Inferring Mutual Fund Intra-Quarter Trading

An Application to ESG Window Dressing

Li An, Shiyang Huang, Dong Lou, Xudong Wen, and Mingxin Xu*

First version: July 2023

This version: January 2025

*An: Tsinghua PBC School of Finance. Email: anl@pbcsf.tsinghua.edu.cn. Huang: University of Hong Kong. Email: huangsy@hku.hk. Lou: Hong Kong University of Science and Technology, London School of Economics, and CEPR. Email: dlou@ust.hk. Wen: Hong Kong University of Science and Technology. Email: xwenam@connect.ust.hk. Xu: CITIC Securities. Email: xumingxin@citics.com. We thank Utpal Bhattacharya, Hank Bessembinder, Mehmet Canayaz, James Guo, Mirco Rubin, Clemens Sialm, Stephan Siegel, Kelly Shue, Yinan Su, Guofu Zhou and seminar and conference participants at Bayes Business School, Capital University of Economics and Business, Chinese Academy of Sciences, Hong Kong University of Science and Technology, Johns Hopkins University, Michigan State University, Penn State University, Shanghai Advanced Institute of Finance, Shanghai University of Finance and Economics, Shenzhen University, Universidad de los Andes, University of Washington, Utrecht University, 2023 Guanghua International Symposium on Finance, 2023 Five Star Workshop, 2024 Greater China Area Finance Conference, 2024 Renmin University SFAF, 2024 Western Finance Association Annual Meeting, 2024 European Finance Association Annual Meeting, 2024 CFRN Young Finance Scholars Forum, 2024 Chinese Finance Annual Meeting, Jupiter Asset Management, Korean Securities Association for helpful comments. Bolin Xu and Jianding Zhang provided excellent research assistance. All remaining errors are our own. This paper was previously circulated under the title "ESG Window Dressing."

Inferring Mutual Fund Intra-Quarter Trading

An Application to ESG Window Dressing

Abstract

We develop a novel method to infer intra-quarter trading of individual mutual funds. After a mutual fund executes a trade, its reported portfolio return deviates incrementally from its quarter-end-holdings-based return, which enables us to infer the transaction date and amount using publicly available data. We apply our method to analyze mutual funds' strategic trading of ESG stocks. In the post-2015 period, mutual funds consistently buy (sell) high- (low-) ESG stocks before quarter ends and reverse their trades shortly after. This trading pattern is particularly pronounced among mutual funds near the cutoffs of extreme ESG rating categories. These trades also affect the returns of ESG stocks around quarter ends.

JEL classification: G02, G12, G23, N22

1 Introduction

Investors exhibit a wide range of preferences when it comes to balancing social considerations with financial goals. Some prioritize advancing their social and ethical objectives, even at the expense of financial returns. Others focus solely on maximizing their financial well-being, believing that social issues are best addressed through public policy rather than private investment. Most investors, however, lie somewhere between these extremes, seeking to balance social value with private financial returns. It would be an "easy" trade-off if social value and financial returns were perfectly aligned; that is, by pursuing social value, such as avoiding firms with high carbon emissions, investors also maximize investment performance. However, standard economic theory suggests otherwise (Pastor, Stambaugh, and Taylor, 2021, 2022). Maximizing a combination of financial and ESG (environmental, social, governance) performance is mathematically equivalent to solving a constrained performance optimization problem. As a result, ESG preferences should (weakly) reduce the financial performance of the optimal portfolio (measured by, for example, the Sharpe ratio).[1]

This presents delegated asset managers with a thorny dilemma. On one hand, asset managers have an incentive to cater to their clients' growing environmental and social awareness.[2] On the other hand, these managers have a fiduciary duty to maximize clients' financial returns. Nearly all mutual funds state in their prospectus that their primary goal is to maximize risk-adjusted returns, with only a few openly expressing any willingness to sacrifice financial returns for social impact. Put simply, the asset management industry has made an explicit promise to their clients to maximize financial performance, while also making an implicit promise to advance social goals.

Monitoring asset managers' financial performance is relatively straightforward: mutual funds, for instance, are legally required to disclose their net asset value (NAV) and portfolio returns at the end of each trading day, and these figures are audited by independent third parties. In contrast, measuring asset managers' ESG performance is far more challenging. A common approach – adopted by most investors and regulators – is to rely on the portfolio-weighted average ESG ratings of each manager's publicly disclosed holdings. However, these disclosures are infrequent. Even for mutual funds, which are among the most tightly regulated, detailed portfolio information is only reported on a quarterly basis. It therefore seems plausible, even natural, that asset managers, who aim to optimize daily-tracked financial

[1] Moreover, recent research (e.g., Bolton and Kacperczyk (2021)) has found that firms with negative externalities have, in recent years, delivered higher average returns compared to firms with positive externalities. Consequently, investing in high-ESG stocks may doubly impact financial performance—yielding lower average returns while also limiting portfolio diversification.

[2] Indeed, Bloomberg Intelligence estimates that as of 2022, asset managers overseeing more than $40 trillion have committed to global sustainable investment initiatives.

returns as well as quarterly-measured ESG performance, engage in "ESG window dressing." That is, they increase their portfolio ESG ratings shortly before the disclosure date, while maximizing financial returns during the rest of the quarter.

In this paper, we take the ESG-window-dressing hypothesis to the data. To this end, we develop a novel method to infer the timing and amount of intra-quarter trading of each mutual fund. Although mutual funds only disclose their portfolio holdings at quarter-ends, they must report their portfolio returns everyday within the quarter. After a mutual fund executes a trade, its reported fund return deviates incrementally from the buy-and-hold return of the quarter-beginning portfolio. Based on this intuition, we derive an identity that relates a fund's daily return gap (i.e., the difference between the reported fund return and the buy-and-hold return of the quarter-beginning portfolio) to individual stock returns. We then reverse-engineer unobservable intra-quarter trading by leveraging observable daily returns of funds and individual stocks.

To illustrate, imagine a mutual fund that invests 100% in stock A on March 31st. Further imagine that on May 10th, the fund sells half of A and buys the same amount of stock B, and holds the resulting portfolio till June 30th. From April 1st to May 9th, the fund's reported return should match its buy-and-hold portfolio return. After May 10th, there occurs a significant jump in the daily return gap that is determined by the return differential of stocks A and B. This shift in the daily return gap reveals both the timing and magnitude of the transaction.

The challenge to our approach is a classic overfitting problem: we have far more parameters to estimate than data points available. We mitigate this problem in two ways: a) by selecting an appropriate stock universe for each fund-quarter and b) by imposing estimation penalties (so to achieve parameter smoothing). First, to select the stock universe, we categorize mutual funds' intra-quarter trading into two types: (1) directional trades, which involve either an increase or decrease in holdings for which the changes can be inferred from consecutive quarter-end reports, and (2) round-trip trades, involving buying and selling in the same quarter that offset each other. To enhance estimation accuracy, we perform a two-step procedure, where we start by focusing solely on directional trades via reported holdings. This is because estimating round-trip transactions is much more difficult, yet they constitute only a small fraction (around 20%) of total fund trading (Puckett and Yan (2011)). Next, we estimate round-trip trades by analyzing the residual fund returns. Here we restrict the universe to the set of stocks held by the fund either at the beginning or end of the quarter.[3]

Second, rather than imposing penalties on daily variations in trading, we adopt a more restrictive approach that significantly reduces the number of parameters. Specifically, we

[3]The results remain similar if we include all stocks held by the fund over the past three to five years.

divide each quarter into W windows and assume that the fund trades a given stock with a constant speed within each window (e.g., buying one million shares per day for five days). This method effectively imposes an infinitely large penalty on variation in trading within a window, while allowing flexibility across windows. We set the window length to one week, motivated by prior evidence that mutual funds typically spread out their trades over several days (Han, Kim, and Nanda (2019)).

We evaluate the accuracy of our proposed method using three datasets: (1) the TFN/CRSP sample with simulated intra-quarter trades, (2) the Ancerno sample with actual trades and trade-implied fund returns, and (3) the merged Ancerno-TFN/CRSP sample with actual trades and actual fund returns. With the first two evaluation samples, our method successfully matches over 50% of the weekly trades during the first and last weeks of a quarter and approximately 40% in the remaining weeks. By comparison, random guessing would yield a matching rate of only 8.3% = 1/12. For funds trading fewer than 50 stocks per quarter – representing roughly half of the sample – the matching rate increases to over 60%. It is worth noting that our method performs particularly well for the first and last weeks of the quarter (the focus of our ESG window-dressing analysis), as these periods are tightly disciplined by the boundary conditions set by quarter-end holdings.

Additionally, by comparing the results from the second and third evaluation samples (i.e., using trade-implied versus actual fund returns), we can assess the impact of unobserved factors in fund returns – e.g., IPO allocations, securities lending, derivatives holdings, and trading costs – on the performance of our proposed method. Our findings indicate that unobserved factors in fund returns have a relatively small effect, reducing the accuracy of our method by less than 5 percentage points in the first and last weeks and around 8 percentage points in the remaining weeks.

As further validations, we apply our trade-inference method to replicate previously-known mutual funds' trading patterns, such as performance window-dressing (e.g., Lakonishok, Shleifer, Thaler, and Vishny, 1991), portfolio pumping (e.g., Hu, McLean, Pontiff, and Wang, 2013), and trading around M&A announcements (e.g., Fich, Lantushenko, and Sialm, 2024). Reproducing these stylized facts – previously identified using the Ancerno or lower-frequency data – provides external validation for our method. Importantly, these replication exercises do not rely on any assumptions or choices regarding the evaluation metrics. These tests also highlight our method's potential for broader applications to the study of mutual funds' trading behavior and economic incentives.

To test the ESG-window-dressing hypothesis, we examine how mutual funds trade stocks with high vs. low ESG scores around the turn of each quarter. Given the low correlations – less than 0.5 for most pairs – across different ESG ratings (Berg, Kölbel, and Rigobon, 2022), we take the average of three major ESG ratings – Morningstar Sustainalytics, MSCI, and Refinitiv – in our empirical analyses. Our results are qualitatively similar if we focus instead on any one or two of these ratings. Since passive mutual funds do not engage in frequent trading, we focus on actively managed mutual funds. Our sample starts in 2015, the year in which over 190 countries joined the Paris Climate Accord, and in which Morningstar started publishing ESG ratings for mutual funds based on quarter-end holdings.

Our baseline finding is that US actively managed mutual funds significantly increase (reduce) investment in high- (low-)ESG stocks in a short window (e.g., a week) before each quarter end. They then quickly reverse these trades at the beginning of the following quarter. In terms of the economic magnitude, the difference in net trading between high-ESG and low-ESG stocks in the one week surrounding each quarter end accounts for 1.2% of mutual funds' total trading volume in the same window. For reference, this effect accounts for roughly 14% of average weekly trading of mutual funds in high- and low-ESG stocks.

Since mutual funds are legally required to report their holdings only at quarter ends, we conduct a placebo test surrounding month-ends other than quarter-ends (e.g., end of January, February), when mutual funds are less strictly monitored for ESG performance.[4] Consistent with our ESG-window-dressing hypothesis, we see no significant change in mutual funds' high- vs. low-ESG stock holdings around non-quarter-end month-ends.

To further validate our ESG-window-dressing hypothesis, we exploit discontinuities in the payoffs to ESG-rating manipulation, employing an empirical design inspired by Chevalier and Ellison (1997). In particular, Hartzmark and Sussman (2019) report discontinuities in capital flows around extreme ESG-globe cutoffs: funds in the five-globe category experience a marked increase in capital flows compared to those in the four-globe category, while funds in the one-globe category see a notable decrease relative to those in the two-globe category. Building on these findings, we examine trading of high- vs. low-ESG stocks around quarter-ends as a function of the fund's ESG percentile ranking (reported by Morningstar). Our results reveal an M-shaped pattern in the magnitude of ESG window dressing, with more pronounced manipulation observed for funds near the one-versus-two-globe and four-versus-five-globe thresholds relative to mutual funds in other parts of the ESG rating distribution.

Consistent with prior results on inelastic demand (e.g., Lou, 2012), we also observe a strong return effect associated with these window-dressing trades. High-ESG stocks con-

[4] Around 20% of our sample funds regularly report monthly holdings in Morningstar. This monthly reporting practice, however, does not invalidate the placebo test around month ends. First, while Morningstar may use monthly as well as quarterly holdings to evaluate fund ESG performance, MSCI and Refinitiv do not have legal access to Morningstar monthly data so cannot evaluate and monitor funds' ESG performance at a monthly frequency. Second and more importantly, since monthly reporting is not mandatory, mutual funds may strategically file their month-end reports: for example, only mutual funds that do not engage in month-end holding manipulation choose to report their monthly holdings to Morningstar.

sistently outperform low-ESG stocks just before quarter-ends, only to underperform at the beginning of the following quarter. Specifically, the return difference between high-ESG and low-ESG stocks exceeds 1% in the week leading up to quarter-ends compared to the week immediately after. In a placebo test, consistent with our earlier findings, we observe no significant differences in returns between high- and low-ESG stocks around month-ends that do not coincide with quarter-ends (e.g., the end of January or February).

We also observe significant variation in mutual funds' ESG window-dressing behavior, both over time and across funds/stocks. For instance, ESG window-dressing tendencies are notably stronger among self-declared ESG funds, funds with better past performance, and funds headquartered in Democratic-leaning states. Additionally, mutual funds are more likely to engage in ESG window-dressing during periods of heightened capital flows to ESG funds and when investor attention to environmental and social issues increases, as proxied by Google search volume. Moreover, these window-dressing trades are concentrated in stocks with higher liquidity and lower idiosyncratic volatility, so less costly to trade.

Finally, we show that ESG window dressing brings significant benefits to mutual funds. Specifically, there is a strong positive correlation between ESG window-dressing activity and subsequent fund inflows. This suggests that mutual fund investors are either unable (or simply choose not) to distinguish between genuine and manipulated ESG performance. However, ESG window dressing is not without its costs. Beyond the direct trading costs incurred, mutual funds also face the cost of deviating from their unconstrained optimal portfolios. Using a simple back-of-the-envelope calculation, we estimate that the net benefit (in terms of attracting capital flows) of a one-standard-deviation increase in ESG window dressing is roughly equivalent to the effect of a 1% increase in fund annual alpha.

In the context of the ESG-window-dressing hypothesis, an alternative approach to our method is to estimate changes in mutual funds' ESG betas around quarter-ends. This requires a) selecting an arbitrary ESG index and b) computing portfolio ESG betas over short time periods. This approach has two significant limitations. First, any long-only ESG index is highly correlated with the market, making it nearly impossible to separate the market beta from ESG beta. Second, estimating portfolio betas over a few days produces very noisy estimates. Our novel method addresses these issues. First, by focusing on the difference between reported fund returns and hypothetical buy-and-hold returns, our approach effectively removes the influence of common risk factors, including the market factor and other systematic exposures. Second, unlike the alternative method which treats ESG betas as free parameters, our estimates strictly adheres to the constraints imposed by the fund's reported holdings, ensuring that the estimated exposures align with the actual portfolio composition at the beginning and end of each quarter. Third, our method is more flexible, enabling detailed heterogeneity analyses across stocks and styles.

Related Literature Our paper contributes to the extensive literature on mutual funds and the asset management industry. A substantial body of empirical research has examined mutual funds' trading behavior and its influence on asset prices using quarterly reported holdings. Another line of work has analyzed mutual funds' daily trading patterns using the Ancerno data, which cover only a small subset of mutual funds over a relatively short time period (e.g., Goldstein, Irvine, Kandel, and Wiener, 2009; Hu, McLean, Pontiff, and Wang, 2013; Puckett and Yan, 2011). In contrast, our novel method infers intra-quarter trading using only publicly available data, enabling researchers to analyze previously unobservable trading behaviors for the entire mutual fund universe over an extended sample period. While this paper focuses on ESG window dressing, our method is highly adaptable and can be applied to a wide range of settings and research questions.

Our study also adds to the growing literature on ESG investment, which integrates environmental, social, and governance factors into portfolio choice.[5] Our findings shed new light on the complexities of ESG investment, particularly the incentives asset managers face in balancing the dual objectives of maximizing financial and ESG performance. Related to our study, Chen and Dai (2023) show that mutual funds allocate less to ESG stocks when facing higher flow-performance sensitivities or when managers' compensation is tied more closely to financial performance. Gantchev, Giannetti, and Li (2024), using quarterly holdings data, examine the behavior of mutual fund managers and investors when the trade-off between financial and sustainability performance becomes salient. Two contemporaneous studies, Parise and Rubin (2024) and Chen, Chen, and You (2024), analyze intra-quarter variation in mutual funds' ESG beta, and report similar findings: mutual funds tend to increase their exposures to high-ESG stocks right before quarter ends.

Finally, our study contributes to the literature on portfolio manipulation (e.g. Bollen and Pool, 2009; Harris, Hartzmark, and Solomon, 2015). Chevalier and Ellison (1997) show that fund managers may act strategically to attract capital inflows (e.g., by increasing portfolio volatility near year-ends), which may come at the cost of the portfolio's risk-adjusted returns. Similarly, our evidence of ESG window-dressing suggests that fund managers cater to

[5]Despite the growing interest, there are many lingering problems with the current framework of ESG investment. For example, a common approach to ESG investment is to exclude firms with low ESG ratings from the investable universe. Cohen, Gurun, and Nguyen (2022) argue that such negative screening may not be optimal because these firms are key innovators in the green patent landscape. Relatedly, Hartzmark and Shue (2023) show that negative screening may even be counter-productive as it deprives brown firms – which have more scope for improvement on carbon emissions – from much-needed capital to carry out these improvements. Furthermore, Amel-Zadeh and Serafeim (2018) based on survey data show that relevance to investment performance is the most frequently-mentioned motivation for the use of ESG ratings in portfolio construction.

investors who value sustainability as a means of increasing assets under management. Our findings also underscore the importance of greater transparency and accountability in ESG investment. By revealing how fund managers' incentives can influence trading behavior, our study highlights challenges in aligning ESG performance metrics with genuine sustainable investments.

2 Inference Method and Evaluation

2.1 Inference Method

In this subsection, we describe in detail our novel approach to infer mutual funds' intra-quarter trading.

2.1.1 Fund return identity

We infer intra-quarter trading for each fund during each quarter. Trading days within a quarter are labeled from 0 (representing the beginning of the quarter or the end of the previous quarter) to T (representing the end of the quarter). At each end of the day t, let $S_{k,t}$ denote the number of shares of a stock $k \in \{1, 2, \ldots, K\}$ held by an equity fund. We further let $R_{k,t}$, $P_{k,t}$, and $V_{k,t} = S_{k,t} \times P_{k,t}$ denote the stock return, stock price, and holding dollar value, respectively. We start with a simple accounting identity: the total capital gain of a fund's equity portfolio is equal to the sum of the capital gains from each individual stock within the portfolio. This can be expressed as follows:

$$\left(\sum_{k=1}^{K} V_{k,t-1} \right) R_t^{equ} \quad = \quad \sum_{k=1}^{K} V_{k,t-1} R_{k,t}, \tag{1}$$

where R_t^{equ} represents the fund's daily return based on its equity holdings. Our method can accommodate cases in which funds hold cash or borrow cash to use leverage, as we will elaborate in Section 2.1.3.

According to disclosure requirements, funds report their holdings every quarter. Let S_k^B denote the reported holding shares of stock k at the beginning of the quarter, and let S_k^E denote the shares at the end of the quarter. Thus, $V_{k,t}^B = S_k^B P_{k,t}$ represents the holding value based on shares at the beginning of the quarter. It is important to note that $V_{k,t}^B$ is time-varying due to fluctuations in $P_{k,t}$. We define the change in shares relative to the beginning of the quarter as $\Delta S_{k,t} = S_{k,t} - S_k^B$, and the corresponding change in dollar value as $\Delta V_{k,t} = V_{k,t} - V_{k,t}^B = (S_{k,t} - S_k^B) P_{k,t} = \Delta S_{k,t} P_{k,t}$. By substituting the relation

$V_{k,t} = V_{k,t}^B + \Delta V_{k,t}$ into Eq.(1), we obtain:

$$\begin{aligned}\left(\sum_{k=1}^{K}[V_{k,t-1}^B + \Delta V_{k,t-1}]\right) R_t^{equ} &= \sum_{k=1}^{K}\left(V_{k,t-1}^B + \Delta V_{k,t-1}\right) R_{k,t} \\ \left(\sum_{k=1}^{K} V_{k,t-1}^B\right) R_t^{equ} - \sum_{k=1}^{K} V_{k,t-1}^B R_{k,t} &= \sum_{k=1}^{K} \Delta V_{k,t-1}\left(R_{k,t} - R_t^{equ}\right). \end{aligned} \tag{2}$$

Next, by dividing both sides of Eq.(2) by $\left(\sum_{k=1}^{K} V_{k,t-1}^B\right)$, we derive the following identity:

$$R_t^{equ} - R_t^B = \sum_{k=1}^{K} \delta_{k,t-1}\left(R_{k,t} - R_t^{equ}\right), \tag{3}$$

where R_t^B denotes the return on a hypothetical portfolio that consists of the fund's quarter-beginning holdings:

$$R_t^B = \frac{\sum_{k=1}^{K} V_{k,t-1}^B R_{k,t}}{\sum_{k=1}^{K} V_{k,t-1}^B} = \frac{\sum_{k=1}^{K}(S_k^B P_{k,t-1}) R_{k,t}}{\sum_{k=1}^{K}(S_k^B P_{k,t-1})}, \tag{4}$$

and $\delta_{k,t-1}$ is the variable associated with intro-quarter trading:

$$\delta_{k,t-1} = \frac{\Delta V_{k,t-1}}{\sum_{k=1}^{K} V_{k,t-1}^B} = \frac{(S_{k,t-1} - S_k^B) P_{k,t-1}}{\sum_{k=1}^{K}(S_k^B P_{k,t-1})}. \tag{5}$$

The derived fund return identity, Eq.(3), has a clear economic interpretation. The left-hand side represents the daily return gap, as defined in Kacperczyk, Sialm, and Zheng (2008). The right-hand side represents the trading-implied return deviation. Intuitively, in the absence of trading during the quarter, $\delta_{k,t-1}$ will be zero for all stocks on all days. As a result, the fund's equity portfolio return, R_t^{equ}, will be identical to the portfolio return based on the fund's quarter-beginning holdings, R_t^B. However, when trading occurs, such as purchasing stock k on day $t-1$, $\delta_{k,t-1}$ will increase as $S_{k,t-1}$ rises relative to S_k^B. Consequently, the return gap will consist of a larger component attributed to the difference between the stock return R_t^k and the portfolio return R_t^{equ}. Therefore, the trading process, as captured by $\delta_{k,t-1}$, influences the relative weights of each stock in contributing to the daily return gap.

2.1.2 Empirical choices to address the overfitting challenge

Our inference approach is grounded in the accounting identity presented in Section 2.1.1 and can be characterized as a reverse-engineering exercise. It utilizes observable data – including fund daily returns, fund quarterly holdings, and stock daily prices and returns – to estimate the unobservable daily holdings of the fund in each stock (i.e., $S_{k,t}$). A big challenge in the problem is the high dimensionality: with T observations of fund daily returns, we need to estimate $K \times T$ parameters. The high dimensionality often leads to overfitting issues and makes estimation unreliable. To tackle this challenge, we take two empirical steps: a) selecting an appropriate stock universe for each fund quarter and b) imposing penalties (i.e., parameter smoothing) in the estimation procedure.

The stock universe We begin by classifying mutual funds' intra-quarter trading into two categories: (1) directional trades, which involve either increasing or decreasing holdings during a quarter and whose quantities can be directly inferred from quarterly changes in reported holdings, and (2) round-trip trades, which involve both buying and selling within a quarter, resulting in zero net changes between two consecutive quarter-ends. Since round-trip trades are not constrained by reported holdings, we must consider a broader set of stocks when inferring these trades. In this regard, the presence of round-trip trades exacerbates the high-dimensionality issues and complicates our estimation.

To address the challenges posed by round-trip trades, we implement a two-step estimation procedure.[6] In a nutshell, we begin by estimating directional trades, focusing exclusively on the set of stocks with non-zero changes in reported holdings between two consecutive quarters. This empirical choice is intentional for two reasons: (a) it is much easier to estimate directional trades than round-trip trades, as the number of stocks involved in directional trading is smaller,[7] and (b) round-trip trades constitute only a small portion of the total trading volume.[8]

In the second step, we estimate round-trip trades using residual fund returns after purging out directional trades. Here, we restrict our universe to the set of stocks that are held by the fund either at the beginning or the end of the quarter.[9] We further impose the constraint that the total amount of round-trip trading must be smaller than the fund's quarterly trading volume minus directional trading volume. In addition, for each fund-quarter, we cap the number of stocks in our universe at 500. In rare cases where the number exceeds 500, we select the top 500 stocks based on holding dollar value.

[6] We discuss the robustness and convergence properties of our estimation approach in Section 2.1.3.

[7] In our sample, the median number of stocks involved in directional trading is 67.

[8] Prior studies utilizing the Ancerno data find that round-trip trades account for less than 20% of a typical mutual fund's intra-quarter trading (Puckett and Yan, 2011).

[9] Results are similar if we include all stocks ever held by the fund over the past three years.

Parameter smoothing As discussed previously, we have more parameters to estimate than daily returns and thus face overfitting problems that can lead to volatile and unreliable estimations. To address the overfitting issues, one simple approach is to impose $L1$ or $L2$ penalties on daily trading, i.e., $S_{k,t} - S_{k,t-1}$. In the extreme case of imposing an infinitely large penalty on trading, it will result in linear changes in $S_{k,t}$, which implies a constant trading speed throughout the quarter.

We adopt a slightly more restrictive approach with significantly fewer parameters by assuming that $S_{k,t}$ follows a piecewise linear function over the days within the quarter. Specifically, we split a quarter into W windows and assume that $S_{k,t}$ is linear in days within each window. Figure 1 illustrates the construction of a piecewise linear function using a combination of basis functions. This piecewise linear parameterization can be interpreted as imposing an infinitely large penalty within each window while allowing flexibility across windows. To balance the tradeoff between overfitting issues and estimation frequency, we set the window length to one week (i.e., five trading days). This choice is motivated by earlier findings using the Ancerno data, which suggest that mutual funds often spread their trades over several days (Han, Kim, and Nanda, 2019).

2.1.3 The full inference problem

We now present the full inference problem and provide a detailed description of the estimation method. The objective of our estimation is to minimize the mean squared error between the observed fund daily return gap and the model-fitted return gap:

$$\min \sum_{t=1}^{T} \left[R_t^{equ} - R_t^B - \sum_{k=1}^{K} \delta_{k,t-1} \left(R_{k,t} - R_t^{equ} \right) \right]^2, \tag{6}$$

where $\delta_{k,t-1}$ is defined as $\frac{(S_{k,t-1} - S_k^B) P_{k,t-1}}{\sum_{k=1}^{K} (S_k^B P_{k,t-1})}$ (see Eq.(5)), and it contains parameters that need to be estimated. It is important to note that S_k^B and $P_{k,t}$ are observable, suggesting that $\delta_{k,t}$ is determined by daily shares $S_{k,t}$. Given that our parameter smoothing approach assumes $S_{k,t}$ to be a piecewise linear function, we can express $S_{k,t}$ as follows:

$$S_{k,t} = S_k^B + \sum_{w=1}^{W} b_{k,w} p_w(t), \tag{7}$$

where $b_{k,w}$'s are the parameters to be estimated and $p_w(t)$'s are a set of pre-determined basis functions. Each $p_w(t)$ is continuous with a unit slope within the w-th window and flat in other windows, as illustrated in Figure 1. To align with quarterly reported holdings, we set

$p_w(0) = 0$, ensuring $S_{k,0} = S_k^B$, and we impose the following constraint:

$$\sum_{w=1}^{W} b_{k,w} p_w(T) = S_k^E - S_k^B, \tag{8}$$

which ensures that $S_{k,T} = S_k^E$.

As discussed before, we employ a two-step sequential estimation: we first estimate directional trades and then infer round-trip trades based on the residuals. The final estimation $b_{k,w}$ is the sum of two components:

$$b_{k,w} = b_{k,w}^{(1)} + b_{k,w}^{(2)}, \tag{9}$$

where $b_{k,w}^{(1)}$ and $b_{k,w}^{(2)}$ are obtained from the first- and second-stage estimation, respectively. For directional trades in the first stage, $b_{k,w}^{(1)}$ by definition changes monotonically in time and must satisfy the following condition:

$$\sum_{w=1}^{W} b_{k,w}^{(1)} p_w(T) = S_k^E - S_k^B, \quad \text{and} \quad b_{k,w}^{(1)} \begin{cases} \geq 0, & \text{if } S_k^E > S_k^B, \\ \leq 0, & \text{if } S_k^E < S_k^B. \end{cases} \tag{10}$$

In the second stage, for round-trip trades, $b_{k,w}^{(2)}$ must satisfy

$$\sum_{w=1}^{W} b_{k,w}^{(2)} p_w(T) = 0. \tag{11}$$

To better discipline our estimation, we impose both a turnover constraint and a sparsity constraint on the round-trip trading estimation. First, we require that the total amount of round-trip trading should be less than the fund's quarterly trading volume minus directional trading volume. The turnover constraint for round-trip trades is given by:

$$\frac{T}{W} \sum_{k=1}^{K} \sum_{w=1}^{W} |b_{k,w}^{(2)}| \bar{P}_{k,w} \leq \hat{\tau}, \tag{12}$$

where $\bar{P}_{k,w}$ is the average price for stock k in window w, and $\hat{\tau}$ is the difference between the fund-reported dollar trading volume and the total value of holding changes.[10] Furthermore, we impose a sparsity constraint to mitigate overfitting that may arise from multiple estimated round-trip trades for a single stock. Specifically, we limit the number of non-zero coefficients

[10] In case the dollar volume reported by the fund is smaller than the total value of the holding changes, we set $\hat{\tau} = 0$.

of $b_{k,w}^{(2)}$'s for each stock k to no greater than two:

$$\sum_{w=1}^{W} \mathbb{I}(b_{k,w}^{(2)} \neq 0) \leq 2. \tag{13}$$

The description above outlines the objective function and the constraints imposed for trading inference. Additionally, we extend our framework to incorporate the following realistic features:

- Cash holding or borrowing. We allow funds to hold cash or use leverage. In this case, the fund's reported return R_t^{fund} is expressed as the equity return R_t^{equ} multiplied by a leverage ratio l_t, which also requires estimation. A leverage ratio $l_t < 1$ indicates cash holding, while $l_t > 1$ indicates borrowing.

  Without loss of generality, we assume l_t remains constant within a quarter. This simplification reduces the number of parameters we need to consider and allows us to focus on estimating equity trading. However, in other applications, such as liquidity management, our method could be naturally extended by introducing a parametric structure for l_t, analogous to our approach with $S_{k,t}$.

- Fund flows. We consider fund flows through the parameter l_t. For example, after an inflow, a fund can either hold more cash (reflected in a change in l_t) or purchase stocks (captured by the trading parameter $b_{k,w}$). Under the current assumption of a constant l_t within each quarter, we implicitly assume that when flows occur, funds trade to maintain a target leverage ratio (or cash holding ratio) throughout the quarter. This assumption is supported by prior studies (e.g., Lou, 2012), which show that mutual funds trade proportionally in response to fund flows.

- Short selling constraint. Given that the majority of mutual funds do not engage in short selling (An, Huang, Lou, and Shi, 2023), we impose a short selling constraint $S_{k,t} \geq 0$ in our trading inference model.

Taken together, the full inference problem can be expressed as follows:

$$\min_{\{l, b_{k,w}^{(1)}, b_{k,w}^{(2)}\}} \sum_{t=1}^{T} \left[R_t^{equ} - R_t^{B} - \sum_{k=1}^{K} \delta_{k,t-1} \left(R_{k,t} - R_t^{equ} \right) \right]^2 \tag{14}$$

$$
\begin{aligned}
\text{s.t.}\quad & R_t^{fund} = l \times R_t^{equ}, && \text{(cash holding or borrowing)}\\
& \delta_{k,t} = \frac{(S_{k,t} - S_k^B)P_{k,t-1}}{\sum_{k=1}^K S_k^B P_{k,t-1}}, && \text{(definition of } \delta)\\
& S_{k,t} = S_k^B + \sum_{w=1}^W b_{k,w} p_w(t), && \text{(piecewise linear representation of daily shares)}\\
& S_{k,t} \geq 0, && \text{(short selling constraint)}\\
& b_{k,w} = b_{k,w}^{(1)} + b_{k,w}^{(2)}, && \text{(two-step estimation: directional and round-trip)}\\
& \sum_{w=1}^W b_{k,w}^{(1)} p_w(T) = S_k^E - S_k^B, && \text{(boundary condition for directional trades)}\\
& \text{sign}(b_{k,w}^{(1)}) = \text{sign}(S_k^E - S_k^B), && \text{(monotonicity constraint for directional trades)}\\
& \sum_{w=1}^W b_{k,w}^{(2)} p_w(T) = 0, && \text{(boundary condition for round-trip trades)}\\
& \sum_{w=1}^W \mathbb{I}(b_{k,w}^{(2)} \neq 0) \leq 2, && \text{(sparsity constraint for round-trip trades)}\\
& \frac{T}{W} \sum_{k=1}^K \sum_{w=1}^W |b_{k,w}^{(2)}| \bar{P}_{k,w} \leq \hat{\tau}. && \text{(turnover constraint for round-trip trades)}
\end{aligned}
$$

We first estimate $b_{k,w}^{(1)}$ by setting $b_{k,w}^{(2)} = 0$, and then estimate $b_{k,w}^{(2)}$ while holding $b_{k,w}^{(1)}$ constant. After obtaining the estimated parameters $\{\hat{b}_{k,w}\}$, the inferred daily holdings are calculated as $\hat{S}_{k,t} = S_k^B + \sum_{w=1}^W \hat{b}_{k,w} p_w(t)$. We consider a 5-day window (i.e., a trading week) for piecewise linear representation and impose a maximum limit on the number of stocks considered, setting $K \leq 500$ in each step of the estimation process. We solve the problem by decomposing it into a sequence of smaller quadratic programming problems using block Gauss-Seidel algorithms (Judd, 1998). This algorithm exhibits nice convergence properties, which are formally proved in Li, Sun, and Toh (2019). In untabluated exercises, we also confirm that our estimation is robust to the choice of initial values, which is unsurprising given the proof in Li, Sun, and Toh (2019).

2.1.4 Further discussions

To make the inference task manageable, we have inevitably abstracted away from certain real-world complexities that may contribute to the return gap (as discussed in Kacperczyk, Sialm, and Zheng (2008)). Below, we outline these simplifying assumptions and discuss their implications.

The timing and costs of actual trades. We assume that all transactions occur at the end of each trading day. While this assumption simplifies the estimation, it overlooks two important features: (a) intra-day trading and (b) trading costs, which include trading commissions and price impact. Although technically, we can incorporate intra-day trading into our framework using finer windows within a day (e.g., hourly intervals), we choose to focus on daily trading a) to reduce the number of parameters and b) to align with daily fund return observations.

Other factors contributing to the return gap. Funds may also engage in various hidden actions that contribute to the return gap. For example, the reported returns are net of expenses; we therefore adjust the reported returns by adding back the expenses. Other hidden activities include IPO allocations, security lending, and derivative holdings, among others. In Section 2.2, we show that trading costs and other unobserved fund actions have a modest impact on the accuracy of our proposed method.

2.2 Accuracy of Our Method

In this subsection, we assess the performance of our trade-inference method. We begin by describing the evaluation metrics. We then evaluate the accuracy of our method using three distinct samples.

2.2.1 Evaluation metrics

We use a heatmap to depict the performance of our proposed method. The heatmap H is a 12×12 matrix, where each element (i, j) indicates the probability that the algorithm classifies a trade from week i (the actual trading week) to week j (the inferred trading week). A perfect classification would result in all diagonal elements being one and off-diagonal elements being zero. More specifically, in each fund-quarter, for a given stock and a given trading direction (either buy or sell), we use $\{v_i\}_{i=1}^{12}$ and $\{\hat{v}_i\}_{i=1}^{12}$ to denote the actual and inferred dollar trading quantities in the stock in each of the 12 weeks, respectively. Each diagonal element of matrix H represents the matched component between the actual and inferred trades; this is determined by taking the smaller of v_i and $\hat{v}_i$. Formally, in each fund-quarter, we calculate the diagonal elements of H by iterating through all stocks and all buys and sells:

$$H_{ii} \;\leftarrow\; \min(v_i, \hat{v}_i), \quad 1 \leq i \leq 12.. \tag{15}$$

We next consider the off-diagonal elements, which represent mismatched trading quantities. Since the constraints we impose ensure that the total amount of inferred trades must equal that of the actual trades, our trading-inference method effectively "reallocates" trade quantities across weeks.[11] We begin by considering the simplest case where both actual and inferred trades occur in a single week. Suppose that an actual buy occurs in week i_0 and the inferred buy in week j_0, where $i_0 \neq j_0$. In this case, the off-diagonal elements are determined by $H_{ij} = \mathbf{I}_{\{i=i_0\}} \times \max(\hat{v}_j - v_j, 0)$; that is, the overestimated trade quantities in each week j are assigned to the element (i_0, j), where i_0 corresponds to the week with underestimation.

We then generalize this calculation to allow actual and inferred trades to span multiple weeks. For example, suppose that the fund buys 40 and 60 shares of a stock in the 1st and 2nd weeks; our method, however, indicates that the fund buys 30 and 70 shares in the 1st and 3rd weeks. As a result, trade quantities in the 1st and 2nd weeks are underestimated by 10 and 60 shares, respectively. Reciprocally, the trade quantity in the 3rd week is overestimated by 70 shares. We then assign the dollar value of the 10 shares (out of the 70 shares) to off-diagonal element (1,3) and the dollar value of the remaining 60 shares to element (2,3), to match the underestimation in weeks 1 and 2. More formally, we obtain the following expression for off-diagonal elements:

$$H_{ij} \;\leftarrow\; \left[\frac{\max(v_i - \hat{v}_i, 0)}{\sum_k \max(v_k - \hat{v}_k, 0)}\right] \times \max(\hat{v}_j - v_j, 0), \quad i \neq j, \quad 1 \leq i, j \leq 12. \tag{16}$$

Again, the intuition for Eq.(16) is that the overestimation in one week arises from underestimation in other weeks; we then distribute the overestimated quantity to multiple off-diagonal elements according to the degree of underestimation in those weeks. By iterating through all stocks and all buys and sells, Eq.(16) provides the total misclassified trading quantity for each (i, j) pair.

To facilitate comparison across fund-quarters, we normalize each element in H by the total dollar trading volume of the fund-quarter, so that the sum of all elements in the matrix is one. We then average the H matrix across all fund-quarters. Finally, we normalize each row of the H matrix to have a sum of one (i.e., within each actual trading week), so that the elements in each row represent conditional probabilities. That is, conditioning on the actual trading week being i, element (i, j) represents the probability that the inferred trading week is j.[12]

[11] For each stock and each trading direction, in the absence of round-trip trades, the inferred and actual quarterly trading quantities must be equal, because both need to match the reported shares in two consecutive quarter-ends. However, when round-trip trades occur, discrepancies may arise. For example, a fund may buy and sell 100 shares of a stock within a quarter, while our method estimates the buy and sell quantities for that stock to be 150 shares. Despite such differences at the stock level, the constraint on the total trading volume ensures that the total actual and inferred trading quantities (in dollars) across all stocks for the fund-quarter remain equal.

[12] Alternatively, we can perform the normalization within each column, so that element (i, j) represents the probability that the actual trading week is i, conditional on the estimated trading week being j.

2.2.2 Three evaluation samples

We construct three distinct samples to evaluate our method. First, we use simulated trades and fund returns, which allows us to fully control the data-generating process and evaluate our method for specific purposes. Second, we evaluate our method using actual trades from the Ancerno database combined with trade-implied fund returns. Finally, we use the matched sample between the Ancerno and TFN/CRSP databases to have both actual trades and actual fund returns.

Simulated trades and returns. We start by obtaining quarterly trades (based on holding changes) of a randomly selected sample of 10,000 fund-quarters from the TFN mutual fund database. We then simulate intra-quarter trades by randomly choosing trading days within the quarter. Specifically, for each fund-quarter and each stock, we randomly select one trading day on which the fund's holding changes from the quarter-beginning shares, S_k^B, to the quarter-end shares, S_k^E. We focus on one-day trades as these are the most difficult to estimate with our smooth-trading framework. Our results are similar (in many cases get better) if we allow simulated trades to span multiple days. With these simulated intra-quarter trades, we then generate the corresponding trade-implied fund daily returns.

Actual trades and trade-implied returns. We obtain detailed intra-quarter fund trades from the Ancerno sample, which spans the period 1997 to 2011 and includes 1,039 unique funds. We then calculate fund holdings by accumulating past trades for at least eight quarters.[13] To form the evaluation sample, we require each fund to hold at least 10 stocks with a total holding value greater than $10 million. After applying these filters, we have 425 funds in the sample. Consistent with prior studies (Anand, Irvine, Puckett, and Venkataraman, 2011; Jame, 2018; Puckett and Yan, 2011), our Ancerno funds are, on average, similar to TFN/CRSP funds in terms of stock holdings and return characteristics. For example, the median number of holdings is 74 in the Ancerno sample and 72 in the TFN/CRSP sample.

Our final sample period is 2001-2011. This is because the Ancerno sample has limited coverage in the first two years (1997–1998, (Hu, Jo, Wang, and Xie, 2018)), and we require additional two years to calculate fund holdings. We use a random sample of 1,000 fund-quarters per year, resulting in 11,000 fund-quarters in total. Finally, we generate trade-implied fund daily returns based on the actual trades reported in Ancerno.

[13] If the sum of trades is negative, it is set to zero. This is to address the possibility that funds may have purchased a stock before they start reporting to Ancerno.

Actual trades and actual returns In our third evaluation sample, we use actual fund returns, by matching the Ancerno sample with the TFN/CRSP mutual fund sample based on fund names. To ensure matching quality, we follow the procedure of Puckett and Yan (2011) and compare the accumulated trades from Ancerno to the holding changes from TFN/CRSP, requiring a minimum of 90% matching rate. This procedure results in 37 matched funds and 158 qualified fund-quarters. Notably, the matched funds hold and trade significantly more stocks than the typical fund in TFN/CRSP. For example, the median number of holdings for matched funds is 137, nearly double that of TFN/CRSP funds at 72.

2.2.3 Evaluation results

Figure 2 presents the evaluation results. Panels A and B display our method's performance based on the TFN/CRSP sample. Panel A shows that for the entire TFN/CRSP sample, the accuracy rates in the first and last weeks of each quarter are 54% and 51%, respectively. Panel B further shows that for funds that trade fewer than 50 stocks per quarter – representing approximately half of the sample, the accuracy rates increase to 69% and 65%. For the remaining weeks, our method produces an average accuracy rate of 30% in the full sample and 44% in the sub-sample of funds trading fewer than 50 stocks per quarter. For comparison, randomly guessing the trading week would result in an accuracy rate of 8.3% (= 1/12). It is worth noting that our method is particularly effective in identifying trades during the first and last weeks of the quarter, the key periods for our ESG window-dressing analysis, since these weeks are better disciplined by the quarter-end boundary conditions.

Panels C and D illustrate our method's performance using the Ancerno sample, with actual trades and trade-implied returns. The accuracy is similar to that reported in Panels A and B; accuracy rates for the first and last weeks are 54% and 57% in the full sample and 63% and 67% in the sub-sample of funds with fewer than 50 trades per quarter. For the remaining weeks, the average accuracy rate is 37% in the full sample and 46% in the sub-sample. These results suggest that our trade-inference method is effective with not only simulated trades but also when applied to actual trades.

Finally, we assess the performance of our method using the Ancerno-TFN/CRSP matched sample, incorporating actual trades and actual returns. As shown in Panel E, The accuracy rates for the first and last weeks of each quarter drop to 33% and 30%, respectively. This decline, however, is primarily due to changes in fund composition — matched funds hold and trade substantially more stocks than funds in Ancerno or TFN/CRSP. (The median matched fund holds 137 and trades 131 stocks per quarter, while the median fund in the TFN/CRSP sample holds 72 and trades 67 stocks.) In Panel F, we evaluate our proposed method using the same set of matched funds, utilizing actual trades but trade-implied fund returns. The results suggest that the choice of actual vs. trade-implied fund returns accounts for less than 5% of the accuracy change in the first and last weeks, and less than 10% for other weeks. In other words, unobserved factors in fund returns (e.g., trading costs and hidden fund actions) only have a modest impact on the performance of our proposed method.

2.2.4 Robustness: unobserved factors in fund returns

Next, we further examine the impact of unobserved factors in fund returns on the accuracy of our proposed method, by introducing varying levels of noise into trade-implied fund returns. To this end, we focus on the second evaluation sample (i.e., the Ancerno sample with actual trades and trade-implied returns); we then add noise – from a normal distribution with mean zero and standard deviation ranging from 1-5 bps – to trade-implied fund returns.[14] The results are shown in Figure 3, where the x-axis is the standard deviation of the noise added to fund returns, and the y-axis is the accuracy rate of our proposed method. The solid (dashed) line corresponds to accuracy in the first and last weeks (remaining weeks). We again see a moderate impact on the accuracy of our method. For example, as we increase the standard deviation of noise from 0 to 2 bps, the accuracy rate in the first and last weeks drops by 7.2 percentage points.

2.2.5 Robustness: fund subsamples

Finally, we explore the performance of our method in various fund subsamples. For brevity, we only report results based on the second evaluation sample. We specifically consider three fund characteristics that likely influence the performance of our method: (1) the number of holdings, (2) the fund turnover ratio, and (3) the proportion of round-trip trading volume.

Panel A of Figure 4 shows that our method performs better for funds holding fewer stocks (so fewer parameters to estimate). Importantly, even for the group of funds with the largest number of holdings (over 200 stocks), the accuracy rate of our method is above 40% in the first and last weeks of the quarter, and is around 30% for other weeks, far surpassing the benchmark rate of 8.3%.[15] Panel B shows that our method's performance improves when funds trade less. Importantly, even for the group with the highest turnover ratio (over 1.5 per year), the accuracy rate is above 45% in the first and last weeks of the quarter and is around 30% in other weeks. Finally, Panel C shows that our method performs better with fewer round-trip trades. For the group of fund-quarters with the lowest proportion of round-

[14] The range of standard deviations is comparable to the discrepancy between actual and trade-implied returns in the Ancerno-TFN/CRSP matched sample, which is approximately 2 basis points per day.

[15] This result is consistent with the comparison between Panels A and B of Figure 2 and can explain the difference between Panel E and Panels A-D of Figure 2.

trip trading (below 0.01), the accuracy rate is 72% in the first and last weeks and 52% in other weeks. In the group with the highest proportion of round-trip trades (exceeding 0.3), the accuracy rate is 48% in the first and last weeks and 30% in other weeks. Collectively, these results suggest that our method works well in a wide range of subsamples.

3 ESG Window Dressing

In this section, we apply our method to infer mutual funds' intra-quarter trading and investigate how they cater to the dual objectives of maximizing both financial and ESG performance.

3.1 Data

Mutual fund daily returns are obtained from the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund database, while quarterly holdings come from Thomson Reuters's CDA/Spectrum Mutual Fund Holdings Database. We use the MFLinks file to merge CDA/Spectrum with the CRSP mutual fund database. Our focus is on US active equity mutual funds, which we select based on the following criteria: (1) the investment objective code reported by CDA/Spectrum must be aggressive growth, growth, growth and income, balanced, unclassified, or missing; (2) the ratio of equity holdings to total net assets is greater than 0.75; and (3) the fund is not an index fund. To further ensure data quality, we also require a minimum fund size of $10 million and at least 10 holdings. Our sample covers the period from 2015Q1 to 2022Q2 as ESG investing gained increasing attention during this period. After applying these filters, our final sample includes 3,529 unique funds and 58,790 fund-quarters. Summary statistics of fund characteristics are provided in Table 1 Panel A.

Our stock-level ESG ratings come from three major rating providers: Morningstar Sustainalytics, MSCI, and Refinitiv Asset4 (Lipper). Since the literature has documented a low correlation across these ratings, we combine ratings from all three sources to minimize noise. Specifically, we take the average rank-normalized ESG scores in the last quarter for each stock across three ratings and identify the top (bottom) 200 stocks as high-ESG (low-ESG) stocks. We require stocks to have at least two non-missing ESG scores from the three rating agencies to be included in the list. Figure 5 shows the time series of Pearson correlations among the three ESG ratings. Consistent with previous literature, the average correlation is 0.357, indicating modest agreement among the ratings.[16]

[16] A notable drop in correlation with Sustainalytics in 2019Q3 is due to Morningstar's change in the ESG rating scheme at that time.

3.2 Trading around Quarter-ends

We begin by examining how mutual funds trade stocks with high- versus low- ESG scores around the turn of each quarter. For each fund each week, we calculate the fraction of trading in high-ESG (low-ESG) stocks by taking the ratio between the trading volume in these stocks and the total trading volume. We calculate this ratio for buy-trades, sell-trades, and net-trading (buy minus sell), respectively. To investigate how trading of ESG stocks evolves over different time windows, we regress the fraction of trading in high- or low- ESG stocks on a series of dummy variables indicating different weeks within a quarter, specifically in the following form:

$$y_{i,t,l} = b_0 + \left(\sum_{j=1}^{3} b_{E,j} \times \mathbf{I}_{E,j} + b_{B,j} \times \mathbf{I}_{B,j} \right) + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}, \tag{17}$$

$y_{i,t,l}$ is the fraction of (buy, sell, or net) trading volume in high- or low- ESG stocks for fund i in quarter t and week l. $\mathbf{I}_{E,j}$ ($\mathbf{I}_{B,j}$) is a dummy variable indicating the j-th week from quarter end (beginning). To control for the overall patterns in buy-sell imbalance possibly due to fund flows or other reasons, we control for the fraction of total buy volume out of total trading volume for fund i in quarter t and week l ($buy\_ratio_{i,t,l}$). We also include fund-year-quarter level fixed effects, $\alpha_{i,t}$. With this specification, the coefficients of the dummy variables indicate the percentage of abnormal trading in each week around the turn of the quarter.

Table 2 Panel A shows the regression results of Eq.(17). We report the buy, sell, and net trading of high- versus low- ESG stocks in weeks proceeding and subsequent to the quarter ends respectively, as well as the difference between these stocks. All estimated coefficients are expressed in percent. We find that mutual funds buy high-ESG stocks and sell low-ESG stocks right before the quarter ends: their abnormal net buy of high- (low-) ESG stocks accounts for 0.61% (−0.17%) of total trading volume in the first week prior to the quarter ends, and the difference 0.78% is statistically significant with a t-stat of 3.98. The magnitude of abnormal trading diminishes in the prior weeks: the difference in net trading between high- and low-ESG stocks becomes 0.33% (t-stat = 2.05) and 0.17% (t-stat = 0.92) in the second and third week before the quarter ends. On the other hand, mutual funds reverse these trades at the beginning of the next quarter: the difference in net trading between high- and low-ESG stocks is −0.43% (t-stat = -1.91) in the first week in the next quarter. Taking the difference between net trading in high- minus low- ESG stocks in the first week before and after the quarter ends, the overall effect accounts for 1.2% (0.78% + 0.43%, t-stat = 3.48) of mutual funds' total trading volume. For reference, this effect is equivalent to 14%

of the sample average of the weekly trading volume for these high- and low-ESG stocks ($\frac{0.615-(-0.130)}{\text{Avg.High.Buy+Avg.Low.Buy}} - \frac{(-0.485)-(-0.019)}{\text{Avg.High.Sell+Avg.Low.Sell}}$, see Table 1 Panel C for sample averages). In the Online Appendix Table A-1, we consider a decomposition analysis and show that directional trades contribute 62% and round-trip trades contribute 38% to the overall effect. We also find consistent evidence from stock-level trading volume (see the Online Appendix Table A-2), which validates our findings in an estimation-free manner.

To strengthen the argument of window dressing at the quarter end, we conduct a placebo test by focusing on the non-quarter-ending month ends (e.g., the end of January and February). The rationale for this placebo test is based on how ESG rating agencies evalulate funds' ESG performance. Since mutual funds only have mandates to report their holdings only at quarter ends, their quarterly holdings are most reliable sources for ESG rating agencies to evaluate and monitor ESG performance. Specifically. we repeat our exercises around these month ends and report the results in Table 2 Panel B. In contrast to the patterns reported in Panel A, the net trading in high- and low- ESG stocks as well as their difference around month ends that require no reporting are all insignificantly different from zero.

3.3 ESG Window Dressing around Extreme Rating Cutoffs

To further sharpen our identification, we leverage the fact that mutual funds with different sustainability percentile rankings may have varying incentives to manipulate their ESG performance due to the discontinuity in investors' flow responses near the thresholds of extreme rating categories. Specifically, we zoom in on Morningstar ratings, for which we have detailed data. Morningstar introduced its sustainability ratings in March 2016, evaluating over 20,000 mutual funds through a percentile system.

The classification of funds is determined by assessing the sustainability of funds' underlying holdings, with each holding assigned a sustainability score derived from Sustainalytics' analysis of public documents. This rating is related to how a firm scores on environmental, social, and governance issues (ESG). At the end of each quarter, Morningstar calculates a fund-specific sustainability score by taking the weighted average of these holding scores. Funds are then ranked within their Morningstar category based on their sustainability scores, and are rated on a five-globe scale based on their percentile ranking. A "High" rating (five globes) is given to the top 10%, "Above Average" (four globes) for 10%-32.5%, "Average" (three globes) for 32.5%-67.5%, "Below Average" (two globes) for 67.5%-90%, and "Low" (one globe) for the bottom 10% in each fund category. The globe ranking is prominently reported using pictures of one to five globes as well as the descriptive label (e.g., "High") on each fund's Morningstar page. The globes are a discrete rating system of five categories, although Morningstar also released each fund's sustainability score and the percentile ranking underlying the ratings.

Important for our identification strategy, investors' flow responses to ESG globe ratings are a) disproportionately strong at the extreme globe categories, and b) exhibit discontinuity at the one-globe and five-globe cutoff points (Hartzmark and Sussman, 2019).[17] This means a fund that is ranked at the 91 (9) percentile, which would receive five globes (one globe), could receive more net inflows (net outflows) due to its ESG ratings than a fund ranked at the 89 (11) percentile, which would receive four globes (two globes), although both funds have similar underlying sustainability characteristics. We conjecture that funds whose ESG scores are around the five-globe or one-globe cutoff point would have the strongest incentives to manipulate their quarter-end holdings, due to the sharp changes in potential payoffs. Here our approach is reminiscent of the research design used by Chevalier and Ellison (1997), who study how mutual fund managers manipulate their risk profiles in response to the flow incentives.

Taking this hypothesis to data, we examine how funds trade high- versus low- ESG stocks around the turn of the quarters as a function of their one-month-lagged sustainability percentage rankings, e.g., for the trades around March 31st, the rankings are from the end of February. In this particular test, since our focus is on funds' incentives to manipulate Morningstar ratings, we classify stocks as high- or low-ESG based solely on Morningstar Sustainalytics, rather than using the average of the three ESG stock ratings as in the other tests. To estimate the incentive variation within each fund category, we further split each rating category into the "Lower half" and the "Upper half" using the underlying percentiles. We adapt our baseline regression and include dummy variables indicating a fund's percentile group ($\mathbf{I}_{\{\text{category } x\}}$) and the last or first week in a quarter ($\mathbf{I}_E$ or $\mathbf{I}_B$). Specifically, we run the following regression:

$$y_{i,t,l} = b_0 + \left[\sum_{x=1L,\ldots,5U} \mathbf{I}_{\{\text{category } x\}} \times (b_{E,x} \times \mathbf{I}_E + b_{B,x} \times \mathbf{I}_B) \right] + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l},$$

where $y_{i,t,l}$ is fund i's high-ESG-minus-low-ESG net trading (buy minus sell) in quarter t week l, defined in the same way as before.

Table 3 reports the results. For funds in each of the two halves within the five categories, we report the ESG window dressing intensity, $b_{E,x} - b_{B,x}$, respectively. We see that the ESG window dressing intensity has an M-shaped pattern with respect to the funds' sustainability percentage rankings. The effect is most pronounced for funds ranked around the cutoff

[17]See Ben-David, Li, Rossi, and Song (2022) for additional evidence on mutual fund investors rating-chasing behavior.

between four and five globes. Specifically, abnormal ESG trading accounts for 5.7% (t-stat = 6.35) of the total trading volume in the four globes upper half (group 4U) and 4.0% (t-stat = 2.97) in the five globes lower half (group 5L). A similar effect is observed at the one-globe cutoff, primarily driven by funds on the left side, with abnormal ESG trading representing 4.7% (t-stat = 3.24) of total trading volume in the upper half of one globe (group 1U). In comparison, the magnitudes of ESG window dressing is 1.0% in the one globe lower half (group 1L), 1.6% in the middle groups (the average of 3L and 3U), and 3.1% in the five globes upper half (group 5U). The stronger effect observed on the right side of the "M" aligns with the notion that funds with higher sustainability ratings are more sensitive to ESG issues and exert greater efforts to attract ESG-related flows. The last column in Table 3 reports the difference in the window dressing intensity between the upper and lower half within each rating category. Consistent with our hypothesis, funds in groups 4U and 1U demonstrate a significantly higher intensity of ESG window dressing compared to groups 4L and 1L, respectively.

An alternative way to gauge the incentive variation is to estimate the magnitude of ESG window dressing as a piecewise linear function (instead of using group dummies) of the funds' sustainability percentile rank. Specifically, we estimate the following regression

$$y_{i,t,l} = b_0 + f_E(p) \times \mathbf{I}_E + f_B(p) \times \mathbf{I}_B + g(p) \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l},$$

where $f_E(p)$, $f_B(p)$, and $g(p)$ are piecewise linear functions partitioned on the rating categories, and p is the percentile rank from 0% (low-ESG) to 100% (high-ESG). We plot $f_E(p) - f_B(p)$ in Figure 6. Again, we observe an M-shaped function where the magnitude of ESG window dressing is strongest around the one-globe and five-globe cutoffs.

3.4 Return Patterns

One may expect that these window-dressing trades can potentially exert price pressure and generate return impact. To empirically assess this conjecture, we next investigate the return patterns of high- versus low-ESG stocks in a short window around the turn of each quarter.

Table 4 Panel A reports cumulative risk-adjusted returns for equal-weighted and value-weighted portfolios of high- versus low-ESG stocks in 1-, 3-, and 5-day windows at the quarter end and beginning over the period from 2015Q1 to 2022Q2. Risk-adjusted returns are calculated based on the Fama-French three-factor model, and we estimate betas using a 60-month rolling window of monthly returns. We see that high-ESG stocks outperform low-ESG stocks before the quarter ends, and the pattern reverses at the beginning of the next quarter. Take the value-weighted portfolio for instance, the high-ESG stocks generate an abnormal return of 0.17% (t-stat=4.06) in the five days before quarter ends and then experience a negative return of -0.17% (t-stat=-1.96) in the next five days at the quarter beginning. Returns of low-ESG stocks show the reverse pattern. Since low-ESG stocks are typically smaller, their abnormal return tends to be larger but more volatile, potentially making returns less statistically significant. To show the differences of return patterns between high- and low-ESG stocks, we consider a long-short portfolio that longs high-ESG stocks and shorts low-ESG stocks and we find that this portfolio generates a five-day return of 0.78% (t-stat=2.07) and -0.28% (t-stat=-1.20) at the end and the beginning of each quarter, respectively. Finally, the difference in long-short portfolios' returns between the quarter end and beginning is 1.07% with a t-stat of 2.04.

Like Table 2, we conduct a placebo test of ESG return patterns in the non-quarter-ending month-ends. As we show previously that mutual funds do not engage in ESG window dressing around month-ends (Table 2 Panel B), we expect no significant difference in returns between high- and low-ESG stocks around these month ends. This is indeed what we find in Table 4 Panel B.

4 Additional Tests on ESG Window Dressing

In this section, we conduct additional tests to corroborate our evidence presented in Section 3. Specifically, we take advantage of our approach and conduct subsample analyses in Section 4.1 and discuss the costs and benefits of ESG window dressing in Section 4.2.

4.1 Subsample Analyses

In this subsection, we conduct sub-sample analyses to strengthen our argument. We first examine what types of funds conduct more ESG window dressing and when funds are likely to conduct such a specific one. Afterward, we exploit how funds select ESG stocks in their window dressing.

4.1.1 Fund Heterogeneity

To exploit fund heterogeneity in ESG window dressing, we run the following panel regression:

$$y_{i,t,l} = b_0 + \left[\sum_x \mathbf{I}_{\{\text{category } x\}} \times (b_{E,x} \times \mathbf{I}_E + b_{B,x} \times \mathbf{I}_B)\right] + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \varepsilon_{i,t,l}.$$

The dependent variable, $y_{i,t,l}$, represents fund i's net trading of high-ESG stocks minus low-ESG stocks (buys minus sells) as a proportion of total trading volume in quarter t and week

l. The variables $\mathbf{I}_{\{\text{category } x\}}$, $\mathbf{I}_E$, and $\mathbf{I}_B$ are binary indicators for the fund category x, the last week of the quarter, and the first week of the quarter, respectively. The key variable of interest is the ESG window dressing intensity,, $b_{E,x} - b_{B,x}$.

We consider three fund-level variations for construction of $\mathbf{I}_{\{\text{category } x\}}$: a) ESG funds, b) past fund performance, and c) state political leanings. For ESG funds, we use the Morningstar classification and identify ESG funds with the filter "Sustainable Investment Overall = Yes," based on the fund prospectus or other regulatory filings. For fund performance, we utilize the information ratio, such as the Fama-French 3-factor alpha divided by residual volatility, and categorize funds into High- (top 20%), Middle- (middle 60%), and Low- (bottom 20%) performance groups. For state political leanings, we assess the political vote of the state where the fund's headquarters is located, defining Democratic States (Republican States) as those that voted for the Democratic Party (Republican Party) in all of the 2004, 2008, 2012, 2016, and 2020 presidential elections.

Table 5 reports the results and reveals some novel findings. First, we observe that ESG window dressing is more pronounced among ESG funds and those located in Democratic states. In the conventional view, these funds are expected to consistently implement an ESG strategy due to their mandates or the preferences of their investors. However, these findings suggest that rather than prioritizing social impact over financial returns, funds may be employing a low-cost strategy, such as ESG window dressing, to meet their clients' growing environmental and social expectations.

We also find that well-performing funds are more likely to engage in ESG window dressing. This likely stems from the funds' optimal balance between social value and financial returns. Intuitively, since well-performing funds have already achieved superior financial performance, they face diminishing returns to scale, and it is difficult for them to improve their financial performance further. As a result, these funds may focus on enhancing their non-financial metrics—such as their social profile—to attract additional fund flows.

4.1.2 Time-Series Variation

Next, we analyze the time-series variation in ESG window dressing. We hypothesize that ESG window dressing becomes more prevalent as investor awareness of environmental and social issues increases. To test this hypothesis, we run the following regressions:

$$y_{i,t,l} = b_0 + \left[\sum_p \mathbf{I}_{\{\text{period } p\}} \times (b_{E,p} \times \mathbf{I}_E + b_{B,p} \times \mathbf{I}_B) \right] + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \varepsilon_{i,t,l}.$$

The dependent variable, $y_{i,t,l}$, represents fund i's high-ESG-minus-low-ESG net trading (buy minus sell), divided by the total trading volume in quarter t and week l. $\mathbf{I}_{\{\text{period } p\}}$, $\mathbf{I}_E$, and $\mathbf{I}_B$ are zero-one indicators for the time period p, the last week of the quarter, and the first week of the quarter, respectively. The variable of interest is the ESG window dressing intensity, $b_{E,x} - b_{B,x}$.

We use two proxies to measure investor ESG awareness: (a) ESG fund flows and (b) Google search index for the keyword "sustainability". For ESG fund flows, we use two steps to identify periods with high ESG fund flows. In the first step, we calculate ESG fund flow as the difference between the aggregate percentage flows to high sustainability rating funds (e.g., five-globe funds) and low sustainability rating funds (e.g., one-globe funds). In the second step, we identify months with "High ESG fund flow" as those in the top decile of the past six-month moving average of ESG fund flows. Similarly, we define "High Google index" months as those in the top decile of the past six-month moving average of Google search index values for the keyword "sustainability".

Table 6 reports the results. Consistent with our conjecture, we find that funds engage in more ESG window dressing when investor awareness of environmental and social issues is higher, as measured by both ESG fund flows and Google search interest in "sustainability".

4.1.3 Stock Heterogeneity

We now examine how funds select stocks in window dressing. Intuitively, there is heterogeneity among ESG stocks. When faced with stocks of similar ESG scores, funds prefer purchasing those with lower transaction costs or lower risks, as trading such stocks incurs lower costs and has less impact on financial performance. To test this hypothesis, we exploit variations in stock characteristics and run the following regression:

$$y^s_{i,t,l} = b_{0,s} + b_{E,s} \times \mathbf{I}_E + b_{B,s} \times \mathbf{I}_B + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \varepsilon_{i,t,l}.$$

The dependent variable, $y^s_{i,t,l}$, represents fund i's high-ESG-minus-low-ESG net trading (buy minus sell) in stock category s, divided by the total trading volume in quarter t and week l. $\mathbf{I}_E$ and $\mathbf{I}_B$ are zero-one indicators for the last week and the first of a quarter, respectively. The variable of interest is the ESG window dressing intensity, $b_{E,x} - b_{B,x}$.

We focus on two stock characteristics: a) bid-ask spread, which serves as a measure of transaction costs, and b) idiosyncratic volatility, which captures the unintended risk associated with such trading. The bid-ask spread is calculated as the average of daily (Ask-Bid)/((Ask+Bid)/2) over the previous month. Idiosyncratic volatility is measured as the residual standard deviation from a Fama-French 3-factor model via a rolling-window regression. Each quarter, we sort high-ESG and low-ESG stocks into two equal groups based on either bid-ask spread or idiosyncratic volatility. We then estimate the coefficient difference, $b_{E,s} - b_{B,s}$, for each group. Table 7 reports the results and confirms our conjecture that ESG window dressing is indeed more prevelant among stocks with higher liquidity and lower idiosyncratic volatility.

4.2 Costs and Benefits of ESG Window Dressing

In this subsection, we explore the costs and benefits associated with ESG window dressing. The primary goal of delegated asset managers is to maximize assets under management (AUM), which directly influences their compensation. We conjecture that ESG window dressing help to attract fund flows, especially given the growing awareness of environmental and social issues among the investor base.

To test this hypothesis, we conduct panel regressions using fund-quarter observations. The dependent variable is the percentage of fund flows in the quarter $t + 1$, and the key independent variable is the fund's *Net ESG trading* at quarter t. *Net ESG trading* is calculated as the net (buy minus sell) ESG trading volume divided by the total trading volume in the last week of quarter t. Consistent with our prior, Table 8 column (1) shows that ESG window dressing significantly increases fund flows (t-stat=7.01). This effect remains highly significant even after controlling for common fund flow determinants, such as past fund flows and fund performance (see column 3). The effect is also economically important, as a one-standard-deviation increase in *Net ESG trading* leads to 21 basis points more in fund flows ($1.214\% \times 0.17$). For reference, the effect of *Net ESG trading* is comparable to the impact of a 108 basis point (21 bps/0.194) increase in annualized alpha.

ESG window dressing may also come at a cost, especially when there is a misalignment between social value and private financial returns. Engaging in ESG window dressing can potentially negatively impact fund performance, as it often involves transaction costs and potential performance trade-offs.

To empirically assess the trade-off between ESG and financial performance, we conduct panel regressions of fund performance at the fund-quarter-week level. The dependent variables are cumulative excess return (column 1) and cumulative CAPM-adjusted return (column 2) for each 5-day window. CAPM β is estimated using monthly returns from month $t - 60$ to month $t - 1$. The main independent variables are an indicator for the last week of the quarter and its interaction with the current-quarter ESG window dressing measure, *Net ESG trading* (defined as in Table 8). We include high-dimensional fixed effects at the fund-year-quarter level to control for unobserved fund characteristics.

Table 9 reports the results and reveals two interesting observations. First, the coefficient for the last-week indicator is not significantly related to fund performance, whether measured by excess return or CAPM α. However, the interaction between this indicator and *Net ESG trading* is negatively and significantly associated with fund performance. A one-standard-deviation increase in *Net ESG trading* in the last week of the quarter results in 21 basis points lower returns annually ($-0.315\% \times 0.17 \times 4$). These results confirm that ESG window dressing indeed incurs some costs for mutual funds. Meanwhile, the insignificant coefficient of the last-week indicator comforts us that our results are not driven by some mechanical pure quarter-end effects.

5 Other Applications

In this section, we present additional applications based on our inference method. First, we apply our method to several settings in which mutual funds exhibit known trading patterns, serving as validation tests. These include mutual funds' performance window-dressing and portfolio-pumping around quarter ends. See evidence of performance window-dressing behavior from Agarwal, Gay, and Ling (2014); He, Ng, and Wang (2004); Lakonishok, Shleifer, Thaler, and Vishny (1991); Meier and Schaumburg (2006); Ng and Wang (2004), and portfolio-pumping behavior from Ben-David, Franzoni, Landier, and Moussawi (2013); Bernhardt and Davies (2005); Bhattacharyya and Nanda (2013); Carhart, Kaniel, Musto, and Reed (2002); Hu, McLean, Pontiff, and Wang (2013). We also examine how mutual funds trade around informational events, such as around M&A announcements (Fich, Lantushenko, and Sialm, 2024). Second, we consider a new application of return gap decomposition. We decompose the return gap of Kacperczyk, Sialm, and Zheng (2008) into components that arise from directional trading and that from round-trip trading. The decomposition provides new insights on the source of fund return predictability.

5.1 Performance Window-Dressing, Portfolio Pumping, and Trading around M&A Announcements

Performance window-dressing is a behavior pattern where fund managers buy past winner stocks and sell past losers as the end of the quarter approaches. Fund managers do so to make their holdings appear more impressive in the quarterly reports. Another strategic behavior, known as portfolio pumping, involves excessive buying of stocks that mutual funds already heavily own. The purpose of portfolio pumping is to inflate the funds' closing net asset value, thereby exaggerating the funds' performance.

We test for these two strategic behaviors based on a similar regression setting as in Section 3.2 Eq.(17). We regress the trading volume of a group of stocks on quarter-ending/beginning indicators in a fund-quarter-week panel regression. To test for performance window dressing, we define winner (loser) stocks as the top 10% (bottom 10%) sorted by past cumulative returns from month $t-12$ to $t-2$. For portfolio pumping, we first sort stocks in descending order by holding values for each fund quarter. We then define heavily-owned (under-owned) stocks as the top (bottom) stocks that collectively account for 10% of total holding values. Using our method, we confirm that mutual fund managers engage in performance window dressing (see the Online Appendix Table A-3) and portfolio pumping (see the Online Appendix Table A-4)). A detailed discussion of these results is presented in the Online Appendix.

As an additional validation, we examine how mutual funds trade acquisition targets around M&A announcements. Based on quarterly holdings, Fich, Lantushenko, and Sialm (2024) document that mutual funds reduce their equity holdings in impending targets, while hedge funds trade in the opposite direction. We replicate and extend their findings by examining trading on a daily basis. Consistent with Fich, Lantushenko, and Sialm (2024), we find that the aggregate daily holdings of mutual funds in targets show a significant decrease surrounding M&A announcements. Furthermore, our daily-level analysis indicates that during a one-quarter window, i.e., $[t-30, t+30]$, approximately 32% of the decrease in ownership occurs before M&A announcements (see the Online Appendix Figure A-1). A detailed discussion of these results is presented in the Online Appendix.

In all three exercises, we confirm and extend prior empirical evidence on mutual fund trading patterns by either expanding the coverage of the fund sample or increasing the frequency of observations. These exercises not only provide the external validity of our trading inference method but also highlight its potential for broader applications.

5.2 Decomposing the Return Gap

We extend our study by exploring the return predictability of the return gap, as introduced by Kacperczyk, Sialm, and Zheng (2008). The return gap, defined as the difference between a fund's reported return and the return on a portfolio based on the fund's previously disclosed holdings, has been widely shown to be a reliable predictor for fund performance. However, the underlying drivers of this predictability are not fully understood. In this subsection, we apply our method to estimate mutual funds' intra-quarter directional and round-trip trading, which directly contribute to the return gap. We then decompose the return gap into components attributable to directional and round-trip trading separately. By examining the return predictability of these components, we aim to better understand the factors driving the return predictability of the return gap.

To decompose the return gap, we first estimate each fund's intra-quarter directional and round-trip trading using the method outlined in Section 3. Next, we compute the fund's returns from the quarter beginning based on these estimates. This allows us to break down the return gap from Kacperczyk, Sialm, and Zheng (2008) into three components: the directional-trade return gap, which reflects the impact of directional trading; the round-trip-trade return gap, which captures the effect of round-trip trading; and the residual component, representing the difference between the total return gap and the sum of the other two components. Finally, we perform portfolio sorting to analyze the return predictability associated with each of these components.

Table 10 presents our findings. Panel A replicates the results of Kacperczyk, Sialm, and Zheng (2008) using our sample, confirming that the return gap can significantly and positively predict future fund performance. Panels B to D show portfolio sorts based on the directional-trade return gap, round-trip-trade return gap, and residual return gap, respectively. Notably, Panel C demonstrates that the round-trip-trade return gap is a significant predictor of future performance. Across various performance measures, funds in the top decile of the past 12-month round-trip-trade return gap outperform those in the bottom decile. For example, when evaluated against the Fama-Frech-Five Factors plus the moment factor, funds in the top decile outperform those in the bottom decile by 13 basis points per month (t-stat=2.35), which is economically significant. For comparison, the difference in fund performance between the top and bottome decile based on the original past 12-month return gap, as identified by Kacperczyk, Sialm, and Zheng (2008), is 17.5 basis points per month (t-stat=2.71). In contrast, the directional-trade return gap shows weak or insignificant return predictability across various performance metrics (as shown in Panel B). Additionally, Panel D reveals that the residual return gap can also significantly and positively predict future fund performance, although the drivers of residual trading remain unclear and warrant further investigation.

The stark contrast between the return predictability of the directional-trade and round-trip-trade return gaps sheds light on the sources of return predictability. One possible explanation is that the return gap predicts future performance because it reflects the funds' liquidity provision. This argument is similar to Da, Gao, and Jagannathan (2011). Da, Gao, and Jagannathan (2011) decompose a mutual fund's stock selection skill into liquidity-absorbing impatient trading and liquidity provision and show that funds with higher return gaps add value through liquidity provision. Since liquidity provision is often associated with short-term round-trip trading, the significant predictability observed in Panel C supports this interpretation.

6 Conclusion

In this paper, we develop a novel, general method to infer detailed intra-quarter trading of individual mutual funds. Although mutual funds report their holdings once every quarter, they are required to report their returns every day. After a mutual fund executes a trade on day t, its reported portfolio return deviates incrementally from its quarter-end-holdings-based return. This sudden jump in return deviation enables us to infer the transaction date and amount. While this paper focuses on a particular application, our method is highly adaptable and can be applied to a wide range of settings and research questions.

We employ our method to study strategic trading of ESG stocks by mutual funds around the turn of each quarter. We find strong evidence of quarter-end ESG-rating manipulation in the post-2015 period: mutual funds buy high-ESG stocks and sell low-ESG stocks right before quarter ends, and reverse their trades immediately at the beginning of the next quarter. This trading pattern is particularly pronounced among mutual funds near the cutoffs of extreme ESG rating categories, who have the strongest incentives to boost ESG ratings. These trades also affect prices: high-ESG stocks outperform low-ESG stocks right before quarter ends, yet underperform at the beginning of the next quarter.

References

Agarwal, Vikas, D. Gerald Gay, and Leng Ling, 2014, Window Dressing in Mutual Funds, *The Review of Financial Studies* 27, 3133–3170.

Amel-Zadeh, Amir, and George Serafeim, 2018, Why and how investors use esg information: Evidence from a global survey, *Financial Analysts Journal* 74, 87–103.

An, Li, Shiyang Huang, Dong Lou, and Jiahong Shi, 2023, An anatomy of long-short equity funds, *Working Paper*.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2011, Performance of institutional trading desks: An analysis of persistence in trading costs, *The Review of Financial Studies* 25, 557–598.

Ben-David, Itzhak, Francesco Franzoni, Augustin Landier, and Rabih Moussawi, 2013, Do Hedge Funds Manipulate Stock Prices?, *The Journal of Finance* 68, 2383–2434.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song, 2022, Ratings-driven demand and systematic price fluctuations, *The Review of Financial Studies* 35, 2790–2838.

Berg, Florian, Julian F Kölbel, and Roberto Rigobon, 2022, Aggregate Confusion: The Divergence of ESG Ratings, *Review of Finance* 26, 1315–1344.

Bernhardt, Dan, and Ryan J Davies, 2005, Painting the Tape: Aggregate Evidence, *Economics Letter* 89, 306–311.

Bhattacharyya, Sugato, and Vikram Nanda, 2013, Portfolio Pumping, Trading Activity and Fund Performance, *Review of Finance* 17, 885–919.

Bollen, Nicolas P.B., and Veronika K. Pool, 2009, Do hedge fund managers misreport returns? evidence from the pooled distribution, *The Journal of Finance* 64, 2257–2288.

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

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

Carhart, Mark M, Ron Kaniel, David K Musto, and Adam V. Reed, 2002, Leaning for the Tape: Evidence of Gaming Behavior , *The Journal of Finance* 57, 661–693.

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

Chen, Yong, and Wenting Dai, 2023, Seeking green? mutual fund investment in esg stocks, *Working Paper*.

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

Cohen, Lauren, Umit G Gurun, and Quoc H Nguyen, 2022, The esg-innovation disconnect: Evidence from green patenting, *Working Paper*.

Da, Zhi, Pengjie Gao, and Ravi Jagannathan, 2011, Impatient Trading, Liquidity Provision, and Stock Selection by Mutual Funds , *The Review of Financial Studies* 24, 675–720.

Fich, Eliezer M., Viktoriya Lantushenko, and Clemens Sialm, 2024, Institutional trading around m&a announcements, *Working Paper*.

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

Goldstein, Michael A., Paul Irvine, Eugene Kandel, and Zvi Wiener, 2009, Brokerage Commissions and Institutional Trading Patterns, *The Review of Financial Studies* 22, 5175–5212.

Han, Munhee, Sanghyun (Hugh) Kim, and Vikram K. Nanda, 2019, Splitting and shuffling: Institutional trading motives and order submissions across brokers, *Working Paper*.

Harris, Lawrence E., Samuel M. Hartzmark, and David H. Solomon, 2015, Juicing the dividend yield: Mutual funds and the demand for dividends, *Journal of Financial Economics* 116, 433–451.

Hartzmark, Samuel M., and Kelly Shue, 2023, Counterproductive sustainable investing: The impact elasticity of brown and green firms, *Working Paper*.

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

He, Jia, Lilian Ng, and Qinghai Wang, 2004, Quarterly Trading Patterns of Financial Institutions, *Journal of Business* 77, 493–509.

Hu, Gang, Koren M. Jo, Yi Alex Wang, and Jing Xie, 2018, Institutional trading and abel noser data, *Journal of Corporate Finance* 52, 143–167.

Hu, Gang, R. David McLean, Jeffrey Pontiff, and Qinghai Wang, 2013, The Year-End Trading Activities of Institutional Investors: Evidence from Daily Trades, *The Review of Financial Studies* 27, 1593–1614.

Jame, Russell, 2018, Liquidity provision and the cross section of hedge fund returns, *Management Science* 64, 3288–3312.

Judd, Kenneth L., 1998, *Numerical Methods in Economics* . , vol. 1 of *MIT Press Books* (The MIT Press).

Kacperczyk, T. Marcin, Clemens Sialm, and Lu Zheng, 2008, Unobserved Actions of Mutual Funds, *The Review of Financial Studies* 21, 2379–2416.

Lakonishok, Josef, Andrei Shleifer, Richard Thaler, and Robert Vishny, 1991, Window Dressing by Pension Fund Managers, *The American Economic Review* 81, 227–231.

Li, Xudong, Defeng Sun, and Kim-Chuan Toh, 2019, A block symmetric gauss-seidel decomposition theorem for convex composite quadratic programming and its applications, *Mathematical Programming* 175, 395–418.

Lou, Dong, 2012, A flow-based explanation for return predictability, *Review of Financial Studies* 25, 3457–3489.

Meier, Iwan, and Ernst Schaumburg, 2006, Do Funds Window Dress? Evidence for U.S. Domestic Equity Mutual Funds, *Working Paper*.

Ng, Lilian, and Qinghai Wang, 2004, Institutional Trading and The Turn-of-The-Year Effect, *Journal of Financial Economics* 74, 343–366.

Parise, Gianpaolo, and Mirco Rubin, 2024, Green window dressing, *Working Paper*.

Pastor, Lubos, Robert F. Stambaugh, and Lucian A. Taylor, 2021, Sustainable investing in equilibrium, *Journal of Financial Economics* 142, 550–571.

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

Puckett, Andy, and Xuemin (Sterling) Yan, 2011, The interim trading skills of institutional investors, *The Journal of Finance* 66, 601–633.

Table 1: Summary statistics. This table presents summary statistics for our sample of mutual funds over the period from 2015Q1 to 2022Q2. Panel A provides summary statistics for fund characteristics. TNA is the quarter-end total net assets of the fund in millions of dollars. Monthly return is net fund return in percentage. Monthly flow is calculated as $[TNA_{i,m} - TNA_{i,m-1} \times (1 + RET_{i,m})]/TNA_{i,m-1}$ for fund i in month m and is expressed in percentage. Panel B provides the summary statistics of estimated weekly fund trading from our method of inferring intra-quarter trading. For each fund-quarter-week, we calculate the ratio of Total/High-ESG/Low-ESG buy or sell volume divided by total trading volume and express these ratios as percentages. For each quarter t, High-ESG (Low-ESG) stocks are defined as the top (bottom) 200 stocks sorted by the average rank-normalized ESG scores from Morningstar Sustainalytics, MSCI, and Refinitiv in quarter $t-1$.

Panel A: Number of funds and fund-quarters					
Total number of funds	3,529				
Total number of fund-quarters	58,790				
Panel B: Summary statistics of fund characteristics					
	Mean	SD	P5	P50	P95
TNA ($ million)	2,530	11,228	22	407	9,578
Age (years)	17.2	11.7	3	15	38
Number of stocks held	149.3	261.2	25	72	538
Monthly return (%)	0.76	5.52	-8.72	1.10	8.76
Monthly flow (%)	0.29	20.46	-6.21	-0.56	7.34
Expense (%)	0.89	0.41	0.20	0.91	1.53
Turnover	0.68	1.23	0.07	0.44	1.77
Panel C: Summary statistics of estimated weekly trading					
	Mean	SD	P5	P50	P95
Total number of fund-quarter-weeks	692,669				
Total buy / Total trading volume (%)	48.68	30.54	0	48.1	100
Total sell / Total trading volume (%)	51.32	30.54	0	51.9	100
High-ESG buy / Total trading volume (%)	6.64	13.23	0	0	33.27
High-ESG sell / Total trading volume (%)	7.60	14.39	0	0	36.79
Low-ESG buy / Total trading volume (%)	1.65	6.36	0	0	9.75
Low-ESG sell / Total trading volume (%)	1.61	6.34	0	0	9.63

Table 2: High-ESG vs. Low-ESG trading around the turn of the quarters. This table reports mutual fund trading of high-ESG and low-ESG stocks at quarter end and beginning over the period from 2015Q1 to 2022Q2. We present abnormal trading in the 1st, 2nd, and 3rd weeks of the quarter end or beginning, represented by the coefficients $\{b_{E,j}, b_{B,j}\}_{j=1}^{3}$ from the following regression: $y_{i,t,l} = b_0 + \sum_{j=1}^{3} b_{E,j} \times \mathbf{I}_{E,j} + \sum_{j=1}^{3} b_{B,j} \times \mathbf{I}_{B,j} + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}$. The dependent variable, $y_{i,t,l}$, is the trading volume of high-ESG or low-ESG stocks divided by the total trading volume for fund i in quarter t and week l expressed in percentage. The trading volume of high-ESG or low-ESG stocks is estimated using our proposed method. High-ESG (low-ESG) stocks are defined as the top (bottom) 200 stocks, sorted by the average rank-normalized ESG scores from Sustainalytics, MSCI, and Refinitiv in the previous quarter. $\mathbf{I}_{E,j}$ or $\mathbf{I}_{B,j}$ is a 0/1 indicator for the jth week of the quarter end or beginning. To control for the imbalance between buy and sell volumes, $buy\_ratio_{i,t,l}$ is defined as the total buying volume divided by the total trading volume for fund i in quarter t and week l. $\alpha_{i,t}$ is fund × year × quarter fixed effects. In Panel B, we conduct placebo tests around non-quarter month ends, i.e., month ends except March, June, September, and December. t-statistics, shown in brackets, are double clustered at both the fund and year-quarter levels. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: high-ESG vs. low-ESG trading around quarter ends

	High-ESG trading			Low-ESG trading			High - Low
	Net	Buy	Sell	Net	Buy	Sell	Net
Ending							
End. 1st week	0.611***	0.460***	-0.151**	-0.169**	-0.205***	-0.036	0.779***
	[3.64]	[3.35]	[-2.16]	[-2.23]	[-3.49]	[-0.76]	[3.98]
End. 2nd week	0.251	0.173	-0.078	-0.083	-0.105**	-0.022	0.334**
	[1.68]	[1.58]	[-1.02]	[-1.44]	[-2.07]	[-0.50]	[2.05]
End. 3rd week	0.190	0.075	-0.115	0.023	-0.043	-0.067	0.167
	[1.19]	[0.70]	[-1.35]	[0.39]	[-0.88]	[-1.68]	[0.92]
Beginning							
Beg. 1st week	-0.489**	-0.155	0.334**	-0.057	-0.074	-0.017	-0.432*
	[-2.32]	[-1.45]	[2.06]	[-1.06]	[-1.59]	[-0.25]	[-1.91]
Beg. 2nd week	-0.196	-0.055	0.142	-0.146**	-0.101**	0.045	-0.050
	[-0.99]	[-0.51]	[1.03]	[-2.46]	[-2.48]	[0.59]	[-0.23]
Beg. 3rd week	-0.184	-0.108	0.075	-0.063	0.015	0.078	-0.120
	[-1.08]	[-1.34]	[0.52]	[-0.79]	[0.27]	[0.90]	[-0.54]
Ending - Beginning							
End. 1st -	1.100***	0.615***	-0.485***	-0.112	-0.130*	-0.019	1.211***
Beg. 1st	[3.60]	[3.08]	[-3.02]	[-1.11]	[-1.84]	[-0.25]	[3.48]

Panel B: high-ESG vs. low-ESG trading around non-quarter month ends (placebo)

	Net high-ESG trading	Net low-ESG trading	High - Low
Ending - Beginning			
End. 1st -	-0.080	-0.052	-0.028
Beg. 1st	[-0.57]	[-1.10]	[-0.19]

Table 3: ESG window dressing and sustainability ratings. This table reports mutual funds' ESG window dressing intensity grouped by their one-month-lagged sustainability rating. Due to rating data availability, the sample period is from 2019Q1 to 2022Q2. We obtain fund sustainability rating data from Morningstar, where funds are classified into 5 categories based on the percentile ranking of their sustainability scores calculated from underlying fund holdings. A fund is marked 5 globes and rated as "High" if percentage ranking is above 90%; 4 globes and rated as "Above Average" if percentage ranking is between 67.5% and 90%; 3 globes and rated as "Average" if percentage ranking is between 32.5% and 67.5%; 2 globes and rated as "Below Average" if percentage ranking is between 10% and 32.5%; 1 globe and rated as "Low" if percentage ranking is below 10%. To examine within-category variation, we further split each rating category into "Lower half" and "Upper half" based on the percentage ranking. Since our focus is on funds' incentives to manipulate Morningstar ratings, we classify stocks as high- or low-ESG based solely on Morningstar Sustainalytics for this particular test. We define high-ESG (low-ESG) stocks as the top 20% (bottom 20%) stocks in the cross-section. We report ESG window dressing intensity, which is the coefficient difference, $b_{E,x} - b_{B,x}$, for each fund sustainability rating category x, from the following regression: $y_{i,t,l} = b_0 + \left[\sum_{x=1L,\ldots,5U} \mathbf{I}_{\{\text{category } x\}} \times (b_{E,x} \times \mathbf{I}_E + b_{B,x} \times \mathbf{I}_B)\right] + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}$. The dependent variable, $y_{i,t,l}$, is fund i's high-ESG-minus-low-ESG net trading (buy minus sell) divided by total trading volume in quarter t and week l. $\mathbf{I}_E$ and $\mathbf{I}_B$ are dummies indicating the last week and the first week of a quarter, respectively. The results are expressed in percentage points. t-statistics, shown in brackets, are calculated via bootstrap with 500 replications. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Split each rating category		
Sustainability rating	Lower half (L)	Upper half (U)	U - L
1 Globe	1.008	4.708***	3.700*
	[0.66]	[3.24]	[1.76]
2 Globes	0.563	1.173	0.611
	[0.72]	[1.54]	[0.58]
3 Globes	1.333**	1.904***	0.572
	[2.21]	[3.14]	[0.69]
4 Globes	1.848**	5.660***	3.813***
	[2.35]	[6.35]	[3.15]
5 Globes	4.003***	3.058*	-0.945
	[2.97]	[1.91]	[-0.46]

Table 4: High-ESG vs. Low-ESG stock returns. This table reports cumulative risk-adjusted returns of high-ESG vs. low-ESG stocks at quarter ending and beginning. The sample period is from 2015Q1 to 2022Q2. At each quarter end, high-ESG (low-ESG) stocks are defined as the top (bottom) 200 stocks sorted by the average rank-normalized ESG scores from Sustainalytics, MSCI, and Refinitiv in the previous month. Risk-adjusted returns are calculated based on the Fama-French three-factor model. We estimate beta using monthly returns in a 60-month rolling window. A valid beta estimation requires at least 20 observations and we cross-sectionally winsorize beta at the 1st and 99th percentiles. In Panel A, columns are grouped by high-ESG stocks' returns, low-ESG stocks' returns, and their differences. Among each stock category, we construct portfolios using an equally-weighted (labeled as EW) or value-weighted (labeled as VW) scheme. Rows are grouped by quarter ending, beginning, and their differences. Among each period category, we report cumulative risk-adjusted returns with different windows. Specifically, let d denote the last trading day of a quarter. The quarter ending corresponds to the window of $[d-D+1, d]$, and the quarter beginning corresponds to the window of $[d+1, d+D]$, where window length $D \in 1, 3, 5$ day(s). In Panel B, we conduct a placebo test around non-quarter month end, i.e., month end except Mar/Jun/Sep/Dec. All returns are expressed in percentage points. t-statistics, shown in brackets, are computed based on standard errors with Newey-West corrections of 8 lags (quarters). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: high-ESG vs. low-ESG stock returns around quarter ends

	High-ESG stocks		Low-ESG stocks		High - Low	
	EW	VW	EW	VW	EW	VW
Ending						
1-day window	-0.018	0.040	-0.034	-0.046	0.016	0.087
	[-0.39]	[1.58]	[-0.84]	[-0.74]	[0.22]	[1.06]
3-day window	0.159***	0.098**	-0.191	-0.268	0.351**	0.366
	[4.78]	[2.72]	[-1.33]	[-1.01]	[2.42]	[1.30]
5-day window	0.162***	0.167***	-0.516	-0.616	0.677**	0.783**
	[4.35]	[4.06]	[-1.64]	[-1.67]	[2.26]	[2.07]
Beginning						
1-day window	-0.131**	-0.103***	0.129*	0.200*	-0.260**	-0.303**
	[-2.11]	[-3.44]	[1.92]	[2.02]	[-2.60]	[-2.70]
3-day window	-0.179**	-0.140***	0.169	0.121	-0.349**	-0.261
	[-2.66]	[-3.38]	[1.49]	[0.79]	[-2.28]	[-1.50]
5-day window	-0.205*	-0.171*	0.045	0.112	-0.249	-0.283
	[-1.97]	[-1.96]	[0.19]	[0.57]	[-1.09]	[-1.20]
Ending - Beginning						
1-day window	0.113*	0.143***	-0.163**	-0.246*	0.276***	0.389**
	[1.84]	[4.23]	[-2.42]	[-1.85]	[2.82]	[2.66]
3-day window	0.339***	0.238***	-0.361*	-0.388	0.700***	0.627
	[3.80]	[4.29]	[-1.92]	[-1.09]	[2.99]	[1.61]
5-day window	0.366***	0.338***	-0.560	-0.729	0.927**	1.067*
	[3.87]	[4.64]	[-1.64]	[-1.49]	[2.37]	[2.04]

(continued)

(continued)

Panel B: high-ESG vs. low-ESG stock returns around non-quarter month ends (placebo)						
	1-day window		3-day window		5-day window	
	EW	VW	EW	VW	EW	VW
Ending - Beginning						
High - Low	-0.160	-0.162	-0.001	0.036	0.031	-0.009
	[-1.19]	[-1.08]	[-0.01]	[0.20]	[0.22]	[-0.04]

Table 5: Fund heterogeneity in ESG window dressing. This table shows the abnormal ending-minus-beginning net high-ESG-minus-low-ESG trading for each fund category x during the sample period 2015Q1 to 2022Q2. We conduct the following regression: $y_{i,t,l} = b_0 + \left[\sum_x \mathbf{I}_{\{\text{category } x\}} \times (b_{E,x} \times \mathbf{I}_E + b_{B,x} \times \mathbf{I}_B)\right] + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \varepsilon_{i,t,l}$, and report the coefficient difference $b_{E,x} - b_{B,x}$ for each fund category x. The dependent variable, $y_{i,t,l}$, is net (i.e., buy minus sell) high-ESG-minus-low-ESG trading volume divided by total trading volume for fund i in quarter t and week l. High and low ESG stocks are defined the same as before. We consider three aspects of fund heterogeneity: ESG funds, past fund performance, and state political leanings. ESG funds are defined using Morningstar's classification "Sustainable Investment Overall = Yes" based on fund prospectus or other regulatory filings. Fund past performance is measured by information ratio, i.e., Fama-French-3-factor alpha divided by residual volatility, estimated using monthly returns and a 60-month rolling window. High performance (low performance) indicates cross-sectionally top 20% (bottom 20%) funds, and middle performance indicates the remaining 60% funds. For state political leanings, we define Democratic states (Republican states) as the states that voted for the Democratic Party (Republican Party) in all of the 2004, 2008, 2012, 2016, and 2020 presidential elections. According to the definition, 17 states are classified as Democratic States: California, Connecticut, D.C., Delaware, Hawaii, Illinois, Massachusetts, Maryland, Maine, Minnesota, New Hampshire, New Jersey, New York, Oregon, Rhode Island, Vermont, Washington; and 20 states are classified as Republican States: Alaska, Alabama, Arkansas, Idaho, Kansas, Kentucky, Louisiana, Missouri, Mississippi, Montana, North Dakota, Nebraska, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, West Virginia, Wyoming. The dependent variable is multiplied by 100. t-statistics, shown in brackets, are double clustered at the fund and year-quarter levels. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dep. variable: net high ESG - net low ESG trading			
	ESG funds	Others		ESG funds - Others
End. - Beg.	5.713*** [3.56]	1.074*** [3.14]		4.639*** [3.01]
	High performance	Middle performance	Low performance	High - Low
End. - Beg.	2.373*** [4.38]	1.093*** [3.18]	0.323 [0.58]	2.050*** [3.16]
	Democratic states	Neutral states	Republican states	Dem. - Rep.
End. - Beg.	1.500*** [3.88]	0.873** [2.12]	0.012 [0.02]	1.487** [2.60]

Table 6: Time series variation in ESG window dressing. This table shows the abnormal ending-minus-beginning net high-ESG-minus-low-ESG trading for each period p during the sample period 2015Q1 to 2022Q2. We conduct the following regression: $y_{i,t,l} = b_0 + \left[\sum_p \mathbf{I}_{\{\text{period } p\}} \times (b_{E,p} \times \mathbf{I}_E + b_{B,p} \times \mathbf{I}_B)\right] + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \varepsilon_{i,t,l}$, and report the coefficient difference $b_{E,p} - b_{B,p}$ for each period p. The dependent variable, $y_{i,t,l}$, is net (i.e., buy minus sell) high-ESG-minus-low-ESG trading volume divided by total trading volume for fund i in quarter t and week l. High and low ESG stocks are defined the same as before. We consider two aspects of time series variations: ESG fund flow and Google index with the keyword "sustainability". ESG fund flow is defined as aggregate percentage flow to high sustainability rating funds (i.e., five-globe funds) minus aggregate percentage flow to low sustainability rating funds (i.e., one-globe funds). Aggregate percentage flows grouped by sustainability rating (i.e., total flow divided by lagged total TNA for funds in the same rating group) are from Morningstar. In the table, the "high ESG fund flow" indicates the top decile periods ranked by past-6-month-moving-average ESG flow. The "high Google index" indicates the top decile periods ranked by the past-6-month-moving-average Google index. The dependent variable is multiplied by 100. t-statistics, shown in brackets, are double clustered at the fund and year-quarter levels. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dep. variable: net high ESG - net low ESG trading		
	High ESG fund flow	Other periods	High - Others
End. - Beg.	3.225***	1.056***	2.169**
	[4.00]	[3.00]	[2.49]
	High Google index (keyword: sustainability)	Other periods	High - Others
End. - Beg.	2.514***	1.048***	1.466*
	[3.50]	[2.92]	[1.96]

Table 7: Stock heterogeneity in ESG window dressing. This table shows the abnormal ending-minus-beginning net high-ESG-minus-low-ESG trading for each stock category s during the sample period 2015Q1 to 2022Q2. We conduct the following regression for each stock category s: $y^s_{i,t,l} = b_{0,s} + b_{E,s} \times \mathbf{I}_E + b_{B,s} \times \mathbf{I}_B + \gamma^s \times buy\_ratio_{i,t,l} + \alpha^s_{i,t} + \varepsilon^s_{i,t,l}$, and report the coefficient difference $b_{E,s} - b_{B,s}$ for each stock category s. The dependent variable, $y^s_{i,t,l}$, is net (i.e., buy minus sell) high-ESG-minus-low-ESG trading volume in stock category s divided by total trading volume for fund i in quarter t and week l. High and low ESG stocks are defined the same as before. To define the stock category s, at each cross-section, we sort high ESG or low ESG stocks into two equal groups based on stock characteristics (e.g., bid-ask spread), and obtain four sets of stocks (e.g., high-ESG-high-baspread, etc.) We consider two stock-level variations: bid-ask spread and idiosyncratic volatility. The bid-ask spread is calculated using the average of daily (Ask-Bid)/((Ask+Bid)/2) in the previous month. Idiosyncratic volatility is the residual standard deviation of the Fama-French 3 factor model, estimated using weekly returns and a three-year rolling window. The dependent variable is multiplied by 100. t-statistics, shown in brackets, are double clustered at the fund and year-quarter levels. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dep. variable: net high ESG - net low ESG trading		
	Low bid-ask spread	High bid-ask spread	Low - High
End. - Beg.	1.183***	0.005	1.178***
	[5.65]	[0.02]	[4.39]
	Low idio. volatility	High idio. volatility	Low - High
End. - Beg.	1.412***	-0.217	1.630***
	[6.50]	[-1.16]	[7.29]

Table 8: ESG window dressing and future fund flows. This table reports the results of fund-quarter panel regressions of future fund flows on ESG trading. The sample period is from 2015Q1 to 2022Q2. Fund monthly flow is calculated by $[TNA_{i,m} - TNA_{i,m-1} \times (1 + RET_{i,m})]/TNA_{i,m-1}$ for fund i in month m. The dependent variable is cumulative fund flow in quarter $t + 1$. The key explanatory variable, net ESG trading, is buy-minus-sell ESG trading volume divided by total trading volume in the last week of quarter t. ESG stocks are defined the same as before. Past fund flow is fund flow in the last month of quarter t. Past fund α is estimated based on the Fama-French three-factor model using monthly returns and a 60-month rolling window. We cross-sectionally winsorize fund flow and fund α at the 1st and 99th percentiles. Fund flow is expressed in percentage points. Fund α is annualized and also in percentage points. The regression includes fund fixed effects and year-quarter fixed effects. t-statistics, shown in brackets, are double clustered at the fund and year-quarter levels. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dep. variable: fund flow in the next quarter		
Net ESG trading	2.438***	1.234***	1.214***
	[7.01]	[3.88]	[3.81]
Past fund flow		0.328***	0.324***
		[8.90]	[8.87]
Past fund α			0.194***
			[4.18]
Observations	56,461	56,461	56,461
R-squared	0.212	0.234	0.235
Fund FE	Yes	Yes	Yes
Year $\times$ Qtr FE	Yes	Yes	Yes

Table 9: ESG window dressing and fund performance. This table reports the results of fund-quarter-week panel regressions of fund performance on ESG trading. The sample period is from 2015Q1 to 2022Q2. The dependent variable is the cumulative excess return (column 1) or cumulative CAPM adjusted return (column 2) in each week. CAPM β is estimated using monthly returns from a five-year rolling window. A valid estimation requires at least 20 observations. Otherwise, we set the CAPM β to be one. The right-hand variables are the indicator of the last week $\mathbf{I}_{\{\text{last week}\}}$ in that quarter, and its interaction with net ESG trading. Net ESG trading is the buy-minus-sell ESG trading volume divided by total trading volume in the last week. ESG stocks are defined the same as before. The dependent variables are expressed in percentage points. t-statistics, shown in brackets, are double clustered at the fund and year-quarter-week levels. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Dependent variable	
	Excess return	CAPM adjusted return
$\mathbf{I}_{\{\text{last week}\}}$	0.208	0.106
	[0.48]	[1.38]
$\mathbf{I}_{\{\text{last week}\}}$ × Net ESG trading	-0.315*	-0.147*
	[-1.83]	[-1.67]
Observations	673,961	673,961
R-squared	0.111	0.087
Fund × Year × Qtr FE	Yes	Yes

Table 10: Return gap decomposition and fund return predictability. This table shows the portfolio sorting results using different components of the return gap. The sample period is from 2000 to 2022Q2 and we form fund portfolios starting from 2001. For each month t, let R_t^{fund}, R_t^B, $\hat{R}_t^d$, and $\hat{R}_t^r$ denote fund actual return, beginning-portfolio hypothetical return, fitted return from a model only allows directional trades, and fitted return from a model allows both directional and round-trip trades, respectively. The sorting variables are the KSZ (2008) return gap (i.e., past 12-month average of $R_t^{fund} - R_t^B$) in Panel A, directional-trade return gap (i.e., past 12-month average of $\hat{R}_t^d - R_t^B$) in Panel B, round-trip trades return gap (i.e., past 12-month average of $\hat{R}_t^r - \hat{R}_t^d$) in Panel C, and residual return gap (i.e., past 12-month average of $R_t^{fund} - \hat{R}_t^r$) in Panel D. We sort funds into 10 groups at the end of each quarter based on the return gap, with a lag of at least 3 months to ensure that the information is publicly available. That is, at the end of quarter t, each component of the return gaps is averaged from month $t-14$ to $t-3$. In each panel, we report the average excess return and alpha relative to factor models of CAPM, FF3, CH4, FF5, and FF5 + MOM. t-statistics, shown in brackets, are computed based on standard errors with Newey-West corrections of 6 lags (months). *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

Panel A: Sort by past 12-month return gap (KSZ 2008)

	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	10-1
Average	0.557*	0.658**	0.699**	0.697**	0.659**	0.675**	0.696**	0.701**	0.740**	0.747**	0.190**
	[1.71]	[2.04]	[2.22]	[2.23]	[2.11]	[2.16]	[2.22]	[2.23]	[2.25]	[2.11]	[2.16]
CAPM	-0.12	-0.022	0.03	0.029	-0.004	0.011	0.03	0.026	0.054	-0.002	0.118
	[-1.43]	[-0.30]	[0.37]	[0.45]	[-0.07]	[0.25]	[0.60]	[0.58]	[0.81]	[-0.02]	[1.30]
FF3	-0.143**	-0.042	0.009	0.015	-0.018	-0.002	0.018	0.011	0.033	-0.033	0.109*
	[-2.20]	[-0.93]	[0.18]	[0.33]	[-0.54]	[-0.07]	[0.43]	[0.28]	[0.53]	[-0.53]	[1.69]
CH4	-0.124*	-0.045	-0.003	0.011	-0.025	-0.005	0.024	0.017	0.037	-0.028	0.096
	[-1.90]	[-0.96]	[-0.05]	[0.22]	[-0.70]	[-0.15]	[0.56]	[0.44]	[0.61]	[-0.48]	[1.43]
FF5	-0.089	-0.043	-0.032	0.001	-0.035	-0.003	0.04	0.045	0.099*	0.094*	0.183***
	[-1.38]	[-0.95]	[-0.77]	[0.03]	[-1.16]	[-0.10]	[1.05]	[1.44]	[1.74]	[1.80]	[2.88]
FF5 + MOM	-0.084	-0.044	-0.035	0	-0.037	-0.004	0.042	0.046	0.098*	0.092*	0.175***
	[-1.29]	[-0.96]	[-0.84]	[0.00]	[-1.21]	[-0.13]	[1.07]	[1.46]	[1.74]	[1.74]	[2.71]

(continued)

(continued)

Panel B: Sort by past 12-month directional-trade return gap

	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	10-1
Average	0.609*	0.675**	0.705**	0.695**	0.678**	0.665**	0.670**	0.718**	0.707**	0.706**	0.097
	[1.87]	[2.10]	[2.25]	[2.19]	[2.20]	[2.13]	[2.14]	[2.23]	[2.16]	[2.00]	[1.09]
CAPM	-0.071	0.001	0.042	0.025	0.012	-0.001	0.001	0.043	0.018	-0.038	0.033
	[-0.82]	[0.02]	[0.44]	[0.43]	[0.23]	[-0.02]	[0.02]	[0.70]	[0.33]	[-0.46]	[0.35]
FF3	-0.096	-0.02	0.02	0.01	-0.002	-0.014	-0.012	0.028	0	-0.068	0.028
	[-1.60]	[-0.52]	[0.32]	[0.27]	[-0.05]	[-0.46]	[-0.34]	[0.46]	[-0.01]	[-0.96]	[0.40]
CH4	-0.076	-0.02	0.009	0.007	-0.007	-0.014	-0.006	0.031	0.002	-0.068	0.009
	[-1.25]	[-0.49]	[0.15]	[0.17]	[-0.20]	[-0.45]	[-0.18]	[0.53]	[0.05]	[-1.01]	[0.13]
FF5	-0.044	-0.029	-0.028	-0.006	-0.018	-0.007	0.01	0.066	0.061	0.07	0.114*
	[-0.72]	[-0.70]	[-0.66]	[-0.14]	[-0.57]	[-0.25]	[0.34]	[1.15]	[1.38]	[1.18]	[1.84]
FF5 + MOM	-0.038	-0.029	-0.031	-0.007	-0.019	-0.007	0.012	0.067	0.06	0.066	0.104*
	[-0.63]	[-0.68]	[-0.73]	[-0.15]	[-0.61]	[-0.25]	[0.38]	[1.15]	[1.34]	[1.10]	[1.68]

Panel C: Sort by past 12-month round-trip-trade return gap

	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	10-1
Average	0.600*	0.630*	0.649**	0.654**	0.680**	0.683**	0.692**	0.771**	0.729**	0.773**	0.173**
	[1.77]	[1.91]	[2.01]	[2.11]	[2.19]	[2.22]	[2.19]	[2.45]	[2.29]	[2.30]	[2.48]
CAPM	-0.094	-0.068	-0.03	-0.012	0.019	0.023	0.024	0.102	0.045	0.058	0.152**
	[-1.23]	[-1.16]	[-0.55]	[-0.23]	[0.39]	[0.36]	[0.43]	[1.17]	[0.78]	[0.77]	[2.07]
FF3	-0.116	-0.088	-0.046	-0.028	0.005	0.008	0.009	0.082	0.026	0.031	0.147**
	[-1.56]	[-1.59]	[-0.98]	[-0.71]	[0.15]	[0.17]	[0.27]	[1.32]	[0.71]	[0.65]	[2.07]
CH4	-0.116	-0.088	-0.048	-0.033	0.005	0.006	0.012	0.078	0.031	0.048	0.165**
	[-1.59]	[-1.59]	[-1.02]	[-0.81]	[0.13]	[0.14]	[0.33]	[1.32]	[0.79]	[1.03]	[2.42]
FF5	-0.024	-0.031	-0.024	-0.027	0.015	0.007	0.008	0.052	0.028	0.099**	0.123**
	[-0.35]	[-0.62]	[-0.57]	[-0.76]	[0.41]	[0.17]	[0.26]	[1.03]	[0.80]	[2.18]	[2.07]
FF5 + MOM	-0.027	-0.033	-0.025	-0.029	0.014	0.007	0.009	0.051	0.03	0.104**	0.131**
	[-0.39]	[-0.64]	[-0.60]	[-0.81]	[0.40]	[0.16]	[0.29]	[1.02]	[0.83]	[2.34]	[2.35]

(continued)

(continued)

Panel D: Sort by past 12-month residual return gap

	1 (Low)	2	3	4	5	6	7	8	9	10 (High)	10-1
Average	0.536	0.603*	0.637**	0.692**	0.698**	0.692**	0.705**	0.737**	0.712**	0.816**	0.280***
	[1.62]	[1.85]	[2.01]	[2.21]	[2.22]	[2.17]	[2.22]	[2.32]	[2.23]	[2.43]	[3.26]
CAPM	-0.145*	-0.086	-0.038	0.019	0.026	0.018	0.03	0.067	0.04	0.102	0.247***
	[-1.95]	[-1.45]	[-0.69]	[0.32]	[0.46]	[0.37]	[0.59]	[0.79]	[0.72]	[1.19]	[2.63]
FF3	-0.163**	-0.103*	-0.055	0.003	0.01	0	0.013	0.045	0.021	0.076	0.239***
	[-2.27]	[-1.94]	[-1.19]	[0.06]	[0.23]	[0.01]	[0.40]	[0.80]	[0.63]	[1.26]	[2.71]
CH4	-0.144**	-0.105**	-0.06	0.003	0.003	-0.003	0.014	0.047	0.027	0.077	0.221***
	[-2.07]	[-1.98]	[-1.30]	[0.07]	[0.06]	[-0.07]	[0.39]	[0.88]	[0.77]	[1.28]	[2.63]
FF5	-0.061	-0.053	-0.035	0.013	0.013	0.006	0.038	0.019	0.044	0.092	0.152**
	[-0.95]	[-1.03]	[-0.82]	[0.27]	[0.33]	[0.17]	[1.21]	[0.48]	[1.34]	[1.60]	[2.07]
FF5 + MOM	-0.057	-0.055	-0.037	0.013	0.011	0.004	0.037	0.02	0.046	0.092	0.148**
	[-0.88]	[-1.08]	[-0.88]	[0.26]	[0.26]	[0.13]	[1.18]	[0.51]	[1.36]	[1.59]	[2.00]

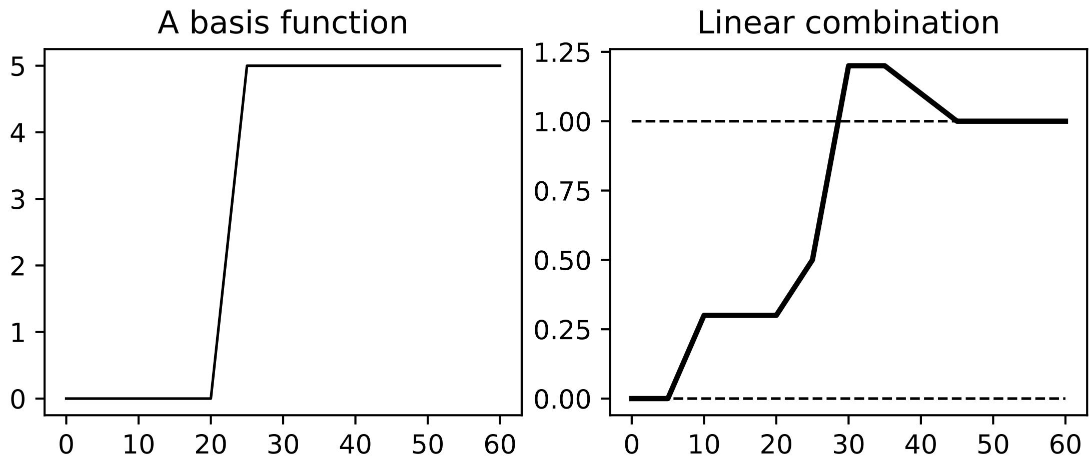

Figure 1: Piecewise linear parameterization. This figure illustrates a piecewise linear function representing the dynamics of daily holdings. The left panel shows an example of a basis function, while the right panel presents a piecewise linear function constructed from a linear combination of these basis functions.

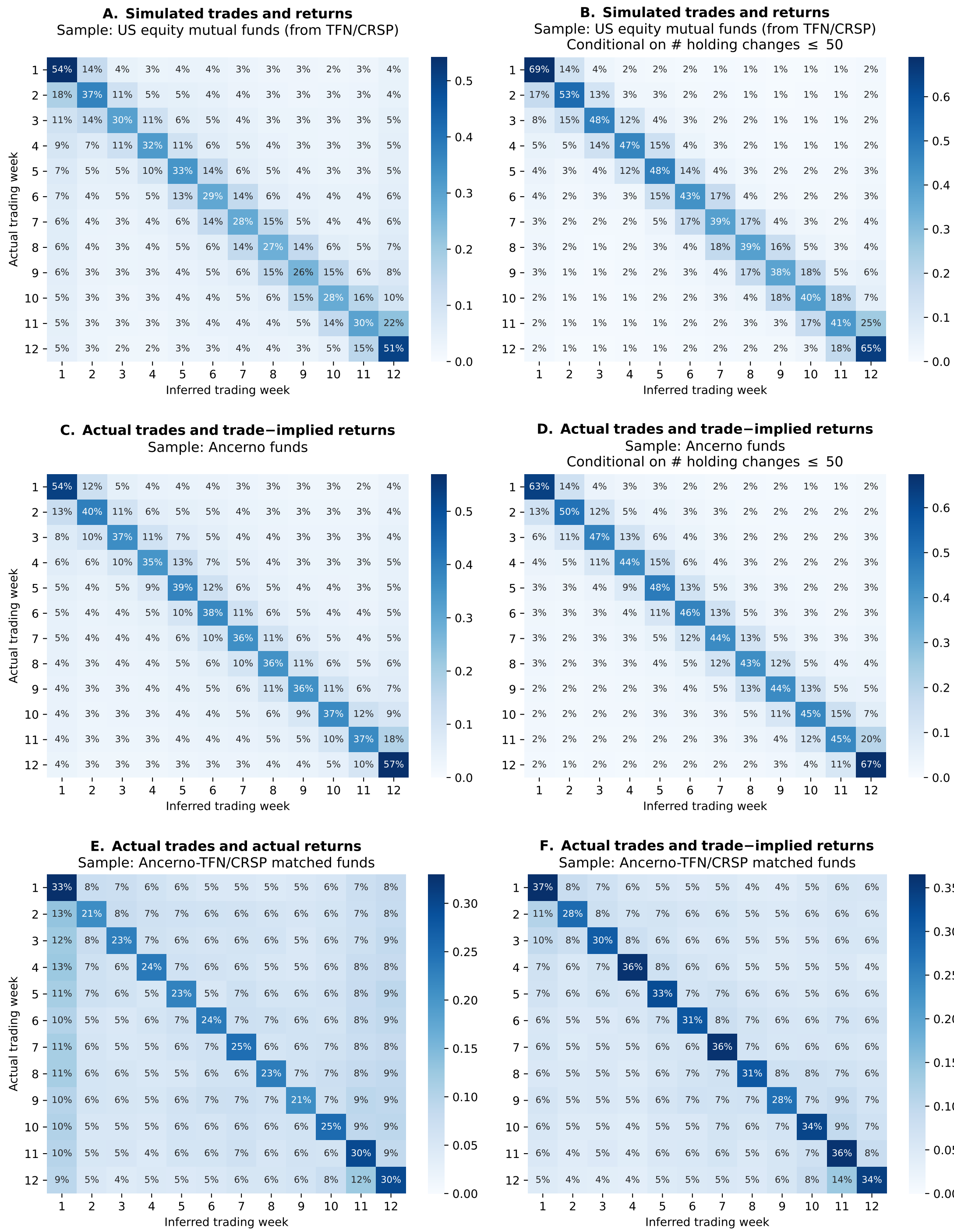

Figure 2: Performance evaluation. We use heatmaps to depict the performance of our trading inference method. The heatmap H is a 12×12 matrix, where each element (i, j) indicates the probability that the algorithm classifies a trade from week i (the actual trading week) to week j (the inferred trading week). Darker colors indicate higher probabilities and lighter colors indicate lower probabilities. A perfect estimation would result in all diagonal elements being one and all off-diagonal elements being zero. As a benchmark, a random-guessing estimation would result in all elements being $1/12 = 8.3\%$.

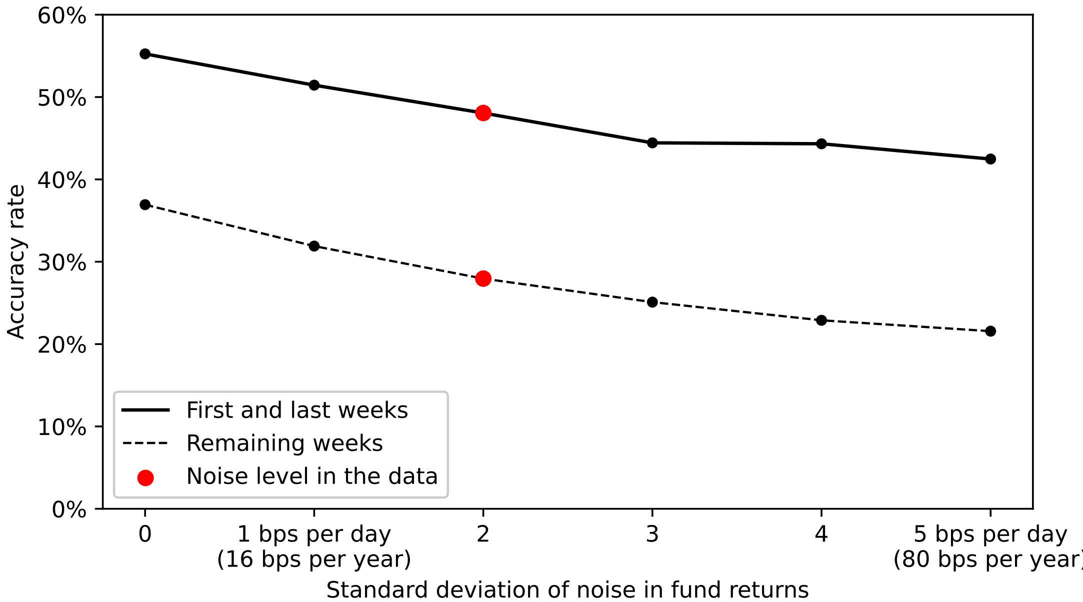

Figure 3: Performance evaluation and unobserved factors in fund returns. We examine the impact of unobserved factors in fund returns on the accuracy of our proposed method by adding simulated noise. We simulate daily noise from a normal distribution with mean zero and standard deviation ranging from 1-5 bps per day (or 16-80 bps per year). As a reference, the level of noise in the data is around 2 bps per day (or 32 bps per year), calculated using the deviation between the actual and trade-implied return from the Ancerno-TFN/CRSP matched sample. The x-axis represents the standard deviation of noise in fund returns, and the y-axis represents the accuracy rate (i.e., the average of diagonal elements in a heatmap). For each standard deviation of noise, we calculate the accuracy rates by randomly drawing 1,000 fund-quarters from our second evaluation sample (i.e., the Ancerno sample with actual trades and trade-implied returns), adding simulated noise to fund returns, inferring trades, and then evaluating. The solid (dashed) line corresponds to the performance in the first and last weeks (remaining weeks). The red dots correspond to the noise level in the data.

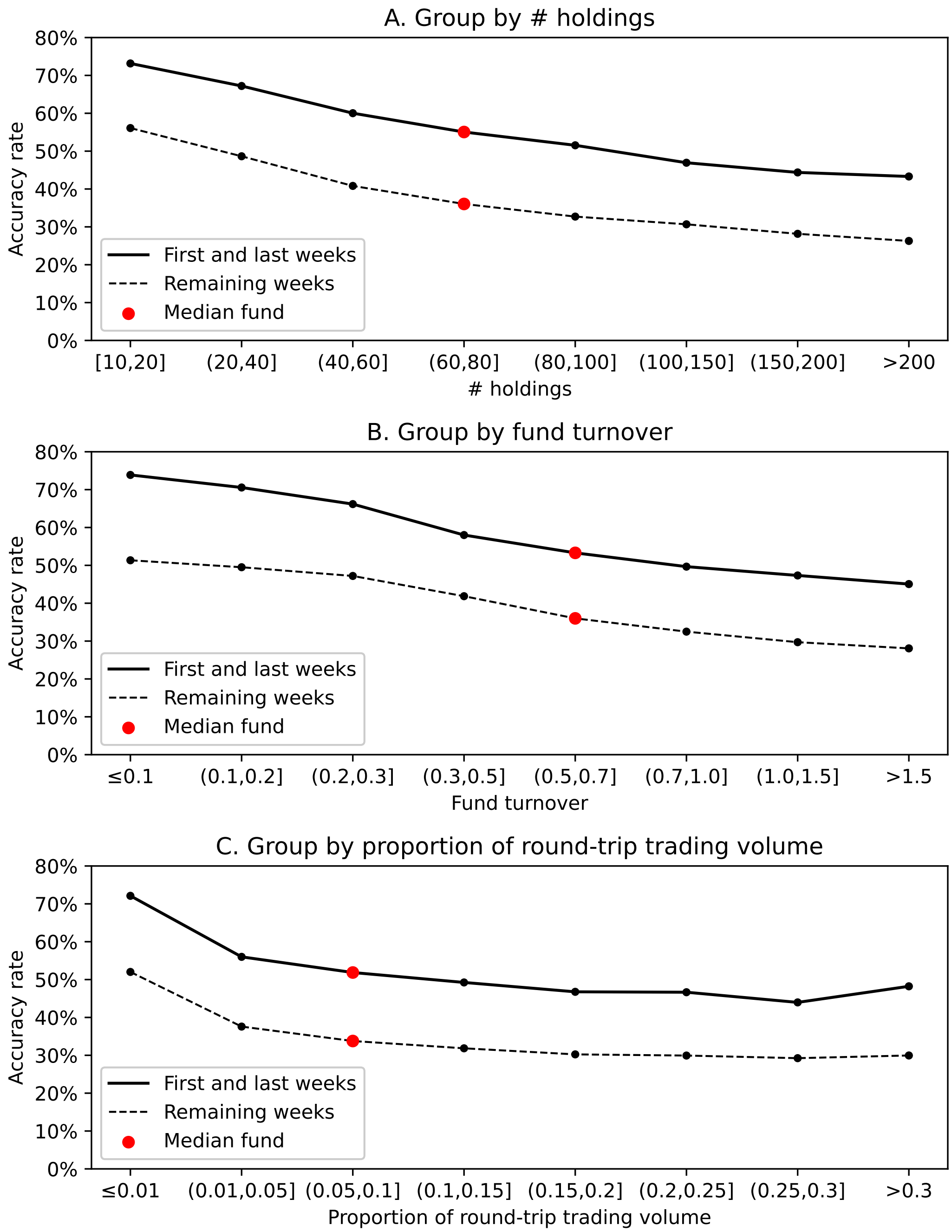

Figure 4: Performance evaluation and fund characteristics. We examine how fund characteristics affect performance by sorting on characteristics and evaluating within each group. The x-axis represents the range for fund characteristics, and the y-axis represents the accuracy rate (i.e., the average of diagonal elements in a heatmap). We conduct the analysis using our second evaluation sample (i.e., the Ancerno sample with actual trades and trade-implied returns). The solid (dashed) lines correspond to the performance in the first and last weeks (remaining weeks). The red dots correspond to the median value of characteristics.

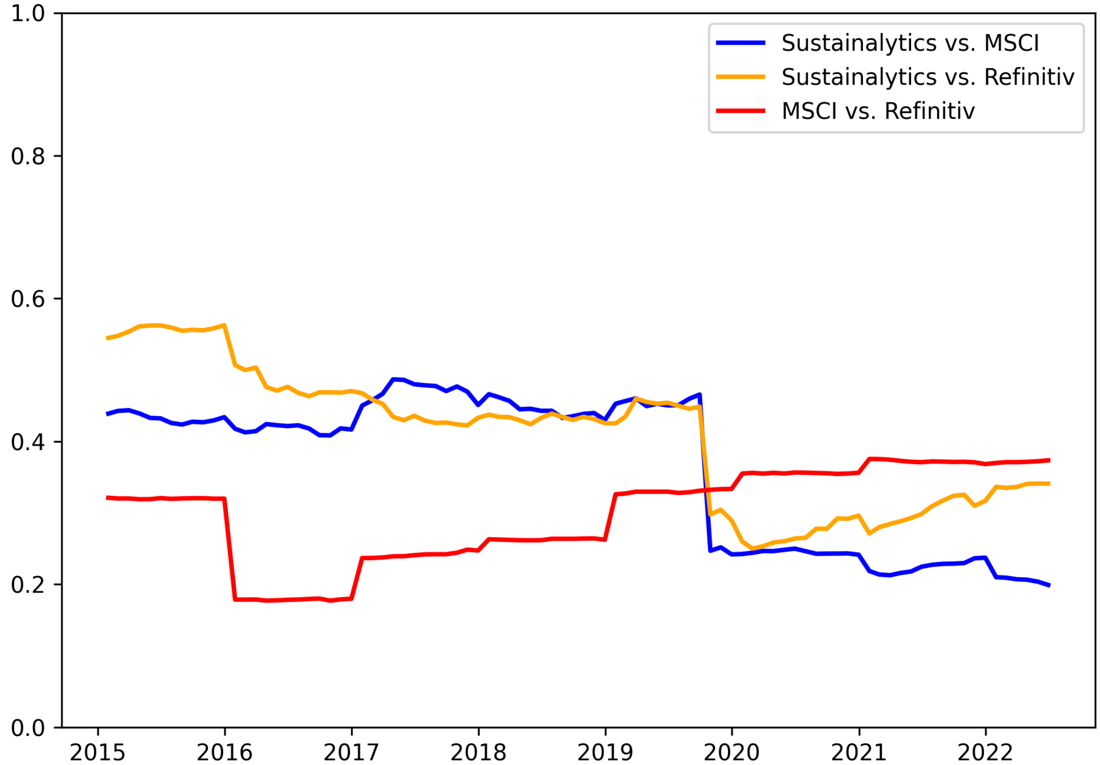

Figure 5: Correlations among different ESG ratings. This figure displays the time series of Pearson correlations among three stock ESG ratings—Morningstar Sustainalytics, MSCI, and Refinitiv—from 2015Q1 to 2022Q2. The correlations are based on percentage rankings of ESG scores for comparability. The blue, orange, and red lines represent the correlations between Sustainalytics and MSCI, Sustainalytics and Refinitiv, and MSCI and Refinitiv, respectively. The average correlation across all pairs and the entire period is 0.357.

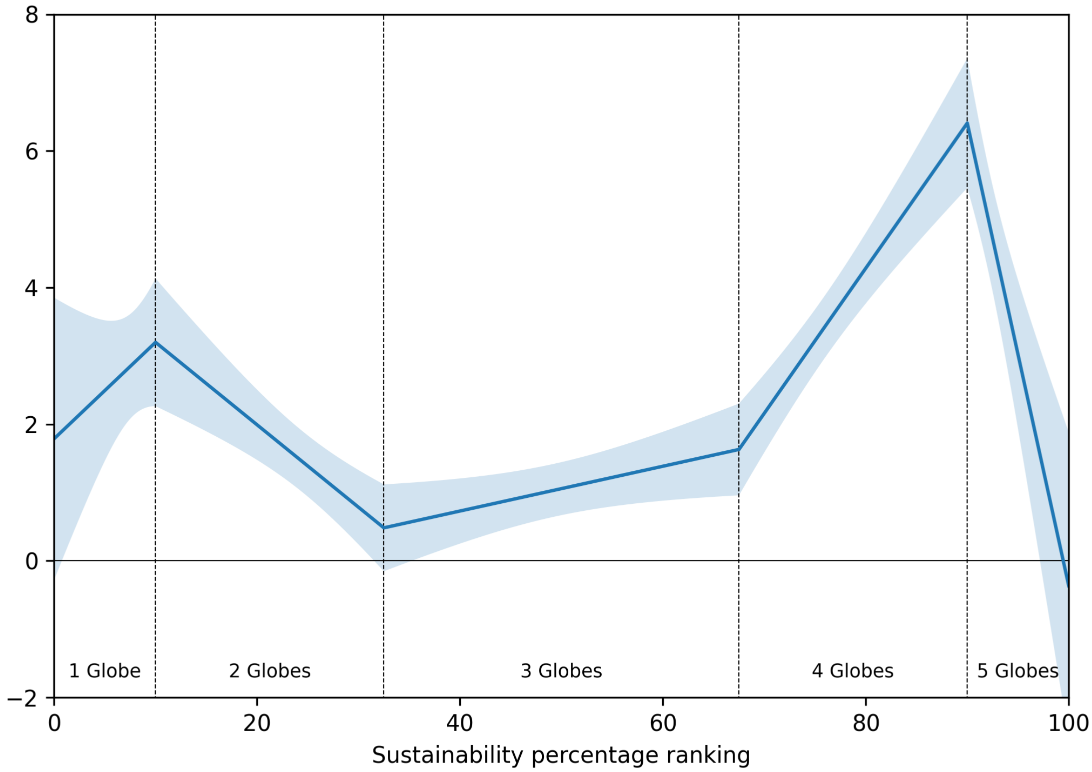

Figure 6: ESG window dressing as a function of sustainability rating. This figure shows the ESG window dressing intensity, $f_E(p) - f_B(p)$, plotted against the one-month-lagged sustainability percentage ranking p. The values of $f_E(p) - f_B(p)$ are estimated from the regression $y_{i,t,l} = b_0 + f_E(p) \times \mathbf{I}_E + f_B(p) \times \mathbf{I}_B + g(p) \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}$, where $f_E(p)$, $f_B(p)$, and $g(p)$ are piecewise linear functions partitioned by rating categories. Sample period, fund sustainability rating, and definition of high- and low- ESG stocks are defined in the same way as in Table 3. The shaded area represents a one-standard-deviation error band, with standard errors computed via bootstrap with 500 replications.

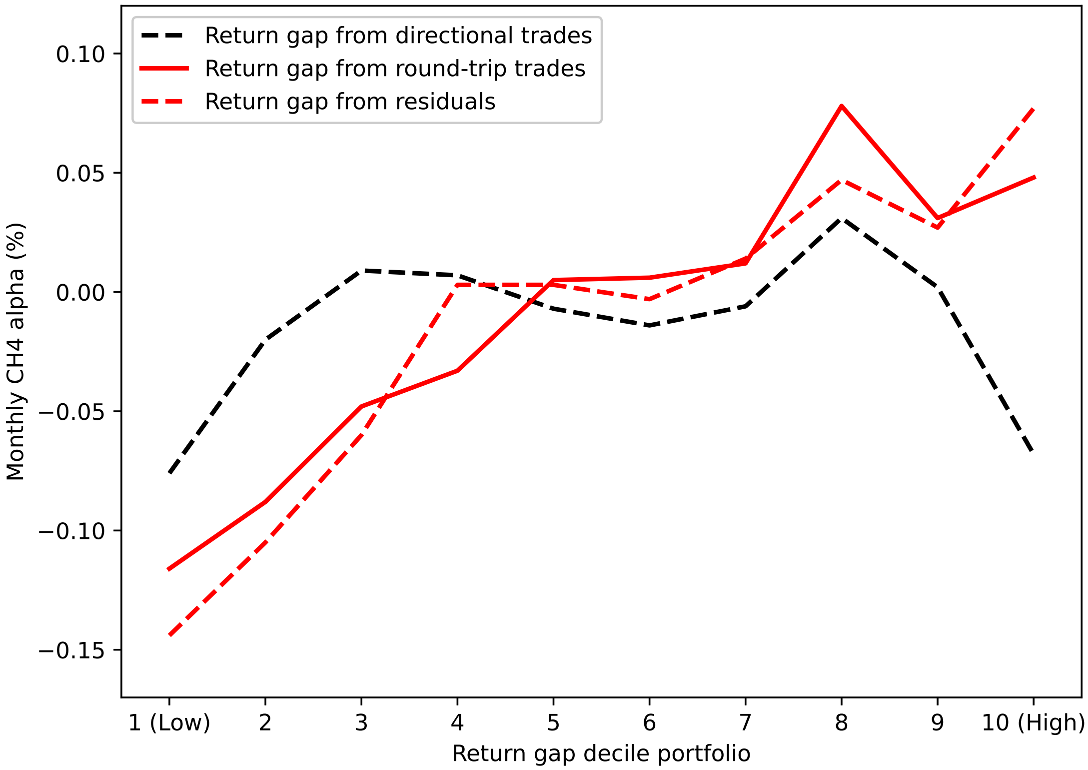

Figure 7: Decomposition of return gap predictability. This figure plots the mutual fund monthly risk-adjusted returns against fund groups sorted by each component of the return gap from 2001 to 2022Q2. Using our trading detection method, we decompose the return gap in Kacperczyk, Sialm, and Zheng (2008) into three components: those attributed to directional trades, round-trip trades, and the residuals, respectively. Funds are sorted into 10 groups at the end of each quarter based on each return gap component, with a lag of at least 3 months to ensure that the information is publicly available. Monthly returns are adjusted using the Carhart (1997) model.