WHEN SIMILARITY SELLS AND DIVERSITY DELIVERS

THE DUAL PREFERENCES DRIVING MUTUAL FUND FLOWS

SIMON ALLERGREN

GUSTAV TROLIN

Bachelor Thesis

Stockholm School of Economics

2026

When similarity sells and diversity delivers: the dual preferences driving mutual fund flows

Abstract:

Which characteristics shape investor demand for ESG mutual funds? Beyond performance, fees, and adherence to conventional investment strategies, we examine whether investors reward funds that adhere to the ESG strategies described in their prospectuses. Using SEC prospectus text, CRSP mutual fund data, and portfolio holdings for actively managed U.S. equity mutual funds from 2016 to 2024, we classify funds by ESG language in their Principal Investment Strategy sections and measure portfolio adherence to both conventional strategy peer groups and ESG-strategy peer groups. We show that fund flows increase when funds remain close to the average holdings of their conventional strategy peers, but not when ESG funds remain close to the average holdings of their ESG peers. Using idiosyncratic industry return shocks as instruments, we find suggestive evidence that investors may even prefer ESG funds that differentiate themselves from their ESG peer-group consensus. Our results uncover dual investor preferences in mutual fund flows: similarity sells in conventional strategy space, while diversity delivers in ESG strategy space.

Keywords:

ESG, Mutual funds, Fund flows, Strategy adherence, ESG adherence

Authors:

Simon Allergren (26038)
Gustav Trolin (25954)

Tutor:

Marieke Bos, Associate Professor, Department of Finance

Examiner:

Ramin Baghai, Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics

1. Introduction

The mutual fund literature has traditionally identified past performance and investor costs as central drivers of fund flows. Capital flows respond strongly and convexly to past performance (Chevalier and Ellison, 1997; Sirri and Tufano, 1998), while fees and other investor costs also shape allocation decisions (Barber, Odean and Zheng, 2005). More recently, however, the literature has shown that investors also respond to the content of funds' promised investment strategies. Abis and Lines (2024) demonstrate, using the Principal Investment Strategy (PIS) sections of mandatory SEC prospectus filings, that investors exhibit a clear preference for fund managers who adhere to the strategies they describe. Capital flows respond negatively when funds deviate from the average holdings of their text-based strategy peer groups, and are sensitive to relative performance within those peer groups. This establishes strategy adherence as a novel, economically meaningful dimension of investor demand that survives the inclusion of standard risk-adjusted performance controls.

At the same time, investor demand for environmental, social, and governance (ESG) investing has grown rapidly. Google searches for the term "ESG" grew at an average rate of fifty percent per year, quintupling over the period 2018-2023 (Andrikogiannopoulou, Krueger, Mitali and Papakonstantinou, 2025), and global sustainable fund assets under management reached \$3.9 trillion by end-2025 (Morningstar, 2026). This surge in demand has driven a parallel rise in discretionary ESG disclosure, as fund managers incorporate ESG language into their prospectuses to signal commitment to sustainable investment principles.

Andrikogiannopoulou et al. (2025) show that funds discussing ESG investing in their PIS attract additional annual inflows averaging 3.6% of total net assets, effectively doubling their capital inflows relative to non-ESG funds. This effect is stronger than that of fundamental-based ESG measures such as portfolio ESG scores or Morningstar globe ratings, confirming that narrative-based information in prospectuses serves as a key channel for investor capital allocation. Critically, however, a significant fraction of funds engage in greenwashing: they overstate their ESG commitment in the prospectus relative to the actual ESG characteristics of their holdings. Andrikogiannopoulou et al. (2025) found that greenwashing grew almost exponentially after 2016, reaching approximately 5.6% of actively managed funds by 2020, with an estimated \$18 billion misallocated to greenwashing equity funds in the United States in that year alone. Investors thus appear to respond more strongly to narrative disclosures than to underlying portfolio characteristics and may have difficulty distinguishing genuinely green funds from greenwashers. This asymmetry motivates our central research question:

> *Do investors penalise ESG-labelled mutual funds that deviate from their stated ESG strategies, as measured by subsequent fund flows?*

In the conventional setting, Abis and Lines (2024) document that investors discipline managers who depart from their promised strategies through negative flow responses. Whether this disciplinary mechanism extends to the ESG dimension is ex ante unclear. Adherence to an ESG mandate may be harder to evaluate than adherence to a conventional style mandate. Style drift is readily audited from widely available market data, such as whether a global fund holds international equities or whether a small-cap fund holds small-cap securities, and investors can identify mandate breaches with limited effort. Meanwhile, ESG mandate breaches are harder to verify because ESG-relevant information is often fragmented and buried in qualitative disclosures rather than standardised metrics (OECD, 2025 and Krueger et al., 2024), and standard third-party ratings provide noisy and partial signals of portfolio-level ESG content. Even sophisticated investors must therefore rely on a mix of correlated cues, including third-party ratings, holdings disclosures, media coverage, and regulatory scrutiny, rather than a clean portfolio-versus-mandate comparison. The investor base for ESG funds may also be driven by non-pecuniary motives, where the mandate is decoupled from expected risk-adjusted returns. Unlike conventional factor exposures, this lack of a direct link to financial outcomes may weaken the disciplinary mechanism that typically penalises strategy deviation. If investors do penalise ESG-strategy deviation, the mechanism documented by Abis and Lines (2024) generalises and ESG labels carry enforceable meaning; if they do not, the ESG label functions as a marketing device that investors reward on its own terms, independent of portfolio-level follow-through.

Our hypothesis concerns whether the disciplinary mechanism of Abis and Lines (2024) extends to the ESG dimension. If it does, our baseline flow regression should detect it as a positive and significant coefficient for funds that more closely resemble the average ESG fund within their strategy. Given the verification frictions noted above, we expect any such effect to be weaker than the analogous within-SPG effect.

We combine SEC prospectus filings with mutual fund data from the Center for Research in Security Prices (CRSP) to assemble a sample of 3,690 actively managed U.S. equity mutual funds between January 2016 and December 2024, of which 881 are classified as ESG-oriented in at least one month. Funds are classified as ESG using the keyword-based approach of Andrikogiannopoulou et al. (2025) applied to the Principal Investment Strategy section of each prospectus, with an independent classifier based on a generative AI model serving as a robustness check. We partition funds into Strategy Peer Groups (SPGs) based on Lipper Objective Codes and into ESG-Strategy Peer Groups (ESPGs) by splitting each SPG on its ESG classification, and measure adherence as the cosine similarity between each fund's portfolio weight vector and the corresponding peer-group centroid. Cosine similarity is invariant to differences in fund size and captures the directional tilt of a portfolio without being confounded by its scale, properties we argue make it better suited than the Euclidean distance measure of Abis and Lines (2024) to high-dimensional, sparse holdings vectors. We show that ESG funds hold portfolios that diverge systematically from their non-ESG counterparts within the same Lipper class, validating the premise that ESG commitment as expressed in the prospectus constitutes a distinct portfolio dimension.

Estimating the flow response to adherence reveals a stark asymmetry. In our OLS baseline, within-SPG adherence is rewarded with subsequent inflows, accumulating to 4% over twelve months, consistent with the disciplinary mechanism documented by Abis and Lines (2024). Within-ESPG adherence shows no flow response in the baseline specification. To address concerns that the OLS estimates may be confounded by unobserved manager attributes, reverse causality, or simultaneity, we instrument adherence with idiosyncratic industry return shocks following Abis and Lines (2024). The IV specification preserves the positive SPG response but reveals an active negative response to within-ESPG adherence, with a one-standard-deviation increase in fitted ESPG adherence corresponding to cumulative outflows of approximately 5.9% over nine months. We treat the IV evidence as suggestive rather than conclusive: with two endogenous regressors, the conventional first-stage F-statistic is not a sufficient test of instrument strength, and the supplementary diagnostics we report indicate the instruments are at the weaker end of what is conventionally accepted. Read together, the OLS and IV evidence are consistent with a setting in which investors at minimum do not discipline ESG funds for portfolio-level deviation, and possibly actively prefer divergence from the ESG peer-group consensus.

We interpret these patterns through the lens of investor differentiation. The OLS null on within-ESPG adherence indicates that investors, at minimum, do not reward funds whose portfolios drift towards the ESG peer-group consensus. Taking the negative IV coefficients at face value suggests a stronger interpretation: investors actively prefer divergence from that consensus. Within the ESG segment, close alignment with the ESPG centroid may signal passive ESG indexing, while distinctive holdings serve as an implicit signal of genuine sustainability conviction or specialist insight.

This suggests that the disciplinary mechanism governing conventional mandates appears to invert when applied to the ESG dimension: investors reward strategy conformity but value ESG-level distinctiveness. In Section 5, we evaluate the plausibility of this differentiation channel, discussing how it offers diversification benefits and how informational opacity leads investors to view divergence as a sign of managerial effort. Finally, given the tension between this effect and a primary competing channel, where sceptical investors might instead penalise non-adherence, we conclude by rigorously evaluating our 2SLS framework. This allows us to discuss whether the shift in our estimates reflects a true causal effect or is a consequence of identification challenges inherent in our empirical setting.

Our paper contributes to the literature in three ways. First, we document a novel asymmetry in how investors discipline mutual fund managers. The flow-based mechanism that ties managers to their stated strategy mandates, established by Abis and Lines (2024), does not extend uniformly to the ESG dimension; if anything, the evidence suggests it inverts. This qualifies the Abis-and-Lines disciplinary mechanism, showing it to be mandate-specific rather than universal. Second, we refine the greenwashing narrative of Andrikogiannopoulou et al. (2025). The information asymmetry between investors and managers operates not only at the boundary between

ESG and non-ESG funds, but also within the ESG segment, in a direction that complicates the standard greenwashing story: the funds that most closely resemble their ESG peers are also those from which capital flows away. Third, we propose a methodological framework that combines a binary text-based ESG classification with a continuous portfolio-level adherence measure based on cosine similarity, transparent and replicable without reliance on the proprietary methodologies of commercial rating agencies. The framework is directly applicable to disclosure-regulation analysis: it provides a way to identify funds whose portfolio behaviour departs systematically from their ESG narrative and to study how capital flows respond.

The remainder of the paper is organised as follows. Section 2 describes the data sources, sample construction, and key variables. Section 3 details the ESG and strategy classification methodologies, including both the keyword-based baseline and the generative AI robustness check. Section 4 presents the empirical analysis: Section 4.1 develops the cosine similarity adherence measure and validates it against portfolio-level differences; Section 4.2 estimates the baseline flow regressions; and Section 4.3 implements the instrumental variable strategy and presents the second-stage results. Section 5 interprets the results and potential underlying mechanisms. Section 6 concludes.

1.1. Related literature

This thesis draws on and contributes to four strands of literature: the role of investment mandate adherence in the mutual fund industry, text-based analysis of ESG disclosure and greenwashing in mutual funds, the broader determinants of mutual fund flows, and the measurement of institutional portfolio similarity.

Abis and Lines (2024) show that investors penalise mutual fund managers for departing from their stated investment strategies, documenting a significant flow-based disciplinary mechanism that binds fund managers to their promised mandates. Using k-means clustering applied to tf-idf encoded PIS descriptions filed with the SEC over 2000-2017, they construct 17 strategy peer groups and show that portfolio divergence from a fund's peer group average is penalised by capital outflows, with the effect persisting for more than twelve months and reverting faster under outflow pressure. We take this strategy peer group framework as our organisational backbone and extend it by introducing ESG commitment as an additional dimension of promised strategy.

The second foundational reference is Andrikogiannopoulou, Krueger, Mitali and Papakonstantinou (2025), who construct novel text-based measures of mutual fund ESG commitment by analysing the PIS sections of fund prospectuses using both a keyword-based classifier and a machine learning approach. They find that text-based ESG measures have a stronger effect on fund flows than fundamentals-based measures such as holdings-based ESG scores or commercial ratings, and document a growing and economically significant greenwashing problem in the actively managed fund industry. We adopt their keyword-based identification methodology to classify funds by ESG classification within the strategy peer group structure of Abis and Lines (2024), and extend their analysis by benchmarking funds against ESG peers within their specific peer group rather than against broad investment-category averages, yielding a more precise measure of ESG strategy adherence that controls for strategy-level heterogeneity in portfolio composition.

A broader literature documents that mutual fund capital flows respond strongly, and asymmetrically, to past performance. Chevalier and Ellison (1997) and Sirri and Tufano (1998) establish the canonical result that investors chase top performers while disciplining laggards only weakly, producing a convex flow-performance relationship that in turn shapes managerial risk-taking incentives. Barber, Odean and Zheng (2005) extend this line of work by showing that visible, salient fund characteristics, most notably expense ratios and front-end loads, exert an independent influence on flows beyond realised returns. This literature matters for our setting because funds that deviate from their ESG peer group may also underperform their peers on standard risk-adjusted metrics, in which case any relationship between ESG-strategy adherence and flows would merely restate the classical flow-performance result in a different guise. We address this concern directly by including Fama-French six-factor alpha (Fama and French, 2018) and SPG-adjusted returns (Abis and Lines, 2024) among the controls in our main flow regressions, so that the coefficient on adherence captures a response to mandate fidelity that is orthogonal to standard performance-chasing behaviour. Our contribution to this literature is therefore to document ESG-strategy adherence as a second, non-performance dimension along which investors allocate capital, once the classical flow-performance channel is accounted for.

The methodological approach we use to measure portfolio similarity is drawn from Girardi, Hanley, Nikolova, Pelizzon and Getmansky Sherman (2021), who develop a general framework for measuring institutional portfolio overlap using cosine similarity. Although their setting is the insurance industry, their central insight that cosine similarity captures the direction of portfolio tilt and is invariant to differences in institution size transfers directly to the mutual fund context. We adapt this approach to measure the directional alignment between each fund's portfolio and the within-strategy ESPG centroid and argue that it is better suited to this purpose than the Euclidean distance measure employed by Abis and Lines (2024).

2. Data

2.1. Data Collection and Description

Our dataset is assembled from multiple sources, which we outline in the following subsections.

2.1.1. Fund Prospectuses

We collect prospectus data covering an exhaustive list of filings made by each mutual fund from the Mutual Fund Prospectus Risk/Return Summary Data Sets, which are published quarterly on the SEC website. The dataset contains both numerical and textual information, including fund and filing identifiers (e.g., fund name, the filing company's Central Index Key (CIK), the fund portfolio's Series ID, and the Accession Number identifying the specific filing), a complete set of text blocks included in the summary section of the prospectus (e.g. describing business risks, investment restrictions, and fund strategy), as well as the filing date of the document.

For the purposes of our analysis, we focus on the text block labelled "StrategyNarrativeTextBlock," which summarises the fund's Principal Investment Strategy. The required content of this section is specified in Form N-1A published by the U.S. Securities and Exchange Commission, which states:

"Based on the information given in response to Item 9(b), summarise how the Fund intends to achieve its investment objectives by identifying the Fund's principal investment strategies (including the type or types of securities in which the Fund invests or will invest principally) and any policy to concentrate in securities of issuers in a particular industry or group of industries."

Our regression sample covers the period from January 2016 to December 2024, but we download prospectus filings over a wider window from January 2015 to December 2025. The 2015 filings let us populate the earliest months of the regression sample using the most recent prospectus a fund had on file at the time, rather than backward-filling from a later 2016 filing. The 2025 filings serve as an attrition signal at the end of the sample: combined with 2025 return data, they confirm whether a fund remains active. For active funds, this allows us to forward-fill the most recent 2024 prospectus through December 2024; for funds that exit the sample, the absence of 2025 filings and returns ensures we do not extend their observations beyond the point at which they ceased to exist.

Funds are obligated to file their prospectuses at a yearly frequency, but sometimes funds file at shorter intervals due to updates to the prospectus. Our study is conducted at the fund-month level, and we therefore forward-fill the prospectus data at a monthly frequency using the most recent filing.

2.1.2. Fund characteristics

Share-class-level fund information comes from the Center for Research in Security Prices Survivorship-Bias-Free U.S. Mutual Fund Database for the period January 2016 to December 2024. The initial sample is restricted to funds that filed at least one prospectus during the sample period and includes 17,308 mutual funds.

Funds are matched across the SEC and CRSP datasets by linking the fund company and portfolio identifiers in the SEC data (CIK + Series ID) to the share-class identifier crsp_fundno in CRSP using the crsp_cik_map provided by Wharton Research Data Services. In addition to linking the two databases, this matching procedure serves two further purposes: (i) it expands the panel to the share-class level, as the same CIK + Series ID pairs may correspond to multiple share classes; and (ii) funds without prospectus filings between January 2016 and December 2024 do not produce valid matches and are therefore excluded from the sample.

We further restrict the sample to actively managed U.S.-based mutual funds investing in both domestic and foreign assets using observable fund characteristics from the CRSP database. Specifically, we exclude exchange-traded funds, money market funds, fixed-income funds, variable annuity funds, mortgage funds, currency funds, and other miscellaneous categories using keyword searches in fund names, CRSP style codes, and database flags, in addition to applying a common-stock holdings threshold.[1] Finally, we exclude funds whose monthly observations cover less than twelve months and funds whose share-class identifiers find no valid match in the CRSP database.

2.1.3. Fund flows

Monthly returns and total net assets (TNA) are obtained from the CRSP Mutual Fund Monthly Returns dataset at the share-class level. This data is aggregated to the fund level using CRSP's portfolio identifier (crsp_portno), which we link to the share-class identifier (crsp_fundno) using the crsp_portno_map provided by WRDS.

Fund-level TNA is calculated as the sum of the TNAs of all associated share classes. Fund returns are computed as the weighted average of share-class returns, where weights are based on beginning-of-month TNA.

Following the mutual fund literature, monthly fund flows are defined as the fraction of the month-to-month change in TNA that is not explained by returns. Formally,

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

where $\text{TNA}_{i,t}$ denotes total net assets and $r_{i,t}$ denotes net return of fund i over month t.

To reduce the impact of potential outliers related to incorrect data handling, we winsorize fund flows at the 99% and 1% levels.

2.1.4. Holdings

Fund-level portfolio holdings data are obtained from the CRSP Mutual Fund Holdings dataset for the period January 2015 to December 2025. When observations are missing or reported only at quarterly frequency, we forward-fill holdings at a monthly frequency for up to twelve months.

---

[1] We exclude index funds using the **index_fund_flag** variable in CRSP and by searching for the keywords "DOW 30," "NASDAQ," "INDEX," and "S&P." ETFs and variable annuity funds are excluded using the **et_flag** and **vau_fund** variables, as well as keyword searches for "ETF." Fixed-income, currency, mortgage, and other miscellaneous funds are excluded by removing all funds with CRSP style codes (**crsp_obj_cd**) beginning with "I" (Fixed Income) or "O" (Other). Finally, we impose a minimum common-stock holdings threshold of 50%, excluding observations below this cutoff using the **per_com** variable.

Inspection of the holdings data indicates that 63.2% of holding-month observations correspond to original reports. Moreover, 98.3% of observations are not forward-filled for more than one quarter, and 99.6% are not forward-filled for more than two quarters.

2.1.5. Additional Controls and Final Sample

Additional control variables, including fund size, age, turnover ratio, and fees, are obtained from the CRSP Mutual Fund Database and aggregated to fund level by taking the mode, sum or weighted average depending on the nature of the variable.

Since our empirical design relies on peer-group benchmarks, we impose two final sample restrictions based on ESG fund availability within Lipper Objective Code groups. Specifically, after classifying funds as ESG or non-ESG and assigning Lipper Objective Codes (see Sections 3.1 and 3.2), we first retain only those objective code groups that contain at least 15 ESG funds for a minimum of 12 months. We then further exclude fund-month observations in which the corresponding objective code contains fewer than 15 ESG funds in that specific month. These restrictions ensure sufficiently populated peer groups and reduce noise in the benchmark estimates, thereby increasing the plausibility of our identifying assumption that the cross-sectional average portfolio provides a meaningful representation of the stated ESG mandate rather than idiosyncratic holdings of a small number of funds.

Our final sample covers 3,690 funds over the time period January 2016 to December 2024, out of which 881 mention ESG in their prospectus in at least one month during our sample period. We follow each fund for 76 months on average. Table 1 presents descriptive summary statistics for the matched sample. Table 2 in Appendix A illustrates the sample construction process, outlining the number of funds lost at each filtering step.

Table 1: Summary statistics

Note: Panel A reports summary statistics including the number of observations (fund months where the data is available), mean, standard deviation, minimum, maximum, and 25th, 50th and 75th percentiles for a series of variables in the data. Panel B limits the sample to only months where the PIS section of the fund's prospectus included an ESG related keyword.

Panel A: Full Sample

				Percentiles		
	N	**Mean**	**SD**	**25th**	**50th**	**75th**
TNA ($M)	266572	2858.21	11294.66	62.70	333.30	1461.90
Fund Age	266566	16.89	13.97	6.51	14.19	23.85
Expense Ratio (%)	256735	0.87	0.44	0.65	0.90	1.09
Turnover Ratio (%)	257542	51.89	70.25	20.00	37.00	63.32
Fund Flows (%)	261209	-0.04	4.14	-1.26	-0.41	0.61
Raw Return (Monthly) (%)	261436	0.89	5.06	-2.12	1.29	3.84
FF6 Alpha (Monthly) (%)	250895	-0.19	0.47	-0.39	-0.16	0.03

Panel B: ESG in PIS

				Percentiles		
	N	**Mean**	**SD**	**25th**	**50th**	**75th**
TNA ($M)	41086	1550.57	5732.93	49.20	254.85	1089.38
Fund Age	41086	16.13	14.29	5.26	13.01	24.44
Expense Ratio (%)	40349	0.87	0.30	0.75	0.90	1.03
Turnover Ratio (%)	40313	50.12	42.19	23.00	39.00	63.00
Fund Flows (%)	40664	0.18	4.32	-1.20	-0.31	0.86
Raw Return (Monthly) (%)	40710	0.87	5.33	-2.63	1.30	4.20
FF6 Alpha (Monthly) (%)	37527	-0.23	0.55	-0.47	-0.19	0.04

3. ESG-Strategy Peer Groups

3.1. ESG Classification

3.1.1. Dictionary based method

Following Andrikogiannopoulou et al. (2025), we classify funds using a predefined list of ESG-related keywords applied at the sentence level to the cleaned PIS text of each fund. The keyword list is drawn directly from Andrikogiannopoulou et al. (2025) and covers terms associated with environmental, social, and governance investing, including references to responsible investment, ESG criteria, and related concepts. We split each fund's PIS into individual sentences, search for keywords within each sentence, and assign each fund a binary ESG indicator equal to one if at least one keyword appears in its PIS in a given year and zero otherwise.

Although a continuous intensity measure could in principle be constructed from the relative frequency of keyword matches across sentences, we opt for a binary classification. Our analysis is structured around adherence within the ESG segment, and a binary classifier is what allows us to define that segment cleanly: each fund is either ESG or non-ESG, and ESG funds are then partitioned into ESG-Strategy Peer Groups against which portfolio-level adherence is measured. A continuous intensity measure would blur the boundary between ESG and non-ESG funds and complicate the construction of peer-group centroids without offering a clear empirical gain. Andrikogiannopoulou et al. (2025) show that a binary indicator of whether a fund mentions ESG in its PIS drives fund flows in a similar way to continuous intensity measures, suggesting that investors respond to the presence of an ESG commitment rather than its depth. This is consistent with the broader literature on investor limited attention, which holds that salient and easy-to-process signals drive capital allocation decisions more strongly than granular differences in the intensity of disclosure (Barber, Odean and Zheng, 2005). A binary classification also avoids introducing noise from variation in keyword frequency that may reflect differences in writing style rather than genuine differences in ESG commitment.

In constructing this measure, we follow the conservative approach of Andrikogiannopoulou et al. (2025), deliberately prioritising the limitation of false positives over the avoidance of false negatives. This approach accepts that some genuinely ESG-oriented funds may be missed in exchange for a cleaner and less biased classification. Applying the keyword list to our data revealed ambiguities of this kind. For instance, the abbreviation SRI matches fund names referencing Sri Lanka, causing geographically focused funds to be incorrectly classified as ESG, which we correct manually. Similarly, terms such as "sustainable" are absent from the keyword list as they are too context-dependent to reliably signal ESG commitment, meaning that funds whose ESG language centres on such terms could go undetected. While these are illustrative examples, they point to a broader limitation of any keyword-based approach: conservative classification may leave a subset of ESG-oriented funds unclassified. This motivates the Generative AI method described in Section 3.1.2, which serves to complement the dictionary-based approach and strengthen the robustness of our overall classification.

3.1.2. Generative AI method

Recent advances in large language models have made it possible to apply generative AI to the classification of complex financial texts in a way that goes beyond the mechanical pattern matching of keyword-based approaches. To complement the dictionary-based method described in Section 3.1.1, we employ Claude Opus 4.7, Anthropic's most capable model at the time of writing, to classify each fund's PIS as ESG or non-ESG. Unlike keyword matching, which relies on the presence of specific terms, a large language model (LLM) can assess the broader context and intent of a prospectus, making it better suited for capturing ESG commitment expressed through varied or unconventional language that falls outside a predefined keyword list.

The classification is conducted by feeding the raw PIS text of multiple funds in batches directly to the model, without prior cleaning or pre-processing, so that the model can assess the text as a typical investor would encounter it. We adapt the prompt of Andrikogiannopoulou et al. (2025), modifying it to ask directly whether a reasonable investor would interpret the fund's stated strategy as ESG-oriented. The full prompt is presented in Appendix B.

The prompt is deliberately framed around the SEC's reasonable investor standard, which reflects the legal threshold applied in regulatory assessments of fund disclosure adequacy. This framing ensures that the classification is not based on incidental mentions of ESG-adjacent language but on whether a reasonably informed investor would interpret the fund's stated strategy as genuinely ESG-oriented. The model produces a binary output of 1 for ESG and 0 for non-ESG for each fund, consistent with the classification scheme of the dictionary-based method. The reasonable-investor standard aligns the classifier with the economic agent whose behaviour we measure: since our outcome is investor flow response, a fund should count as ESG only if a typical investor reading the prospectus would recognise it as such.

The keyword-based classification serves as the baseline throughout the paper. The AI classifier complements this baseline by capturing funds whose ESG commitment is expressed through language that a predefined keyword list cannot anticipate, for example through terms too context-dependent to include in a fixed vocabulary, or through phrasing that signals ESG intent without invoking standard ESG terminology. To assess whether our main results are robust to this broader notion of ESG commitment, we compare the two classifications across our full sample of 4,745 funds (before the last filtering step which relies on the classifications themselves; see step 8 in Appendix A). The two classifiers agree on 99.0% of classifications, with 46 disagreements: 16 funds classified as ESG by the keyword approach but not by the AI, and 30 classified as ESG by the AI but not by the keyword approach. Treating the AI classification as a benchmark, the keyword classifier achieves a precision of 94.8% and a recall of 90.7% (F1 = 0.93). The high cross-classifier agreement supports the robustness of our ESG classification. The disagreements are concentrated in values-based exclusionary funds (e.g., faith-based screens based on religious principles) that the AI captures but the keyword list does not, and in sector funds where ESG-adjacent terms appear incidentally rather than as part of the investment process.

3.2. Strategy Classification

To increase the baseline homogeneity between individual funds and their benchmark portfolios, we cluster funds based not only on their prospectus mentioning ESG, but also on their stated strategy. This secondary classification is necessary to prevent mandate confounding, where the idiosyncratic signals of ESG-driven portfolio construction are statistically overshadowed by broader determinants of asset allocation, such as market capitalisation or style tilts. By benchmarking small-cap ESG funds against their small-cap non-ESG peers, rather than the aggregate ESG universe, we effectively control for unobserved strategy-specific factors that dictate baseline holdings. This granular approach minimises estimation noise and ensures that our adherence measures isolate the marginal impact of the ESG mandate, distinct from the commonalities inherent in the fund's primary investment style.

To implement this classification, we utilise Lipper Objective Codes from the CRSP database. Crucially, these codes align with our ESG classifier in that both are based on ex-ante prospectus language rather than ex-post portfolio characteristics. This approach ensures our benchmarks reflect the fund's stated mandate rather than its realised style, thereby avoiding the mechanical endogeneity inherent in holdings-based clustering and isolating the investor's reaction to the implementation of a stated promise.

Figure 1 shows that the number of funds mentioning ESG in their prospectus, and the total net assets (TNA) invested in them, increased substantially over the course of our sample period. As a result, our results are likely to more precisely reflect behaviour observed at the later stage of our sample, as the individual fund-months are more likely to clear the 15 ESG fund per Lipper Objective Code threshold. This concentration enhances the external validity of our findings, ensuring that they capture the dynamics of the modern ESG landscape rather than nascent trends during its maturation stage.

Figure 2 presents the distribution of fund-month observations across Lipper Objective Codes, alongside the share of funds with explicit ESG prospectus mandates. The ESG penetration appears relatively uniform across categories, with a mean share of 18.36% (ranging from 11.6% to 25.8%). This increases our confidence that our results are not limited to select types of funds but generalise across the larger universe of Lipper Objective Codes included in our sample. One limitation is that our results may not extend to highly specialised or niche strategies. In these categories, there are too few ESG-mandated funds to build a reliable peer benchmark. Therefore, our analysis focuses on broader market segments where the number of funds is high enough to identify clear patterns in portfolio construction.

Figure 1: Number of ESG Funds and TNA Over Time

Note: This figure displays the cumulative TNA invested in ESG funds between 2016 and 2024 as well as the number of funds mentioning ESG in their prospectus during the same time period.

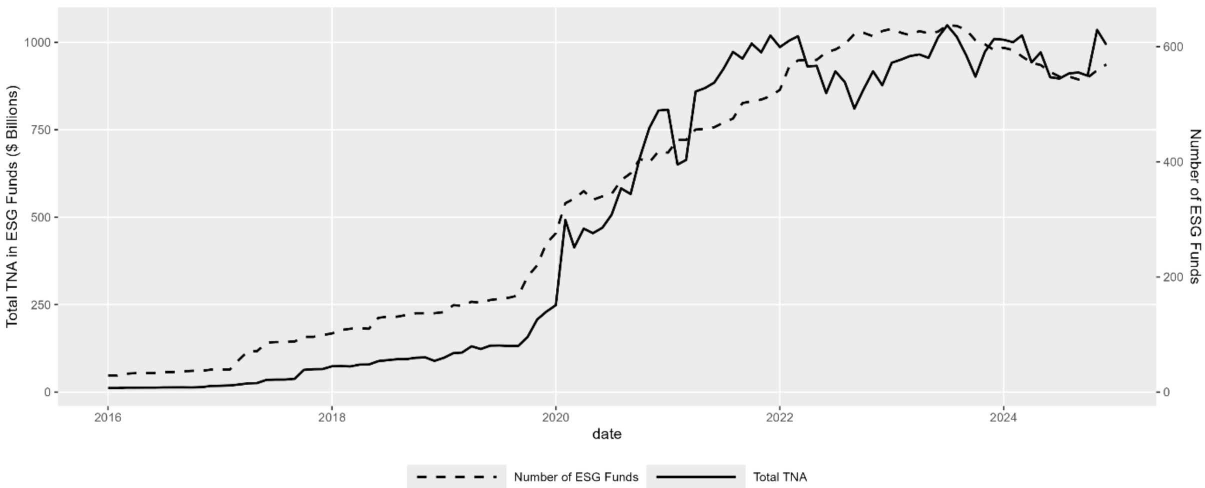
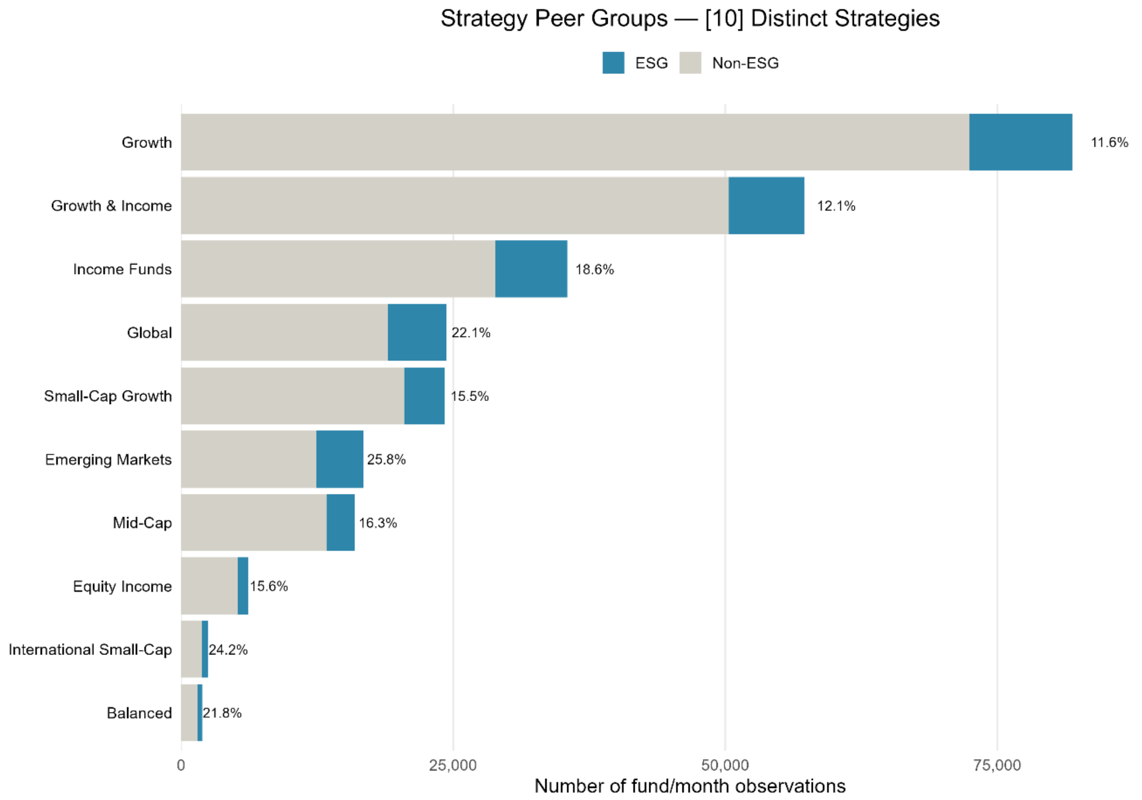


Figure 2: Observations per SPG Assignment

Note: This figure displays the number of fund/month observations assigned to each SPG, with observations simultaneously classified as ESG highlighted by a blue bar. The number to the right of each bar represents the percentage share of observations for each strategy where an ESG keyword is mentioned in the prospectus.



4. Empirical Analysis

4.1. Adherence to Promised Strategies

In this section, we create a measure of ESPG adherence that combines our ESG-classification methodology with portfolio holdings data to create a measure of ESG-Adherence that is aligned with our narrative-based approach and independent of investment research company ratings (e.g., Morningstar and MSCI) as well as quantitative sustainability metrics (e.g., total emissions and carbon intensity). The identifying assumption is that for any ESPG and any SPG there exists a core portfolio that accurately captures the stated mandate, and that any deviations within individual portfolios are purely idiosyncratic. The mean portfolio is then computed by expressing each portfolio's holdings as a vector, taking the union of all vectors within each ESPG and calculating the average holdings in each security. Adherence is subsequently measured as the cosine of the angle between the individual portfolio vector and the core portfolio vector, referred to as the cosine similarity. This adherence measure is selected in favor of Euclidean distance, despite its previously documented use within this particular context by Abis and Lines (2024). The decision to use cosine similarity in our primary specification is driven by two key considerations:

(i) Ease of interpretation: Cosine similarity is bounded between -1 and 1. This offers an intuitive scale, a cosine similarity equal to 1 means the portfolios are identical, a cosine similarity equal to 0 means they are orthogonal and a cosine similarity of -1 means they are taking opposite positions in the same assets.

(ii) Superior handling of high-dimensional and sparse data: This is critical given that our mean portfolio vectors include an average of 7,449 securities. In such high-dimensional spaces, distance metrics often suffer from the "curse of dimensionality," where distances converge and become less discriminatory. Furthermore, while the ESPG and SPG averages are dense, individual funds hold only 351 securities on average. In this sparse-to-dense comparison, Euclidean distance penalises concentration, as every security a fund does not hold adds to its distance from the average. Cosine similarity, by contrast, focuses on the angular overlap of the vectors, ensuring that a fund's decision to hold a concentrated subset of the market does not obscure its strategic alignment with the benchmark.

For our analysis, we are interested in four types of adherence.

(1) Within ESPG Adherence.

This measure captures the extent to which a fund adheres to the core portfolio allocation of funds within its own ESG-strategy group. It serves as our primary independent variable of interest, reflecting the degree to which a fund delivers on its stated sustainability mandate.

(2) Within SPG Adherence.

This measure captures the extent to which a fund's portfolio aligns with the core allocation of its full Lipper strategy peer group, which contains both ESG and non-ESG funds. It serves as a control variable, isolating adherence to sustainability objectives from general adherence to investment style.

(3) Within SPG Outside ESPG Adherence

This measure captures the extent to which a fund's portfolio aligns with the core allocation of only the non-ESG subset of its strategy peer group. Unlike (2), the benchmark here excludes ESG funds, providing a validation check for the economic relevance of the ESG classification: if ESG funds adhere more closely to their ESPG than to the non-ESG subset of their SPG, the ESG classification captures meaningful portfolio-level differences within strategies.

(4) Outside SPG Adherence.

This measure captures the extent to which a fund's portfolio aligns with the core allocations of other strategy groups. It provides a validation check for the economic relevance of the strategy classification, as low similarity would indicate meaningful differentiation across groups.

Formally:

$$\text{Adherence}_{j,t}^{K_{j,t}} = \cos(\theta)_{j,t}^{K_{j,t}} = \frac{\sum_{i=1}^{n} w_{i,t}^{j} \overline{w}_{i,t}^{K_{j,t}}}{\sqrt{\sum_{i=1}^{N_t^j} {w_{i,t}^{j}}^2} \sqrt{\sum_{i=1}^{N_t^j} {\overline{w}_{i,t}^{K_{j,t}}}^2}} \tag{2}$$

for K = [ESPG, SPG, $\widetilde{ESPG}$, $\widetilde{SPG}$] where $ESPG_{j,t}$ is the ESPG of fund j at time t, $SPG_{j,t}$ is the SPG, $\widetilde{ESPG}_{j,t}$ is the non-ESG group within $SPG_{j,t}$ and $\widetilde{SPG}_{j,t}$ denotes all SPGs other than $SPG_{j,t}$. $w_{i,t}^{j}$ is the weight in fund $j's$ portfolio in security i at time t and $\overline{w}_{i,t}^{K_{j,t}}$ is the average weight in security i at time t for all funds belonging to group $K_{j,t}$. N_t^j is equal to the length of the security vector for group $K_{j,t}$.

Next, we perform two sanity checks to confirm that our ESG and strategy classifiers extend to real differences in portfolio holdings, and that our adherence measure successfully captures these differences. If the classifications map into genuine portfolio-level differences, funds should adhere more closely to their own ESPG and SPG than to benchmarks outside these groups; a positive and significant α^{ESPG} and α^{SPG} in the regressions below would be consistent with this prediction. We run two regressions:

$$\text{Adherence}_{j,t}^{ESPG_{j,t}} - \text{Adherence}_{j,t}^{\widetilde{ESPG}_{j,t}} = \alpha^{ESPG} + \lambda^{ESPG'} X_{j,t} + \varepsilon_{j,t}^{ESPG} \tag{3}$$

and

$$\text{Adherence}_{j,t}^{SPG_{j,t}} - \text{Adherence}_{j,t}^{\widetilde{SPG}_{j,t}} = \alpha^{SPG} + \lambda^{SPG'} X_{j,t} + \varepsilon_{j,t}^{SPG} \tag{4}$$

We focus on the coefficients α^{ESPG} and α^{SPG} which represent the difference in adherence within ESPG and SPG, respectively, relative to outside these groups. $X_{j,t}$ denotes a vector of control variables, including the logarithm of total net assets (TNA), the logarithm of fund age, the expense ratio and the turnover ratio. All controls are demeaned, allowing us to interpret the intercept (α^{ESPG} and α^{SPG}) as the difference in adherence when all controls are equal to their mean. We expect the error terms to be correlated within funds over time and within months across funds, and therefore we use two-way clustered standard errors at the fund and month level.

Table 3 reports the results of the regressions. The first two columns report the average adherence within and outside the ESPG (SPG) in panel A (panel B) along with t-statistics. Column 3 reports the estimated coefficients α^{ESPG} and α^{SPG}. Taking the relative difference between column 1 and 2 for ease of interpretation, we see that for ESG funds adherence to the conventional group is 4% lower than to the ESPG and funds in general adhere 79% less to the average of other strategy groups than to their own, with both differences being statistically significant.

Note that these results are purely descriptive: while we document differences in portfolio composition across classifications, we cannot make any causal inference that these differences are driven by the classifications themselves.

Nevertheless, our results are consistent with the interpretation that the classifiers capture meaningful differences in portfolio composition and provide evidence that mutual fund prospectuses contain economically relevant information. Investment strategies and ESG commitments as described in the prospectus are associated with systematic differences in realised holdings, and our approach successfully captures this alignment between stated objectives and portfolio behaviour. While this relationship may partly reflect underlying fund characteristics that jointly shape both disclosures and holdings, the results nonetheless suggest that prospectus language is informative about cross-sectional differences in portfolios.

Table 3: ESG and Strategy Classifiers Map into Portfolio Differences

Note: Panel A reports estimates from Equation (3) using the ESPG classifier, while Panel B reports estimates from Equation (4) using the SPG classifier. Adherence is defined as the cosine similarity between a fund's portfolio weight vector and its peer-group centroid. Columns (1) and (2) report average within- and outside-group adherence, respectively; Column (3) reports the difference. The unit of observation is the fund-month.

Panel A. ESPG Classification

	Within ESPG	**Outside ESPG**	**Difference**
	(1)	**(2)**	**(3)**
Cosine similarity	0.371***	0.356***	0.015***
	(46.731)	(41.651)	(4.912)
Num. of Obs.	40290	40290	40290
Controls	Yes	Yes	Yes
SE Cluster	Fund + Month	Fund + Month	Fund + Month

t statistics in parenthesis; $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Controls include log fund age, log TNA, expense ratio and turnover ratio, all demeaned.

Panel B. SPG Classification

	Within SPG	**Outside SPG**	**Difference**
	(1)	**(2)**	**(3)**
Cosine similarity	0.320***	0.066***	0.252***
	(76.082)	(52.331)	(62.642)
Num. of Obs.	255123	240302	240302
Controls	Yes	Yes	Yes
SE Cluster	Fund + Month	Fund + Month	Fund + Month

t statistics in parenthesis; $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Controls include log fund age, log TNA, expense ratio and turnover ratio, all demeaned.

4.2. Fund flows and ESPG Adherence

Having established that our adherence measure captures economically meaningful differences in portfolio composition, we now examine whether investors respond to variation in ESPG adherence through their capital allocation decisions. We estimate the following regression at multiple cumulative flow horizons:

$$Flow_{j,t+1:t+k} = \alpha + \beta^{ESPG} Adherence_{j,t}^{ESPG} + \beta^{SPG} Adherence_{j,t}^{SPG} + \lambda' X_{j,t} + \eta_{s,t} + \varepsilon_{j,t}, \quad (5)$$

where $Flow_{j,t+1:t+k}$ denotes cumulative percentage fund flows from month $t+1$ to $t+k$ as defined in Equation 1, for $k \in \{1,3,6,12\}$. $Adherence_{j,t}^{ESPG}$ and $Adherence_{j,t}^{SPG}$ are within-ESPG and within-SPG cosine similarity as defined in Equation 2. $X_{j,t}$ denotes a vector of control variables, including fund characteristics in the form of the logarithm of total net assets (TNA), the logarithm of fund age, the expense ratio, and the turnover ratio.

To account for the possibility that skilled managers might seek opportunities that reduce adherence but enhance performance, we include two performance controls: risk-adjusted returns, specifically Fama-French 6 factor alphas (Fama and French, 2018), and performance relative to SPG peers. Abis and Lines (2024) emphasise the importance of the latter, and in Appendix C we show that its significance persists under our alternative strategy classification based on Lipper objective codes. $\eta_{s,t}$ denotes fund and month fixed effects. Standard errors are two-way clustered at the fund and month level. As a falsification test, we also estimate the specification using lagged cumulative flows as the dependent variable ($k \in \{-3,-1\}$). Adherence should not affect past flows, and significant results would indicate reverse causality or spurious correlation, undermining the causal interpretation of our primary specification's results.

Figure 3 reports the estimated coefficients $\hat{\beta}^{ESPG}$ at each horizon. The point estimates are economically small and statistically indistinguishable from zero throughout. ESG funds that hold portfolios closely aligned with their ESG peers attract no more capital than those that deviate.

Figure 4 reports the corresponding coefficients $\hat{\beta}^{SPG}$. The lagged estimates are insignificant as seen by their confidence intervals overlapping zero, satisfying the no-pre-trends condition. This result is necessary, but insufficient, for ruling out reverse causality. Beginning at the one-month horizon, however, the coefficient is positive and statistically significant. The effect accumulates monotonically over time, maintaining significance at the 5% level through the nine-month horizon (with an estimate of 0.021). The cumulative effect peaks at 0.040 at twelve months, albeit with a slight reduction in statistical precision as the significance level drops to the 10% level. A one standard deviation increase in adherence results in 4% additional inflows in the following year. This growing cumulative pattern closely mirrors the dynamics documented in Abis and Lines (2024), where the flow response to a strategy divergence shock builds over a twelve-month window as investors gradually observe and act on deviations from the promised strategy.

The two figures together reveal a clear asymmetry. Investors allocate capital toward funds whose portfolio remains close to the core holdings of their SPG, consistent with the disciplinary mechanism of Abis and Lines (2024). The same mechanism does not extend to the ESG dimension. Variation in how closely an ESG fund's portfolio resembles those of its ESG peers generates no detectable flow response at any horizon. The contrast with Andrikogiannopoulou et al. (2025) is apparent: they show that the presence of ESG language in a fund's prospectus doubles capital inflows relative to non-ESG funds. Together, these findings suggest that while the ESG label acts as a primary catalyst for entry into the ESG segment, investors do not differentiate among funds within this domain based on actual portfolio-level adherence.

The asymmetry between SPG and ESPG adherence suggests a fundamental disparity in investor discipline. While strategy drift is readily detectable and penalised, the null OLS results for ESG point toward higher observability costs and a potential lack of investor willingness to enforce strict portfolio-level adherence. This implies the ESG label may function as a static marketing signal rewarded at entry rather than a mandate that investors actively monitor. The factors enabling greenwashing at the market level thus appear to characterise the internal dynamics of the ESG segment, where either information barriers or investor passivity limit the degree of oversight.

Figure 3: The Flow Response to ESG-Strategy Peer Group Adherence

Note: Estimated coefficients $\hat{\beta}^{ESPG}$ from Equation (5), regressing cumulative fund flows from month *t+1* to *t+k* on within-ESPG adherence, for $k \in \{-3, -1, 0, 1, 3, 6, 12\}$ months. Negative horizons correspond to the falsification test using lagged cumulative flows as the dependent variable. Adherence is measured as the cosine similarity between each fund's portfolio and its ESPG centroid. Vertical bars denote 95% confidence intervals.

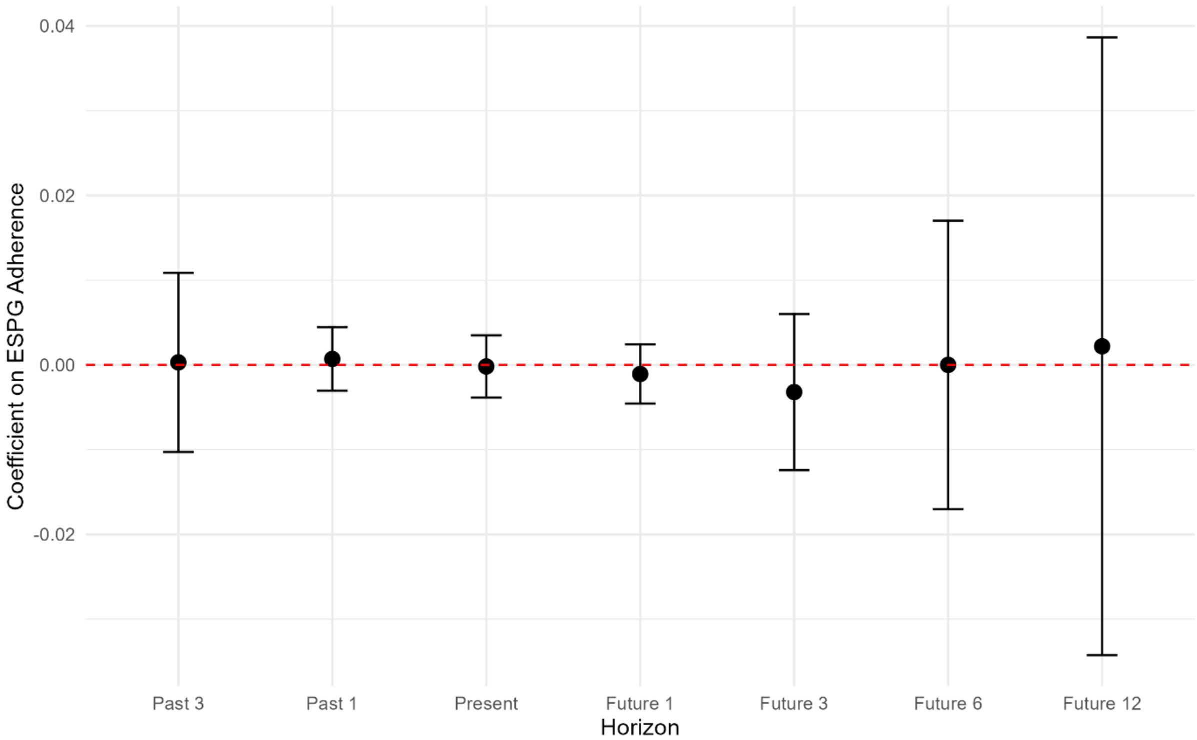
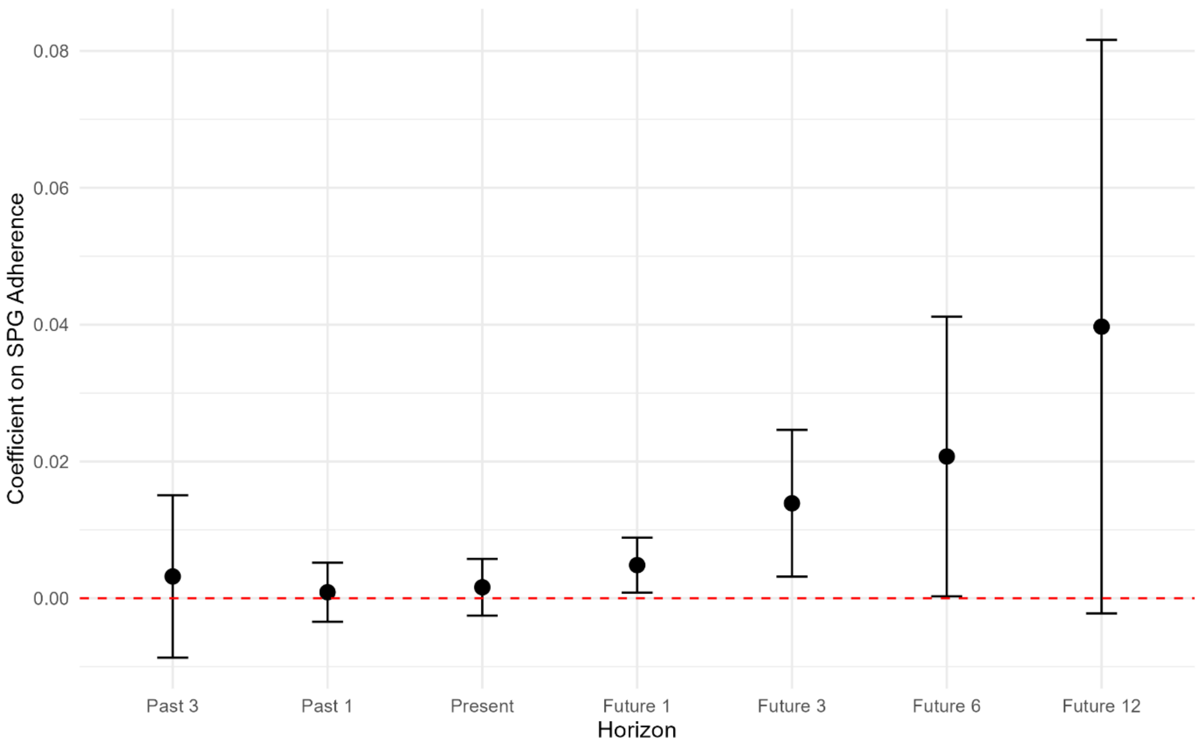


Figure 4: The Flow Response to Strategy Peer Group Adherence

Note: Estimated coefficients $\hat{\beta}^{SPG}$ from Equation (5), regressing cumulative fund flows from month $t+1$ to $t+k$ on within-SPG adherence, for $k \in \{-3, -1, 1, 3, 6, 12\}$ months. Negative horizons correspond to the falsification test using lagged cumulative flows as the dependent variable. Adherence is measured as the cosine similarity between each fund's portfolio and its SPG centroid. Vertical bars denote 95% confidence intervals.



4.3. Identification

4.3.1. Instrumental variable

To strengthen the causal interpretation of our results, we employ an instrumental variable approach previously proposed by Abis and Lines (2024). The main concern is that the relation between peer group adherence and fund flows may be endogenous due to omitted variables, reverse causality, or simultaneity bias arising from forward-looking behaviour by managers. In particular, unobserved manager characteristics, such as skill or reputation, may simultaneously influence adherence decisions and investor allocations, thereby inducing omitted variable bias.

Reverse causality may arise if managers adjust their trading behaviour in response to past flows. For instance, managers who experience significant inflows may view it as a feedback mechanism and subsequently take on more risky (less adherent) positions due to increased confidence. Finally, simultaneity may arise if managers adjust adherence in anticipation of future fund flows. A manager expecting large institutional backing in the near future may adopt a more adherent position due to the increased stakes.

The approach relies on separating variation in adherence due to managerial decisions, possibly endogenous, from variation in adherence due to abnormal returns in the underlying portfolio, plausibly exogenous. We use the same instrument as Abis and Lines (2024), absolute abnormal sector returns. The rationale is that since individual portfolios do not hold the exact same sector weights as the peer mean, sector returns and the subsequent re-weighting of portfolios should mechanically change adherence through its differential effect on individual funds and their peer mean.

To calculate abnormal sector returns, we regress the 10 industry portfolios, downloaded from Kenneth French's website, on the Fama French 6 factor model. We use a rolling window of 36 months to estimate each sector's betas and alpha, and our measure of sectoral shock is then the absolute value of the residual monthly return after subtracting compensation for exposure to the six risk factors and the sector's 36-month alpha.

The validity of our instrument hinges on two conditions:

(1) Relevance: Sectoral shocks as measured by absolute industry residual returns must significantly predict differences in SPG and ESPG adherence.

(2) Exogeneity: Absolute industry residual returns must not affect fund flows through any channel other than its effect on ESPG-Adherence and SPG-Adherence once we control for performance measures (Fama-French 6 alphas and SPG-adjusted returns) and fund characteristics (Total Net Assets, Age, Expense- and Turnover ratio).

While the exogeneity condition is inherently untestable, Abis and Lines (2024), who introduce the instrument, offer credible arguments for its plausibility. First, projecting returns onto a well-specified factor model purges them of systematic macroeconomic variation, rendering the residuals orthogonal to broad market movements. Second, conditional on the contemporaneous impact on overall returns, sectoral exposures in parts of the underlying portfolio should not be driving fund flows through any other channels.

We find their argument reasonable and agree that even if the exogeneity assumption is violated through some unobserved channel, the instrumental variable approach still contributes with a closer to randomly assigned measure of adherence and removes one channel of omitted variable bias associated with the OLS estimate, the fund manager.

We formulate a two-stage least squares (2SLS) model to estimate the effect of changes in ESPG-Adherence on fund flows. Our instrumental variable approach does not escape the potential confounding variable SPG-Adherence, as industry shocks will have a similar mechanical effect on this variable. Therefore, we use the 10 abnormal industry returns to instrument not only ESPG-Adherence, but also SPG-Adherence, and run the second stage on the predicted values from both regressions. As a result, we have two first stages, formally:

$$\Delta Adherence_{j,t}^{ESPG_{j,t}} = \sum_{n=1}^{10} \delta_n I_{n,t} + \gamma' X_{j,t} + \tau_j + \varepsilon_{j,t} \quad (6)$$

and

$$\Delta Adherence_{j,t}^{SPG_{j,t}} = \sum_{n=1}^{10} \delta_n I_{n,t} + \gamma' X_{j,t} + \tau_j + \varepsilon_{j,t} \quad (7)$$

where $\Delta Adherence_{j,t}$ is the first difference of adherence for fund j at time t, $I_{n,t}$ is the absolute residual return for industry n at time t, $X_{j,t}$ is a control vector including performance measures and fund characteristics and τ_j captures fund specific intercepts (fixed effects). Notably, since the instrument is the same for all funds, we have no cross-sectional variation within months, which means we cannot include month fixed effects. Another key distinction from the OLS specification is that we model the first difference of adherence rather than its level. This choice is economically intuitive, as idiosyncratic industry returns are expected to influence changes in adherence relative to the previous month, rather than deviations from a fixed baseline. In econometric terms, adherence is persistent and behaves approximately as a random walk, with each observation highly correlated with its previous value. Portfolio composition, and by extension adherence, is inherently path-dependent rather than independent and identically distributed (i.i.d), as it reflects the joint impact of discretionary managerial rebalancing and passive asset price drift on the portfolio in the previous period.

This stands in contrast to our earlier dependent variable, fund flows, which typically exhibits mean reversion.

For identification, both the order and rank conditions must be satisfied. The order condition holds since the number of instruments (10) is at least as large as the number of endogenous regressors (2). The rank condition requires that the instruments provide sufficiently independent variation to identify each endogenous variable separately, implying that the industry shocks must not affect ESPG-Adherence and SPG-Adherence in perfectly collinear ways.

The plausibility of this condition follows from heterogeneity in funds' initial sector exposures. Although the same industry shocks enter the instrument set for all funds, their impact on ESPG-Adherence and SPG-Adherence differs mechanically because these adherence measures are defined relative to different peer benchmarks. As funds start from different sector allocations, identical shocks reshape their distance to each benchmark in distinct ways, generating non-collinear variation in the two endogenous regressors.

We then estimate the second stage with the predicted values from the two first stages:

$$Flow_{j,t+1:t+k} = \beta^{ESPG} \Delta Adh\widehat{erence}_{j,t}^{ESPG} + \beta^{SPG} \Delta Adh\widehat{erence}_{j,t}^{SPG} + \zeta' X_{j,t} + \tau_j + \varepsilon_{j,t}. \quad (8)$$

where $Flow_{j,t+1:t+k}$ denotes cumulative percentage fund flows from month $t+1$ to $t+k$, for $k \in \{1,3,6,12\}$, $\Delta \widehat{Adherence}_{j,t}^{ESPG}$ and $\Delta \widehat{Adherence}_{j,t}^{SPG}$ denote the predicted values from the first stages, $X_{j,t}$ is the same control vector as in the first stages. We again include only fund fixed effects and cluster standard errors at the fund level.

4.3.2. Results

Table 4 presents the first stage results. As an initial check of the relevance condition, we report the F statistics for each first stage in isolation, yielding values of 10.69 and 12.40 in Columns 1 and 3. These values exceed the conventional single regressor threshold of 10 (Staiger and Stock, 1997). However, because our model identifies two endogenous variables simultaneously, these baseline statistics provide only an upper bound on instrument strength. We provide a more rigorous assessment of joint identification and potential weak instrument bias, including multivariate diagnostics, within our broader discussion in Section 5.

Notably, the sign of the industry-specific coefficients captures the degree of cross-sectional dispersion in fund positioning. A positive coefficient indicates relative homogeneity in sector weights, where increased magnitude of return shocks mechanically shifts individual portfolios and the benchmark in tandem, thereby increasing cosine similarity on average across funds. Conversely, a negative coefficient suggests significant heterogeneity in holdings, where absolute return shocks exacerbate existing weight disparities and reduce portfolio alignment. These results imply that while ESG funds maintain relatively uniform exposures in sectors such as High Tech, they exhibit marked divergence in their positioning within the Manufacturing and Healthcare sectors.

Importantly, we do not impose or test any prior on the sign of these coefficients. From an identification perspective, both increases and decreases in adherence induced by absolute sectoral shocks are informative, as long as investor responses are symmetric and the exogeneity assumption holds, as discussed in Section 4.3.1.

Finally, we observe considerable differences in both the significance levels and the sign of the estimated coefficients across the two first-stage regressions. This is a necessary (but insufficient) feature for identification, as it implies that the instruments generate distinct variation in ESPG-Adherence and SPG-Adherence, supporting the plausibility of the rank condition.

Table 5 reports the second stage estimates across varying horizons of cumulative fund flows. Variation in SPG and ESPG adherence, induced solely by idiosyncratic industry shocks, yields no significant contemporaneous effect on capital allocation in Column 1. Consistent with a delayed investor recognition of portfolio drift, the impact remains statistically insignificant in the first month immediately following the shock in Column 2. However, at the three-month horizon, the cumulative effect of SPG (ESPG) adherence becomes significant at the 10% (5%) level with a positive (negative) sign. For SPG, this significance increases to the 5% level and persists through nine months, while the ESPG pattern persists for twelve months as shown in Columns 3 through 6.

To assess economic magnitude, we scale the peak 9-month coefficients by the standard deviations of the fitted adherence measures, which are 0.004 for SPG and 0.008 for ESPG. A one standard deviation increase in SPG (ESPG) adherence corresponds to cumulative inflows of approximately 2.5% (-5.9%) in the following 9 months. These magnitudes are modest in absolute terms, reflecting the limited variation generated by the instrument. Because these shock-induced shifts are likely less salient to investors than deliberate, large-scale strategy changes, we interpret these results as conservative lower bound estimates of the true causal effect of portfolio adherence on capital flows.

The asymmetry in the signs of these coefficients suggests a fundamental decoupling in investor expectations: while investors demand style consistency and reward adherence to broad strategy benchmarks, they prize divergent positioning within the ESG space and penalise funds that mechanically drift toward the ESG-consensus centroid. We examine the potential underlying mechanisms behind these findings in greater detail in Section 5.

Table 4: 2SLS First-Stage Results

Note: This table reports coefficient estimates, t-statistics, and F-statistics from the first stage of the 2SLS model (Equations (6) and (7)), in which SPG and ESPG adherence are regressed on idiosyncratic industry returns. Columns (1) and (3) present baseline specifications used to assess the relevance condition and therefore exclude control variables. Columns (2) and (4) include controls for performance measures and fund characteristics and produce the fitted values which are used in the second stage. All four specifications include fund fixed effects; standard errors are clustered at the fund level.

	ESPG Adherence		**SPG Adherence**	
	(1)	**(2)**	**(3)**	**(4)**
Consumer Durables	0.007**	0.007**	0.012***	0.011***
	(2.214)	(2.111)	(4.568)	(3.825)
Consumer NonDur	0.014	0.017	-0.009	-0.005
	(1.332)	(1.448)	(-0.927)	(-0.514)
Manufacturing	-0.037**	-0.029*	-0.015	-0.010
	(-2.127)	(-1.662)	(-0.880)	(-0.654)
Energy	-0.006	-0.008**	-0.002	-0.004
	(-1.465)	(-2.012)	(-0.883)	(-1.493)
High Tech	0.173***	0.163***	0.141***	0.131***
	(8.758)	(8.158)	(8.415)	(7.929)
Telecoms	-0.002	-0.002	-0.009*	-0.013**
	(-0.265)	(-0.216)	(-1.704)	(-2.392)
Wholesale/Retail	-0.001	-0.006	-0.038***	-0.034***
	(-0.055)	(-0.557)	(-3.662)	(-3.252)
Healthcare	-0.033*	-0.029	-0.013	-0.012
	(-1.836)	(-1.515)	(-1.292)	(-1.211)
Utilities	0.005	0.007	0.005	0.009
	(0.481)	(0.706)	(0.820)	(1.322)
Other	-0.037	-0.045*	0.007	-0.002
	(-1.485)	(-1.712)	(0.431)	(-0.109)
Controls	No	Yes	No	Yes
Observations	35847	32188	35847	32188
F-Statistic	10.69	10.32	12.40	12.98

t statistics in parenthesis; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: 2SLS Second Stage Results

Note: This table reports the estimated coefficients and t statistics from regressing cumulative fund flows over multiple horizons on fitted values of ESPG and SPG instrumented with absolute idiosyncratic sector returns (Equation 8). All presented specifications include the same control variables including performance measures and fund characteristics, include fund fixed effects and cluster standard errors at the fund level. The only difference between columns is the cumulative horizon used as dependent variable. Coefficients are reported for one-unit changes in fitted adherence; for comparability with OLS estimates where adherence is standardised, rescale effects using the standard deviations of the fitted adherence measures reported in Section 4.3.2.

	Cumulative Fund Flow Horizon (Months)					
	t	**t+1**	**t+3**	**t+6**	**t+9**	**t+12**
ESPG Adherence	0.150	-0.249	-1.250**	-4.909***	-7.205***	-6.701**
	(0.548)	(-0.885)	(-1.968)	(-3.055)	(-3.236)	(-2.549)
SPG Adherence	-0.194	-0.044	1.193*	4.542***	5.744***	4.777*
	(-0.720)	(-0.158)	(1.794)	(2.854)	(2.589)	(1.682)
Log(TNA)	0.007***	-0.008***	-0.029***	-0.071***	-0.122***	-0.177***
	(3.724)	(-5.517)	(-7.182)	(-9.066)	(-10.302)	(-11.035)
Log(Fund Age)	-0.027***	-0.012***	-0.033***	-0.055***	-0.058**	-0.042
	(-8.356)	(-4.101)	(-3.810)	(-3.311)	(-2.340)	(-1.267)
Turn. Ratio	-0.006***	-0.006***	-0.014**	-0.022**	-0.032**	-0.038*
	(-3.068)	(-3.120)	(-2.413)	(-2.007)	(-2.001)	(-1.773)
Exp. Ratio	-0.350	-1.527***	-3.852**	-5.247	-6.193	-7.520
	(-0.539)	(-2.589)	(-2.236)	(-1.469)	(-1.130)	(-1.022)
FF6 Alpha	0.654***	0.803***	2.560***	4.966***	6.911***	8.623***
	(3.128)	(3.238)	(3.185)	(3.135)	(3.053)	(2.915)
Rel. SPG Perf.	0.039*	0.103***	0.249***	0.459***	0.692***	0.948***
	(1.785)	(5.017)	(5.654)	(5.904)	(6.440)	(7.003)
Fixed Effects	Fund	Fund	Fund	Fund	Fund	Fund
SE Cluster	Fund	Fund	Fund	Fund	Fund	Fund
Observations	32159	32037	31730	31079	30236	28387

t statistics in parenthesis; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5. Discussion

The result in Section 4.3.2 deserves further attention. The prior, as suggested by previous literature, implies that the estimated coefficient of ESPG adherence in a regression of fund flows on ESPG adherence should be positive, in line with the story that investors monitor and value upheld ESG promises (Abis and Lines, 2024), or insignificant, in line with the story that investors either fail to monitor or do not care about the portfolio-level implementation of ESG promises enough to alter their capital allocation behaviour (OECD, 2025; Krueger et al., 2024; Berg et al., 2022).

The negative estimated coefficient instead suggests that investors reward funds that, ceteris paribus, deviate from their ESPG when that deviation is driven by idiosyncratic sectoral shocks. Under our identifying assumption that the ESPG mean accurately captures the promised ESG mandate, this pattern can be interpreted as evidence that ESG funds are rewarded for departing from their stated commitments. As this finding runs counter to prior expectations, however, we evaluate the plausibility of two alternative explanations.

First, investors may view ESG as a dimension across which they desire differentiation rather than conformity. Under this interpretation, an ESG fund that closely mirrors its peers might be perceived not as adherent, but as failing to fulfil its specialised mandate, signalling managerial passivity or "closet-indexing." Our findings suggest that investors instead value deviation as a credible signal of rigorous, proprietary research.

Beyond the implicit signals of competence that differentiation conveys, investors may also value the tangible diversification benefits offered by less adherent funds. This is particularly salient in the ESG space, where reliance on ubiquitous third-party ratings often leads to high portfolio overlap and "crowded" holdings. For ESG-insistent investors facing a sparse opportunity set, a fund that identifies "hidden gems", securities with high ESG integrity but low consensus coverage, offers a scarce and valuable diversification tool.

This story becomes especially plausible if we assume investors struggle to evaluate ESG integrity directly. As documented by the OECD (2025) and Krueger et al. (2024), the information required to assess ESG performance is often fragmented and buried in qualitative disclosures rather than standardised metrics. Berg et al. (2022) further show that even when investors turn to third-party ratings, those ratings disagree substantially across major providers, exacerbating the difficulty of assessing ESG performance. Crucially, the holdings that contribute to lower peer adherence are frequently the ones least likely to be covered by major third-party ratings. If funds tend to chase ratings, the peer mean will be mechanically biased towards covered firms, leaving uncovered holdings systematically underrepresented and their ESG shortcomings harder for investors to detect.

This creates a sharp contrast in how investors monitor different fund mandates. In traditional strategy dimensions, such as market capitalisation or value/growth, a “broken promise” is easily audited using readily available market data, and style drift is typically punished. However, in the more opaque ESG domain, the difficulty of verifying a security's ESG credentials prevents investors from definitively identifying a mandate breach. In the absence of standardised proof to the contrary, investors may grant managers the benefit of the doubt, interpreting differentiation driven by sectoral shocks as a proxy for rigorous research rather than a deviation from the fund’s stated mandate. This disparity in observability may help explain the discrepancy in our findings: while strategy-level divergence is easily identified and penalised, ESG-level divergence is viewed through a more optimistic lens of managerial effort.

While this channel is a plausible driver of investor behaviour, it likely represents only one of several mechanisms linking portfolio adherence to fund flows. A primary competing channel, which we initially expected to predominate, posits that investors may interpret deviations from the ESPG mean as a breach of mandate. Under this view, whether driven by superior information or inherent scepticism, investors perceive decreases in adherence as failures in delivering on sustainability promises and penalise the fund through subsequent outflows.

We therefore assess the validity of the 2SLS specification with particular attention to the trade-off between consistency and finite-sample precision. Relative to prior implementations of similar instruments, identification is more demanding than in Abis and Lines (2024), who consider a single endogenous regressor, as our more granular fund sorting along two dimensions generates two endogenous variables requiring separate identification. Accordingly, credible estimation requires that the excluded instruments provide sufficiently strong and distinct variation for each endogenous variable. In the first stage, the instruments exhibit some differential predictive power across the endogenous regressors, as reflected in differences in coefficients and statistical significance across the two first-stage equations, providing support for their role in separately identifying the structural parameters. Whether this differentiation is sufficient for strong identification overall is less clear, however, as the first-stage F-statistics of 10.69 and 12.40 remain close to conventional weak-instrument thresholds in a single-regressor setting, and are therefore less informative in our two-endogenous-regressor framework. This suggests that the instruments, while not without separate predictive content, may fall short of providing the distinct and robust variation that credible identification in this setting demands.

Given this concern, we complement the baseline first-stage diagnostics with additional tests tailored to multivariate IV settings. In particular, the Sanderson-Windmeijer conditional first-stage F-statistics, which assess each endogenous regressor conditional on the other, yield values of 7.2139 and 10.3108 for ESPG Adherence and SPG Adherence respectively. This points to uneven conditional relevance across the two endogenous variables, with at least one first stage falling below conventional benchmarks for instrument strength. We further report the Cragg-Donald Wald F-statistic of 4.46907, which provides a joint measure of instrument strength under the assumption of i.i.d. errors and is typically evaluated against Stock-Yogo critical values to assess the extent of potential weak-instrument bias. Taken together, these diagnostics suggest that while the instruments are relevant and contain predictive content, the strength of identification is limited, and weak-instrument concerns cannot be ruled out. This warrants caution in finite-sample inference and with regard to causal interpretation.

Against this backdrop, the shift from an insignificant OLS estimate to a statistically significant negative 2SLS coefficient should be interpreted with caution. One interpretation is that OLS is upward biased due, for example, to omitted variables or attenuation bias, and that IV estimation uncovers an underlying negative causal relationship. However, given that our instruments exhibit only moderate strength, 2SLS estimates may be sensitive to small violations of the exclusion restriction and can display substantial finite-sample instability. We therefore view the IV results primarily as suggestive evidence of a negative relationship rather than as a precise estimate of economic magnitude, and we present them as a robustness check complementing our baseline OLS results.

6. Conclusion

In this paper, we investigate whether U.S. mutual fund investors discipline ESG-oriented funds for departing from the ESG strategies described in their prospectus documents. We use ESG-related language in the Principal Investment Strategy sections of fund prospectuses to classify funds as ESG-oriented and combine this information with portfolio holdings and monthly fund flows. We then construct ESG-Strategy Peer Groups (ESPGs) by separating ESG and non-ESG funds within conventional strategy categories and use these groups to infer average portfolios that correspond to different ESG-strategy promises.

To answer the question of whether investors respond to ESG-strategy adherence, we compute fund-level similarity to the core portfolio of each fund's conventional strategy peer group and ESG-strategy peer group. We first show that ESG funds hold portfolios that differ systematically from otherwise comparable non-ESG funds, suggesting that ESG language in prospectuses captures a distinct dimension of product differentiation. We then document a market discipline effect for conventional strategies. Controlling for performance and other fund characteristics, flows are higher for funds that remain closer to the core strategy of their conventional peer group. This evidence is consistent with investors rewarding funds that adhere to their stated investment mandates.

However, we find no corresponding disciplinary mechanism in the ESG dimension. In our baseline specification, flows do not respond positively when ESG funds remain closer to the core portfolio of their ESG-strategy peer group. In our instrumental variables specification, based on idiosyncratic industry return shocks to portfolio composition, the estimated response to ESG-strategy adherence is negative. Although this evidence should be interpreted cautiously given the identification challenges in our setting, it does not support the view that investors penalise ESG funds for deviating from the ESPG consensus.

Finally, we document a plausible economic interpretation for these results. Adherence to a stated mandate elicits two countervailing preferences: investors reward funds that deliver on their promises, but they also reward funds that differentiate themselves from peers, since distinctive holdings can signal active management and offer diversification in a crowded peer space. In the conventional strategy dimension, the promise-keeping channel dominates: conventional mandates promise specific factor exposures and an associated risk-adjusted return profile, and deviations from those mandates are both easily detected and directly costly in risk-adjusted return terms. Consequently, the balance tips toward similarity. In the ESG dimension, however, the link between the non-pecuniary mandate and expected risk-adjusted returns is less direct, and verification frictions further weaken the promise-keeping channel. While the OLS null suggests that these competing flow dynamics effectively neutralise one another, our suggestive IV evidence, interpreted with the caveats noted earlier, points to a possible inversion of the conventional result. This reveals a dual preference structure that is mandate-specific: while similarity sells in conventional strategy space, diversity may deliver more than conformity in the ESG dimension.

Appendix A: Sample Construction

Table 2: Sample Construction Table

Note: This table reports the sequential filters applied to construct our sample. The initial sample comprises all U.S. mutual funds available in the CRSP Mutual Fund Database between January 2016 and December 2024. Each row reports the filter applied at that step and the number of unique funds remaining. The final sample, after Step 8, comprises 3,690 funds, of which 881 mention ESG in their prospectus in at least one month.

Step	Filter Applied	Funds left
0	Initial Sample	17308
1	Remove Index Funds	13939
2	Remove Variable Annuity Funds	11878
3	Remove ETFs	9983
4	Remove non-equity funds	7725
5	Remove funds with average common stock holdings $< 50\%$	6616
6	Remove funds that found no match in the CRSP data	5468
7	Remove funds with less than 24 months between first and last observation	4745
8	Remove funds belonging to Lipper classes with less than 12 months with at least 15 ESG funds and monthly observations from funds in a Lipper class that had less than 15 ESG funds that month	3690
9	Remove funds that never mentioned ESG in their prospectus	881

Appendix B: AI Prompt and Disclosure

B.1 Prompt for ESG classification (Section 3.1.2)

The following prompt was provided to Claude Opus 4.7 to classify each fund's Principal Investment Strategy (PIS) section as ESG or non-ESG:

> "*A set of mutual fund prospectus excerpts will be provided to you. For each excerpt, assess the following: would a reasonable investor, as understood under SEC guidance, based solely on the investment process described in the text, infer that this fund incorporates environmental, social, and governance (ESG) factors as an element of its investment strategy? Your answer should be based solely on the provided text and should take the form of a binary output: 1 if yes and 0 if no.*"

B.2 AI tool disclosure

We have used Claude (Anthropic) to support the coding process underlying this paper.

Appendix C: OLS Results Table

	Cumulative Fund Flow Horizon (Months)						
	t-3	**t-1**	**t**	**t+1**	**t+3**	**t+6**	**t+12**
ESPG Adherence	0.000	0.001	-0.000	-0.001	-0.003	-0.000	0.002
	(0.054)	(0.362)	(-0.102)	(-0.603)	(-0.682)	(-0.000)	(0.117)
SPG Adherence	0.003	0.001	0.002	0.005**	0.014**	0.021**	0.040*
	(0.525)	(0.401)	(0.761)	(2.361)	(2.529)	(1.982)	(1.850)
Log(TNA)	0.027***	0.007***	0.005***	-0.011***	-0.037***	-0.085***	-0.204***
	(4.740)	(3.688)	(2.917)	(-6.723)	(-8.770)	(-10.306)	(-11.315)
Log(Fund Age)	-0.085***	-0.025***	-0.024***	-0.005	-0.006	0.007	0.080*
	(-5.904)	(-5.020)	(-4.737)	(-1.127)	(-0.495)	(0.297)	(1.754)
Turn. Ratio	-0.020***	-0.006***	-0.006***	-0.006***	-0.017***	-0.034***	-0.056***
	(-3.503)	(-2.924)	(-2.743)	(-2.745)	(-2.858)	(-2.992)	(-2.624)
Exp. Ratio	-4.398**	-1.357*	-0.292	-1.778***	-5.391***	-10.397***	-18.289**
	(-2.160)	(-1.886)	(-0.424)	(-2.630)	(-2.771)	(-2.790)	(-2.408)
FF6 Alpha	1.861**	0.668**	0.679***	0.763***	2.260***	4.294***	7.106**
	(2.484)	(2.470)	(2.690)	(2.744)	(2.690)	(2.674)	(2.434)
Rel. SPG Perf.	-0.049	-0.012	0.037	0.093***	0.266***	0.492***	0.825***
	(-0.891)	(-0.541)	(1.502)	(4.285)	(5.495)	(5.668)	(5.317)
Fund & Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SE Cluster	Fund	Fund	Fund	Fund	Fund	Fund	Fund
Observations	31523	31958	32159	32037	31730	31079	28387

t statistics in parenthesis; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

References

Abis, Simona, and Anton Lines. Broken promises, competition, and capital allocation in the mutual fund industry. *Journal of Financial Economics*, 162:103948, 2024. doi: 10.1016/j.jfineco.2024.103948.

Andrikogiannopoulou, Angie, Philipp Krueger, Shema Mitali, and Filippos Papakonstantinou. Empty ESG promises: A text analysis of mutual fund prospectuses. Swiss Finance Institute Research Paper No. 25-43, 2025. doi: 10.2139/ssrn.4082263.

Barber, Brad M., Terrance Odean, and Lu Zheng. Out of sight, out of mind: The effects of expenses on mutual fund flows. *Journal of Business*, 78(6):2095-2120, 2005. doi: 10.1086/497042.

Berg, Florian, Julian F. Kölbel, and Roberto Rigobon. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6):1315-1344, 2022. doi: 10.1093/rof/rfac033.

Chevalier, Judith, and Glenn Ellison. Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105(6):1167-1200, 1997. doi: 10.1086/516389.

Fama, Eugene F., and Kenneth R. French. Choosing factors. *Journal of Financial Economics*, 128(2):234-252, 2018. doi: 10.1016/j.jfineco.2018.02.012.

Girardi, Giulio, Kathleen W. Hanley, Stanislava Nikolova, Loriana Pelizzon, and Mila Getmansky Sherman. Portfolio similarity and asset liquidation in the insurance industry. *Journal of Financial Economics*, 142(1):69-96, 2021. doi: 10.1016/j.jfineco.2021.05.050.

Krueger, Philipp, Zacharias Sautner, Dragon Yongjun Tang, and Rui Zhong. The effects of mandatory ESG disclosure around the world. *Journal of Accounting Research*, 62(5):1795-1847, 2024. doi: 10.1111/1475-679X.12548.

Morningstar. ESG funds: 2025 closes with continued outflows amid persistent headwinds. Morningstar Sustainable Investing, 2026. URL https://global.morningstar.com/en-eu/sustainable-investing/esg-funds-2025-closes-with-continued-outflows-amid-persistent-headwinds.

OECD. *Behind ESG Ratings: Unpacking Sustainability Metrics*. OECD Publishing, Paris, 2025. URL https://www.oecd.org/en/publications/behind-esg-ratings_3f055f0c-en.html.

Sirri, Erik R., and Peter Tufano. Costly search and mutual fund flows. *Journal of Finance*, 53(5):1589-1622, 1998. doi: 10.1111/0022-1082.00066.

Staiger, Douglas, and James H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557-586, 1997. doi: 10.2307/2171753.