification task in the training sample but that do not make much sense, hence are likely to be overfitted (see the Internet Appendix for details on our iterative procedure). For example, we remove features such as "target", "region', and "research". With this finalized set of features, we train the algorithm on the entire training sample, and in Figure 4 we show the most important features for this algorithm.

In Figure 5, we illustrate how the ML algorithm works by showing, for five specific PISs, how the various features present in or absent from a PIS contribute to the algorithm's prediction of class allocation. For example, in Panel (a) the PIS is classified as ESG with probability 0.98; the feature that contributes the most to this classification is "esg" followed by "environment" and then "social". We see that the tf-idf weight of "esg" (0.04) is smaller than that of "social" (0.28), because "esg" is more common across PISs, but the presence of "esg" in the PIS is much more informative about classifying it as ESG. In Panel (b), the PIS is classified as ESG with probability 0.90; the features that contribute the most are "tobacco", followed by "ethic", while the absence of the features "esg" and "sustain" from the PIS lowers its predicted probability of belonging to the ESG class.

Using the testing subsample, we calculate multiple performance metrics for the ML as well as for the keyword approach. We find that the keyword method has better overall performance than the ML algorithm, albeit by a small margin: across both classes (ESG and non-ESG), we obtain 98% accuracy with the former vs. 97% with the latter. For the non-ESG class, the keyword method is marginally superior, as it has higher recall (99% vs. 98%) and about the same precision (98%). For the ESG class, the keyword method trades off a substantially higher precision (100% vs. 85%), meaning no false positives, for a marginally lower recall (73% vs. 77%), meaning more false negatives.[19] Indeed, this reflects our original intention when choosing the keywords, which

[18] A larger number of trees increases accuracy, but only up to a certain point. Higher values for the other hyper-parameters generally lead to more powerful individual trees, but may lead to overfitting. As we discuss in detail in the Internet Appendix, we use stratified k-fold cross-validation to pick the hyper-parameters that yield the highest "out-of-sample" accuracy metrics.

[19] Accuracy is the proportion of correctly predicted classifications across both classes, recall is the accuracy within each class, and precision is a classification's predictive value within each class. In the Internet Appendix, we also report

was to limit false classifications into the ESG class at the expense of more false omissions from this class. See the Internet Appendix for detailed performance metrics for ESG classification and for the more granular E, S, and G classifications.

Finally, we use the trained random forest model to classify *all* PISs. As expected, in Table 3 we see that the ML method is more likely to classify a PIS as ESG than the keyword method (6.6% vs. 4.3% of all fund-months). But we also see that the correlation between two dummies indicating if a PIS is classified as ESG according to each method is a high 0.74. Furthermore, we see that the ML method classifies a PIS as E, S, and G for 2%, 3.4%, and 2.3% of all fund-months, respectively.

3.2 Fundamentals-based ESG measures

Holdings-based ESG measure. Our primary measure for assessing the ESG intensity of fund fundamentals is based on the ESG scores of the fund's most recent portfolio holdings. This holdings-based ESG measure is defined as the value-weighted average of the stock-level ESG scores of the stocks included in a fund's portfolio. Formally, we define the measure for fund i in month t as

$$ESG^H_{i,t} := \frac{\sum_{s \in S_t} V_{i,s,t} ESG_{s,t}}{\sum_{s \in S_t} \left| V_{i,s,t} \right|}, \tag{2}$$

where $V_{i,s,t}$ is the value of fund i's most-recently reported holdings in stock s, $ESG_{s,t}$ is the ESG score for stock s in month t, and S_t is the universe of stocks for which we have an ESG score in month t.[20] As noted previously, in our baseline calculation we use stock-level ESG scores from MSCI but, for robustness, we also use a combined stock-level ESG score that averages the standardized z-scores from MSCI, Sustainalytics, and Refinitiv. Furthermore, we repeat the same calculation replacing stock-level ESG scores with stock-level carbon emissions levels. Finally, we use the fund-level globe ratings which are calculated according to Morningstar's methodology using Sustainalytics' stock-level ESG and controversy scores.

more sophisticated metrics like the class-specific F_1 score which symmetrically takes into account both recall (so false negatives) and precision (so false positives) and the P_4 metric which is the harmonic mean of the F_1 scores across classes.

[20]Note that $V_{i,s,t}$ is positive for long and negative for short positions. We divide by the sum of the absolute value of the holdings in each stock, because otherwise the measure is not well defined (e.g., if the total value of longs equals that of shorts). In any case, short positions are very rare in the funds we study, with about 0.5% (0.2% by value) of all positions being shorts.

Returns-based ESG measure. We also calculate a returns-based ESG measure using an approach inspired by the returns-based style analysis of Sharpe (1992). While the holdings-based measure discussed above has the advantage that it provides a more direct measure of a fund's level of ESG investment, the returns-based measure we discuss here has the advantage that it is immune to potential window dressing in fund holdings (see, e.g., Kacperczyk, Sialm and Zheng, 2008 and Agarwal, Gay and Ling, 2014) and that it is based on a fund's entire portfolio rather than just the part for which we have stock-level ESG scores.

To calculate the returns-based ESG measure, we use the asset-class factor model

$$r_{i,d} = \alpha_i + \mathbf{F}_d'\mathbf{w}_i + \varepsilon_{i,d}, \tag{3}$$

where $r_{i,d}$ is the net rate of return of fund i in day d, $\mathbf{F}_d$ is a vector of returns of different asset classes on day d, $\mathbf{w}_i$ is the vector of fund-specific style-weights that are positive and required to sum to 1, and $\varepsilon_{i,d}$ is the residual component that represents the return attributable to selection rather than style. The model we use contains several asset classes, similar to Sharpe (1992): U.S. stocks, emerging markets stocks, developed markets stocks, government bonds, corporate bonds, and mortgage-backed securities. U.S. stocks in particular are partitioned into quintiles based on their MSCI ESG scores; specifically, the return of each ESG quintile is the market-capitalization value-weighted return of the stocks allocated to this quintile.

For each fund, we estimate this model on a monthly rolling basis using 12 months of daily returns, and thus we estimate the time-varying portfolio weight of each asset class. For each fund-month, we then use the style weights of the ESG quintiles to calculate a measure of ESG intensity. This measure is $ESG^R_{i,t} = \sum_{q=1}^{5} q \cdot w^q_{i,t}$, where $w^q_{i,t}$ is the style weight of the q^{th} ESG quintile for fund i in month t, with 1 (5) corresponding to stocks with the lowest (highest) ESG scores.

In Table 3, we present summary statistics for all our fundamentals-based ESG measures (Panel B) and correlations between the text-based and fundamentals-based ESG measures (Panel C). For our fundamentals-based ESG measures, we see that their average value is near the midpoint of their range, and their spread is quite narrow (standard deviation is about a fifth that of the *firm*-level

ESG scores). Furthermore, all correlations are positive and statistically significant at the 1% level. The correlation of our baseline text-based ESG measure with the fundamentals-based measures is low, while the fundamentals-based measures are more highly correlated with each other.

3.3 Measures of greenwashing

Finally, we use our text-based and fundamentals-based ESG measures to devise multiple measures of greenwashing for each fund in each month that capture whether funds overstate in their PIS their level of actual ESG investing. Our first measure of greenwashing is an indicator variable that indicates funds that talk about ESG in their PIS—placing themselves in the extreme top 5% of funds in terms of what they say about ESG—but the ESG measure based on their most-recently reported holdings is below the median across all funds in the same CRSP investment category in the same month.[21]

In setting a high discrepancy threshold when classifying funds as greenwashers—top 5% versus bottom 50%—we essentially focus on relatively egregious greenwashing. We also consider alternative measures, where we set the discrepancy threshold even higher but the trade-off is that doing so restricts the sample of greenwashers and so lowers the statistical power of our analysis. Specifically, we classify as greenwashers funds that talk about ESG in their prospectus (i.e., they are in the top 5% in terms of what they say) but (i) their holdings-based ESG score is in the bottom quartile (i.e., they are in the bottom 25% in terms of what their holdings indicate) or (ii) both their holdings-based and their returns-based ESG score are below the median. In the preceding greenwashing measures, we use our baseline holdings-based ESG scores based on MSCI, but we also construct a measure that uses holdings-based scores based on multiple data providers (MSCI, Sustainalytics, and Refinitiv). Results are similar in all cases.

In robustness checks, we repeat the construction of our greenwashing measures using our text-based ESG measures derived from the ML algorithm rather than the keyword approach. We

[21]To avoid misclassifying as greenwashers funds that have just started making ESG investments, updated their PIS to include ESG mentions, but have not disclosed their updated portfolio holdings yet, our greenwashing dummy is not defined for all months after the PIS change until the fund discloses its new portfolio holdings.

also construct measures that capture greenwashing in the E, S, and G dimensions separately. Finally, we address the concern that our measures may erroneously classify as greenwashers funds that talk about ESG in their prospectus but invest in firms with low current ESG scores in order to subsequently improve them via shareholder activism.

In Panel D of Table 3, we present summary statistics for our main greenwashing measures. According to our baseline measure of ESG greenwashing, on average 1.5% of all funds greenwash across the years. More specifically, Figure 2d shows that the proportion of greenwashers starts at about 1.2% in 2011, drops to a low of about 0.8% in 2015, and then rises almost exponentially year-by-year until it reaches a high of 5.6% at the end of our sample in 2020.[22]

Next, we use the various measures we have constructed to analyze investor reaction to funds' ESG commitment and, subsequently, to greenwashing.

4 Do investor flows respond to text-based ESG measures?

We start by studying how investors respond to a fund's ESG commitment—stated or actual. Do investors pay attention to information contained in fund prospectuses? Or do they pay attention to fund fundamentals such as information revealed by portfolio holdings?

Our baseline specification is

$$Flows_{i,t} = \alpha_{s,t} + \beta_T ESG^T_{i,t-1} + \beta_H ESG^H_{i,t-1} + \beta' \mathbf{x}_{i,t-1} + \varepsilon_{i,t}, \tag{4}$$

where $Flows_{i,t}$ are the flows of fund i in month t defined in Equation 1, $ESG^T_{i,t-1}$ is the fund's text-based ESG intensity score in month $t-1$ based on its most recent prospectus (see Section 3.1), and $ESG^H_{i,t-1}$ is the fund's holdings-based ESG score in month $t-1$ based on its most recently disclosed portfolio (see Equation 2). $\alpha_{s,t}$ are category-by-month fixed effects that control for omitted time-varying heterogeneity at the category level. In our baseline analysis, we control for

[22]This pattern is likely because, while the number of funds increases for most of this period and peaks in 2018 (see Figure 2a), the interest in, hence the number of, ESG funds is almost flat until 2015 and then rises exponentially (see Figure 2b), and so the proportions of both ESG funds and greenwashers (see Figures 2c and 2d) initially dip and then sharply increase.

fund investment category using the CRSP objective codes which are stable over time. To allow for time-varying investment styles, in the Internet Appendix we report similar results when fund category is inferred from the exposure of the past 24 months of fund returns to the Fama and French (1993) factors. Finally, $\mathbf{x}_{i,t-1}$ is a vector of fund characteristics that might affect fund flows, such as age, size, expense ratio, marketing expenses, and past performance. Standard errors are heteroskedasticity-consistent and clustered by fund and year-month.

In Table 4, we report the results from estimating Equation 4. Across all specifications, our text-based ESG measures have a consistent strong positive effect on fund inflows, which is robust to controlling for a variety of holdings-based ESG measures. Specifically, in Columns 1–3, the text-based ESG measure we use is the frequency of ESG keywords in the PIS; in Column 4 we use the relative length of the ESG-related portion of the PIS; and in Columns 5–7 we replace the continuous ESG measure with a dummy indicating if the PIS contains at least one ESG keyword or a dummy indicating if the frequency of ESG keywords is high (above median across funds whose PIS contains ESG keywords).[23] We control for the ESG intensity of fund holdings using the raw ESG score (Columns 2 and 4), the percentile rank of the ESG score within the fund's investment category-month (Column 3), and dummies indicating if the fund's ESG score is high (above median) or exceptionally high (above the 90$^{\text{th}}$ percentile) across funds in the same investment category-month (Columns 5–7).[24]

The coefficients on our text-based ESG measures (β_T) are economically large and statistically significant at the 1% level across all specifications. Specifically, since the mean (standard deviation) of ESG-keyword frequency among funds that talk about ESG investments in their PIS is about 1% (1%), a coefficient of about 0.3 on ESG-keyword frequency in Columns 2–3 suggests that (i) funds that talk about ESG attract additional inflows of about 0.3% of their net asset value per month (i.e., 3.6% per year)—translating to an additional 4.8 (0.7) million U.S. dollars per month for the mean (median) fund—and (ii) funds with ESG-keyword frequency a standard deviation above the mean get another 0.3% additional inflows per month. The estimated effect of the ESG text's relative

[23]We obtain similar results if we use variations of these measures, such as the frequency of unique (rather than all) ESG keywords or a simple count (rather than the frequency) of ESG keywords.

[24]The ESG percentile ranks are defined for months in which there are at least 100 funds in the fund's CRSP objective code.

length in Column 4 is consistent with these results. Similarly, in Columns 5–6, we see that funds that have at least one ESG keyword in their PIS attract about 0.3%–0.4% higher monthly flows than funds that do not talk about ESG at all, essentially doubling their inflows as the mean monthly flows are about 0.39% (see Table 1A). In Column 7, we see that monthly flows are 0.6% higher for funds with above-median ESG-keyword frequency than for those with low or zero ESG-keyword frequency. Overall, our results indicate that investors direct their investments to funds that discuss their ESG strategy in their PIS, and even more so to those that discuss ESG extensively.

In Table 5, we repeat our analysis using alternative measures for fundamentals-based ESG intensity. In Column 1, we use the Morningstar globe ratings that have previously been shown to have a significant positive effect on fund inflows (see Hartzmark and Sussman, 2019); in Column 2, we use a holdings-based measure calculated from multiple data providers of stock-level ESG scores; in Column 3, we use a holdings-based measure calculated from stock-level carbon emissions; and in Column 4, we use a returns-based measure calculated from a "style" analysis of fund returns over the preceding 12 months (see Section 3.2). The estimate of the coefficient on ESG-keyword frequency remains highly significant in all these specifications.

Our analysis also shows that fund flows significantly react to the ESG intensity of fund holdings, but only for exceptionally high levels of holdings-based ESG measures (see Columns 6–7 of Table 4 and Column 1 of Table 5), and the effect is economically and statistically less significant than that of text-based ESG measures. Specifically, we estimate that funds with holdings-based ESG score above the 90^{th} percentile are associated with additional inflows of about 0.2% per month. Our finding that extreme holdings-based ESG measures matter is consistent with that of Hartzmark and Sussman (2019), who show that funds with a five-globe Morningstar rating experience greater inflows right after the introduction of the globe ratings.

In Appendix C, we present some additional results. In Panel A of Table C.1, we repeat our analysis for the period after Spring 2016 (coinciding with the signing of the Paris Agreement and the publication of the Morningstar globe ratings), which is when the media and investors started paying more attention to ESG investing. We see that, in more recent years, the effect of text-based ESG measures becomes even stronger, while the effect of holdings-based measures becomes weaker and loses its statistical significance. In Panel B of Table C.1, we show that the results of our baseline analysis continue to hold if we include fund fixed effects in addition to the category-by-month fixed effects. As expected, the results of this analysis are strongest when focusing on large within-fund changes in ESG mentions in the PIS (see column 4 in Table C.1B).[25] In Table C.2, we add an interaction term to test whether the effect on fund flows is different for funds targeted to institutional versus retail investors. We find that flows to the two types of funds do not respond differently to the ESG content of the prospectus, suggesting that our results are not driven by one type of investor that has specific preferences or level of sophistication.[26]

Overall, our finding that text-based ESG measures matter and that they matter more than fundamentals-based ESG measures is consistent with the idea that investors have limited attention and pay attention to more salient information (see Barber, Odean and Zheng, 2005; Hartzmark and Sussman, 2019; Kostovetsky and Warner, 2020 for related evidence on mutual fund investing and the survey by Barber and Odean, 2013 for related evidence on other investment decisions). Also consistent with this idea, in Table 6 we find that the readability of the ESG text in the PIS has a positive and significant effect on fund flows. In the same table, we also show that tone uncertainty has a marginally significant negative effect on flows, while the uniqueness and positioning of the ESG text do not significantly affect flows. The uniqueness of the ESG text may have no effect because an ESG focus is sufficient to differentiate a fund's investment strategy, hence draw investors' attention, in the first place, while the positioning may not matter because the PIS text is in any case quite short at an average of 400 words so investors are likely to read its entirety. Throughout this analysis, we control for the writing style (readability, tonality, and uniqueness) of the entire PIS and find that higher readability and more positive tonality of the entire PIS have a weakly positive effect on fund flows.

[25] This analysis accounts for omitted fund heterogeneity, but has the disadvantage that it is much less efficient as it identifies the effect of text-based ESG measures on fund flows through within-fund time-series variation only, which is rather limited (only 6% of funds have switched from none to at least one ESG mention in their PIS during our 10-year sample period). Comparing with Table 4, we see that the estimated coefficients on our text-based ESG measures have similar magnitudes but, as expected, lower statistical significance with fund fixed effects. Yet, in several specifications, the coefficients are significant at conventional levels, particularly so when focusing on large within-fund time series variation.

[26] In unreported results, we have also interacted past performance with our ESG measures, and find no evidence that the sensitivity of flows to past performance differs between ESG and non-ESG funds.

5 Analysis of greenwashing

In this section, we first study (i) which fund characteristics are associated with greenwashing, and (ii) the relationship between greenwashing and fund performance. This analysis is not intended to provide causal evidence linking fund characteristics to greenwashing and, in turn, to fund performance. Rather, the intention is to corroborate whether the measures we construct in Section 3.3 capture greenwashing behavior, by checking if funds we identify as greenwashers are observationally different from non-greenwashers in meaningful and predictable ways.

Subsequently, we study how investors react to greenwashing, i.e., do they recognize the discrepancy between words and actions, or do they respond by rewarding greenwashing funds with additional capital? We also study fund greenwashing behavior and investors' reaction to such behavior at the E, S, and G dimensions separately. Then, we discuss the economic implications of our findings on funds' greenwashing behavior and investors' response to it. And, finally, we check whether our results are robust to accounting for ESG-related activism.

5.1 Fund characteristics and greenwashing

In this section, we study the fund characteristics of greenwashers. Our hypothesis is that funds are more likely to greenwash if they have (i) a potential benefit from doing so, (ii) a strong incentive to do so, and/or (iii) the opportunity to do so without being detected.

First, greenwashing is beneficial as long as there is a demand for ESG investing, so we would expect that funds are more likely to greenwash in the more recent years, during which investors' interest in ESG investing is higher. Second, the incentive to attract capital flows by overstating a fund's ESG commitment is likely higher when the fund's past capital flows have been low. Finally, the opportunity to greenwash and go undetected exists when oversight is low; following Gil-Bazo and Ruiz-Verdú (2009) who show that funds with weaker governance charge higher expense ratios, we use a fund's expense ratio as a proxy for managerial oversight. To study these hypotheses, we estimate the model

$$GW_{i,t} = \alpha_s + \beta_1 PastFlows_{i,t-1} + \beta_2 ExpenseRatio_{i,t-1} + \beta_3 After2016_t + \gamma' \mathbf{x}_{i,t-1} + \varepsilon_{i,t}, \quad (5)$$

where $GW_{i,t}$ is an indicator variable equal to 1 if fund i starts greenwashing in month t and 0 in all preceding months,[27] $PastFlows_{i,t-1}$ is the mean fund flow of fund i over the 12-month period ending in month $t-1$, $ExpenseRatio_{i,t-1}$ is the logarithm of the expense ratio charged by fund i in month $t-1$ adjusted for amortized loans, and $After2016_t$ indicates the period after the signing of the Paris Agreement and the publication of the Morningstar ratings in Spring 2016. $\mathbf{x}_{i,t-1}$ is a vector of other lagged fund characteristics that are potential determinants of greenwashing such as fund age, size, turnover ratio, marketing expenses, skill (the average fund alpha over the past 12 months based on the 4-factor model of Carhart, 1997), and an institutional-fund dummy. We also include fixed effects for funds' investment styles based on the CRSP objective codes.

In Table 7, we report the results from estimating Equation 5 using four alternative greenwashing measures.[28] We see that the estimated coefficient on past fund flows (β_1) is significantly negative, the coefficient on fund expense ratio (β_2) is significantly positive, and the coefficient on the after-2016 dummy (β_3) is significantly positive.[29] Taken together, these findings are consistent with our hypotheses that, as investors' interest in ESG investing rises, funds that have experienced fund outflows over the preceding year and/or have potentially weaker oversight are more likely to exhibit greenwashing. These results provide support that our measures capture fund greenwashing behavior.

[27]As the decision to greenwash is likely rather sticky, for this analysis we do not use a fund's observations from months after it decides to greenwash for the first time.

[28]All OLS regression results we report in this table are qualitatively identical to those from a logistic regression model.

[29]Specifically, a standard deviation (i.e., 0.1) decrease in the mean flows of the past 12 months is associated with a 0.01% increase in a fund's likelihood to start greenwashing in any given month (so a 0.12% increase in the likelihood to do so in any given year); a standard deviation (i.e., 0.5) increase in the log expense ratio is associated with a 0.02% increase in a fund's likelihood to start greenwashing in any given month (so a 0.24% increase in the likelihood to do so in any given year); and, after 2016, a fund is 0.06% more likely to start greenwashing in any given month (so 0.70% more likely to do so in any given year). These effects are economically substantial, since the proportion of greenwashing funds is relatively small (e.g., 5.6% in year 2020).

5.2 Greenwashing and fund performance

In this section, we study whether funds that greenwash are likely to have worse or better fund performance. We posit that, since a fund that greenwashes inserts ESG keywords in its prospectus without altering its investment strategy, there should be no effect on its subsequent performance. As a result, as long as our measures of greenwashing are valid, we should find that greenwashing is not related to subsequent performance. Essentially, this estimation serves as a further sanity check for our greenwashing measures.

Specifically, we estimate the model

$$Performance_{i,t} = \alpha_{s,t} + \gamma_T^{GW} ESG^T_{i,t-1} \times GW_{i,t-1} + \gamma_T^{NGW} ESG^T_{i,t-1} \times \left(1 - GW_{i,t-1}\right) + \gamma_H ESG^H_{i,t-1} + \gamma' \mathbf{x}_{i,t-1} + \varepsilon_{i,t}, \tag{6}$$

where $Performance_{i,t}$ is fund i's performance in month t measured as fund alpha based on the Carhart (1997) 4-factor model, $ESG^T_{i,t-1}$ is our text-based ESG measure in month $t-1$, and $GW_{i,t-1}$ and $1 - GW_{i,t-1}$ indicate whether fund i greenwashes or not in month $t-1$. If greenwashing is not related to subsequent performance, then γ_T^{GW} should not be statistically different from zero.

In Table 8 we present the results from estimating Equation 6, restricting the sample to domestic equity funds. We see that the estimated coefficient for greenwashing funds (γ_T^{GW}) has a low t-statistic in all specifications, indicating that greenwashing does not have a significant effect on fund alpha.[30] On the other hand, the estimated coefficient for non-greenwashing funds (γ_T^{NGW}) is positive and mostly significant, indicating that funds that engage with ESG investing (and truthfully reveal this in their PIS) have superior performance. These findings further corroborate that the funds we identify as greenwashers are observationally different from the non-greenwashers in predictable ways. We also note that our finding of higher alpha for funds that truly engage with ESG investing supports recent evidence suggesting a positive relationship between ESG investing

[30]It could be argued that if, as we have shown above, greenwashing increases subsequent fund inflows, hence increases fund size, then it may in fact decrease subsequent performance; see for example Pástor, Stambaugh and Taylor (2015). While it is not clear that greenwashing has a sufficiently large effect on fund size to reduce performance (especially during our sample period), in several of our specifications we do estimate a negative, though insignificant, effect of greenwashing on fund alpha.

and realized stock returns (e.g., Pástor, Stambaugh and Taylor, 2022).

5.3 Do investor flows respond to greenwashing?

Next, we modify our baseline model of fund flows in Equation 4 so that we estimate the effect of talking about ESG in the prospectus separately for greenwashers and for non-greenwashers. That is, we estimate the model

$$Flows_{i,t} = \alpha_{s,t} + \beta_T^{GW} ESG^T_{i,t-1} \times GW_{i,t-1} + \beta_T^{NGW} ESG^T_{i,t-1} \times \left(1 - GW_{i,t-1}\right) + \beta_H ESG^H_{i,t-1} + \beta' \mathbf{x}_{i,t-1} + \varepsilon_{i,t}, \quad (7)$$

where we interact our text-based ESG measure $ESG^T_{i,t-1}$ with the variables $GW_{i,t-1}$ and $1-GW_{i,t-1}$ indicating whether fund i greenwashes or not in month $t-1$, hence β_T^{GW} and β_T^{NGW} are the effects on flows of talking about ESG in the PIS deceptively and truthfully, respectively. If investors are rational, i.e., they pay attention to and correctly process the information available to them, then β_T^{GW} should not be statistically different from zero.[31]

In Table 9, we present the results from estimating Equation 7 for four alternative greenwashing measures. These are dummies indicating funds that talk about ESG in their PIS but (i) their holdings-based MSCI ESG score is below the median (Column 1), (ii) their holdings-based ESG score is in the bottom quartile based on MSCI (Column 2) or a combined ESG database (Column 4), and (iii) both their holdings- and returns-based MSCI ESG score are below the median (Column 3). Our results show that, contrary to what rationality would imply, investors direct their flows to funds that talk about ESG in their prospectus regardless of whether these funds back their words with actions by actually investing in ESG stocks.[32]

[31] We note that there may be a lag between when a fund changes its PIS to include ESG keywords and when these changes are reflected in its investment strategy hence its holdings. Nonetheless, this lag should be small, as the PIS is meant to describe the fund's actual investment strategy rather than its long-term aspirations. To account for this possible lag, and for the possibility that investors may invest in a fund in anticipation of impending changes in investment strategy that are not already reflected in a fund's holdings at the time of the change in the PIS, in a robustness check we exclude from our analysis the first year following a change in a fund's text-based ESG measure. Results from this robustness check are reported in Section 5.6 and are similar to those reported here.

[32] To further alleviate concerns that our holdings-based ESG measure may be inaccurate, in the Internet Appendix we also report similar results from an analysis that excludes fund-months for which the portfolio coverage of the stock-level ESG scores that we use to construct our fund-level holdings-based ESG measure is below 50%. Furthermore, the results

Specifically, we see that our estimate of β_T^{GW} is consistently positive and statistically significant. We note that, while the coefficient β_T^{GW} on the greenwashing funds appears to be almost twice as high as the coefficient β_T^{NGW} on the non-greenwashing funds, it is estimated using a smaller sample (as fewer funds greenwash) and so has substantially wider confidence intervals. Indeed, while the point estimates differ, their difference is not statistically significant.

In Table 10, we repeat this analysis using the corresponding measures calculated from the ML algorithm. Specifically, we measure text-based ESG intensity $ESG_{i,t-1}^T$ using the estimated probability, according to the ML algorithm, that the PIS is ESG-related, and we measure greenwashing $GW_{i,t-1}$ using a set of dummies indicating funds whose PIS is classified as ESG-related by the ML algorithm but whose fundamentals-based ESG scores are low. We see that, across all specifications, our results are qualitatively the same as in Table 9.

This evidence indicates that (at least some) investors do not uncover the discrepancy between what funds say about ESG investing in their prospectus and how much ESG investing they actually do. These results extend to the context of fund prospectuses and ESG investing the finding by Cooper, Dimitrov and Rau (2001), who study fund name changes during the dot com bubble and find that investors irrationally respond to cosmetic changes.

5.4 E-, S-, and G-related greenwashing

In this section, we study fund greenwashing behavior and investors' response to such behavior at the E, S, and G pillar levels separately. This analysis allows us to identify greenwashing behavior at a more granular level, but also addresses the concern that it is difficult to measure greenwashing using multi-dimensional ESG scores because different funds may have different objectives. For example, a fund that cares about and discusses social issues in its PIS would be classified as an ESG fund, but if it ignores environmental and governance issues it would likely have a low overall ESG score, and so it could be erroneously classified as a greenwasher.

in Columns 2 and 4 are unchanged if we amend the definition of greenwashing to be more symmetric: conditional on having ESG keywords in its PIS, a fund is a greenwasher (non-greenwasher) if its holdings-based ESG score is in the bottom (top) quartile, otherwise it is excluded from the analysis.

To conduct this analysis, we first construct greenwashing measures that capture discrepancies between text-based and fundamentals-based measures of funds' E, S, and G commitments separately. Our measure of E-, S-, and G-related greenwashing is an indicator variable that indicates funds whose PIS is classified by our ML algorithm as E-, S-, and G-related respectively, but the value-weighted mean of their investments' respective E, S, and G scores is below the median within their investment category for that month. In Table 3, we present summary statistics for these greenwashing measures. We find that, across the years, on average 0.5% of all funds overstate in their PIS description their E-related investments, 1.6% of funds overstate their S-related investments, while 0.9% of funds overstate their G-related investments.

Then, we examine whether investor flows respond to greenwashing behavior at each of these three levels. In Columns 1–3 of Table 11, we present results from estimating variations of Equation 7, where we interact our text-based E measure E^T with the variables GW^E and $1-GW^E$ indicating E-related greenwashing, and similarly for S and G. We see that our estimate of β_T^{GW} is positive and statistically significant in Columns 1–2, indicating that investors direct their flows to funds whose PIS descriptions have E- and S-related content regardless of their actual investments in these categories. The coefficient remains positive but loses its significance in Column 3, indicating weaker statistical evidence that investors direct their flows to funds that greenwash at the G level. However, this lack of statistical significance should be interpreted with caution; our algorithm's performance in the E and S classifications is stronger and comparable to that for the ESG classification, but its performance in the G classification is relatively diminished (for detailed performance results, see the Internet Appendix), most likely because terms that are commonly used in governance contexts are also used in other contexts.

In Column 4 of the table, we repeat the analysis from Column 1 but use an alternative measure of E-related greenwashing. Specifically, in our construction of the fundamentals-based E score we replace the stock-level E scores with stock-level carbon emissions levels, and we estimate a similar greenwashing effect.

Finally, in Column 5, we show results from an alternative analysis aimed at further alleviating the ESG multi-dimensionality concerns discussed above. We replace our baseline greenwashing measures from Section 3.3 with a more extreme version that classifies a fund as a greenwasher only if it scores low on *all* three pillars (E, S, and G) separately. Reassuringly, we find that our results remain the same as in our baseline analysis.

5.5 Economic implications of greenwashing

Here, we discuss the economic implications of our findings on funds' greenwashing behavior and investors' response to it.

Using the most conservative coefficient estimate in Table 9 (0.468 from Column 1) we calculate that, through greenwashing, the average fund can get an additional inflow of \$7.5 million per month or \$90 million per year.[33] Considering that, in year 2020, we find that about 200 funds greenwash, this translates to approximately \$18 billion per year in the aggregate actively managed U.S. equity mutual fund industry. We note that this is our estimate of the amount of money that is newly misallocated in 2020 in this industry. The amount that *has* been misallocated and the amount that *could* be misallocated if additional funds were to greenwash are both substantially larger.[34] The amount of funds that could be misallocated increases much further if we extrapolate outside the U.S. and to other professionally managed assets. We note that the U.S. market which we study here constitutes less than 20% of the \$3 trillion global market of sustainable funds, and that as much as \$35 trillion may be invested globally in ESG-labeled assets (Walther and Schildbach, 2022).

The finding that large sums of money are miscallocated to greenwashers has several important implications. First, investors do not get to invest in what they intend to invest in. For example, in 2020, we find that the average holdings-based ESG measure for greenwashing ESG funds was similar to that for the average fund, but one standard deviation below that for genuinely green

[33]Specifically, since the mean ESG-keyword frequency among funds with ESG keywords in their PIS is about 1% and a fund's mean net asset value is about \$1,604 million, an estimated coefficient of 0.468 suggests that, through greenwashing, the average fund would attract additional inflows of $0.468 \cdot 1\% \cdot \$1,604$ million $\approx$ \$7.5 million per month.

[34]For example, in 2020, investors delegated about \$400 billion to actively managed U.S. equity mutual funds with ESG keywords in their PIS. Taking this as an estimate of the wealth of U.S. equity mutual fund investors with ESG preferences, then this \$400 billion could conceivably be captured by greenwashers since we have shown that their deception cannot be recognized by investors.

funds. Martin and Moser (2016) and Riedl and Smeets (2017) show that individual investors value the societal benefits associated with sustainable investments and are willing to sacrifice financial benefits in return. Unfortunately, quantifying the value investors place on these societal benefits and hence the welfare cost of greenwashing for investors is complicated by the fact that the evidence on the returns of green firms is mixed.[35] But what we can say is that, classifying in 2020 about 40% of ESG funds as greenwashers means that investors with a preference for ESG investments miss out on 40% of the societal benefits they would expect to get from their investments.

Second, greenwashing affects society's financing of the transition toward sustainability. According to the United Nations' Intergovernmental Panel on Climate Change (IPCC) 6$^{\text{th}}$ Assessment Report published in 2022, global climate finance would need to grow by a factor of 3 to 6 to keep track with the Paris Agreement's targets. Greenwashing and the resultant misallocation—and mismeasurement—of sustainable financing makes it harder to meet this goal or even to know where we truly stand relative to this goal. This risks missing the Paris Agreement's targets at a cost of trillions of dollars for the U.S. and the global economy (Burke, Davis and Diffenbaugh, 2018).

Third, greenwashing could have important implications for the effect of ESG investing on asset prices hence firms' cost of capital. As first shown by Heinkel, Kraus and Zechner (2001), socially responsible investors can drive a wedge between the cost of capital for responsible and non-responsible firms (also see Pástor, Stambaugh and Taylor, 2021; Pedersen, Fitzgibbons and Pomorski, 2021; Berk and van Binsbergen, 2021; van der Beck, 2022; Hartzmark and Shue, 2023). Intuitively, if responsible investors boycott certain firms, then the unrestricted investors need to hold the boycotted firms and their risk, driving up the required risk premium and hence the cost of capital. The larger the fraction of responsible financing and the more difficult it is to substitute boycotted for non-boycotted firms, the larger this effect on the cost of capital. Estimates for the effect of responsible investing on the cost of capital vary across studies from almost negligible to quite high,[36] but in any case the implication of greenwashing is that it reduces the fraction of truly

[35] As Goldstein et al. (2022) show, while the demand of green assets by green investors may lower green assets' required returns, it may also increase informational risks for traditional investors hence increase these assets' required returns, which potentially explains this mixed empirical evidence.

[36] Focusing on the sin stocks (tobacco, alcohol, and coal), Luo and Balvers (2017) estimate that their cost of capital is

responsible financing and hence the effectiveness of responsible investing. Our estimates imply that the effect of responsible investing on the cost of capital is almost half of what it could be, which in turn reduces its effect on corporate policy.

We conclude that it is important that regulators tackle greenwashing, build trust, and improve transparency in the complex landscape of ESG investing. Indeed, regulators around the world are deeply concerned about the possibility that investors may receive inaccurate information relating to ESG investing and are in the process of implementing new regulation in an effort to tackle greenwashing. Our findings and methodology could be useful in this effort.

5.6 Robustness to engagement

Finally, we note that a potential concern with our greenwashing measures is that activist funds may discuss ESG considerations in their prospectus but invest in firms that have low ESG scores with the intention of subsequently engaging with these firms to improve their ESG performance. Though mutual funds are rarely activist, our methodology would indeed be prone to misclassifying ESG-related activists as greenwashers.[37] In this section, we address this concern in two ways.

First, we analyze funds' PISs to identify funds that engage with their portfolio firms; as direct engagement with portfolio firms would be a major component of an activist fund's strategy, we would expect this to be discussed in their PIS. Specifically, first we classify our subsample of 1,500 PISs (see Section 3.1.2) as engagement-related or not. Subsequently, we observe that, in this subsample, all PISs classified as engagement-related contain one of the following keywords—"engage", "proxy voting", and "shareholder advocacy"—corresponding to the common modes of activism. Then, we search for these engagement keywords in all PISs in our data, and finally we read the

as much as 16% higher per year, which could be at least partially attributed to a boycott effect. On the other hand, more widely considering the boycott of firms that do not pass an ESG screening procedure, specifically those excluded from FTSE USA 4 Good Select Index, Berk and van Binsbergen (2021) calculate that their cost of capital would only be about a basis point higher. The difference in these findings is consistent with the idea that the smaller the set of boycotting investors and the easier it is to substitute boycotted for non-boycotted firms, the smaller the effect on the cost of capital.

[37] Unlike hedge funds, mutual funds often face regulatory and structural constraints (e.g., diversification requirements, lack of incentives, and conflicts of interest) that prevent them from engaging in explicit forms of activism (see Brav, Jiang and Li, 2023). Heath et al. (2023) also find little direct evidence that socially responsible investment funds impact firm behavior by using shareholder proposals.

flagged PISs to eliminate false positives. As expected, we find that a very small proportion of the funds under study (0.4%, on average, across the years) undertake shareholder activism. In Panel A of Table 12, we repeat the estimation of Equation 7 excluding activist funds from the sample. Reassuringly, the estimated coefficients remain statistically significant and of similar magnitude to the main analysis, suggesting that our results are not driven by these funds.

Second, recent evidence suggests that the largest reduction in corporate pollution occurs immediately following a fund activist campaign (Akey and Appel, 2019). So, if funds invest in companies with the intention of improving their environmental behavior through activist campaigns, we should observe an increase in fund ESG scores relatively quickly. Motivated by this evidence, we account for the possibility of fund activism by allowing for a one-year lag between when a fund starts discussing ESG considerations in its PIS and when these changes are reflected in its holdings. So, in Panel B of Table 12, we repeat the estimation of Equation 7 excluding the first year following a change in a fund's text-based ESG measure. Our results are similar to those from our main analysis, hence provide additional evidence that our conclusions are not driven by ESG-related engagement.

6 Conclusion

In this paper, we develop novel measures of a mutual fund's ESG commitments that are based on discretionary information provided by the mutual fund itself. These measures are based on text analysis of the principal investment strategy section of the prospectus that each mutual fund must submit to the SEC. We observe that, over time, an increasing number of funds discuss ESG topics in their prospectuses. More specifically, at the end of our sample period in 2020, about ten percent of prospectuses contain ESG-related language. We find that fund flows respond more strongly to text-based measures of ESG intensity than to fundamentals-based measures constructed from disclosed fund holdings or realized fund returns. Comparing our two sets of measures, we find that there is often a discrepancy between the discretionary ESG information provided in fund prospectuses and the objective ESG information contained in fund fundamentals, implying that some funds greenwash.

Moreover, we find that, as investors tend to pay more attention to text-based measures, they do not distinguish between funds that greenwash and those that truthfully reveal in the prospectus their commitment to ESG investing. We also show that greenwashing behavior is more likely to be observed in the last five years of our sample period, during which investors' interest in ESG is higher, and that it is associated with lower past flows and potentially weaker fund oversight. Providing further support for our greenwashing measures, we find that greenwashing funds do not have better performance, while funds that follow through on their ESG commitments do perform better. Our results provide valuable insights for mutual fund managers wishing to communicate their ESG commitment to potential investors. They should also be useful for investors and for regulators who seek to use text analysis and machine learning tools to identify cases of greenwashing from fund documentation and to evidence the effectiveness of any new regulation they introduce to tackle such practices.

References

Abis, S. 2022. "Man vs. machine: Quantitative and discretionary equity management." Unpublished Paper.

Abis, S, A. M Buffa, A Javadekar, and A Lines. 2022. "Learning from prospectuses." Unpublished Paper.

Abis, S, and A Lines. 2022. "Do mutual funds keep their promises?" Unpublished Paper.

Agarwal, V, G. D Gay, and L Ling. 2014. "Window dressing in mutual funds." *Review of Financial Studies*, 27(11): 3133–3170.

Akey, P, and I Appel. 2019. "Environmental externalities of activism." Unpublished Paper.

Ammann, M, C Bauer, S Fischer, and P Müller. 2019. "The impact of the Morningstar Sustainability Rating on mutual fund flows." *European Financial Management*, 25(3): 520–553.

Barber, B. M, and T Odean. 2013. "The behavior of individual investors." In *Handbook of the Economics of Finance*, Vol. 2, ed. George M. Constantinides, Milton Harris and Rene M. Stulz, Chapter 22, 1533–1570. Amsterdam, the Netherlands:Elsevier.

Barber, B. M, T Odean, and L Zheng. 2005. "Out of sight, out of mind: The effects of expenses on mutual fund flows." *Journal of Business*, 78(6): 2095–2120.

van der Beck, P. 2022. "Flow-driven ESG returns." Unpublished Paper.

Berg, F, J. F Koelbel, and R Rigobon. 2022. "Aggregate confusion: The divergence of ESG ratings." *Review of Finance*, 26(6): 1315–1344.

Berg, F, J. F Kölbel, A Pavlova, and R Rigobon. 2021. "ESG confusion and stock returns: Tackling the problem of noise." Unpublished Paper.

Berk, J, and J. H van Binsbergen. 2021. "The impact of impact investing." Unpublished Paper.

Bolton, P, and M Kacperczyk. 2021. "Do investors care about carbon risk?" *Journal of Financial Economics*, 142(2): 517–549.

Brav, A, W Jiang, and R Li. 2023. "Governance by persuasion: Hedge fund activism and market-based shareholder influence." In *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.

Burke, M, W. M Davis, and N. S Diffenbaugh. 2018. "Large potential reduction in economic damages under UN mitigation targets." *Nature*, 557(7706): 549–553.

Carhart, M. M. 1997. "On persistence in mutual fund performance." *Journal of Finance*, 52(1): 57–82.

Chevalier, J, and G Ellison. 1997. "Risk taking by mutual funds as a response to incentives." *Journal of Political Economy*, 105(6): 1167–1200.

Cohen, L, C Malloy, and Q Nguyen. 2020. "Lazy prices." *Journal of Finance*, 75(3): 1371–1415.

Cooper, M. J, H Gulen, and P. R Rau. 2005. "Changing names with style: Mutual fund name changes and their effects on fund flows." *Journal of Finance*, 60(6): 2825–2858.

Cooper, M. J, O Dimitrov, and P. R Rau. 2001. "A rose.com by any other name." *Journal of Finance*, 56(6): 2371–2388.

Fama, E. F, and K. R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33(1): 3–56.

Flesch, R. 1948. "A new readability yardstick." *Journal of Applied Psychology*, 32(3): 221–233.

Gantchev, N, M Giannetti, and R Li. 2021. "Sustainability or performance? Ratings and fund managers' incentives." Unpublished Paper.

Gibson Brandon, R, P Krueger, and P. S Schmidt. 2021. "ESG rating disagreement and stock returns." *Financial Analysts Journal*, 77(4): 104–127.

Gibson Brandon, R, S Glossner, P Krueger, P Matos, and T Steffen. 2022. "Do responsible investors invest responsibly?" *Review of Finance*, 26(6): 1389–1432.

Gil-Bazo, J, and P Ruiz-Verdú. 2009. "The relation between price and performance in the mutual fund industry." *Journal of Finance*, 64(5): 2153–2183.

Goldstein, I, A Kopytov, L Shen, and H Xiang. 2022. "On ESG investing: Heterogeneous preferences, information, and asset prices." Unpublished Paper.

Gunning, R. 1952. *Technique of clear writing.* McGraw-Hill.

Gustafson, P. 2003. *Measurement error and misclassification in statistics and epidemiology: Impacts and Bayesian adjustments.* Chapman and Hall/CRC.

Hartzmark, S. M, and A. B Sussman. 2019. "Do investors value sustainability? A natural experiment examining ranking and fund flows." *Journal of Finance*, 74(6): 2789–2837.

Hartzmark, S. M, and K Shue. 2023. "Counterproductive impact investing: The impact elasticity of brown and green firms." Unpublished Paper.

Heath, D, D Macciocchi, R Michaely, and M C. Ringgenberg. 2023. "Does socially responsible investing change firm behavior?" *Review of Finance, forthcoming*.

Heinkel, R, A Kraus, and J Zechner. 2001. "The effect of green investment on corporate behavior." *Journal of Financial and Quantitative Analysis*, 36(4): 431–449.

Hoberg, G, and G Phillips. 2016. "Text-based network industries and endogenous product differentiation." *Journal of Political Economy*, 124(5): 1423–1465.

Hong, H, and L Kostovetsky. 2012. "Red and blue investing: Values and finance." *Journal of Financial Economics*, 103(1): 1–19.

Intergovernmental Panel on Climate Change. 2022. "Climate change 2022: Mitigation of climate change." Intergovernmental Panel on Climate Change, Working Group III.

Jain, P. C, and J. S Wu. 2000. "Truth in mutual fund advertising: Evidence on future performance and fund flows." *Journal of Finance*, 55(2): 937–958.

Kacperczyk, M, C Sialm, and L Zheng. 2008. "Unobserved actions of mutual funds." *Review of Financial Studies*, 21(6): 2379–2416.

Kim, S, and A Yoon. 2022. "Analyzing active managers' commitment to ESG: Evidence from United Nations Principles for Responsible Investment." *Management Science*, 69(2): 741–758.

Kostovetsky, L, and J. B Warner. 2020. "Measuring innovation and product differentiation: Evidence from mutual funds." *Journal of Finance*, 75(2): 779–823.

Krakow, N. J, and T Schäfer. 2020. "Mutual funds and risk disclosure: Information content of fund prospectuses." Unpublished Paper.

Kumar, A, A Niessen-Ruenzi, and O. G Spalt. 2015. "What's in a name? Mutual fund flows when managers have foreign-sounding names." *Review of Financial Studies*, 28(8): 2281–2321.

Liang, H, L Sun, and M Teo. 2022. "Responsible hedge funds." *Review of Finance*, 26(6): 1585–1633.

Loughran, T, and B McDonald. 2016. "Textual analysis in accounting and finance: A survey."

Journal of Accounting Research, 54(4): 1187–1230.

Lundberg, S. M, and S.-I Lee. 2017. “A unified approach to interpreting model predictions.” In *Proceedings of the Advances in Neural Information Processing Systems*, Vol. 30, 4765–4774.

Luo, H. A, and R. J Balvers. 2017. “Social screens and systematic investor boycott risk.” *Journal of Financial and Quantitative Analysis*, 52(1): 365–399.

Martin, P. R, and D. V Moser. 2016. “Managers’ green investment disclosures and investors’ reaction.” *Journal of Accounting and Economics*, 61(1): 239–254.

Pástor, L, R. F Stambaugh, and L. A Taylor. 2015. “Scale and skill in active management.” *Journal of Financial Economics*, 116(1): 23–45.

Pástor, L, R. F Stambaugh, and L. A Taylor. 2021. “Sustainable investing in equilibrium.” *Journal of Financial Economics*, 142(2): 550–571.

Pástor, L, R. F Stambaugh, and L. A Taylor. 2022. “Dissecting green returns.” *Journal of Financial Economics*, 146(2): 403–424.

Pedersen, L. H, S Fitzgibbons, and L Pomorski. 2021. “Responsible investing: The ESG-efficient frontier.” *Journal of Financial Economics*, 142(2): 572–597.

Porter, M, and R Boulton. 2001. “The English (Porter2) stemming algorithm.” http://snowball.tartarus.org/algorithms/english/stemmer.html.

Riedl, A, and P Smeets. 2017. “Why do investors hold socially responsible mutual funds?” *Journal of Finance*, 72(6): 2505–2550.

Roussanov, N, H Ruan, and Y Wei. 2021. “Marketing mutual funds.” *Review of Financial Studies*, 34(6): 3045–3094.

Sharpe, W. F. 1992. “Asset allocation: Management style and performance measurement.” *Journal of Portfolio Management*, 18(2): 7–19.

Sheng, J, N Xu, and L Zheng. 2022. “Do mutual funds walk the talk? A textual analysis of risk disclosure by mutual funds.” Unpublished Paper.

Walther, U, and J Schildbach. 2022. “Sustainable finance – coming of age.” Deutsche Bank Research, https://tinyurl.com/2emaac2k.

Table 1: Summary statistics of fund characteristics

In Panel A, we present summary statistics for the entire sample of 398,572 fund-months (4,863 funds at an average of 81.96 months each). In Panel B, we present summary statistics for the sample of 17,325 fund-months (503 funds at an average of 34.44 months each) that have at least one ESG keyword in the Principal Investment Strategy (PIS) text block of their prospectus. Our sample period is from January 2011 to June 2020. Fund age is the number of years since the fund's inception. Total net asset value (TNAV) is measured in millions of U.S. dollars. Expense ratio is defined as total annual management, administrative, and 12b-1 fees and expenses divided by year-end TNAV. Turnover ratio is defined as the minimum of aggregate purchases and sales of securities divided by the average TNAV over the calendar year. Marketing expenses (effective 12b-1 fees) are defined similar to Roussanov, Ruan and Wei (2021) as the combination of 12b-1 fees and front loads. Fund inflows are the ratio of monthly fund flows to the beginning-of-month TNAV, winsorized at 1% and 99%. The institutional-fund dummy classifies each fund as targeted to institutional or retail investors using CRSP's institutional class indicator for its largest (by net asset value) share class. Net returns are the monthly fund returns net of fees, expenses, and transaction costs. Fund alpha is the monthly fund alpha estimated from the Carhart (1997) 4-factor model using daily fund and factor returns.

Panel A: Entire sample

				Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
Fund age	398,474	13.25	12.08	1	11	26
Total net asset value	395,029	1,604.37	6,491.72	10.40	223.70	3,112.90
Expense ratio (%)	386,751	1.08	0.49	0.46	1.08	1.59
Turnover ratio (%)	387,228	74.25	195.74	13.43	47.28	132.04
Marketing expenses (%)	388,231	0.48	0.46	0.00	0.25	1.00
Fund inflows (%)	391,768	0.39	5.84	-3.19	-0.33	4.15
Institutional-fund dummy	398,572	0.50	0.50	0	1	1
Fund returns (net) (%)	394,569	0.78	4.51	-4.35	1.00	5.63
Fund alpha (%)	397,799	-0.10	2.20	-2.20	-0.05	2.00

Panel B: Sample with ESG keywords in prospectus

				Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
Fund age	17,325	11.48	10.73	1	9	24
Total net asset value	17,181	632.07	1,701.07	6.70	100.50	1,517.50
Expense ratio (%)	16,403	1.11	0.41	0.61	1.07	1.56
Turnover ratio (%)	16,403	56.61	53.18	14.15	42.23	113.39
Marketing expenses (%)	16,598	0.45	0.46	0.00	0.25	1.00
Fund inflows (%)	17,024	0.67	5.69	-2.83	-0.07	4.56
Institutional-fund dummy	17,325	0.50	0.50	0	1	1
Fund returns (net) (%)	17,101	1.11	5.36	-5.08	1.42	6.91
Fund alpha (%)	17,290	-0.07	2.24	-2.47	-0.05	2.31

Table 2: Summary statistics for ESG keywords

In this table, we present the ESG keywords we use in the text analysis of the PIS in funds' prospectuses, together with their summary statistics across the 37,399 prospectuses we analyze. In the first row we show statistics for all keywords together, and in subsequent rows we show statistics for each keyword separately, from the most frequent to the least frequent. In the first column, we show the number of instances of the ESG keyword across all prospectuses. In the second column, we show the number of prospectuses in which the keyword appears. In the third and fourth columns, we show the frequency counterparts (expressed as percentages) of the first and second columns, that is, the total number of instances of the ESG keyword divided by the total number of words in the PIS across all prospectuses, and the number of prospectuses containing the keyword divided by the number of all prospectuses. In the fifth column, we show the percentage of words corresponding to the keyword (similar to the third column), but here we divide by the total number of words in the PIS across all prospectuses in which at least one ESG keyword appears.

ESG keyword	**# of instances**	**# of prospectuses**	**% of words**	**% of prospectuses**	**% of words (ESG only)**
All ESG keywords	7,725	1,949	0.0508	5.21	0.7873
ESG	3,269	835	0.0215	2.23	0.3332
environmental	2,199	1,262	0.0145	3.37	0.2241
ethic	413	309	0.0027	0.83	0.0421
carbon	351	118	0.0023	0.32	0.0358
SRI	279	278	0.0018	0.74	0.0284
responsible investing	251	204	0.0017	0.55	0.0256
human rights	219	185	0.0014	0.49	0.0223
green	191	78	0.0013	0.21	0.0195
climate change	139	118	0.0009	0.32	0.0142
renewable energy	106	64	0.0007	0.17	0.0108
social responsibility	103	91	0.0007	0.24	0.0105
pollution	63	62	0.0004	0.17	0.0064
sustainable business practice	42	9	0.0003	0.02	0.0043
sustainable development goals	35	23	0.0002	0.06	0.0036
biological	25	25	0.0002	0.07	0.0026
clean energy	21	21	0.0001	0.06	0.0021
SDG	12	4	0.0001	0.01	0.0012
toxic	7	7	0.0000	0.02	0.0007

Table 3: Summary statistics and correlation coefficients of fund ESG measures

In Panel A (B), we present summary statistics for text-based (fundamentals-based) ESG measures, in Panel C we present the correlations of the baseline ESG measures, and in Panel D we present summary statistics for greenwashing measures. *ESG-keyword frequency* is their frequency in the PIS. *ESG in PIS* indicates if the PIS contains ESG keywords. *ESG in PIS from ML* indicates if the ML algorithm classifies the PIS as ESG-related. *E, S, and G in PIS from ML* indicate separate classifications as E-, S-, and G-related. *Holdings ESG score* is the value-weighted mean of the fund investments' ESG scores. *Holdings E, S, and G score* is the separate value-weighted mean of the fund investments' E, S, and G scores. *Returns ESG score* is the fund's returns-based ESG score. *ESG GW* indicates if *ESG in PIS* is 1 and its *Holdings ESG score* is below median within its category-month. *ESG GW from ML* is similar but uses *ESG in PIS from ML* instead. *E GW* indicates if *E in PIS from ML* is 1 but the fund's *Holdings E score* is below median within its category-month. *S GW* and *G GW* are similar. Percentiles in Panels A and D are conditional on fund-months whose PIS contains ESG keywords. The number of observations drops from Panel A to B due to missing portfolio holdings or returns and from B to D because we require at least 100 observations in a category-month to calculate holdings-based ESG rankings.

Panel A: Text-based ESG measures

				Conditional Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
ESG-keyword frequency (as %)	398,572	0.041	0.280	0.156	0.615	2.240
ESG in PIS	398,572	0.043	0.204	1	1	1
ESG in PIS from ML	398,572	0.066	0.248	1	1	1
E in PIS from ML	398,572	0.020	0.139	0	0	1
S in PIS from ML	398,572	0.034	0.181	0	1	1
G in PIS from ML	398,572	0.023	0.151	0	0	1

Panel B: Fundamentals-based ESG measures

				Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
Holdings ESG score	337,623	0.039	0.275	-0.302	0.043	0.403
Holdings E score	337,623	0.086	0.290	-0.286	0.082	0.487
Holdings S score	337,623	-0.041	0.217	-0.310	-0.035	0.226
Holdings G score	337,623	0.121	0.222	-0.130	0.119	0.381
Returns ESG score	377,166	2.243	0.980	0.979	2.198	3.594

Panel C: Correlations

	(1)	(2)	(3)
(1) Text ESG score	1		
(2) Holdings ESG score	0.071	1	
(3) Returns ESG score	0.055	0.302	1

Panel D: Greenwashing measures

				Conditional Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
ESG GW	273,137	0.015	0.121	0	0	1
ESG GW from ML	273,137	0.025	0.156	0	0	1
E GW	264,326	0.005	0.070	0	0	1
S GW	275,258	0.016	0.125	0	0	1
G GW	275,258	0.009	0.094	0	0	1

Table 4: Fund flows and the presence of ESG keywords in the PIS

This table shows how fund flows respond to various definitions of text- and fundamentals-based ESG scores. A fund's text-based ESG score is: the ESG-keyword frequency in its prospectus's PIS text block (columns 1–3); the relative length of the part of the PIS containing ESG keywords (column 4); a dummy indicating if the PIS contains ESG keywords (columns 5–6); a dummy indicating if the ESG-keyword frequency in the PIS exceeds the median conditional on containing ESG keywords (column 7). A fund's fundamentals-based ESG score is: the value-weighted mean of its investments' ESG scores (columns 2 and 4); this score's ranking within the fund's investment category-by-month (column 3); a dummy indicating if this score is in the top 50% within the investment category-by-month (column 5) or in the top 90% (columns 6–7). All specifications include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. */**/*** indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ESG keyword frequency	0.208 ***	0.294 ***	0.297 ***				
	3.000	4.261	4.322				
ESG text relative length				0.012 ***			
				3.570			
ESG in prospectus					0.004 ***	0.003 ***	
					3.084	2.807	
ESG keyword frequency > p50							0.006 ***
							4.564
Holdings ESG score (raw)		-0.000		-0.000			
		-0.765		-0.741			
Holdings ESG score (rank)			-0.000				
			-0.610				
Holdings ESG score > p50					-0.000		
					-0.753		
Holdings ESG score > p90						0.002 ***	0.002 **
						2.786	2.527
log(Fund size)	-0.001 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 **
	-4.899	-2.693	-2.651	-2.699	-2.673	-2.620	-2.591
log(Expense ratio)	-0.459 ***	-0.184 **	-0.177 **	-0.184 **	-0.179 **	-0.173 **	-0.173 **
	-6.222	-2.343	-2.283	-2.350	-2.307	-2.247	-2.241
log(Effective 12b-1 fee)	0.311 ***	0.282 ***	0.276 ***	0.282 ***	0.277 ***	0.280 ***	0.276 ***
	5.464	4.827	4.755	4.837	4.767	4.817	4.749
Prior 1-month return (raw)	0.049 ***	0.045 ***	0.041 ***	0.045 ***	0.041 ***	0.041 ***	0.040 ***
	3.855	3.106	2.843	3.100	2.832	2.834	2.834
Prior 12-month return (rank)	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***	0.000 ***
	24.920	25.658	25.572	25.642	25.586	25.804	25.824
Prior 12-month α < p10	-0.005 ***	-0.005 ***	-0.004 ***	-0.005 ***	-0.005 ***	-0.005 ***	-0.005 ***
	-8.881	-7.818	-7.806	-7.828	-7.849	-7.892	-7.837
Prior 12-month α > p90	0.011 ***	0.012 ***	0.012 ***	0.012 ***	0.011 ***	0.011 ***	0.011 ***
	13.639	14.281	14.235	14.284	14.171	14.126	14.156
log(Fund age)	-0.011 ***	-0.010 ***	-0.010 ***	-0.010 ***	-0.010 ***	-0.010 ***	-0.010 ***
	-24.384	-20.622	-20.594	-20.641	-20.670	-20.722	-20.667
Fund for institutionals	-0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-0.261	0.321	0.214	0.326	0.248	0.278	0.255
log(Prospectus word count)	-0.000	-0.001 ***	-0.001 **	-0.001 ***	-0.001 ***	-0.001 ***	-0.001 ***
	-0.639	-2.766	-2.584	-2.793	-2.724	-2.804	-2.703
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	307,170	261,269	258,212	261,269	258,212	258,212	258,212
Adjusted R^2	0.084	0.081	0.080	0.081	0.080	0.080	0.080

Table 5: Fund flows and the presence of ESG keywords in the PIS – Additional specifications

This table shows how fund flows respond to text- and fundamentals-based ESG measures, focusing on alternative definitions for the fundamentals-based ESG score. A fund's fundamentals-based ESG score is defined as follows: in column 1, it is a dummy indicating if its Morningstar globe rating is 1 or 5 globes, with 5 the top rating; in column 2, it is the value-weighted mean of its investments' standardized ESG scores, which are calculated as the mean of the standardized ESG scores across the MSCI, Sustainalytics, and Refinitiv databases; in column 3, it is the value-weighted mean of its investments' carbon emissions; in column 4, it is the returns-based ESG score calculated using the Sharpe (1992) style analysis (see Section 3.2). In all specifications, a fund's text-based ESG measure is the ESG-keyword frequency in its prospectus's PIS text block. All specifications include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. */**/*** indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG keyword frequency	0.307 ***	0.294 ***	0.279 ***	0.214 ***
	3.350	4.270	4.085	3.089
Morningstar Globe 1	0.001			
	0.842			
Morningstar Globe 5	0.002 **			
	2.269			
Holdings ESG score (combined)		-0.002		
		-1.291		
Holdings carbon emissions			-0.000	
			-1.136	
Returns ESG score				-0.000
				-1.370
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	146,495	262,207	258,670	307,125
Adjusted R^2	0.073	0.081	0.081	0.084

Table 6: Fund flows and the characteristics of the ESG-related text in the PIS

This table shows how fund flows respond to various characteristics of the ESG-related text in the PIS text block of a fund's prospectus. Specifically, the effects of the following ESG text characteristics are presented: readability in columns 1–2; tonality in column 3; uniqueness in column 4; and positioning in columns 5–6. *ESG in PIS* is a dummy indicating if the PIS contains any ESG keywords. *ESG readability* is calculated using the Flesch Reading Ease (Flesch, 1948) and the (sign-flipped) Gunning Fog Index (Gunning, 1952) measures, which are designed to measure how easy a passage in English is to understand. *ESG uncertain word freq* is the frequency of uncertain words as defined in the Loughran-McDonald sentiment word list. *ESG text uniqueness* captures the uniqueness of the ESG-related text in a fund's PIS relative to that of other funds, conditioning on prospectuses submitted in the same calendar year. *Distance to ESG text* is the proportion of the text from the beginning of the PIS to the first sentence containing an ESG keyword. *ESG in first sentence* is a dummy indicating if an ESG keyword appears in the PIS's first sentence. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count, as well as controls for the overall PIS's text style (i.e., readability, tonality, and text uniqueness). t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. */**/*** indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)
ESG in PIS	0.006 ***	0.020 ***	0.005 ***	0.003 ***	0.003	0.003 ***
	4.863	7.205	3.536	2.64	1.551	2.783
ESG Text Readability						
ESG in PIS * ESG readability (Flesch)	0.000 ***					
	6.940					
ESG in PIS * ESG readability (Fog)		0.001 ***				
		7.383				
ESG Text Tonality						
ESG in PIS * ESG uncertain word freq			-0.104 *			
			-1.759			
ESG Text Uniqueness						
ESG in PIS * ESG text uniqueness				-0.001		
				-1.228		
ESG Positioning						
ESG in PIS * Distance to ESG text					0.001	
					0.238	
ESG in PIS * ESG In first sentence						0.001
						0.312
Fund Controls	Yes	Yes	Yes	Yes	Yes	Yes
Overall PIS style controls	Yes	Yes	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	261,269	261,269	261,269	261,269	261,269	261,269
Adjusted R^2	0.082	0.082	0.082	0.081	0.081	0.081

Table 7: Fund characteristics related to greenwashing

This table shows which fund characteristics are associated with greenwashing. The specifications differ in the definition of greenwashing. In column 1, a greenwashing fund is one that includes an ESG keyword in its prospectus's PIS text block but whose value-weighted mean of its investments' ESG scores from MSCI (i.e., whose holdings-based ESG score) is below the 50th percentile within the fund's investment category for that month. In column 2, the holdings-based ESG score cutoff below which a fund is deemed to be greenwashing changes from the 50th to the 25th percentile. In column 3, the holdings-based ESG score cutoff is back at the 50th percentile as in column 1, but for a fund to be deemed a greenwasher it is additionally required that the fund's *returns*-based ESG score calculated using the Sharpe (1992) style analysis is below the 50th percentile within the fund's investment category for that month. In column 4, greenwashing is defined as in column 2 (i.e., a fund greenwashes if the holdings-based ESG score is below the 25th percentile), but the fund's holdings-based ESG score is calculated as the fund's investments' standardized ESG scores averaged across the MSCI, Sustainalytics, and Refinitiv databases. All specifications include investment category fixed effects and the explanatory variables fund age, size, expense ratio, turnover ratio, 12b-1 fees, prior 12-month mean flows, and prior 12-month mean α, a dummy indicating funds targeted to institutional investors, and a dummy indicating months after 2016/03. The analyses include, for each fund, all months until the month the fund greenwashes for the first time. The dependent variable (the greenwashing dummy) is expressed as a percent, i.e., 0% or 100%, so the regression coefficients represent the percent change in a fund's likelihood to greenwash in a given month. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1) discrepancy with holdings	(2) discrepancy with holdings (higher)	(3) discrepancy with holdings & returns	(4) discrepancy with holdings (composite)
log(Fund size)	-0.007 *	-0.006 ***	-0.009 ***	-0.003
	-1.772	-2.681	-3.011	-1.494
log(Expense ratio)	0.039 **	0.027 **	0.030 **	0.029 ***
	2.122	2.573	2.535	2.769
Turnover ratio	0.004	0.001	0.000	-0.001
	0.745	0.194	0.038	-0.361
log(Effective 12b-1 fee)	3.457	1.866	-0.486	2.002
	1.049	0.994	-0.290	1.225
Prior 12-month mean flows	-0.100 ***	-0.069 ***	-0.060 **	-0.052 *
	-2.768	-2.885	-2.542	-1.783
Prior 12-month mean α	0.715	0.862	-0.231	0.941
	0.743	1.154	-0.226	1.064
log(Fund age)	0.001	0.001	0.015 *	-0.006
	0.064	0.091	1.788	-1.075
Fund for institutionals	0.017	-0.009	0.020	0.002
	0.716	-0.616	1.150	0.154
After 2016	0.058 ***	0.017	0.045 ***	0.031 **
	2.736	1.360	3.000	2.575
Category Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	159,725	161,059	156,112	162,328
Adjusted R^2	0.0002	0.0001	0.0002	0.0001

Table 8: Fund performance and greenwashing in the PIS

This table shows the effect on fund performance (alpha) of the inclusion of ESG keywords in the PIS text block of a fund's prospectus, restricting the sample to domestic equity funds. The effect for greenwashing funds is shown in the row presenting the interaction of *ESG-keyword frequency* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds (*non-GW*). The specifications differ in the definition of the greenwashing dummy, so also of the non-greenwashing dummy which is 1 minus the former. In column 1, the greenwashing dummy is 1 for any fund that includes an ESG keyword in its prospectus's PIS but whose value-weighted mean of its investments' ESG scores from MSCI (i.e., whose holdings-based ESG score) is below the 50th percentile within the fund's investment category for that month. In column 2, the holdings-based ESG score cutoff below which the greenwashing dummy equals 1 changes from the 50th to the 25th percentile. In column 3, the holdings-based ESG score cutoff is back at the 50th percentile as in column 1, but for the greenwashing dummy to equal 1 it is additionally required that the fund's *returns*-based ESG score calculated using the Sharpe (1992) style analysis is below the 50th percentile within the fund's investment category for that month. In column 4, the greenwashing dummy is defined as in column 2 (i.e., it equals 1 if the holdings-based ESG score is below the 25th percentile), but the fund's holdings-based ESG score is calculated as the fund's investments' standardized ESG scores averaged across the MSCI, Sustainalytics, and Refinitiv databases. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG keyword frequency * GW	0.016	-0.001	0.044	-0.016
	0.355	-0.017	0.589	-0.289
ESG keyword frequency * Non-GW	0.036 **	0.036 **	0.026	0.041 **
	2.157	2.188	1.510	2.560
Holdings ESG score (raw)	0.000	0.000	0.000	-0.000
	1.099	1.095	0.721	-0.289
log(Fund size)	0.000	0.000	0.000	0.000
	0.944	0.944	0.521	0.977
log(Expense ratio)	-0.072 ***	-0.072 ***	-0.069 ***	-0.075 ***
	-3.469	-3.466	-3.395	-3.713
log(Effective 12b-1 fee)	-0.010	-0.010	-0.009	-0.010
	-0.799	-0.804	-0.731	-0.809
Prior 1-month return (raw)	0.029 **	0.029 **	0.033 **	0.028 **
	2.121	2.121	2.345	2.110
Prior 12-month return (rank)	0.000	0.000	0.000	0.000
	1.102	1.101	1.131	1.186
Prior 12-month α < p10	-0.000	-0.000	-0.000	-0.000
	-1.070	-1.071	-0.873	-1.028
Prior 12-month α > p90	0.000	0.000	0.000	0.000
	0.481	0.481	0.607	0.434
log(Fund age)	-0.000	-0.000	-0.000	-0.000
	-0.791	-0.792	-0.723	-0.847
Fund for institutionals	0.000	0.000	0.000	0.000
	0.672	0.677	1.029	0.512
log(Prospectus word count)	0.000	0.000	0.000	0.000
	0.150	0.150	0.230	0.006
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	185,058	185,058	180,152	186,685
Adjusted R^2	0.083	0.083	0.093	0.083

Table 9: Fund flows and greenwashing in the PIS

This table shows how fund flows respond to the inclusion of ESG keywords in the PIS text block of a fund's prospectus, for funds that greenwash versus those that do not. The effect for greenwashing funds is shown in the row presenting the interaction of *ESG-keyword frequency* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds (*non-GW*). The specifications differ in the definition of the greenwashing dummy, so also of the non-greenwashing dummy which is 1 minus the former. In column 1, the greenwashing dummy is 1 for any fund that includes an ESG keyword in its PIS but whose MSCI holdings-based ESG score is below the 50^{th} percentile within the fund's investment category for that month. In column 2, the holdings-based ESG score cutoff below which the greenwashing dummy equals 1 changes from the 50^{th} to the 25^{th} percentile. In column 3, the greenwashing dummy is 1 for any fund that includes an ESG keyword in its PIS but whose MSCI holdings-based and returns-based ESG score (calculated using the style analysis of Sharpe, 1992) are both below the 50^{th} percentile within the fund's investment category for that month. In column 4, the greenwashing dummy is defined as in column 2 (i.e., it equals 1 if the holdings-based ESG score is below the 25^{th} percentile), but the fund's holdings-based ESG score is calculated as the fund's investments' standardized ESG scores averaged across multiple databases (MSCI, Sustainalytics, and Refinitiv). All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG keyword frequency * GW	0.468 **	0.716 **	0.698 ***	0.795 **
	2.485	2.555	2.891	2.522
ESG keyword frequency * Non-GW	0.268 ***	0.262 ***	0.235 ***	0.255 ***
	3.648	3.755	3.318	3.730
Holdings ESG score (raw)	-0.000	-0.000	-0.000	-0.001
	-0.687	-0.658	-0.805	-1.135
log(Fund size)	-0.000 ***	-0.000 ***	-0.001 ***	-0.000 ***
	-2.666	-2.668	-4.596	-2.850
log(Expense ratio)	-0.182 **	-0.183 **	-0.220 ***	-0.222 ***
	-2.335	-2.34	-2.842	-2.840
log(Effective 12b-1 fee)	0.278 ***	0.279 ***	0.273 ***	0.302 ***
	4.777	4.784	4.676	5.143
Prior 1-month return (raw)	0.041 ***	0.041 ***	0.043 ***	0.042 ***
	2.854	2.856	3.098	2.869
Prior 12-month return (rank)	0.000 ***	0.000 ***	0.000 ***	0.000 ***
	25.617	25.629	25.091	25.497
Prior 12-month α < p10	-0.004 ***	-0.004 ***	-0.004 ***	-0.005 ***
	-7.814	-7.811	-7.471	-8.069
Prior 12-month α > p90	0.012 ***	0.012 ***	0.012 ***	0.012 ***
	14.255	14.265	14.215	14.301
log(Fund age)	-0.010 ***	-0.010 ***	-0.009 ***	-0.010 ***
	-20.616	-20.618	-20.557	-20.745
Fund for institutionals	0.000	0.000	0.000	0.000
	0.203	0.188	0.182	0.079
log(Prospectus word count)	-0.001 **	-0.001 **	-0.001 **	-0.001 **
	-2.606	-2.606	-2.539	-2.344
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	258,099	258,099	251,150	259,863
Adjusted R^2	0.080	0.080	0.081	0.081

Table 10: Fund flows and greenwashing in the PIS – with ML-based measures

Like Table 9, this table shows how fund flows respond to the inclusion of ESG-related words in the PIS, for funds that greenwash versus those that do not. The difference is that, in this table, all text-based ESG measures and the corresponding greenwashing measures are calculated using a machine learning (ML) algorithm as discussed in Section 3.1.2. The effect for greenwashing funds is shown in the row presenting the interaction of *ESG intensity* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds ($non\text{-}GW = 1 - GW$). *ESG intensity* is the estimated probability, according to the ML algorithm, that the PIS is ESG-related; it is scaled to have the same variance as *ESG keyword frequency* so estimated coefficients are comparable with those in Table 9. The specifications differ in the definition of the greenwashing dummy; for details on its definition, see the caption of Table 9. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG intensity * GW	0.358 ***	0.443 **	0.488 ***	0.462 **
	2.627	2.359	3.005	2.450
ESG intensity * Non-GW	0.452 ***	0.417 ***	0.352 ***	0.397 ***
	4.634	4.534	3.651	4.381
Holdings ESG score (raw)	-0.000	-0.000	-0.000	-0.001
	-0.852	-0.775	-0.850	-1.161
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	258,081	258,081	251,132	259,845
Adjusted R^2	0.080	0.080	0.081	0.081

Table 11: Fund flows and greenwashing in the PIS – E vs. S vs. G dimension

This table shows how fund flows respond to the inclusion of E- S- and G-related words in the PIS text block of a fund's prospectus, for funds that greenwash versus those that do not. The effect for greenwashing funds is shown in the row presenting the interaction of E/S/G intensity with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds (*non-GW*). Specifications differ in that *E/S/G intensity* is the estimated probability, according to the ML algorithm, that the PIS is E-related (in columns 1, 4), S-related (in column 2), or G-related (in column 3), or the *ESG keyword frequency* (in column 5); these probabilities are scaled to have the same variance as *ESG keyword frequency* so that estimated coefficients are comparable across specifications. Specifications also differ in the definition of the greenwashing dummy. In columns 1 and 4, the greenwashing dummy relates to the E dimension, i.e., it is 1 for any fund whose PIS is classified as E-related by the ML algorithm but whose value-weighted mean of its investments' E scores (in column 1) or carbon emissions (in column 4) is below the corresponding 50th percentile within the fund's investment category for that month. In columns 2 and 3, the greenwashing dummy is defined as in column 1, with the difference that it relates to the S dimension (in column 2) and the G dimension (in column 3). In column 5, the greenwashing dummy is 1 for any fund that includes an ESG keyword in its PIS but whose value-weighted means of its investments' E, S, and G scores are all below the 50th percentile within the fund's investment category for that month. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1) E-washing vs. E scores	(2) S-washing vs. S scores	(3) G-washing vs. G scores	(4) E-washing vs. GHG	(5) ESG-washing vs. E,S,&G scores
E/S/G intensity * GW	0.314 *	0.241 *	0.191	0.421 ***	1.079 **
	1.735	1.948	1.094	4.014	2.086
E/S/G intensity * Non-GW	0.388 ***	0.406 ***	0.419 ***	0.298 **	0.272 ***
	3.717	4.293	4.248	2.198	4.063
Holdings ESG score (raw)	-0.002	-0.002	-0.002	-0.000	-0.001
	-1.240	-1.342	-1.247	-0.856	-1.174
Fund Controls	Yes	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of Observations	259,845	259,845	259,845	254,840	259,863
Adjusted R^2	0.081	0.081	0.081	0.079	0.081

Table 12: Fund flows and greenwashing in the PIS – Robustness to activism

This table shows how fund flows respond to the inclusion of ESG-related words in the PIS text block of a fund's prospectus, for funds that greenwash versus those that do not, accounting for the possibility of fund activism. The analysis presented in this table is almost identical to that presented in Table 9, with the following differences. In Panel A, we exclude from the analysis all funds that state in their PIS that they engage with the firms they invest in (for details, see Section 5.6). In Panel B, we exclude fund-months for which the most recent date on which the fund reported its holdings (and so, the date on which the fund's holdings-based ESG score is based) is not at least a year after the date on which the fund introduced an ESG keyword to its PIS. In both panels, the effect for greenwashing funds is shown in the row presenting the interaction of *ESG-keyword frequency* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds ($non\text{-}GW = 1 - GW$). The specifications differ in the definition of the greenwashing dummy; for details on its definition, see the caption of Table 9. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

Panel A: Excluding engagement funds

	(1)	(2)	(3)	(4)
ESG keyword frequency * GW	0.627 ***	0.884 ***	0.809 ***	1.133 ***
	3.002	2.794	2.911	3.617
ESG keyword frequency * Non-GW	0.320 ***	0.317 ***	0.299 ***	0.303 ***
	3.596	3.757	3.581	3.700
Holdings ESG score (raw)	-0.000	-0.000	-0.000	-0.001
	-0.701	-0.677	-0.839	-1.157
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	257,007	257,007	250,132	258,771
Adjusted R^2	0.080	0.080	0.081	0.081

Panel B: Excluding first year for ESG funds

	(1)	(2)	(3)	(4)
ESG keyword frequency * GW	0.429 **	0.589 **	0.663 **	0.644 **
	2.172	2.130	2.592	2.018
ESG keyword frequency * Non-GW	0.275 ***	0.272 ***	0.236 ***	0.267 ***
	3.572	3.692	3.176	3.692
Holdings ESG score (raw)	-0.000	-0.000	-0.000	-0.001
	-0.663	-0.649	-0.758	-1.092
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	256,002	256,002	249,296	257,766
Adjusted R^2	0.080	0.080	0.080	0.080

Table 13: Fund flows and greenwashing in the PIS – Controlling for PRI signatories

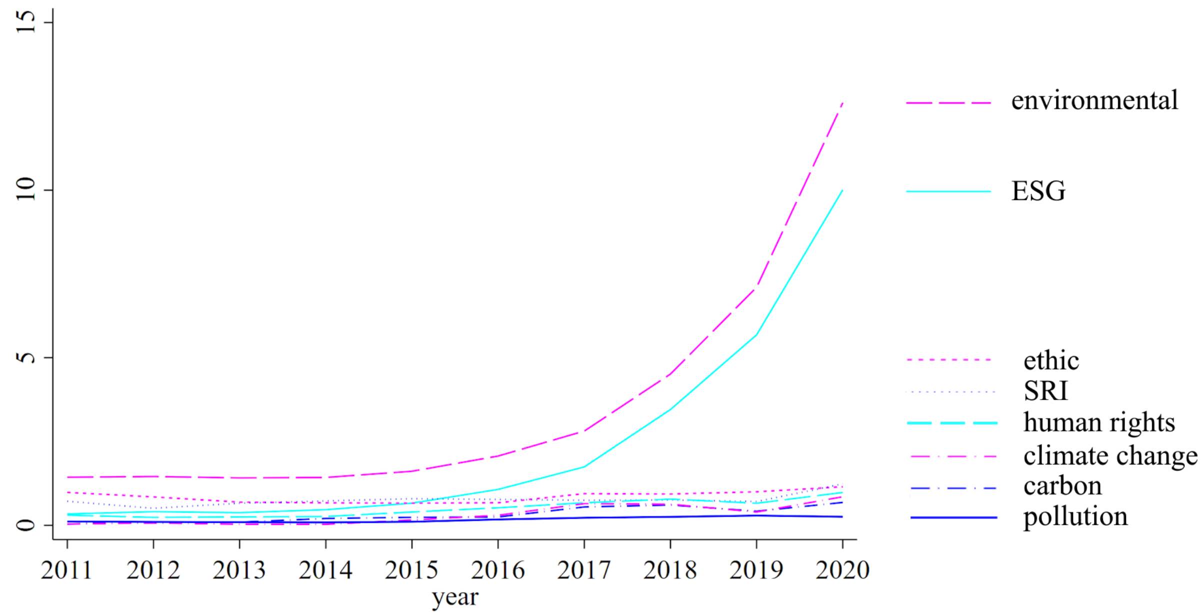
This table shows how fund flows respond to the inclusion of ESG-related words in the PIS text block of a fund's prospectus, for funds that greenwash versus those that do not, controlling for signatories of the United Nation's Principles for Responsible Investing (PRI). The analysis presented in this table is almost identical to that presented in Table 9, with the following differences. In Panel A, we exclude from the analysis all funds that state in their PIS that they engage with the firms they invest in (for details, see Section 5.6). In Panel B, we exclude fund-months for which the most recent date on which the fund reported its holdings (and so, the date on which the fund's holdings-based ESG score is based) is not at least a year after the date on which the fund introduced an ESG keyword to its PIS. In both panels, the effect for greenwashing funds is shown in the row presenting the interaction of *ESG-keyword frequency* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds (*non-GW* $= 1 -$ *GW*). The specifications differ in the definition of the greenwashing dummy; for details on its definition, see the caption of Table 9. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	Panel A: PRI alone	Panel B: With text- and fundamentals-based ESG measures			
	(1)	(1)	(2)	(3)	(4)
PRI signatory	0.002 ***	0.001	0.001	0.001 *	0.001 *
	3.312	1.645	1.654	1.702	1.933
ESG keyword frequency * GW		0.448 **	0.701 **	0.675 ***	0.771 **
		2.373	2.508	2.770	2.420
ESG keyword frequency * Non-GW		0.259 ***	0.251 ***	0.224 ***	0.243 ***
		3.481	3.554	3.130	3.502
Holdings ESG score (raw)		-0.000	-0.000	-0.000	-0.001
		-0.675	-0.643	-0.795	-1.129
Fund Controls	Yes	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of Observations	307,170	258,099	258,099	251,150	259,863
Adjusted R^2	0.084	0.080	0.080	0.081	0.081

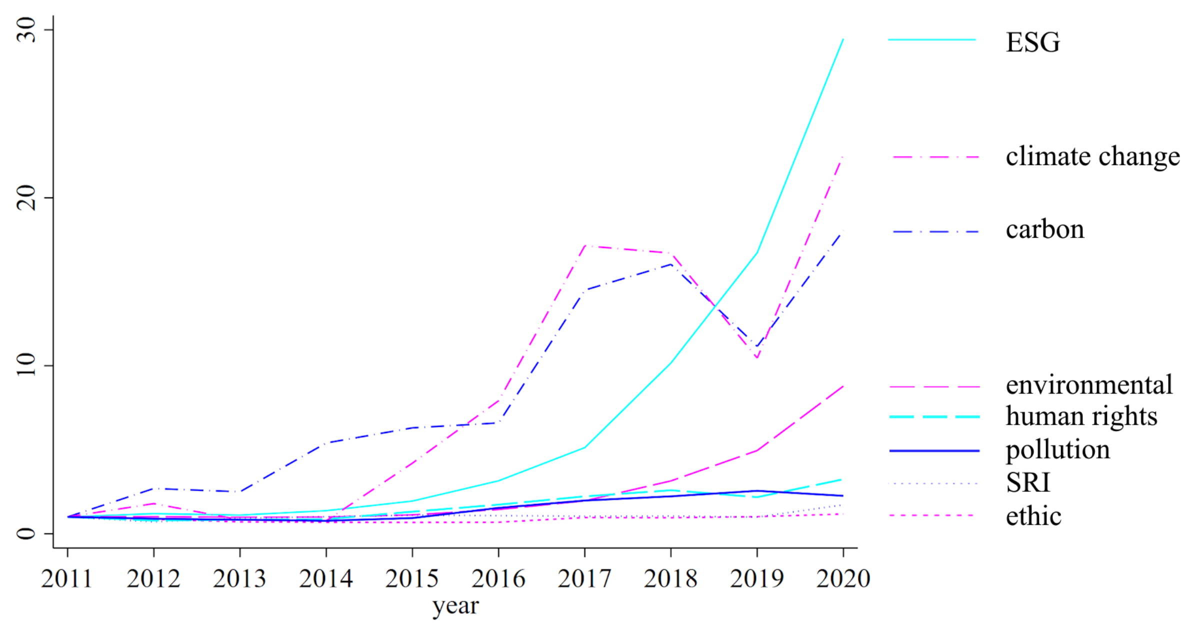
Table 14: Fund flows and greenwashing in the PIS – Controlling for PRI signatories

This table shows how fund flows respond to the inclusion of ESG-related words in the PIS text block of a fund's prospectus, for funds that greenwash versus those that do not, controlling for signatories of the United Nation's Principles for Responsible Investing (PRI). The analysis presented in this table is almost identical to that presented in Table 9, with the following differences. In Panel A, we exclude from the analysis all funds that state in their PIS that they engage with the firms they invest in (for details, see Section 5.6). In Panel B, we exclude fund-months for which the most recent date on which the fund reported its holdings (and so, the date on which the fund's holdings-based ESG score is based) is not at least a year after the date on which the fund introduced an ESG keyword to its PIS. In both panels, the effect for greenwashing funds is shown in the row presenting the interaction of *ESG-keyword frequency* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds ($non\text{-}GW = 1 - GW$). The specifications differ in the definition of the greenwashing dummy; for details on its definition, see the caption of Table 9. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the $10\%/5\%/1\%$ levels.

	(1)	(2)	(3)	(4)	(5)
PRI signatory	0.002 ***	0.001	0.001	0.001 *	0.001 *
	3.312	1.645	1.654	1.702	1.933
ESG keyword frequency * GW		0.448 **	0.701 **	0.675 ***	0.771 **
		2.373	2.508	2.770	2.420
ESG keyword frequency * Non-GW		0.259 ***	0.251 ***	0.224 ***	0.243 ***
		3.481	3.554	3.130	3.502
Holdings ESG score (raw)		-0.000	-0.000	-0.000	-0.001
		-0.675	-0.643	-0.795	-1.129
Fund Controls	Yes	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes
# of Observations	307,170	258,099	258,099	251,150	259,863
Adjusted R^2	0.084	0.080	0.080	0.081	0.081

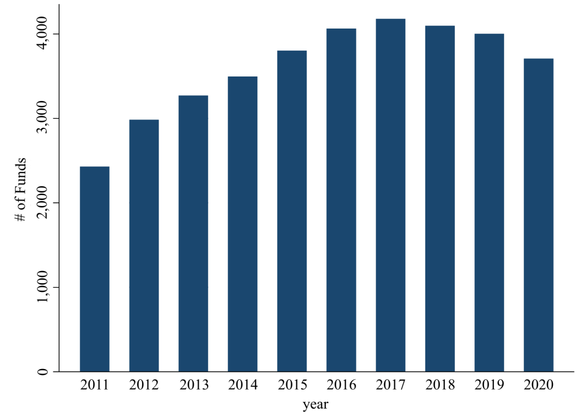


(a) Prevalence in prospectuses, as a percentage.

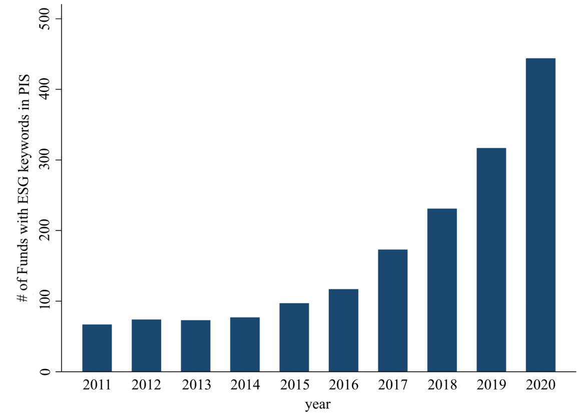


(b) Prevalence in prospectuses, as a multiple of the year-2011 value.

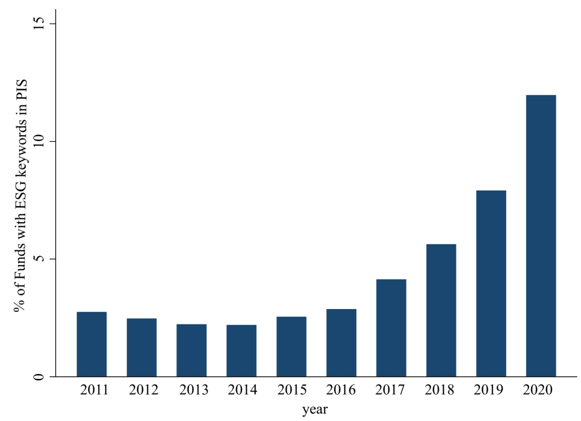
Figure 1: The evolution of ESG keywords' prevalence in funds' prospectuses over time. This figure plots, for our most common ESG keywords, the proportion of prospectuses in which the keyword appears. In Panel (a), the proportion of each keyword is plotted using raw numbers, expressed as percentages. In Panel (b), the proportion of each keyword is plotted as a multiple of the corresponding year-2011 value. The keywords and the line style representing the evolution of each keyword are shown in the legend to the right of each panel, sorted by prevalence in year 2020.



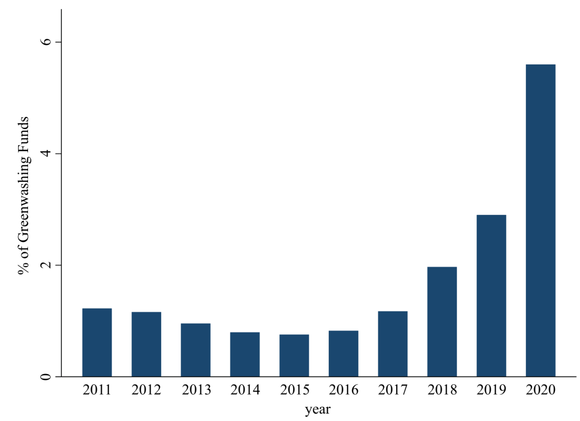
(a) Number of funds.



(b) Number of ESG funds.



(c) Proportion of ESG funds.



(d) Proportion of greenwashing funds.

Figure 2: The evolution over time of the number of funds (in Panel *a*), the number of ESG funds (in Panel *b*), the proportion (expressed as a percent) of ESG funds (in Panel *c*), and the proportion (expressed as a percent) of greenwashing funds (in Panel *d*). In a given year, a fund is classified as ESG if the PIS from its prospectus contains at least one ESG keyword. It is classified as greenwashing if, in addition, its holdings-based ESG measure is below the 50th percentile within the fund's investment category for that year.

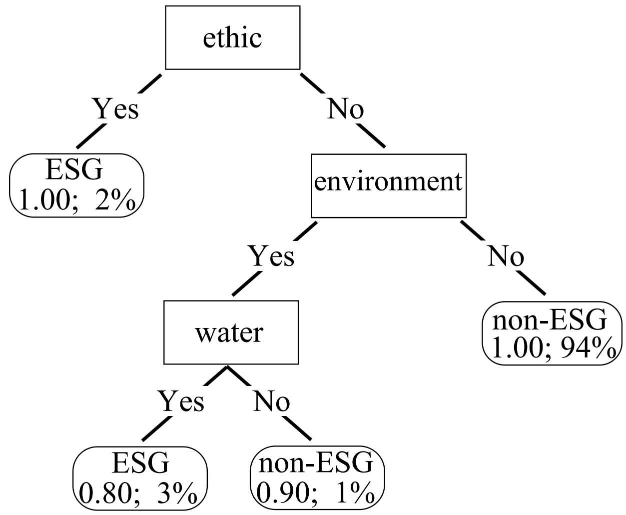


Figure 3: A (fictitious) example of a decision tree for the classification of PIS sections as ESG or non-ESG. Non-leaf nodes are indicated by regular rectangles and contain the feature on which cases are split. Leaf nodes are indicated by rounded rectangles and contain the majority classification (ESG or non-ESG) of the cases in the leaf, the proportion of these cases belonging to this classification, and the percentage of these cases in the sample. For example, the bottom left leaf node indicates that 3% of the cases in the sample do not contain the feature "ethic" but contain the features "environment" and "water"; of these, the majority (80%) are classified as ESG and the rest as non-ESG.

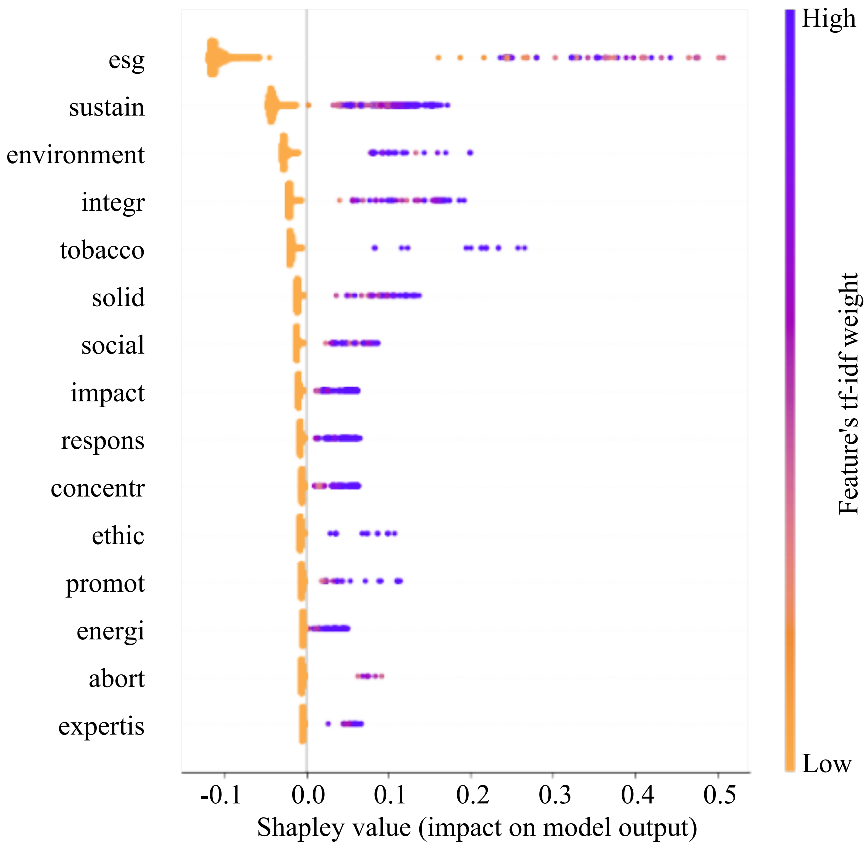
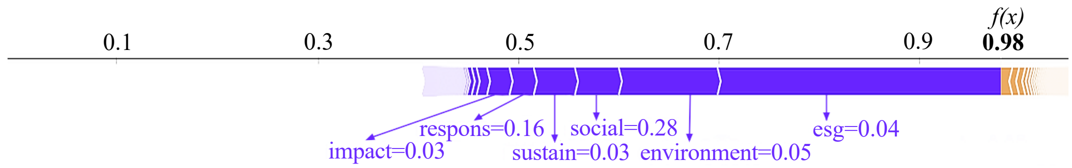
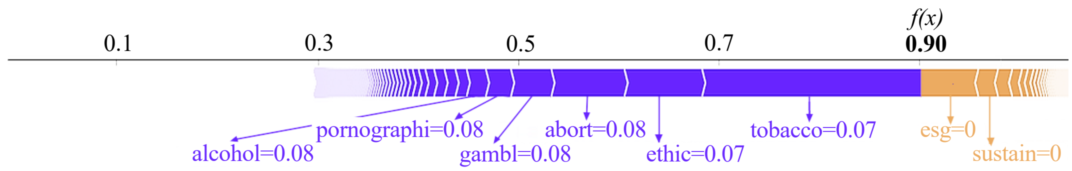


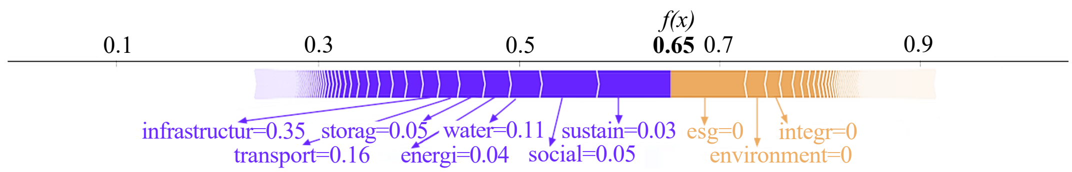
Figure 4: The Shapley values and tf-idf weights for the 15 most important features for random forest classification of PISs as ESG or non-ESG. For each of the most important features, labeled on the left, we plot the Shapley value for each PIS in the training subsample, indicated by a dot whose color ranges from light orange (for cases with a low tf-idf weight for the feature) to deep purple (for cases with a high tf-idf weight for the feature). For example, consider the feature "esg", shown in the top line of the figure. We see that, for PISs that do not contain it hence its tf-idf weight is zero, its Shapley value is negative, meaning that the absence of the "esg" feature makes it less likely that a PIS is classified as ESG. On the other hand, for PISs that contain it hence its tf-idf value is positive (even if relatively low, e.g., because the PIS text is long), its Shapley value is positive, meaning that the presence of the "esg" feature makes it more likely that a PIS is classified as ESG.



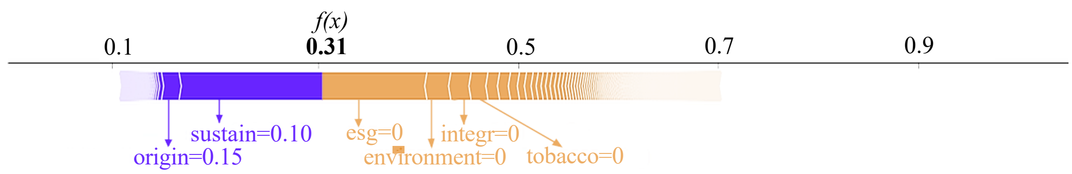
(a) Very high probability of PIS being allocated to the ESG class.



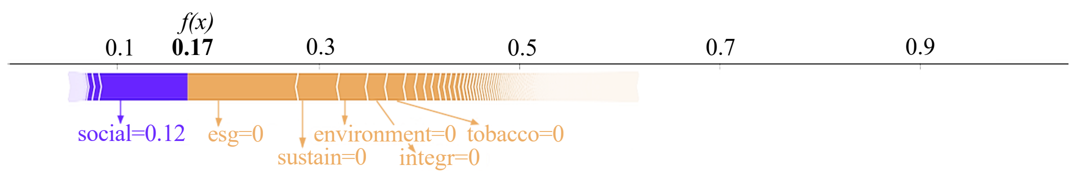
(b) High probability of PIS being allocated to the ESG class.



(c) Medium probability of PIS being allocated to the ESG class.



(d) Low probability of PIS being allocated to the ESG class.



(e) Very low probability of PIS being allocated to the ESG class.

Figure 5: Illustration of how features contribute to the random forest's prediction of a PIS's class allocation. Each panel shows a *force plot*, which shows the feature contribution to the predicted probability of allocation to the ESG class for a specific PIS, with predicted probabilities ranging from very high (in Panel a) to very low (in Panel e). The sequence of boxes in each plot represents the Shapley values of the most important features in the PIS; boxes on the left (right) and colored deep purple (light orange) represent the Shapley values for features with positive (negative) contribution toward classification of the PIS as ESG. The area of each box represents the Shapley value (positive if deep purple, negative if light orange) of the corresponding feature. The value next to each feature is the tf-idf weight of the feature in the PIS, with a 0 value indicating it is absent from the PIS and higher positive values indicating it has higher relative frequency within the PIS than across PISs. The predicted probability of belonging to the ESG class is denoted by $f(x)$ and written in bold.

Appendix

A Sample PIS text blocks

We show examples of funds' PIS text blocks containing ESG keywords, and our text-based measures. Sentences containing ESG keywords are in bold typeface.

1. Fund Name: BMO Small-Cap Growth Fund, Prospectus Date: December 2019

   The Fund invests at least 80% of its assets in growth-oriented common stocks of small-sized U.S. companies similar in size, at the time of purchase, to those within the Russell 2000 Growth Index. The largest company by market capitalization in the Russell 2000 Growth Index was approximately $7.0 billion as of October 31, 2019 and the median market capitalization of companies in the Index as of the same period was $869 million. The Fund may at times focus its investments in one or more sectors.

   The Adviser selects stocks using a unique, growth-oriented approach focusing on high quality companies with sustainable earnings growth that are available at reasonable prices, which combines the use of proprietary analytical tools and the qualitative judgments of the investment team. In general, the Adviser believes companies that are undervalued relative to their fundamentals and exhibit improving investor interest outperform the market over full market cycles. As a result, the Adviser's investment process begins by using tools to rank stocks based on expected returns, construct preliminary portfolios with the use of fundamental factors, and manage risk. **The Adviser also integrates environmental, social, and governance (ESG) considerations into its security selection, portfolio construction, and monitoring processes.** All purchases and sales of portfolio securities, however, are subjected ultimately to the investment team's qualitative judgments developed from their cumulative investment experience. The entire process is designed to focus on company fundamentals through both quantitative and qualitative analysis to balance return generation with risk management.

   From time to time, the Fund maintains a portion of its assets in cash. The Fund may increase its cash holdings in response to market conditions or in the event attractive investment opportunities are not available.

   Total word count: 281, ESG-keyword frequency: 1%, ESG-related word frequency: 7%.

2. Fund Name: Putnam Sustainable Leaders Fund, Prospectus Date: March 2018

   We invest mainly in common stocks of U.S. companies of any size, with a focus on companies that we believe exhibit a commitment to sustainable business practices. Stocks of companies that exhibit a commitment to sustainable business practices are typically, but not always, considered to be growth stocks. Growth stocks are stocks of companies whose earnings are expected to grow faster than those of similar firms, and whose business growth and other characteristics may lead to an increase in stock price. **We may consider, among other factors, a company's sustainable business practices (as described below), valuation, financial strength, growth potential, competitive position in its**

industry, projected future earnings, cash flows and dividends when deciding whether to buy or sell investments. We may also invest in non-U.S. companies.

Sustainable investing. **We believe that companies that exhibit leadership in sustainable business practices also often exhibit more profitable, durable financial returns with lower risk profiles. Accordingly, in selecting investments, we focus on companies that we believe have a demonstrated commitment to sustainable business practices. This commitment may be reflected through environmental, social and/or corporate governance (ESG) policies, practices or outcomes.**

Total word count: 187, ESG-keyword frequency: 9%, ESG-related word frequency: 78%.

B Sample pre-processing of PIS text block for ML algorithm

Here, we show an example of pre-processing a PIS text block to be used in the ML algorithm. Specifically, this text is from the PIS text block in the January 2020 filing of the Highland Socially Responsible Equity Fund. To conserve space, we only show approximately the first half of the text block.

The following is the original text. Words to be removed in the cleaning stage are shown in red bold style. Uninformative words that are to be removed subsequently are shown in blue italic style:[38]

The *Fund seeks* **to** *achieve* **its** *investment objectives* **by** investing **at least 80% of its** *net* **assets (**plus **any** *borrowings* **for** *investment purposes* **) under** *normal circumstances* **in** equity *securities* **, such as** *common* **and** *preferred stocks* **of** socially responsible **companies. This** *investment policy* **may be** *changed* **by the** *Fund* **upon 60** *days* **'** *prior written notice* **to** *shareholders*. **The** *Fund defines* socially responsible **companies as those** contained **in the** *MSCI* KLD **400** Social *Index* **(the "** *Index* **"), which is a** *capitalization weighted index* **of 400** *U.S. securities* **that** *provides exposure* **to companies with** outstanding **Environmental, Social and Governance ("**ESG**")** *ratings* **and** excludes **companies whose** *products* **have** *negative* social **or** environmental impacts. Highland Capital *Management Fund* Advisors, *L.P.* **("**HCMFA**" or the "** *Adviser* **")** *believes* **that using the** *Index* **as its** *investable universe* **is** *consistent* **with its** *obligations* **to** investors **because it** *provides* investors **with** access **to** socially responsible *stocks* **while** applying **the** *Adviser* **'s** *proprietary analysis* **in** *order* **to** *seek* **to** *provide better* risk-*adjusted returns* **than the** *Index* **across a market** *cycle*. **The** *Index* serves **as the** *investable universe* **from which the** *Adviser* chooses *investments* **to** *meet* **its** *policy* **to** *invest* **at least 80% in** socially responsible **companies and against which the** *Adviser applies proprietary* fundamental **and** technical *criteria* **, which** *includes* evaluations *based* **on** growth, value, trend **and** momentum. **The** *Fund*

[38] As mentioned in detail in the text, in the cleaning stage we replace three-word combinations like "environmental, social, and governance" with "esg", and we remove digits, punctuation, and other non-alphabet characters, generic and financial stop-words, dates and numbers, geographic names, and single-letter words. The uninformative words we subsequently remove are those that are used too frequently (in the top 10% of the distribution) in absolute terms across PISs and in relative terms as compared to their usage in fund names.

> **may** *invest* **in** equity *securities* **of** *issuers* **of any market** *capitalization*. *Investment selection* **will be** *based* **on** *proprietary* fundamental **and** technical *criteria*, **which** *includes* evaluations *based* **on** growth, value, trend **and** momentum. **The** *Adviser seeks* **to** *achieve* long-*term* growth **of** capital **across a full market** *cycle* **by** *focusing* **on an** *investable universe* **it** *believes* **has a** sustainable, long-*term competitive* advantage **against the market and the** *potential* **for both** capital appreciation **and** *increasing* dividends **over** *time*.

The following is the same text after cleaning (i.e., removing the words shown in red bold and blue italic style as shown above) and after stemming using the English Porter2 stemmer. Uninformative stemmed words that are to be subsequently removed are shown in blue italic style:

> invest plus equiti social respons social respons contain kld social outstand esg *exclud* social environment impact highland capit *advisor* hcmfa investor investor access social respons *appli* risk *serv choos* social respons fundament technic *evalu* growth valu trend momentum equiti fundament technic *evalu* growth valu trend momentum long growth capit sustain long advantag capit appreci dividend

C Additional tables

Here, we present some supplementary tables.

In Panel A of Table C.1, we repeat our analysis from Section 4, focusing on the period after the signing of the Paris Agreement and the publication of the Morningstar globe ratings in Spring 2016. Comparing with Table 4, we see that, in more recent years, the effect of text-based ESG measures becomes even stronger, while the effect of holdings-based measures becomes weaker and statistically insignificant. In Panel B of Table C.1, we repeat our analysis but include fund fixed effects. Comparing with Table 4, we see that the effects of text-based measures on fund flows have similar magnitudes with and without fund fixed effects; as expected, the statistical significance is lower with fund fixed effects, but still high when focusing on large within-fund time series variation. On the other hand, the effects of holdings-based ESG measures do not seem to carry over; for example, the dummy indicating exceptionally high levels of holdings-based ESG intensity is no longer statistically significant (t-statistic of 0.7 in columns 3–4 of Table C.1B with fund fixed effects vs. about 2.8 in columns 3–4 of Table 4 without fund fixed effects).

In Table C.2, we repeat the analysis from Section 4, adding an interaction term to test whether the effect on fund flows is different for funds targeted to institutional versus retail investors. We find that flows to the two types of funds do not respond differently to the PIS's ESG content, suggesting that our results are not driven by one type of investor.

Table C.1: Fund flows and the presence of ESG keywords in the PIS – 2016/03 onward and fund fixed effects

Like Table 4, this table shows how fund flows respond to text- and fundamentals-based ESG scores, but in Panel A it shows results for the period 2016/03 onward and in Panel B it shows results from a model that includes fund fixed effects. A fund's text-based ESG score is: the ESG-keyword frequency in the PIS text block, in columns 1–2; a dummy indicating if the PIS contains ESG keywords, in column 3; a dummy indicating if the ESG-keyword frequency in the PIS exceeds the median conditional on containing ESG keywords, in column 4. A fund's fundamentals-based ESG score is: the value-weighted mean of its investments' ESG scores, in column 1; this score's ranking within the fund's investment category-by-month, in column 2; a dummy indicating if this score is in the top 90% within the investment category-by-month, in columns 3–4. All specifications include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. */**/*** indicate significance at the 10%/5%/1% levels.

Panel A: For 2016/03 onward

	(1)	(2)	(3)	(4)
ESG keyword frequency	0.345 ***	0.355 ***		
	4.045	4.182		
ESG in prospectus			0.005 ***	
			3.733	
ESG keyword frequency > p50				0.007 ***
				4.831
Holdings ESG score (raw)	0.000			
	0.047			
Holdings ESG score (rank)		-0.000		
		-0.855		
Holdings ESG score > p90			0.001	0.001
			1.550	1.332
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	144,354	144,345	144,345	144,345
Adjusted R^2	0.068	0.068	0.068	0.069

Panel B: With fund fixed effects

	(1)	(2)	(3)	(4)
ESG keyword frequency	0.289 *	0.296 *		
	1.873	1.913		
ESG in prospectus			0.003	
			1.449	
ESG keyword frequency > p50				0.008 ***
				3.158
Holdings ESG score (raw)	-0.000			
	-0.206			
Holdings ESG score (rank)		-0.000		
		-0.149		
Holdings ESG score > p90			0.001	0.000
			0.707	0.653
Fund Controls	Yes	Yes	Yes	Yes
Fund Fixed Effects	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	261,267	258,207	258,207	258,207
Adjusted R^2	0.165	0.163	0.163	0.163

Table C.2: Fund flows and ESG keywords in the PIS – with institutional-fund interactions

This table shows how fund flows respond to text- and fundamentals-based ESG scores, but estimates a different effect for funds targeted to institutional investors and to retail investors by including terms that interact each ESG measure with a dummy indicating institutional-targeted funds. We classify a fund as institutional or retail using CRSP's institutional class indicator for its largest (by net asset value) share class. A fund's text-based ESG score is: the ESG-keyword frequency in its prospectus's PIS text block, in columns 1–2; a dummy indicating if the PIS contains ESG keywords, in column 3; a dummy indicating if the ESG-keyword frequency in the PIS exceeds the median conditional on containing ESG keywords, in column 4. A fund's fundamentals-based ESG score is: the value-weighted mean of its investments' ESG scores, in column 1; this score's ranking within the fund's investment category-by-month, in column 2; a dummy indicating if this score is in the top 90% within the investment category-by-month, in columns 3–4. All specifications include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, and the PIS total word count. In all specifications, the dependent variable is fund flows, expressed as a ratio over a fund's total net asset value and taking values between -1 and 1. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG keyword frequency	0.222 ***	0.218 ***		
	3.212	3.165		
ESG in prospectus			0.002	
			1.464	
ESG keyword frequency > p50				0.005 ***
				3.138
Holdings ESG score (raw)	-0.000			
	-0.091			
Holdings ESG score (rank)		0.000		
		0.609		
Holdings ESG score > p90			0.001 *	0.001
			1.685	1.434
ESG keyword frequency * Institutional Fund	0.137	0.149		
	1.041	1.138		
ESG in prospectus * Institutional Fund			0.002	
			1.141	
ESG keyword frequency > p50 * Institutional Fund				0.003
				1.022
Holdings ESG score (raw) * Institutional Fund	-0.000			
	-0.567			
Holdings ESG score (rank) * Institutional Fund		-0.000		
		-1.515		
Holdings ESG score > p90 * Institutional Fund			0.001	0.001
			1.039	1.061
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	261,269	258,212	258,212	258,212
Adjusted R^2	0.081	0.080	0.080	0.080

Internet Appendix for “Discretionary Information in ESG Investing: A Text Analysis of Mutual Fund Prospectuses”

Angie Andrikogiannopoulou, Philipp Krueger, Shema Mitali, and Filippos Papakonstantinou*

*Angie Andrikogiannopoulou and Filippos Papakonstantinou are at King’s College London, Philipp Krueger is at the University of Geneva, and Shema Mitali is at EPFL.

Here, we present additional information on the machine learning algorithm (in Section IA.A) and tables that provide additional information and support to the paper's results (in Section IA.B).

IA.A Additional information on the machine learning algorithm

First, we summarize the main steps in implementing the ML algorithm, taking the opportunity to add some details that were omitted from the main text of the paper for the sake of brevity.

1. We pre-process the text of all PISs, i.e., turn all characters to lowercase, remove non-alphabet characters, stop words, and uninformative words.

2. We use stratified sampling to select a subsample of PISs to be pre-classified, and then experts pre-classify these as E, S, G, and/or ESG or not.

3. We split the pre-classified subsample into a training and a testing subset. We only perform this split once, and we use the same training subset to train the separate ML algorithms that classify PISs as E, S, G, and/or ESG; this is so that the testing subset is untouched by (any variation of) the ML algorithm. To do this training/testing split, we use stratification with respect to the ESG classification. Regardless, the proportions of the other classifications of interest—E, S, and G—are also very similar across the two subsets.

4. We process the PISs using a bag-of-words approach that converts text into a vector of feature weights. When we compile the list of features, we use up to two-word combinations because we have not found higher-order combinations to be useful; exceptions are "environmental, social, and governance" and "socially responsible investing" which are replaced by "esg" and "responsible investing". As we explain below, we also exclude features that are likely to lead to overfitting.

5. To build the random forest, we need to build a number (or forest) of decision trees that classify PISs based on the presence or absence of certain features. To illustrate the process of building a decision tree, in Figure IA.A.1 (reproduced from the paper) we show a fictitious example of a decision tree for classifying PIS sections as ESG or non-ESG. For simplicity, we consider only three features: "ethic", "environment", and "water". We see that we first split the sample on "ethic", then on "environment", and finally on "water". It is clear that a first split on "ethic" works well because all PISs that contain it are classified as ESG, and a subsequent split on "environment" also works well because PISs that contain neither feature are all classified as non-ESG. The remaining PISs, i.e., those that do not contain "ethic" but contain "environment", can further—though imperfectly—be split on "water"; 80% of those that contain it are classified as ESG while 90% of those that do not contain it are classified as non-ESG.

6. We use cross-validation to optimally select the algorithm's hyper-parameters, specifically the number of decision trees in the forest, the size of the bootstrap subsample used to generate each

tree, the maximum depth of each tree, and the number of features considered at each node split. First, we construct a 4-dimensional grid of possible values, specifically we choose among number of trees in {500, 1000, 2000}, maximum depth in $\{10, 20, \infty\}$, bootstrap subsample size in {60%, 70%, 80%} of the training sample size, and number of features in {5%, 10%, 20%} of the total number of features. Then we use stratified k-fold cross-validation to pick the combination of hyper-parameters that optimizes performance. This scheme splits all the cases in the training sample into k subsamples—called *folds*—that have similar size and similar proportions of each class; the training is then done in k iterations, each time using $k-1$ folds to learn and leaving out one of the folds to test and, essentially, calculate accuracy metrics out-of-sample hence avoid overfitting. Figure IA.A.2 illustrates how the stratified $k-$fold cross-validation scheme works. We repeat this stratified k-fold scheme a number of times, with different randomization in each repetition, and then accuracy metrics (specifically, the F_1 score[1]) are averaged across repetitions. The fine-tuned hyper-parameters are the ones that yield the highest average accuracy metric across all iterations.

7. We remove features that may lead to overfitting. To do this, we identify features that are important in classifying the training sample but do not make much sense so are likely to be overfitted and lead to poor out-of-sample performance. To find the most important features, we use Shapley values; the Shapley value of a given feature for a given case is the weighted average—across all combinations of features that could be included in the model—of the difference in the model's prediction for this case with and without this feature. For each feature, we calculate the mean absolute Shapley value across cases, and then we sort features on this. We see that many important features make sense, but also that some do not; examples of the latter are "target", "equiti", and "research". While these features have a large average contribution in determining classification in the training sample, it is likely they will not work well in unseen cases, so it is preferable to exclude them from the model's training. To remove features, we iteratively train the model and each time we inspect the Shapley values for the 15 most important features and remove those for which there are cases with positive feature weight but zero Shapley value.[2]
8. With this finalized set of features, we fine-tune the algorithm's hyper-parameters again, and we train on the entire training subsample.
9. Finally, we use the trained algorithm model to classify *all* PISs in our data.

In Tables IA.A.1 and IA.A.2, we show how the algorithm performs in the testing subsample.

[1] $F_1 := \frac{2}{1/\text{precision}+1/\text{recall}}$, where precision $:= \frac{\text{\# true positives}}{\text{\# true positives+\# false positives}}$ and recall $:= \frac{\text{\# true positives}}{\text{\# true positives+\# false negatives}}$. Intuitively, the precision (or positive predictive value) is the probability that a PIS classified as ESG is truly that, and recall (or sensitivity) is the probability that an ESG PIS will be classified as such; F_1 is their harmonic mean.

[2] We end up removing the following features: alloc, avoid, develop, distribut, equiti, equiti equiti, financi, focus, fundament, futur, global, growth, incom, opportun, region, research, resourc, risk, servic, target, valu, world.

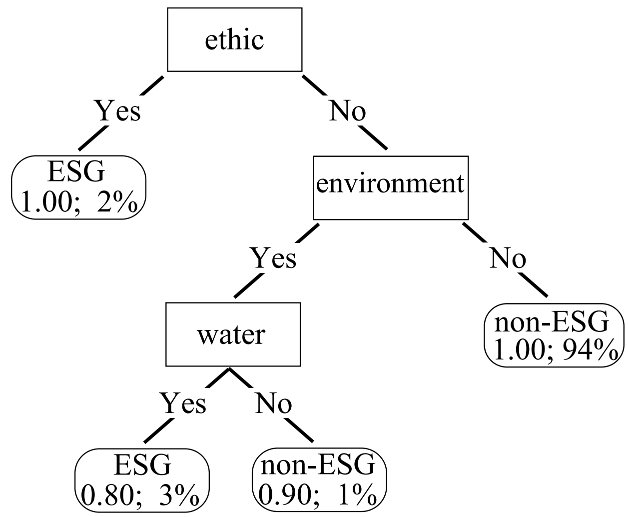


Figure IA.A.1: A (fictitious) example of a decision tree for the classification of PIS sections as ESG or non-ESG. Non-leaf nodes are indicated by regular rectangles and contain the feature on which cases are split. Leaf nodes are indicated by rounded rectangles and contain the majority classification (ESG or non-ESG) of the cases in the leaf, the proportion of these cases belonging to this classification, and the percentage of these cases in the sample. For example, the bottom left leaf node indicates that 3% of the cases in the sample do not contain the feature "ethic" but contain the features "environment" and "water"; of these, the majority (80%) are classified as ESG and the rest as non-ESG.

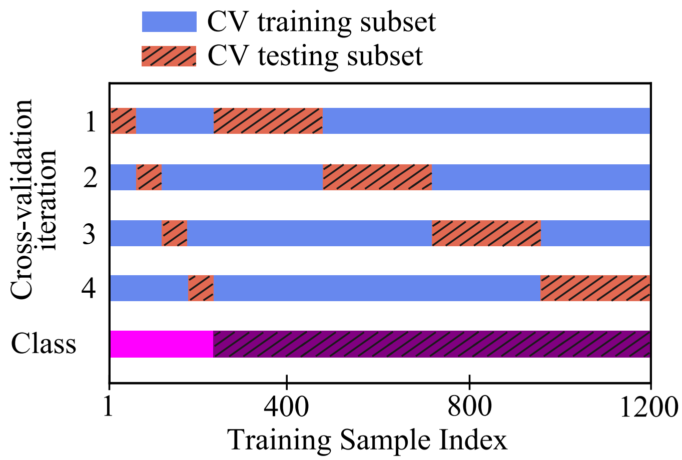


Figure IA.A.2: A (fictitious) example illustrating stratified k-fold cross-validation. The horizontal axis indicates the indices of the cases in the training sample, from case 1 to case 1,200, which are ordered so that those belonging to the first class (indicated, in the bottom row, with the pink-shaded area) appear before those belonging to the second class (indicated, in the bottom row, with the purple-shaded hatched area). For each iteration of the 4-fold cross-validation, the corresponding row shows which cases are used as a training subset (blue-shaded area) and which are used as a testing subset (orange-shaded hatched area).

Table IA.A.1: Performance for ESG classification on testing subsample — keyword-based vs. ML approach

This table shows the performance on the testing subsample of the keyword-based approach (in Panel A) and of the random forest approach (in Panel B) for classifying PISs as ESG or non-ESG. The performance metrics shown are: (i) accuracy, i.e., the proportion of correctly predicted classifications across both classes; (ii) recall, i.e., the accuracy within each class; (iii) precision, i.e., a classification's predictive value within each class; (iv) F_1, i.e., the harmonic mean of recall and precision within each class; and (v) P_4, i.e., the harmonic mean of the F_1 score across the classes.

Panel A: Performance metrics for keyword-based approach

	accuracy	recall	precision	F_1	P_4
non-ESG	98%	100%	98%	99%	91%
ESG		73%	100%	84%	

Panel B: Performance metrics for ML approach

	accuracy	recall	precision	F_1	P_4
non-ESG	97%	99%	98%	99%	89%
ESG		77%	85%	81%	

Table IA.A.2: Performance for E, S, and G classification on testing subsample

This table shows the performance on the testing subsample of the ML approach for classifying PISs as E or non-E (in Panel A), as S or non-S (in Panel B), and as G or non-G (in Panel C). The performance metrics are: (i) accuracy, i.e., the proportion of correctly predicted classifications across both classes; (ii) recall, i.e., the accuracy within each class; (iii) precision, i.e., a classification's predictive value within each class; (iv) F_1, i.e., the harmonic mean of recall and precision within each class; and (v) P_4, i.e., the harmonic mean of the F_1 score across the classes.

Panel A: Performance metrics for E classification

	accuracy	recall	precision	F_1	P_4
non-E	99%	99%	100%	99%	89%
E		86%	75%	80%	

Panel B: Performance metrics for S classification

	accuracy	recall	precision	F_1	P_4
non-S	99%	99%	100%	99%	91%
S		89%	80%	84%	

Panel C: Performance metrics for G classification

	accuracy	recall	precision	F_1	P_4
non-G	98%	100%	98%	99%	80%
G		55%	86%	67%	

IA.B Additional information on the data and results

In Table IA.B.1, we present summary statistics for the text characteristics of the entire PIS (in Panel A) and its ESG portion (in Panel B).

In Table IA.B.2, we repeat the analysis of fund characteristics related to greenwashing presented in Table 7 of the paper, with the following difference. Instead of including investment category fixed effects, here we include investment category-by-time fixed effects to control for time-varying unobserved heterogeneity in each investment category. We also exclude from this analysis the dummy indicating months after 2016/03, since it has no variation within each month.

In Tables IA.B.3 and IA.B.4 we repeat the analysis of fund flows presented in Tables 4 and 9 of the paper but, to allow for time-varying investment styles, fund category is inferred from the exposure of the past 24 months of fund returns to the Fama and French (1993) factors.

In Table IA.B.5, we present an additional specification showing how fund flows respond to greenwashing versus non-greenwashing funds that include ESG keywords in the PIS text block of their prospectus. Specifically, the analysis presented in Table IA.B.5 is similar to that presented in Table 9 of the paper, with the difference that here we exclude from the analysis fund-months for which the portfolio coverage of the stock-level ESG scores that we use to construct our fund-level holdings-based ESG measure is below 50%.

The results in these tables are very similar to those in the corresponding tables presented in the paper.

Table IA.B.1: Summary statistics of text characteristics of PIS text block in fund prospectus

Summary statistics for text characteristics of the entire PIS (Panel A) and of the ESG portion of the PIS, i.e., all sentences containing at least one ESG keyword (Panel B). Word count is the number of words. Text readability is calculated using the Flesch Reading Ease (Flesch, 1948) and the (sign-flipped) Gunning Fog Index (Gunning, 1952) measures, with higher values indicating a passage that is easier to understand. Text uniqueness captures the text's average uniqueness relative to the corresponding text in other funds' prospectuses submitted in the same calendar year (see Section 3.1 for details on the definition). Text tonality is measured using the frequency (expressed as a percent) of positive/negative/uncertain words as defined in the Loughran-McDonald sentiment word list. ESG positioning is measured as the proportion of the text from the beginning of the PIS to the first sentence containing an ESG keyword (*Distance to ESG text*) and as a dummy indicating if an ESG keyword appears in the PIS's first sentence (*ESG in first sentence*). The percentiles presented in Panel B are conditional on fund-months whose PIS contains ESG keywords.

Panel A: Entire PIS

				Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
Total word count	398,572	403.79	275.10	135	347	730
Text readability (Flesch)	398,572	19.13	9.92	7.58	19.58	30.49
Text readability (Fog)	398,572	-22.43	2.94	-25.78	-22.22	-19.19
Text uniqueness	398,558	0.00	0.99	-1.17	0.09	1.13
Text tonality (Uncertain word freq, as %)	398,572	2.25	1.08	0.91	2.23	3.62
Text tonality (Positive word freq, as %)	398,572	1.09	0.91	0.00	0.92	2.27
Text tonality (Negative word freq, as %)	398,572	0.48	0.53	0.00	0.37	1.20

Panel B: ESG portion of PIS

				Conditional Percentiles		
	# Obs	Mean	Std. Dev.	10^{th}	50^{th}	90^{th}
ESG-portion word count	17,325	107.19	119.59	17	65	240
ESG positioning (Distance to ESG text)	17,325	0.42	0.30	0.00	0.42	0.85
ESG positioning (ESG in first sentence)	17,325	0.13	0.33	0	0	1
Text readability (Flesch)	17,325	-15.64	29.53	-58.08	-7.25	10.77
Text readability (Fog)	17,325	-31.90	11.14	-46.08	-28.58	-22.63
Text uniqueness	17,325	0.00	0.96	-1.20	-0.04	1.29
Text tonality (Uncertain word freq, as %)	17,325	1.59	1.94	0.00	0.98	4.55

Table IA.B.2: Fund characteristics related to greenwashing – Alternative specification

This table shows which fund characteristics are associated with greenwashing. The analysis presented in this table is almost identical to that presented in Table 7 of the paper, with the difference that here we include investment category-by-time fixed effects instead of investment category fixed effects to control for time-varying unobserved heterogeneity in each investment category; as a result, we must exclude the dummy indicating months after 2016/03. The specifications presented in the table differ in the definition of greenwashing. In column 1, a greenwashing fund is one that includes an ESG keyword in its prospectus's PIS text block but whose value-weighted mean of its investments' ESG scores from MSCI (i.e., whose holdings-based ESG score) is below the 50th percentile within the fund's investment category for that month. In column 2, the holdings-based ESG score cutoff below which a fund is deemed to be greenwashing changes from the 50th to the 25th percentile. In column 3, the holdings-based ESG score cutoff is back at the 50th percentile as in column 1, but for a fund to be deemed a greenwasher it is additionally required that the fund's *returns*-based ESG score calculated using the Sharpe (1992) style analysis is below the 50th percentile within the fund's investment category for that month. In column 4, greenwashing is defined as in column 2 (i.e., a fund greenwashes if the holdings-based ESG score is below the 25th percentile), but the fund's holdings-based ESG score is calculated as the fund's investments' standardized ESG scores averaged across the MSCI, Sustainalytics, and Refinitiv databases. All specifications include investment category-by-time fixed effects and the explanatory variables fund age, size, expense ratio, turnover ratio, 12b-1 fees, prior 12-month mean flows, and prior 12-month mean α, and a dummy indicating funds targeted to institutional investors. The analyses include, for each fund, all months until the month the fund greenwashes for the first time. The dependent variable (the greenwashing dummy) is expressed as a percent, i.e., 0% or 100%, so the regression coefficients represent the percent change in a fund's likelihood to greenwash in a given month. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1) discrepancy with holdings	(2) discrepancy with holdings (higher)	(3) discrepancy with holdings & returns	(4) discrepancy with holdings (composite)
log(Fund size)	-0.007	-0.006 **	-0.008 ***	-0.003
	-1.649	-2.581	-2.746	-1.303
log(Expense ratio)	0.047 ***	0.030 ***	0.038 ***	0.034 ***
	2.694	2.881	3.201	3.220
Turnover ratio	0.003	-0.000	0.001	-0.002
	0.579	-0.107	0.156	-0.777
log(Effective 12b-1 fee)	3.498	1.783	-0.371	1.953
	1.055	0.946	-0.214	1.193
Prior 12-month mean flows	-0.094 **	-0.066 ***	-0.067 ***	-0.042
	-2.588	-2.737	-2.663	-1.428
Prior 12-month mean α	0.268	1.032	-0.437	0.766
	0.198	1.106	-0.413	0.760
log(Fund age)	-0.003	-0.002	0.011	-0.009
	-0.271	-0.254	1.481	-1.602
Fund for institutionals	0.015	-0.009	0.017	0.001
	0.693	-0.673	1.072	0.078
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	159,725	161,059	156,112	162,328
Adjusted R^2	0.008	0.007	0.007	0.007

Table IA.B.3: Fund flows and the presence of ESG keywords in the PIS – Alternative style categories

This table shows how fund flows respond to various definitions of text- and fundamentals-based ESG scores. The analysis presented in this table is almost identical to that presented in Table 4 of the paper, with the difference that here fund investment category is inferred from the exposure of the past 24 months of fund returns to the Fama and French (1993) factors. A fund's text-based ESG score is: the ESG-keyword frequency in its prospectus's PIS text block (columns 1–3); the relative length of the part of the PIS containing ESG keywords (column 4); a dummy indicating if the PIS contains ESG keywords (columns 5–6); a dummy indicating if the ESG-keyword frequency in the PIS exceeds the median conditional on containing ESG keywords (column 7). A fund's fundamentals-based ESG score is: the value-weighted mean of its investments' ESG scores (columns 2 and 4); this score's ranking within the fund's investment category-by-month (column 3); a dummy indicating if this score is in the top 50% within the investment category-by-month (column 5) or in the top 90% (columns 6–7). All specifications include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. */**/*** indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ESG keyword frequency	0.270 ***	0.329 ***	0.344 ***				
	3.852	4.691	4.882				
ESG text relative length				0.014 ***			
				3.813			
ESG in prospectus					0.004 ***	0.004 ***	
					3.779	3.597	
ESG keyword frequency > p50							0.007 ***
							4.959
Holdings ESG score (raw)		0.000		0.000			
		0.696		0.724			
Holdings ESG score (rank)			-0.000				
			-0.737				
Holdings ESG score > p50					-0.000		
					-0.820		
Holdings ESG score > p90						0.001 *	0.001 *
						1.921	1.677
Fund Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Observations	288,644	248,128	245,576	248,128	245,576	245,576	245,576
Adjusted R^2	0.082	0.082	0.082	0.082	0.082	0.082	0.082

Table IA.B.4: Fund flows and greenwashing in the PIS – Alternative style categories

This table shows how fund flows respond to the inclusion of ESG keywords in the PIS text block of a fund's prospectus, for funds that greenwash versus those that do not. The analysis presented in this table is almost identical to that presented in Table 9 of the paper, with the difference that here fund investment category is inferred from the exposure of the past 24 months of fund returns to the Fama and French (1993) factors. The effect for greenwashing funds is shown in the row presenting the interaction of *ESG-keyword frequency* with the dummy indicating greenwashing funds (*GW*), while for non-greenwashing funds it is shown in the row presenting the interaction with the dummy indicating non-greenwashing funds ($non\text{-}GW = 1 - GW$). The specifications differ in the definition of the greenwashing dummy. In column 1, the greenwashing dummy is 1 for any fund that includes an ESG keyword in its PIS but whose MSCI holdings-based ESG score is below the 50$^{\text{th}}$ percentile within the fund's investment category for that month. In column 2, the holdings-based ESG score cutoff below which the greenwashing dummy equals 1 changes from the 50$^{\text{th}}$ to the 25$^{\text{th}}$ percentile. In column 3, the greenwashing dummy is 1 for any fund that includes an ESG keyword in its PIS but whose MSCI holdings-based and returns-based ESG score (calculated using the style analysis of Sharpe, 1992) are both below the 50$^{\text{th}}$ percentile within the fund's investment category for that month. In column 4, the greenwashing dummy is defined as in column 2 (i.e., it equals 1 if the holdings-based ESG score is below the 25$^{\text{th}}$ percentile), but the fund's holdings-based ESG score is calculated as the fund's investments' standardized ESG scores averaged across multiple databases (MSCI, Sustainalytics, and Refinitiv). All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG keyword frequency * GW	0.510 ***	0.696 **	0.658 **	0.763 **
	2.713	2.523	2.311	2.469
ESG keyword frequency * Non-GW	0.301 ***	0.299 ***	0.276 ***	0.332 ***
	3.998	4.168	3.800	4.640
Holdings ESG score (raw)	0.000	0.000	0.000	-0.004 ***
	0.950	0.960	0.680	-3.893
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	245,468	245,468	238,773	246,952
Adjusted R^2	0.082	0.082	0.082	0.083

Table IA.B.5: Fund flows and greenwashing in the PIS – Better ESG score coverage

This table presents additional specifications showing how fund flows respond to greenwashing versus non-greenwashing funds that include ESG keywords in the PIS text block of their prospectus. The analysis presented in this table is almost identical to that presented in Table 9 of the paper, with the difference that here we exclude from the analysis fund-months for which the portfolio coverage of the stock-level ESG scores that we use to construct our fund-level holdings-based ESG measure is below 50%. The specifications differ in the definition of the greenwashing (*GW*) and non-greenwashing (*non-GW*) dummy; for details on its definition, see the caption of Table 9. *ESG-keyword frequency* is the frequency of ESG keywords in the PIS text block. All specifications control for a fund's fundamentals-based ESG score, measured as the value-weighted mean of its investments' ESG scores. All specifications also include investment category-by-month fixed effects and fund controls for age, size, expense ratio, 12b-1 fees, prior 1-month raw return and 12-month return ranked within investment category-by-month, dummies indicating if prior 12-month α is in the bottom or top 10% for the investment category-by-month, a dummy indicating funds targeted to institutional investors, and the PIS total word count. t-statistics from standard errors clustered two-ways at the fund and year-by-month levels are reported. $^{*}/^{**}/^{***}$ indicate significance at the 10%/5%/1% levels.

	(1)	(2)	(3)	(4)
ESG keyword frequency * GW	0.480 *	0.764 **	0.812 **	0.726 **
	1.726	1.997	2.326	2.125
ESG keyword frequency * Non-GW	0.191 **	0.183 **	0.145 *	0.172 **
	2.371	2.359	1.888	2.356
Holdings ESG score (raw)	-0.001	-0.001	-0.001	-0.001
	-1.232	-1.195	-1.312	-0.803
Fund Controls	Yes	Yes	Yes	Yes
Category-by-Time Fixed Effects	Yes	Yes	Yes	Yes
# of Observations	166,914	166,914	165,320	172,495
Adjusted R^2	0.082	0.082	0.083	0.082

References

Fama, E. F, and K. R French. 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33(1): 3–56.