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: February 2024

*An: Tsinghua PBC School of Finance. Email: anl@pbcsf.tsinghua.edu.cn. Huang: University of Hong Kong. Email: huangsy@hku.hk. Lou: London School of Economics and CEPR. Email: d.lou@lse.ac.uk. Wen: Hong Kong University of Science and Technology. Email: xwenam@connect.ust.hk. Xu: CITIC Securities. Email: xumingxin@citics.com. We thank Hank Bessembinder, James Guo, Stephan Siegel, and seminar participants at Bayes Business School, Capital University of Economics and Business, Hong Kong University of Science and Technology, Johns Hopkins University, Michigan State University, Penn State University, Universidad de los Andes, University of Washington, Utrecht University for helpful comments. Bolin Xu and Jianding Zhang provided excellent research assistance. All remaining errors are our own. This paper was previously circulated with 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. Although mutual funds report their holdings once every quarter, they are required to report their portfolio returns every day. After a mutual fund executes a trade, its reported portfolio returns further deviate from its quarter-end-holdings-based returns (assuming no trading). This sudden jump in return deviation allows us to infer the transaction date and amount. We apply our method to studying strategic trading of ESG stocks by mutual funds around quarter ends. Our evidence suggests that in recent years, mutual funds buy high-ESG stocks and sell low-ESG stocks right before quarter ends, and reverse their trades at the beginning of the next quarter. This trading pattern is concentrated among mutual funds right around the cutoff of four and five ESG rating stars, which have the strongest incentives to boost ESG performance. These trades also affect prices: high-ESG stocks outperform low-ESG stocks right before quarter ends, and underperform at the beginning of the next quarter.

JEL classification: G02, G12, G23, N22

1 Introduction

Investors have diverse preferences. Some have strong ethical and social concerns, and are willing to advance their social objectives at any cost. Some focus exclusively on their financial well-being and believe that all social issues should be resolved in the public domain. Most people are perhaps somewhere in-between these two extremes, aiming to strike a balance between social value and 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 (e.g., shunning firms with high carbon emissions), investors also maximize investment performance. Standard economic theory, however, predicts otherwise (Ľuboš Pástor, Stambaugh, and Taylor, 2021, 2022). Maximizing a weighted sum of financial and ESG performance is equivalent to solving a constrained performance optimization problem. Consequently, regardless of the relative returns of High-ESG vs. Low-ESG stocks, ESG preferences (weakly) reduce the optimal portfolio's financial performance (measured by, for example, the Sharpe ratio).[1]

This presents delegated asset managers with a thorny dilemma. On the one hand, asset managers have a strong incentive to cater to their clients' increasing environmental and social awareness. For example, Bloomberg Intelligence estimates that by 2022, asset managers with more than 40 trillion dollars under management have signed up to global initiatives on sustainable investment. On the other hand, asset managers have a fiduciary duty to maximize the financial performance of their clients. Indeed, nearly all mutual funds state in their prospectus that their objective is to maximize their portfolios' risk-adjusted returns, with very few publicly acknowledging any willingness to forego financial returns for social impact. Put differently, the asset management industry has made an *explicit* promise to their clients to maximize financial performance and at the same time an *implicit* promise to

[1] A volume of recent research (e.g., Bolton and Kacperczyk (2021)) further shows that firms with negative externalities have in recent years had higher average returns than firms with positive externalities. Therefore, investing in high-ESG stocks can doubly hurt financial performance – a lower average return on top of a more concentrated portfolio.

advance social goals.

Monitoring asset managers' financial performance is straightforward: asset managers (mutual funds, for example) are required by law to report their net asset value (NAV) and portfolio returns to the public at the end of each day, which are then audited by independent third parties. Monitoring their Environmental, Social and Governance (ESG) performance is much more difficult. A common practice – adopted by most investors and regulators – is to rely on the portfolio-weighted-average ESG rating of each manager's publicly disclosed holdings. Public portfolio disclosures, however, are infrequent. Mutual funds, the most tightly regulated group of asset managers, are required to report detailed information on portfolio holdings only at a quarterly frequency. It seems plausible, and almost natural, that asset managers – who aim to maximize both portfolio returns (observed daily) and ESG performance (measured once every quarter) – "window-dress" their portfolios. That is, to increase (decrease) their portfolio weights in firms with high (low) ESG ratings right before portfolio disclosure dates and pursue financial performance outside of these windows.

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. Our starting point is that while mutual funds disclose their portfolio holdings only at quarter-ends, they must report their portfolio returns everyday in the quarter. After a mutual fund executes a trade on day t, its reported fund return deviates further from its buy-and-hold portfolio-based-return. To illustrate, imagine a mutual fund that invests 100 dollars in stock A on March 31st. Further imagine that on May 10th, the fund sells 50 dollars of A and buys the same amount of stock B, and holds the resulting portfolio till June 30th. The fund's reported return should be exactly equal to its hypothetical buy-and-hold return (based on the portfolio disclosed on March 31st) from April 1st to May 9th. On May 10th, there emerges a significant divergence between the reported return and its hypothetical portfolio return. This sudden jump in return deviation allows us to infer the transaction date and amount.[2] Further, to reduce the impact of noise in our procedure, we apply a number of methods/penalties to smooth our estimates of daily mutual fund trading. Finally, we sum up trading in high-ESG (low-ESG) stocks by each mutual fund in the few days surrounding quarter-ends to identify its ESG-window-dressing behavior.

An alternative, naive approach would be to estimate changes in mutual funds' ESG betas around quarter-ends. This requires a) picking an arbitrary ESG index and b) computing portfolio ESG betas over short time periods. There are two obvious issues with this approach: a) any long-only ESG index is nearly perfectly correlated with the market so it is almost impossible to distinguish market beta from ESG beta; b) estimating portfolio betas over a few days yields very noisy estimates. Compared to this naive approach, our novel method has two advantages. First, by focusing on the difference between reported fund returns and the hypothetical holdings-based returns, our method helps clean up the impact of common factors (aside from the market factor, any other factor tilt the fund may have). Second, unlike the naive method which treats ESG betas as free parameters, our estimates automatically satisfy the constraints imposed by the fund's reported holdings as we need to match exactly its holdings at the beginning and end 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 (MSCI, Morningstar, Refinitiv) in our empirical analyses. Our results are qualitatively similar if we focus instead on one or two of the three ratings. Since passive mutual funds do not engage in frequent trading, we include in our sample all actively managed mutual funds. Our sample spans the period 2015 to 2022, after ESG became a global concern and Morningstar started to publish ESG ratings for all mutual funds based on their quarter-end holdings.

Our baseline result is that US actively managed mutual funds significantly increase their

[2] Prior studies (e.g., Hu et al. (2013)) have used the Ancerno transaction data to study mutual funds' intra-quarter trading behavior (its drivers and implications). They show that mutual funds tend to buy winner stocks and stocks already in the portfolio (to pump up fund returns) near quarter ends; we confirm both patterns using our novel approach.

investment in high-ESG stocks and reduce their portfolio weights in low-ESG stocks in a short window (e.g., a week) right 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 1.3% of mutual funds' total trading volume in the same window. For reference, combined trading in high- and low-ESG stocks in a typical week accounts for roughly 15% of mutual funds' total trading volume.

Since mutual funds 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 not monitored for ESG performance. 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 buttress our ESG-window-dressing hypothesis, we exploit a discontinuity in the incentives for mutual funds to manipulate their ESG ratings. In particular, Hartzmark and Sussman (2019) show that a) investor flows are a convex function in Morningstar ESG ratings, and b) there is a discontinuity in flows right around the cutoff of four and five ESG rating stars. Consequently, we zoom in on Morningstar ratings, and exploit the threshold between four and five stars. Specifically, we examine trading of high-ESG vs. low-ESG stocks around quarter-ends as a function of Morningstar ratings. Our analysis reveals that a one-standard-deviation change in ESG-window-dressing can bump a mutual fund's percentage ESG ranking by 1.3% (e.g., from 88.7th percentile (4 stars) to 90th percentile (5 stars)).

Perhaps not surprisingly, we also observe a strong return effect associated with these window-dressing trades. High-ESG stocks outperform low-ESG stocks right before quarter ends, and yet underperform at the beginning of the next quarter. The difference in returns between high-ESG and low-ESG stocks is over 1% higher in the week before quarter-ends than the week after. In a placebo test, and consistent with earlier result, we find no significant difference in returns between high- and low-ESG stocks around month-ends that are not quarter-ends (e.g., the end of January or February).

We also find significant variation in mutual funds' ESG window-dressing behavior both in the time-series and in the cross-section. For example, the tendency to ESG window-dress is much stronger among self-declared ESG funds, for mutual funds with better past performance and for funds headquartered in Democratic-leaning states. Mutual funds are also more likely to ESG window-dress in periods of larger capital flows to ESG funds, and when investors pay more attention to environmental and social issues as proxied by Google search volume. In addition, mutual funds ESG window-dressing trades are concentrated in stocks with higher liquidity and lower idiosyncratic volatility (so lower costs of window-dressing).

Finally, we show that mutual funds reap significant benefits from ESG window dressing: there is a strong correlation between ESG window dressing and subsequent fund flows. In other words, mutual fund investors are unable to distinguish true from manipulated ESG performance. At the same time, there are also costs to ESG manipulation: aside from the direct trading cost associated with ESG-window-dressing, mutual funds also have to bear the cost of deviating from the unconstrained optimal portfolio.

Related Literature First, our study contributes to the recent literature on ESG investment. In recent years, there has been a growing interest in socially responsible investing, which has led to the development of ESG investment strategies. These strategies take into account environmental, social, and governance factors when selecting investments. Chen and Dai (2023) examine how equity mutual fund managers make decisions on investing in ESG stocks and show mutual funds with flows highly sensitive to performance (positively sensitive to ESG score) invest less (more) in ESG stocks. Gibson et al. (2020) document that responsible investing does not enhance portfolio returns but reduces risk. Our study aims to contribute to this literature by examining how asset managers perceive ESG firms and strategically make investment decisions.

Second, our paper is also related to issues with ESG investment. Despite the growing interest in ESG investment, there are also issues with this investment strategy. For example, while low ESG firms are often explicitly excluded from ESG funds' investment universe, Cohen, Gurun, and Nguyen (2022) suggest the negative screening may not be optimal because there firms are key innovators in the United States' green patent landscape. Amel-Zadeh and Serafeim (2018) use survey data to show that relevance to investment performance is the most frequent motivation for use of ESG data followed by client demand and product strategy. Our study aims to shed light on these issues and provide insights into asset managers' incentives about ESG investing.

Third, echoing the literature on ESG firm returns, our analysis builds on and provides confirming evidence that asset managers perceive ESG firms to have lower average returns than low-ESG firms. This suggests that there may be a trade-off between investing in socially responsible companies and achieving high returns. Previous studies have shown mixed results on the relationship between ESG factors and financial performance. Bolton and Kacperczyk (2021) find stocks of firms with higher total CO2 emissions (and changes in emissions) earn higher returns. Duan, Li, and Wen (2023) use Trucost and TRACE data and show that bonds issued by carbon-intensive firms have lower returns, which contradicts with the carbon premium hypothesis. Chava, Kim, and Lee (2021) use ES ratings data from MSCI KLD and find no relationship between ES ratings and realized stock returns. Our study contributes to this literature by providing further evidence on the perceived lower returns of ESG firms.

Finally, our study contributes to the literature on rating and portfolio manipulation. As Chevalier and Ellison (1997) mention, a potential agency conflict between mutual fund managers and investors is that managers have an incentive to take actions that increase the inflow of investments rather than fully to maximize risk-adjusted returns for investors' benefits. Hartzmark and Sussman (2019) document being categorized as high (low) sustainability led to net inflows (outflows), which suggests that investors marketwide value sustainability. Our findings about ESG window dressing are consistent with prior results that fund managers have incentives to cater to investors who value sustainability and in the meanwhile to minimize cost on investing on perceived low-return stocks. Our study also highlights the need for greater transparency and accountability in the ESG investment industry.

2 Methodology and Data

2.1 Methodology

Our method of inferring intra-quarter trading is conducted for each fund quarter. Consider a quarter with trading days labeled from 0 (i.e., quarter beginning or the end of last quarter) to T (i.e., quarter ending). On each day t, an equity fund holds $S_{k,t}$ shares of stock $k \in \{1, 2, \cdots, K\}$ with stock return $R_{k,t}$, stock price $P_{k,t}$, and thus a holding value of $V_{k,t} = S_{k,t} \times P_{k,t}$. We start our method with a simple math identity: the total capital gain of the fund equals the sum of capital gains from each individual stock:

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

where R_t^{equ} is the fund's daily return from equity holdings. Our method allows funds to hold equity and cash (or borrow cash and take leverage), which will be discussed later. Because funds report holdings every quarter, let S_k^B denote the reported holding shares of stock k at the beginning of the quarter, and let $V_{k,t}^B = S_k^B \times P_{k,t}$ denote the holding value under beginning shares (note: $V_{k,t}^B$ is time-varying because $P_{k,t}$ is time-varying). We define the change of shares relative to quarter-beginning shares as $\Delta S_{k,t} = S_{k,t} - S_k^B$ and the corresponding change of value as $\Delta V_{k,t} = V_{k,t} - V_{k,t}^B = (S_{k,t} - S_k^B) \times P_{k,t} = \Delta S_{k,t} \times P_{k,t}$.

Substituting the relation $V_{k,t} = V_{k,t}^B + \Delta V_{k,t}$ into Eq.(1) and doing some simple algebra will lead to

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

where R_t^B is the hypothetical beginning portfolio return defined as

$$R_t^B = \frac{\sum_{k=1}^{K} V_{k,t-1}^B R_{k,t}}{\sum_{k=1}^{K} V_{k,t-1}^B}, \qquad (3)$$

and $\delta_{k,t-1}$ is the change of holding value scaled by the total value of the hypothetical beginning portfolio

$$\delta_{k,t-1} = \frac{\Delta V_{k,t-1}}{\sum_{k=1}^{K} V_{k,t-1}^B} = \frac{P_{k,t-1}(S_{k,t-1} - S_k^B)}{\sum_{k=1}^{K} V_{k,t-1}^B}. \qquad (4)$$

The obtained fund return identity, i.e., Eq.(2), plays a central role in our intro-quarter trading detection method. The left hand of Eq.(2) is the daily return gap, which is the difference between the fund actual return and the return on a portfolio that invests in the previously disclosed fund holdings (Kacperczyk, Sialm, and Zheng, 2008). The right hand of Eq.(2) is the weighted sum of stock returns excess of fund return with the weights from the scaled change of holding value, $\delta_{k,t-1}$. Intuitively, if there is no trade during the quarter, $\delta_{k,t-1}$ will be 0 for all stocks and all days, the fund actual return, R_t^{equ}, will coincide with the beginning portfolio return, R_t^B. If there is a trade, for example, to buy stock k, during the quarter, $\delta_{k,t-1}$ will increase as $S_{k,t-1}$ increases relative to S_k^B, and then R_t^{equ} will reflect more variation from stock return R_t^k. Therefore, the trading process implied by $\delta_{k,t-1}$ will change the relative weights of each stock and then contribute to the fund return deviation from the beginning portfolio return.

The central idea of our trading detection is a reverse engineering process, which is to use the identity relation and observable variables (i.e., fund daily returns, fund quarterly holdings, and stock daily prices and returns) to solve the remaining unobservable variable (i.e., fund daily shares $S_{k,t}$). To proceed, we parameterize $S_{k,t}$ by introducing $\theta^d_{k,t}$ and $\theta^r_{k,t}$ in the following way:

$$S_{k,t} = S^B_k + \underbrace{\theta^d_{k,t}(S^E_k - S^B_k)}_{\text{directional trading}} + \underbrace{\theta^r_{k,t} C_k}_{\text{round-trip trading}}, \tag{5}$$

where S^E_k, a similar notation to S^B_k, is the reported shares at the quarter ending, and C_k is a constant for normalization. Specifically, $\theta^d_{k,t}$ is to capture directional trading, i.e., the trades that change shares monotonically from beginning shares S^B_k to ending shares S^E_k, while $\theta^r_{k,t}$ is to capture round-trip trading, i.e., the trades that change shares back and forth. Thus, we restrict $\theta^d_{k,t}$ to be non-decreasing and start at 0 (i.e., $\theta^d_{k,0} = 0$) and end at 1 (i.e., $\theta^d_{k,T} = 1$). Instead, $\theta^r_{k,t}$ has no restriction on monotonicity but needs to satisfy $\theta^r_{k,0} = \theta^r_{k,T} = 0$ to ensure $S_{k,0} = S^B_k$ and $S_{k,T} = S^E_k$. Since our trading detection method is mainly applied to equity mutual funds, which are usually not allowed to short-sell, we impose the non-negative shares constraint $S_{k,T} \geq 0$ to improve estimation precision. Also, by comparing fund holding change and reported actual turnover, we can obtain the round-trip trading volume of a fund, which can be used in the following constraint: $\sum_{k=1}^{K} \sum_{t=1}^{T} |\theta^r_{k,t} - \theta^r_{k,t-1}| \leq \kappa$, where κ is an upper bound calculated from round-trip trading volume.

To estimate the intro-quarter trading process, we need to focus on the dynamics of $\theta^d_{k,T}$ and $\theta^r_{k,T}$. However, a direct estimation is challenging because the dimension in a magnitude of $K \times T$ is large. To overcome this issue, instead of viewing each day t as a trading unit, we assume $\theta^d_{k,T}$ and $\theta^r_{k,T}$ to be piecewise linear functions with a D-day window. For example, $D = 5$ means we treat each week as a trading unit and within the week trading is conducted linearly. This can help us reduce the number of parameters by a factor of D.

To consider the situation of holding or borrowing cash, we introduce a leverage parameter L so that

$$R_t^{fund} = L \times R_t^{equ}, \tag{6}$$

where R_t^{fund} is the reported fund daily return. When $L < 1$, the fund holds L portion of equity and $1 - L$ portion of cash. When $L > 1$, the fund borrows a $L - 1$ portion of cash and invests it to equity.

The optimization problem is to solve L, $\theta_{k,T}^d$ and $\theta_{k,T}^r$ so that the distance between fund reported return and model fitted return is minimized, which is given by

$$\begin{aligned}
& \min_{\{L,\theta_{k,t}^d,\theta_{k,t}^r\}} \sum_{t=1}^{T} \left[R_t^{equ} - R_t^B - \sum_{k=1}^{K} \delta_{k,t-1} \left(R_{k,t} - R_t^{equ} \right) \right]^2 \\
\text{s.t.} \quad & R_t^{equ} = R_t^{fund}/L \\
& \delta_{k,t} = \frac{P_{k,t}(S_{k,t} - S_k^B)}{\sum_{k=1}^{K} V_{k,t}^B}, \\
& S_{k,t} = S_k^B + \theta_{k,t}^d (S_k^E - S_k^B) + \theta_{k,t}^r C_k, \\
& S_{k,t} \geq 0, \\
& \theta_{k,t}^d \text{ and } \theta_{k,t}^r \text{ are piece-wise linear functions,} \\
& \theta_{k,t}^d \text{ is non-decreasing, } \theta_{k,0}^d = 0, \ \theta_{k,T}^d = 1, \\
& \theta_{k,0}^r = \theta_{k,T}^r = 0, \\
& \sum_{k=1}^{K} \sum_{t=1}^{T} |\theta_{k,t}^r - \theta_{k,t-1}^r| \leq \kappa.
\end{aligned} \tag{7}$$

Instead of directly solving for all stocks, we use an iterative approach by updating $\theta_{k,t}^d$ and $\theta_{k,t}^r$ stock by stock. In the following validation and empirical application, we allow K to be as large as 500.

2.2 Validation based on Ancerno data

To validate our trading detection method, we use Ancerno data to construct a testing sample. We randomly draw 10,000 fund quarters from Ancerno data as our testing sample. For each fund quarter, we maintain the entire trading structure without any changes on the trades, i.e., we test our method under a real situation. We consider three evaluation measures: Tracking Error, Holding-based R^2, and Trading Match. Tracking Error is the root of mean squared error between fund actual daily return and fitted return expressed in basis points, which is

$$\text{Tracking Error} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(R_t^{fund} - \hat{R}_t^{fund})^2} \tag{8}$$

Holding-based R^2 is the explained variation of normalized daily excess holding defined as

$$\text{Holding-based } R^2 = 1 - \frac{\sum_{t=1}^{T}\sum_{k=1}^{K}(\delta_{k,t-1} - \hat{\delta}_{k,t-1})^2}{\sum_{t=1}^{T}\sum_{k=1}^{K}(\delta_{k,t-1})^2}, \tag{9}$$

where $\delta_{k,t}$ is defined in Eq.(4), and $\hat{\delta}_{k,t}$ is the estimated value of $\delta_{k,t}$. Trading Match is the overlap between true and estimated trading volume calculated by

$$\text{Trading Match} = \sum_{w=1}^{W}\sum_{k=1}^{K}\sum_{x\in\{buy,sell\}} \min(v_{k,w}^{x}, \hat{v}_{k,w}^{x}), \tag{10}$$

where $v_{k,w}^{x}$ is trading volume on stock k in week w with direction x (i.e., buy or sell) and $\hat{v}_{k,w}^{x}$ is estimated value of $v_{k,w}^{x}$. We report Trading Match by scaling the benchmark of random guessing and it can be interpreted as how many times the trading can be matched relative to random guessing. The evaluation measures are calculated for each fund quarter, and we report the average values over 10,000 fund quarters. To have a comparison, we consider three benchmarks: (1) random guessing with each stock traded from beginning shares to ending shares on a randomly chosen day; (2) assuming all trades to happen at the beginning of the quarter; (3) assuming all trades to happen at the ending of the quarter. For our methods, we report the performance from (1) a constrained model allowing only directional trades, and (2) a full model allowing both directional and round-trip trades.

The performance evaluation is shown in Table 2. First, our method has a better fitness on fund return with daily tracking error smaller than 1 basis point, i.e., 0.47 and 0.83 basis points for models with and without round-trip trades, compared to the benchmarks around 10 basis points. Second, our method can explain over 80% variation on fund daily holding changes (i.e., daily holdings relative to beginning holdings), which is much higher than the 11.1, -51.4%, and 0 for benchmarks of random guessing, assuming beginning trades, and ending trades, respectively. Note that this measure reflects additional explanatory power relative to assuming ending trades, thus zero or negative value means no improvement or worse than assuming ending trades. Finally, our method can match trades with around 5 times of random guessing. Allowing round-trip trades improves this number from 4.77 to 4.86. In summary, our method exhibits much better performance than the benchmarks.

In Table 3, we report the evaluation measures by sorting on fund characteristics such as number of holdings, turnover, and proportion of round-trip trading volume. First, better performance corresponds to the funds with fewer holdings, lower turnover, and a smaller proportion of round-trip trading volume. Second, the detection performance remains relatively good even in the worst cases. For example, the holding-based R^2 is always above 70%, and the trading match is usually above 3 times of random guessing. Finally, in Panel C, as the proportion of round-trip trading volume increases, allowing round-trip for detection brings more improvement in trading matches.

In Figure 1, we evaluate our method under a classification perspective, i.e., whether there is a trade in week w for stock k. The ROC curve shows that at a 10% false positive rate, i.e., type-I error, our method can correctly classify around 70% of trades.

2.3 Data

Our stock-level ESG ratings come from three rating providers: Morningstar Sustainalytics, MSCI, and Refinitiv Asset4 (Lipper). We define our ESG stock list by combining the above three ratings to reduce noise, as literature document much variation on ESG ratings. Specifically, 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 Sustainalytics, MSCI, and Refinitiv in quarter $t-1$. To be included in our high-ESG or low-ESG stock list, a stock must receive at least 2 non-missing ESG scores from the above 3 rating agencies. Figure 1 plots the time series of Pearson correlations among the three stock ESG ratings. Consistent with previous literature, the average correlation is 0.357, which is not high.[3]

Mutual fund daily returns are obtained from the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund database, and quarterly holdings are from Thomson Reuters's CDA/Spectrum Mutual Fund Holdings Database. We use the MFLinks file to merge between CDA/Spectrum and the CRSP mutual fund database. We focus on US active equity mutual funds by requiring (1) the investment objective code reported by CDA/Spectrum to be aggressive growth, growth, growth and income, balanced, unclassified, or missing, (2) the ratio of the equity holdings to total net assets to be between 0.75 and 2, and (3) no index funds. To further ensure data quality, we require a minimum fund size of \$10 million and a minimum number of holdings of 10 stocks. Our sample period is from 2015Q1 to 2022Q2, which is a period that ESG attracted more and more investors' attention. After applying the above filters, we finally obtain a sample with 3,525 unique funds and 58,198 fund-quarters. The summary statistics of fund characteristics are shown in Table 1 Panel A.

To obtain mutual fund intra-quarter trading for the examination of ESG window dressing,

[3] A sudden decrease on the correlation with Sustainalytics at 2019Q3 is because Morningstar changed stock ESG rating scheme at that point.

we apply our detection method to the data. The key variables in our analysis are the ratios of high-ESG or low-ESG trading volume divided by total trading volume in each week. We calculate the numerators and denominators based on the estimation results from our detection method and aggregation of stock-level trading volume. High-ESG and low-ESG stocks are defined by combining three ESG ratings as mentioned before.

3 ESG Window Dressing

In this section, we apply our methodology 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 Trading around the turn of the quarters

We begin by examining how mutual funds trade stocks with high versus low ESG scores around the turn of each quarter. To gauge the ESG performance for each stock, we obtain ESG ratings from the three major ESG rating providers, namely Morningstar Sustainalytics, MSCI, and Refinitiv. We require stocks to have at least two non-missing ESG scores from the three rating agencies. We then take the average rank-normalized ESG scores in the last quarter for each stock across three ratings, and label the top (bottom) 10% stocks as high-ESG (low-ESG) stocks.

For each fund each week, we calculate the fraction of trading in high-ESG (low-ESG) stocks by taking the ratio between 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 dummy variables indicating different weeks within a quarter, specifically in the following form:

$$y_{i,t,l} = b_0 + \sum_{j=1}^{3} b_{E,j} \times I_{l=E,j} + \sum_{j=1}^{3} b_{B,j} \times I_{l=B,j} + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}, \tag{11}$$

$y_{i,t,l}$ 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. $I_{l=E,j}$ ($I_{l=B,j}$) is a dummy variable indicating the jth 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 4 Panel A shows the results based on Eq.(11). 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 multiplied by 100. We see 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 buy 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.

Mutual funds are required to report their holdings only at quarter ends. To test the window dressing hypothesis, we conduct a placebo test by focusing on the non-quarter-ending month ends (e.g., the end of January and February). We repeat our exercises around these month ends and report the results in Table 1 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.

3.2 ESG window dressing around the cutoff of ratings

To further sharpen our identification, we exploit the fact that mutual funds with different percentile ranks may have varying incentives to manipulate their ESG performance because of the discontinuity in investors' flow responses around the rating category edges. 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 ranks underlying the ratings.

Important for our purpose, investors' flow responses to ESG globe ratings are a) disproportionately strong at the extreme globe categories, and b) exhibit discontinuity around the cutoff of four and five rating globes. (Hartzmark and Sussman, 2019). This means a fund who is ranked at the 91 percentile (would receive 5 globes) could receive much higher inflows due to its ESG ratings than a fund ranked at the 89 percentile (would receive 4 globes), although both funds have similar underlying sustainability characteristics. We conjecture that funds whose ESG scores are around the cutoffs between 4 and 5 globes would have the strongest incentives to manipulate their quarter-end holdings, due to the sharp changes in potential payoffs.

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 Morningstar ESG ranks. Specifically, we take the five categories and further split funds in each rating category into the "Lower Half" and the "Upper Half" using the underlying percentiles. To avoid look-ahead bias, we sort funds based on the average percentile ranks in the past two quarters (quarters $t-1$ and $t-2$) and examine net ESG trading around the quarter end t. We adapt our baseline regression and include dummy variables indicating a fund's percentile group (x) and the last or first week in a quarter ($E1$ or $B1$). Specifically, we run the following regression:

$$y_{i,t,l} = b_0 + \sum_{x=1L,1U,\dots,5U} b_{x,E1} \times I_{x,E1} + \sum_{x=1L,1U,\dots,5U} b_{x,B1} \times I_{x,B1} + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l},$$

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

Table 5 reports the results. For funds in each of the two halves within the five categories, we report the difference in net ESG trading in one week before and after the quarter ends

$(b_{x,E1} - b_{x,B1})$, respectively. We see that the magnitude of ESG window dressing has a hump-shaped pattern with respect to the funds' percentile ranks. The effect is strongest for funds ranked on both sides around the cutoff between four and five globes (four globe upper half and five globe lower half), where their abnormal ESG trading accounts for 2.9% (t-stat=4.93) and 3.0% (t-stat=3.29) of the total trading volume, respectively. For comparison, the magnitude of abnormal ESG trading for funds in the next two groups around the cutoff (four globe lower half and five globe upper half) is 1.0% and 1.4% respectively, and the average magnitude for all funds is 1.2% (as shown in Table 4). The last row in Table 5 shows that the difference between the intensity of window dressing of the two groups around the cutoff (4H + 5L) and that of the next two groups (4L + 5U) is 3.5% and statistically significant with a t-stat of 2.03.

An alternative way for gauging the change in magnitude is to estimate the magnitude of net ESG trading as a piecewise-linear function of the funds' sustainability percentile rank. Specifically, we estimate the following regression

$$y_{i,t,l} = b_0 + f_E(p) \times I_{LastWeek} + f_B(p) \times I_{FirstWeek} + g(p) \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l},$$

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% to 100%. We report the difference in slopes between quarter ends and quarter beginnings $(f_E(p) - f_B(p))$ within each rating category in the last column in Table 5, and also plot these slopes as function of rating category in Figure 3. Again, we observe a hump-shaped function where the magnitude of ESG window dressing is strongest around the cutoff between the four and five globes.

3.3 Return impact

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 6 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 using a 60-month rolling window of monthly returns. We see that high-ESG stocks outperform low-ESG stocks before 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 in size, their abnormal return tends to be larger and more volatile. The hedged portfolio that longs high-ESG stocks and shorts low-ESG stocks would experience 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 hedged returns between the quarter end and beginning is 1.07% with a t-stat of 2.04.

Similar to our trading analyses, we run a placebo test of ESG return patterns in the non-quarter-ending month-ends, and report the results in Table 6 Panel B. We show previously that mutual funds do not engage in ESG window dressing around month-ends that are not quarter-ends (Table 4 Panel B); Consequently, we observe no significant difference in returns between high- and low-ESG stocks around these month ends.

3.4 Additional results

To gain further insights on mutual funds ESG window dressing, we conduct two additional sets of analyses. Due to limited space, we briefly outline the results in this subsection and provide a detailed presentation of the analyses in the Appendix.

First, we run sub-sample analyses to exploit the heterogeneity in window dressing intensity across different funds types, stocks characteristics, and various time periods (full details reported in Appendix Section 1.1). Our findsings reveal that ESG window dressing is more-pronounced among funds that explicitly identify as ESG funds, those situated in Democratic-leaning states, and funds demonstrating stronger recent financial performance (Appendix Table A1). The mutual fund industry also exhibits stronger ESG window dressing when investors are more aware of environmental and social issues, proxied by periods with higher aggregate flows into ESG funds and periods with higher google search volume on sustainability topics (Appendix Table A2). Regarding stock choices, funds are more likely to practice ESG window dressing with stocks characterized by lower bid-ask spreads and lower idiosyncratic volatility, so lower costs of window-dressing (Appendix Table A3).

Our second set of analyses aims to further understand the costs and benefits of ESG window dressing (full details reported in Appendix Section 1.2). We show that ESG window dressing indeed helps to attract more fund flows: a one-standard-deviation increase in window dressing leads to 19bp more flows in the next quarter (Appendix Table A4). On the other hand, window dressing trading is financially costly: a one-standard-deviation increase in window dressing leads to 5bp lower returns on a quarterly basis, or 20 bp per year (Appendix Table A5).

4 Other Applications

In this section, we extend our study and employ our methodology to three additional applications in the mutual fund literature. Specifically, we apply our methodology to investigate mutual funds' window dressing behavior around quarter ends in Section 4.1. We study the portfolio pumping behavior of mutual funds in Section 4.2. In Section 4.3, we use our methodology to infer mutual funds' intra-quarter directional and round-trip trading and then decompose the return gap of Kacperczyk, Sialm, and Zheng (2008) into those arising from directional trading and round-trip trading. After that, we study the return predictability of return gaps arising from directional trading and round-trip trading separately. These applications not only broaden the scope of our study but also show the generalization and robustness of our methodology.

4.1 Performance window dressing

Popular wisdom among practitioners is that institutional investors have incentives to "reshuffle" or "window dress" their portfolios in order to make their holdings look impressive in their reports. Prior studies (e.g., Agarwal, Gay, and Ling, 2014; He, Ng, and Wang, 2004; Lakonishok et al., 1991; Meier and Schaumburg; Ng and Wang, 2004) focus on quarterly or semi-annual holding data and find supporting evidence consistent with "window dressing" behavior before the end of the quarter or the year. In this section, we take advantage of our methodology and revisit "window dressing" behavior of mutual funds. Specifically, we apply our methodology to estimate mutual funds' intra-quarter trading and investigate mutual funds' trading on the winner and loser stocks around the end of the quarter, supplementing the literature on "window dressing".

We take the following steps to investigate mutual funds' "window dressing" behavior. First, at each month's end, we sort stocks based on their cumulative returns in the past

12 months (skipping the current month) and define stocks in the top (bottom) decile as the winner (loser) stocks. Second, we follow the methodology in Section 3.1 and examine mutual funds' trading on the winner and loser stocks around the quarter end. That is, for each fund each week, we calculate the fraction of trading in the winner (loser) 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. After that, we re-run the regression of Equation (11).

Panel A of Table 7 reports the results. The patterns in Panel A of Table 7 are like those in ESG window dressing and are consistent with "window dressing". As we can observe, mutual funds buy winner stocks and sell loser stocks right before the quarter ends: their abnormal net buy of the winner (loser) stocks account for 0.19% (−0.64%) of total trading volume in the first week prior to the quarter ends, and the difference 0.83% is statistically significant with a t-stat of 4.42. The magnitude of abnormal trading diminishes in the prior weeks: the difference in net trading between the winner and loser stocks becomes 0.78% (t-stat = 5.05) and 0.39% (t-stat = 3.08) 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 buy between the winner and loser ESG stocks is −0.49% (t-stat = -2.85) in the first week in the next quarter.

4.2 Portfolio pumping

In addition to "window dressing", the other common strategic behavior among institutional investors is "portfolio pumping." "portfolio pumping" refers to the excess buying of stocks that mutual funds heavily own. The purpose of "portfolio pumping" is to inflate the funds' closing net asset value and consequently exaggerate the funds' performance (see evidence from Ben-David et al., 2013; Bernhardt and Davies, 2005; Bhattacharyya and Nanda, 2013; Carhart et al., 2002; Hu et al., 2013). In this section, we take advantage of our methodology again and revisit "portfolio pumping" behavior of mutual funds. Specifically, we apply our methodology to estimate mutual funds' intra-quarter trading and investigate mutual funds' trading on stocks in which mutual funds overweight or underweight.

We take the following steps to investigate mutual funds' "portfolio pumping" behavior. First, for each fund at the beginning of each quarter, we focus on its portfolio stocks and sort the portfolio stocks based on portfolio weights within this fund. For stocks taking account for the top (bottom) 10% of the fund's portfolio, we define them to have top (bottom) positions. Second; we follow the methodology in Section 3.1 and examine mutual funds' trading on stocks with top and bottom positions around the quarter end using the regression of Equation (11).

Panel B of Table 7 reports the results and provides evidence consistent with "portfolio pumping". As we can observe, mutual funds buy stocks with top positions in their portfolio and sell stocks with bottom positions right before the quarter ends: their abnormal net buy of top-position (bottom-position) stocks account for 1.13% (-2.18%) of total trading volume in the first week prior to the quarter ends, and the difference 3.30% is statistically significant with a t-stat of 17.17. Like the patterns in Panel A of Table 7 and Table 4, the magnitude of abnormal trading diminishes in the prior weeks: the difference in net trading between the top-position and the bottom-position stocks becomes 2.50% (t-stat = 16.12) and 1.54% (t-stat = 12.46) in the second and third week before the quarter ends. Again, mutual funds reverse these trades at the beginning of the next quarter.

In sum, applying our methodology, we revisit the common strategic behavior of mutual funds–"window dressing" and "portfolio pumping"–and find evidence consistent with them. This revisit is an external validity of our methodology and has demonstrated the generalization and robustness of our methodology.

4.3 Decomposing return gap

We further extend our study and examine the return predictability of the return gap proposed by Kacperczyk, Sialm, and Zheng (2008). The return gap refers to the difference between the reported fund return and the return on a portfolio that invests in the previously disclosed fund holdings. Kacperczyk, Sialm, and Zheng (2008) find that the return gap can predict fund performance. It is still not completely clear what drives the return predictability of the return gap. In this section, we apply our methodology to estimate mutual funds' intra-quarter directional and round-trip trading, which mechanically contribute to the return gap. Then, we decompose the return gap into components from directional trading and round-trip trading separately and examine different components' return predictability. Such a decomposition can shed light on the driving forces for the return predictability of the return gap.

We take the following steps to conduct our decomposition. First, for each fund at each quarter, we use the methodology in Section 3 and estimate its intra-quarter directional and round-trip trading within this quarter. Second, we compute the fund's returns from the quarter beginning based on the estimated directional and round-trip trading. By doing this way, for each fund at each quarter, we decompose the return gap of Kacperczyk, Sialm, and Zheng (2008) into the component from the directional trading (dubbed as directional-trade return gap), the component from the round-trip trading (dubbed as round-trip-trade return gap), and the residual component (the difference between the return gap and the sum of the directional-trade and round-trip-trade return gaps). After that, we use the portfolio sorting approach and examine the return predictability of different components in the return gap.

Table 8 reports the results. Panel A replicates Kacperczyk, Sialm, and Zheng (2008) using our sample, Panel B reports the results on the return predictability of the directional-trade return gap, Panel C reports the results on the return predictability of the round-trip-trade return gap, and Panel D reports the results on the return predictability of the residual return gap. We have several observations. First, Panel A confirms that the return gap can significantly and positively predict future fund performance, confirming the return predictability of the return gap in our sample. Second, the round-trip-trade return gap can significantly and positively predict future fund performance. That is, regardless of the performance metrics, funds in the top decile of the past 12-month round-trip-trade return gap outperform those in the bottom decile. For example, in terms of Fama-Frech-Five Factors augmented by the moment factor, funds in the top decile of the past 12-month round-trip-trade return gap outperform those in the bottom decile by 13 bps per month (with t-statistics of 2.35). This return predictability is economically significant. For a comparison, funds in the top decile of the past 12-month return gap Kacperczyk, Sialm, and Zheng (2008) outperform those in the bottom decile by 17.5 bps per month (with t-statistics of 2.71). Interestingly, the directional-trade return gap insignificantly or weakly predicts future fund performance (see Panel B).[4]

The sharp contrast of the return predictability of the directional-trade and round-trip-trade return gaps suggests that the return predictability of the return gap of Kacperczyk, Sialm, and Zheng (2008) mainly comes from the round-trip-trade. This finding is consistent with Da, Gao, and Jagannathan (2011). Da, Gao, and Jagannathan (2011) decompose a mutual fund's trading into liquidity-absorbing impatient trading and liquidity provision and find that funds with higher "return gaps"—defined in Kacperczyk, Sialm, and Zheng (2008) to capture the benefit of "unobservable" actions of mutual funds—add value through liquidity provision. The round-trip trade in our study is more like temporary liquidity provision, and in this sense, the results in Panel C are consistent with the findings of Da, Gao, and Jagannathan (2011).

[4]As shown in Panel D, the residual return gap can also significantly and positively predict future fund performance. It is unclear what drives the residual trading, and we leave the study the residual return gap for future study.

5 Conclusion

With the rising popularity of socially responsible investment, asset managers often make dual promises to their clients: to maximize both financial and social performance. While it is straightforward to monitor the former (as asset managers have to report their returns and net asset value on a daily basis), it is difficult to gauge the latter. A common practice is to monitor asset managers' ESG performance based on their quarter-end portfolio disclosures.

We develop a novel method to infer details of 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 further from its quarter-end-holdings-based return. This sudden jump in return deviation allows us to infer the transaction date and amount.

We apply our method to studying the strategic trading behavior of 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 concentrated among mutual funds right around the cutoff of four and five ESG rating stars, which have the strongest incentives to boost their ESG performance. These trades also affect prices: high-ESG stocks outperform low-ESG stocks right before quarter ends, and yet underperform at the beginning of the next quarter.

References

Agarwal, V., D. G. Gay, and L. Ling. 2014. Window Dressing in Mutual Funds. *The Review of Financial Studies* 27:3133–70.

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

Ben-David, I., F. Franzoni, A. Landier, and R. Moussawi. 2013. Do Hedge Funds Manipulate Stock Prices? *The Journal of Finance* 68:2383–434.

Berg, F., J. F. Kölbel, and R. Rigobon. 2022. Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance* 26:1315–44.

Bernhardt, D., and R. J. Davies. 2005. Painting the Tape: Aggregate Evidence. *Economics Letter* 89:306–11.

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

Bolton, P., and M. Kacperczyk. 2021. Do investors care about carbon risk? *Journal of Financial Economics* 142:517–49.

Carhart, M. M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52:57–82. ISSN 00221082, 15406261.

Carhart, M. M., R. Kaniel, D. K. Musto, and A. V. Reed. 2002. Leaning for the Tape: Evidence of Gaming Behavior . *The Journal of Finance* 57:661–93.

Chava, S., J. H. J. Kim, and J. Lee. 2021. Doing well by doing good? risk, return, and environmental and social ratings. *Working Paper* .

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

Chevalier, J., and G. Ellison. 1997. Risk taking by mutual funds as a response to incentives. *Journal of Political Economy* 105:1167–200.

Cohen, L., U. G. Gurun, and Q. H. Nguyen. 2022. The esg-innovation disconnect: Evidence from green patenting. *Working Paper* .

Cremers, K. J. M., and A. Petajisto. 2009. How active is your fund manager? a new measure that predicts performance. *Review of Financial Studies* 22:3329–65.

Da, Z., P. Gao, and R. Jagannathan. 2011. Impatient Trading, Liquidity Provision, and Stock Selection by Mutual Funds . *The Review of Financial Studies* 24:675–720.

Duan, T., F. W. Li, and Q. Wen. 2023. Is carbon risk priced in the cross-section of corporate bond returns? *Journal of Financial and Quantitative Analysis* 1–52.

Gibson, R., S. Glossner, P. Krueger, P. Matos, and T. Steffen. 2020. Responsible institutional investing around the world. *Working Paper* .

Hartzmark, S. M., and A. B. Sussman. 2019. Do investors value sustainability? a natural experiment examining ranking and fund flows. *The Journal of Finance* 74:2789–837.

He, J., L. Ng, and Q. Wang. 2004. Quarterly Trading Patterns of Financial Institutions. *Journal of Business* 77:493–509.

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

Hsu, P.-H., K. Li, and C.-Y. Tsou. 2023. The pollution premium. *The Journal of Finance* 78:1343–92.

Hu, G., R. D. McLean, J. Pontiff, and Q. Wang. 2013. The Year-End Trading Activities of Institutional Investors: Evidence from Daily Trades. *The Review of Financial Studies* 27:1593–614.

In, S. Y., K. Y. Park, and A. Monk. 2019. Is 'being green' rewarded in the market?: An empirical investigation of decarbonization and stock returns. *Working Paper* .

Kacperczyk, T. M., C. Sialm, and L. Zheng. 2008. Unobserved Actions of Mutual Funds. *The Review of Financial Studies* 21:2379–416.

Lakonishok, J., A. Shleifer, R. Thaler, and R. Vishny. 1991. Window Dressing by Pension Fund Managers. *The American Economic Review* 81:227–31.

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

Ng, L., and Q. Wang. 2004. Institutional Trading and The Turn-of-The-Year Effect. *Journal of Financial Economics* 74:343–66.

Pedersen, L. H., S. Fitzgibbons, and L. Pomorski. 2021. Responsible investing: The esg-efficient frontier. *Journal of Financial Economics* 142:572–97.

Ľuboš Pástor, R. F. Stambaugh, and L. A. Taylor. 2021. Sustainable investing in equilibrium. *Journal of Financial Economics* 142:550–71.

———. 2022. Dissecting green returns. *Journal of Financial Economics* 146:403–24.

———. 2023. Green tilts. *Working Paper* .

Table 1: Summary statistics. This table presents summary statistics for our sample of mutual funds. The sample period is from 2015Q1 to 2022Q2. Panel A presents the summary statistics of fund characteristics. TNA is quarter-end total net fund assets in millions of dollars; age is the fund age in years; number of stocks held is the number of stocks in fund quarterly holding disclosure; monthly return is net fund return in percentage; 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 and displayed in percentage; expense is the fund expense ratio in percentage; and turnover is the turnover ratio of the fund. Panel B presents the summary statistics of estimated fund weekly 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 display them in percentage. 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 Sustainalytics, MSCI, and Refinitiv in quarter $t-1$.

Panel A: Number of funds and fund-quarters					
Total number of funds	3,525				
Total number of fund-quarters	58,198				
Panel B: Summary Statistics of Fund Characteristics					
	Mean	SD	P5	P50	P95
TNA ($ million)	2,549	11,282	22	413	9,636
Age (years)	17.17	11.69	3	15	38
Number of stocks held	149.90	261.10	25	73	541
Monthly return (%)	0.75	5.52	-8.72	1.09	8.76
Monthly flow (%)	0.35	27.85	-6.24	-0.56	7.39
Expense (%)	0.88	0.41	0.20	0.91	1.53
Turnover	0.69	1.24	0.08	0.45	1.77
Panel C: Summary Statistics of Estimated Weekly Trading					
	Mean	SD	P5	P50	P95
Total number of fund-quarter-weeks	676,244				
Total buy / Total trading volume	48.71	34.27	0.00	48.06	100.00
Total sell / Total trading volume	51.29	34.27	0.00	51.94	100.00
High-ESG buy / Total trading volume	6.88	15.40	0.00	0.00	39.31
High-ESG sell / Total trading volume	7.90	16.71	0.00	0.00	43.69
Low-ESG buy / Total trading volume	1.61	7.41	0.00	0.00	10.04
Low-ESG sell / Total trading volume	1.59	7.45	0.00	0.00	9.73

Table 2: Overall performance evaluation based on Ancerno data. This table reports the performance evaluation of our intro-quarter trading detection method. We randomly draw 10,000 fund quarters from Ancerno data as our testing sample. For each fund quarter, we maintain the entire trading structure without any changes on the trades, i.e., we test our method under a real situation. We consider three evaluation measures: Tracking Error, Holding-based R^2, and Trading Match. Tracking Error is the root of mean squared error between fund actual daily return and fitted return expressed in basis points. Holding-based R^2 is the explained variation of normalized daily excess holding calculated via $1 - \left[\sum_{t=1}^{T}\sum_{k=1}^{K}(\delta_{k,t-1} - \hat{\delta}_{k,t-1})^2\right] / \left[\sum_{t=1}^{T}\sum_{k=1}^{K}(\delta_{k,t-1})^2\right]$, where $\delta_{k,t}$ is normalized daily excess holding defined in fund return identity, and $\hat{\delta}_{k,t}$ is estimated value of $\delta_{k,t}$. Trading Match is the overlap between true and estimated trading volume calculated via $\sum_{w=1}^{W}\sum_{k=1}^{K}\sum_{x\in\{buy,sell\}}\min(v_{k,w}^{x}, \hat{v}_{k,w}^{x})$, where $v_{k,w}^{x}$ is trading volume on stock k in week w with direction x (i.e., buy or sell) and $\hat{v}_{k,w}^{x}$ is estimated value of $v_{k,w}^{x}$. We report Trading Match by scaling the benchmark of random guessing and it can be interpreted as how many times the trading can be matched relative to random guessing. The evaluation measures are calculated for each fund quarter, and we report the average values over 10,000 fund quarters. To have a comparison, we consider three benchmarks: (1) random guessing with each stock traded from beginning shares to ending shares on a randomly chosen day; (2) assuming all trades to happen at the beginning of the quarter; (3) assuming all trades to happen at the ending of the quarter. For our methods, we report the performance from (1) a constrained model allowing only directional trades, and (2) a full model allowing both directional and round-trip trades.

	Tracking Error	Holding-based R^2	Trading Match
Benchmarks			
Random guessing	8.80	11.1%	1.00
Assume beginning trades	12.51	-51.4%	0.96
Assume ending trades	9.69	0.0%	0.88
Algorithm			
Directional trades only	0.83	82.7%	4.77
Allow round-trip trades	0.47	82.6%	4.86

Table 3: By-group evaluation based on Ancerno data. This table reports the performance evaluation of our intro-quarter trading detection partitioned on fund characteristics. We sort our testing sample, i.e., 10,000 fund quarters from Ancerno data, by fund number of holdings, fund turnover, and proportion of round-trip trading volume in Panel A, B, and C, respectively. The evaluation measures are defined the same as before. We report the performance from (1) a constrained model allowing only directional trades, and (2) a full model allowing both directional and round-trip trades.

Panel A: Group by fund number of holdings							
	# Obs.	Tracking Error		Holding-based R^2		Trading Match	
		Direc. Only	Allow R.Trip	Direc. Only	Allow R.Trip	Direc. Only	Allow R.Trip
[10, 20]	460	3.21	1.93	90.2%	90.3%	7.20	7.52
(20, 40]	1,644	1.54	0.90	89.8%	89.9%	6.42	6.57
(40, 60]	1,918	1.01	0.54	86.1%	86.0%	5.36	5.48
(60, 80]	1,304	0.64	0.35	84.1%	84.0%	4.70	4.77
(80, 100]	1,035	0.50	0.25	81.6%	81.5%	4.22	4.27
(100, 150]	1,807	0.34	0.20	79.0%	79.0%	3.84	3.87
(150, 200]	733	0.27	0.18	76.4%	76.3%	3.54	3.56
$>$ 200	1,099	0.17	0.11	72.4%	72.3%	3.23	3.23
Panel B: Group by fund turnover							
	# Obs.	Tracking Error		Holding-based R^2		Trading Match	
		Direc. Only	Allow R.Trip	Direc. Only	Allow R.Trip	Direc. Only	Allow R.Trip
[0, 0.1]	487	0.19	0.16	90.2%	90.4%	7.22	7.29
(0.1, 0.2]	653	0.29	0.25	91.0%	91.0%	6.65	6.70
(0.2, 0.3]	884	0.43	0.36	90.2%	90.3%	6.36	6.42
(0.3, 0.5]	1,759	0.50	0.37	87.9%	87.9%	5.50	5.56
(0.5, 0.7]	1,469	0.62	0.40	84.5%	84.6%	4.71	4.78
(0.7, 1.0]	1,822	0.73	0.44	82.1%	82.1%	4.18	4.26
(1.0, 1.5]	1,487	1.06	0.53	76.2%	76.2%	3.71	3.81
$>$ 1.5	1,439	2.06	0.93	70.8%	70.4%	3.14	3.29
Panel C: Group by proportion of round-trip trading volume							
	# Obs.	Tracking Error		Holding-based R^2		Trading Match	
		Direc. Only	Allow R.Trip	Direc. Only	Allow R.Trip	Direc. Only	Allow R.Trip
0	2,152	0.55	0.55	93.2%	93.2%	7.44	7.44
(0, 0.05]	1,448	0.42	0.36	87.4%	87.4%	5.32	5.33
(0.05, 0.1]	1,452	0.47	0.32	83.9%	83.9%	4.57	4.61
(0.1, 0.15]	1,284	0.66	0.40	81.2%	81.3%	4.13	4.21
(0.15, 0.2]	951	0.69	0.34	76.7%	76.6%	3.79	3.87
(0.2, 0.25]	793	0.99	0.46	78.5%	78.4%	3.57	3.70
(0.25, 0.3]	573	0.90	0.39	75.2%	75.1%	3.28	3.42
$>$ 0.3	1,347	2.26	0.84	70.7%	70.5%	2.78	3.07

Table 4: High-ESG vs. Low-ESG Trading. This table reports mutual fund trading of high-ESG and low-ESG stocks at quarter ending and beginning over the period from 2015Q1 to 2022Q2. We report abnormal trading on the 1st, 2nd, and 3rd week at quarter ending or beginning, which are the coefficients $\{b_{E,j}, b_{B,j}\}_{j=1}^{3}$ of the following regression: $y_{i,t,l} = b_0 + \sum_{j=1}^{3} b_{E,j} \times \mathbb{I}_{E,j} + \sum_{j=1}^{3} b_{B,j} \times \mathbb{I}_{B,j} + \gamma \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}$. The dependent variable, $y_{i,t,l}$, is high-ESG or low-ESG stocks' trading volume divided by total trading volume for fund i in quarter t and week l. To calculate high-ESG or low-ESG stocks' trading volume, stock-level trading volume in each week is estimated using our proposed method. 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 Sustainalytics, MSCI, and Refinitiv in quarter $t-1$. The key explanatory variable, $\mathbb{I}_{E,j}$ or $\mathbb{I}_{B,j}$, is 0/1 indicator of the jth week of the quarter ending or beginning. To control for the imbalance between buy and sell in each window, $buy\_ratio_{i,t,l}$ is defined as total buying volume divided by total trading volume for fund i in quarter t and week l. $\alpha_{i,t}$ is fund × year × quarter fixed effects. In Panel A, columns are grouped by high-ESG trading, low-ESG trading, and their difference. Among each trading category, we separately report buy, sell, and net trading (i.e., buy minus sell). Rows are grouped by quarter ending, beginning, and their difference. In Panel B, we conduct placebo tests around non-quarter month end, i.e., month end except Mar/Jun/Sep/Dec. The dependant variables are multiplied by 100. 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-ESG - Low-ESG
	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 - Beg. 1st	1.100***	0.615***	-0.485***	-0.112	-0.130*	-0.019	1.211***
	[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	Net High-ESG - Low-ESG
Ending - Beginning			
End. 1st - Beg. 1st	-0.080	-0.052	-0.028
	[-0.57]	[-1.10]	[-0.19]

Table 5: ESG Window Dressing and Sustainability Rating. This table reports mutual funds' ESG trading around quarter ending grouped by fund past sustainability rating. Due to rating data availability, the sample period is from 2019Q1 to 2022Q2. Fund sustainability rating is from Morningstar, where funds are classified into 5 categories sorted by sustainability scores calculated from 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" groups. For example, the lower (upper) half of the 5-globe rating corresponds to a percentage ranking between 90% and 95% (above 95%). In the first 2 columns, we report abnormal ending-minus-beginning net ESG trading for each rating group $x \in \{1L, 1U, \cdots, 5L, 5U\}$, which is the coefficient difference, $b_{x,E1} - b_{x,B1}$, of the following regression: $y_{i,t,l} = b_0 + \sum_{x=1L,\ldots,5U} [b_{x,E1} \times \mathbb{I}_{x,E1} + b_{x,B1} \times \mathbb{I}_{x,B1} + \gamma_x \times buy\_ratio_{i,t,l}] + \alpha_{i,t} + \epsilon_{i,t,l}$. The dependant variable, $y_{i,t,l}$, is ESG stocks' net trading volume (i.e., buy minus sell) divided by total trading volume for fund i in quarter t and week l. ESG stocks are defined the same as before. The key explanatory variables, $\mathbb{I}_{x,E1}$ and $\mathbb{I}_{x,B1}$, are 0/1 indicators of last week and first week, respectively, of a quarter for rating group x. When determining fund rating group x, we use the average of fund sustainability ratings in quarter $t-1$ and $t-2$. In the third column, we estimate abnormal ending-minus-beginning net ESG trading as a piecewise linear function of sustainability percentage ranking. Specifically, we estimate the following regression $y_{i,t,l} = b_0 + f_E(p) \times \mathbb{I}_{E1} + f_B(p) \times \mathbb{I}_{B1} + g(p) \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}$, where $f_E(p)$, $f_B(p)$, and $g(p)$ are piecewise linear functions partitioned on rating categories, and p is percentage ranking from 0% to 100%. We report the slopes of $f_E(p) - f_B(p)$ within each rating category. The dependant variables are multiplied by 100. t-statistics, shown in brackets, are calculated via bootstrap with 500 replications. *, **, *** denote significance at the 10%, 5%, and 1% level, respectively.

	Sort within each rating category		Piecewise-linear Function
Sustainability Rating	Lower Half (L)	Upper Half (U)	Slope
1 Globe	0.676	0.654	0.025
	[0.94]	[0.89]	[0.20]
2 Globes	0.671	1.279***	0.012
	[1.52]	[2.72]	[0.36]
3 Globes	1.105***	0.981***	-0.009
	[3.07]	[2.76]	[-0.46]
4 Globes	1.006**	2.933***	0.099**
	[1.98]	[4.93]	[2.49]
5 Globes	2.956***	1.352	-0.326*
	[3.29]	[1.09]	[-1.67]
Diff.		Slope Diff.	
(5U - 5L) - (4U - 4L)	-3.531**	(5 - 4)	-0.425*
	[-2.03]		[-1.95]

Table 6: 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 over the period 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 percent. 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-ESG - Low-ESG	
	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]

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 ESG - Low-ESG	-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 7: Momentum window dressing and portfolio pumping. We test fund momentum window dressing and portfolio pumping using the same regression specification as Eq.(11) with dependant variable changes accordingly. The sample period is from 2000 to 2022Q2. In Panel A, for each month t, winner (loser) stocks are defined as the top 10% (bottom 10%) stocks sorted by the cumulative return from month $t-12$ to $t-2$. In Panel B, for each fund quarter, we sort stocks by holding value in descending order among the stocks with positive holding values in quarter-beginning holdings. Top-position (bottom-position) stocks are defined as the top (bottom) stocks cumulatively accounting for 10% of total holding values. 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: Momentum window dressing

	Winner Trading			Loser Trading			Winner - Loser
	Net	Buy	Sell	Net	Buy	Sell	Net
Ending							
End. 1st week	0.192	-0.211	-0.403***	-0.635***	-0.461***	0.175***	0.827***
	[1.14]	[-1.46]	[-3.81]	[-9.99]	[-8.95]	[3.85]	[4.42]
End. 2nd week	0.354***	0.077	-0.277***	-0.423***	-0.223***	0.200***	0.777***
	[2.74]	[0.61]	[-2.79]	[-7.17]	[-4.78]	[4.47]	[5.05]
End. 3rd week	0.107	0.019	-0.089	-0.280***	-0.035	0.245***	0.387***
	[0.96]	[0.15]	[-0.91]	[-6.41]	[-0.90]	[5.94]	[3.08]
Beginning							
Beg. 1st week	-0.492***	-0.276**	0.216	0.539***	-0.013	-0.552***	-1.032***
	[-2.85]	[-2.31]	[1.41]	[8.46]	[-0.25]	[-10.98]	[-5.27]
Beg. 2nd week	-0.423***	-0.319***	0.104	0.227***	-0.080*	-0.307***	-0.650***
	[-3.05]	[-3.15]	[0.75]	[5.76]	[-1.96]	[-6.43]	[-4.20]
Beg. 3rd week	-0.307**	0.010	0.316**	0.106**	-0.001	-0.108**	-0.413***
	[-2.22]	[0.10]	[2.36]	[2.04]	[-0.04]	[-2.32]	[-2.72]
Ending - Beginning							
End. 1st - Beg. 1st	0.684***	0.066	-0.619***	-1.175***	-0.448***	0.727***	1.859***
	[2.76]	[0.30]	[-2.98]	[-12.06]	[-6.04]	[10.34]	[6.55]

Panel B: Portfolio pumping

	Top-position Trading			Bottom-position Trading			Top - Bottom
	Net	Buy	Sell	Net	Buy	Sell	Net
Ending							
End. 1st week	1.125***	0.819***	-0.305***	-2.175***	-1.535***	0.640***	3.300***
	[15.76]	[18.23]	[-6.47]	[-15.37]	[-20.11]	[7.16]	[17.17]
End. 2nd week	0.788***	0.567***	-0.221***	-1.713***	-1.281***	0.433***	2.502***
	[12.84]	[14.74]	[-5.44]	[-14.56]	[-19.28]	[5.50]	[16.12]
End. 3rd week	0.514***	0.307***	-0.207***	-1.022***	-0.708***	0.314***	1.536***
	[9.93]	[10.87]	[-5.83]	[-10.80]	[-12.92]	[4.56]	[12.46]
Beginning							
Beg. 1st week	-1.614***	-0.328***	1.287***	1.805***	0.541***	-1.264***	-3.419***
	[-17.92]	[-14.26]	[16.34]	[11.99]	[6.80]	[-9.75]	[-16.50]
Beg. 2nd week	-0.801***	-0.242***	0.558***	1.518***	0.896***	-0.622***	-2.318***
	[-11.00]	[-9.33]	[8.73]	[12.85]	[15.52]	[-6.32]	[-15.02]
Beg. 3rd week	-0.326***	-0.139***	0.187***	0.547***	0.720***	0.173**	-0.873***
	[-4.14]	[-5.76]	[2.70]	[5.09]	[12.73]	[2.27]	[-5.64]
Ending - Beginning							
End. 1st - Beg. 1st	2.739***	1.147***	-1.592***	-3.980***	-2.075***	1.905***	6.719***
	[22.06]	[21.37]	[-18.25]	[-15.84]	[-18.39]	[10.61]	[19.46]

Table 8: Return predictability from the decomposition of return gap. 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 using the return gap with at least 3-month gap to ensure information publicly available, i.e., at quarter-end month t, each component of return gaps takes average 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]

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]

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]

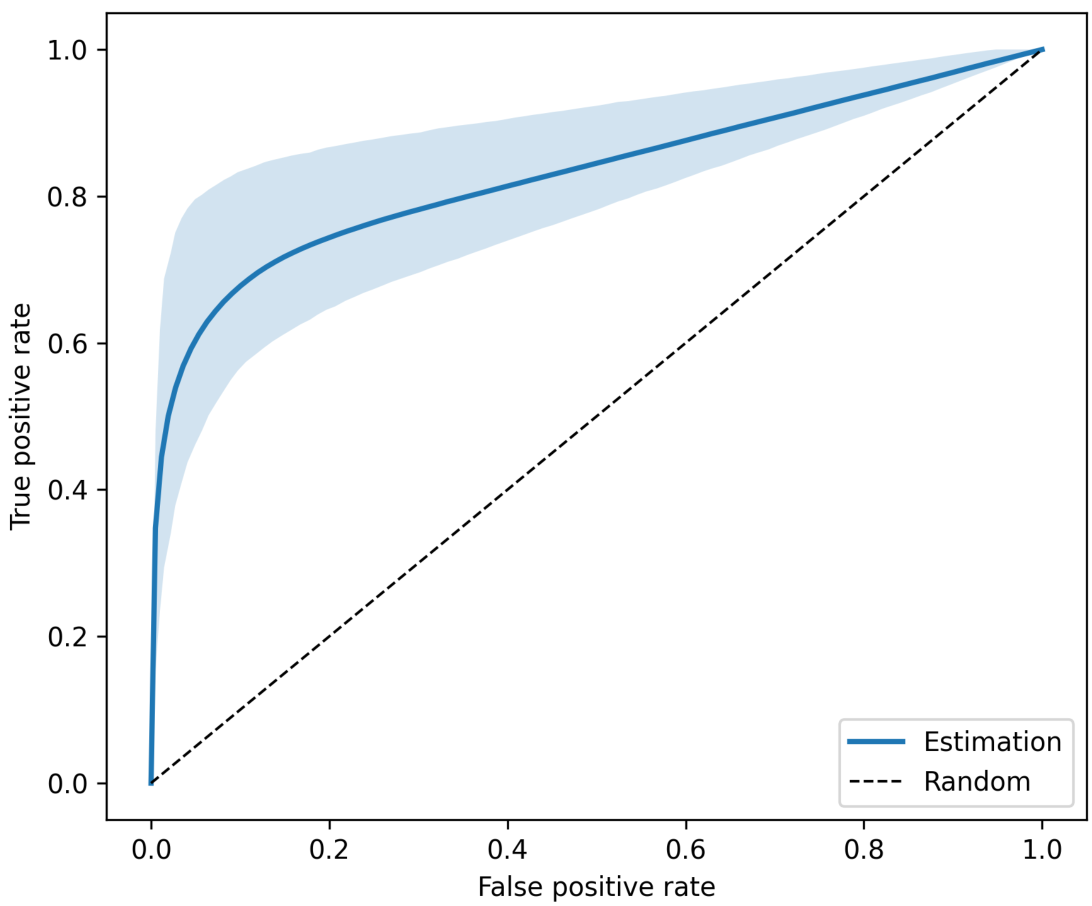

Figure 1: ROC curve of trading detection based on Ancerno data

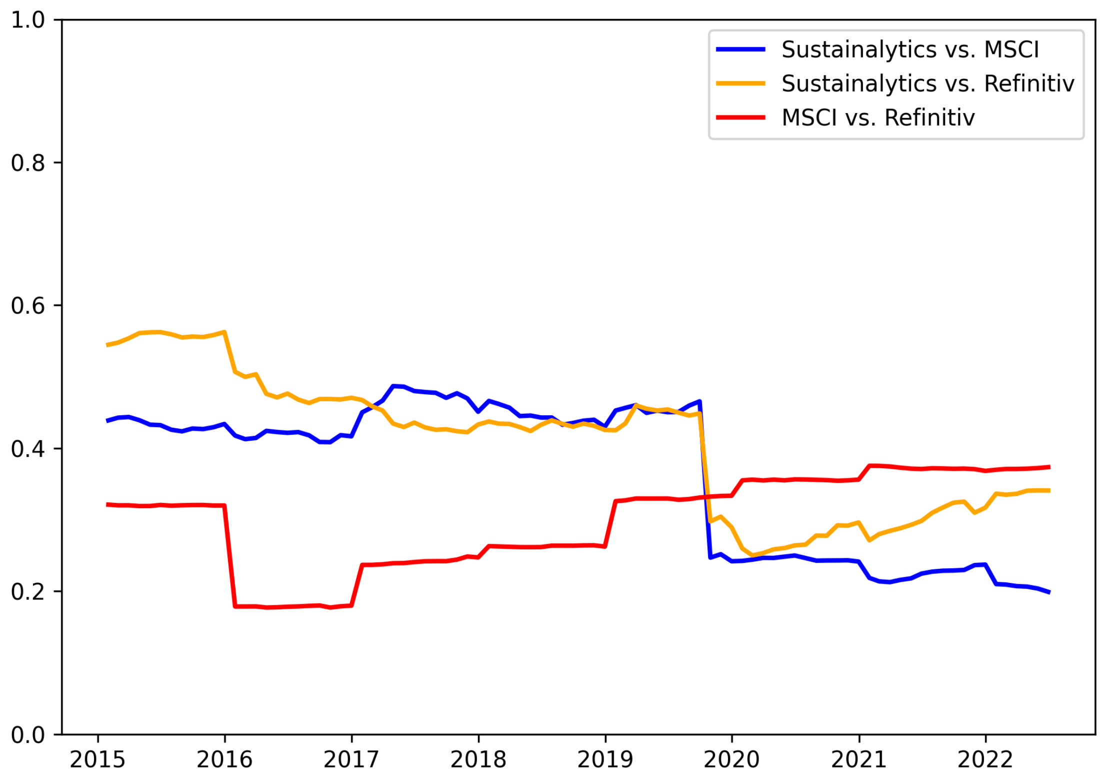

Figure 2: Correlations among different ESG ratings. This figure shows the time series of Pearson correlations among three stock ESG ratings, i.e., Sustainalytics, MSCI, and Refinitiv, over the period from 2015Q1 to 2022Q2. To ensure comparability, the correlations are calculated based on the percentage ranking of ESG scores. Blue, orange, and red lines denote the correlations between Sustainalytics and MSCI, Sustainalytics and Refinitiv, and MSCI and Refinitiv, respectively. The average of these correlations (i.e., across both time series and the three pairs) is 0.357.

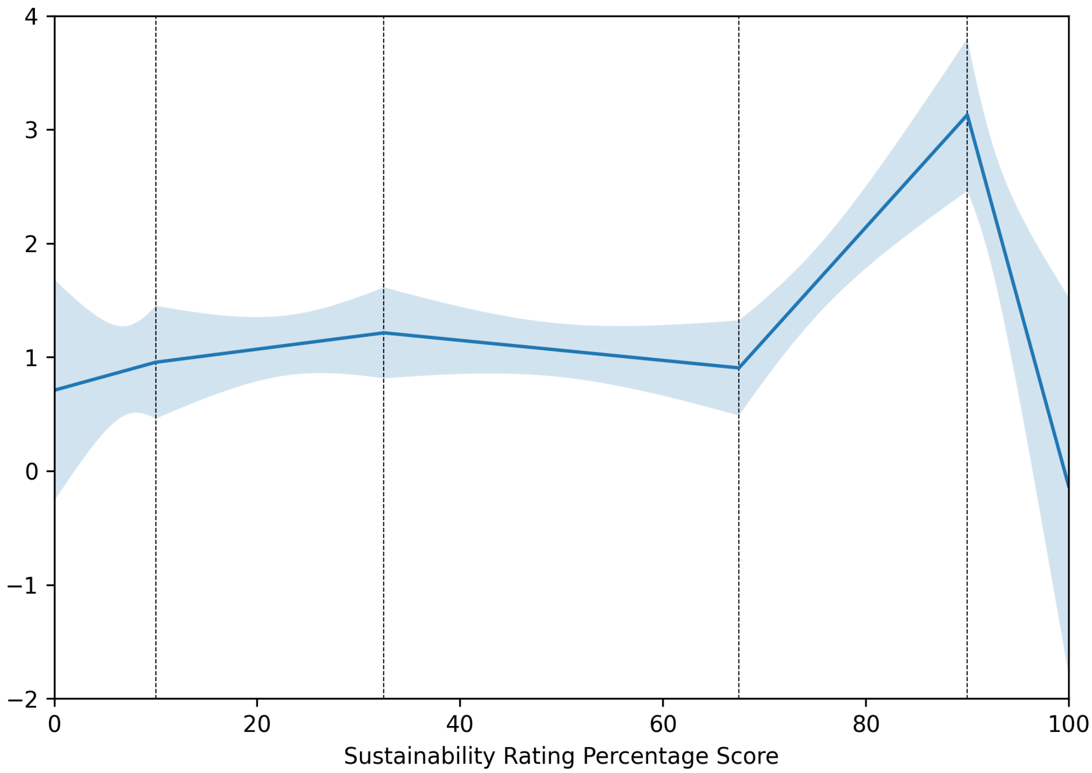

Figure 3: ESG window dressing as a function of sustainability rating. This figure is to visualize the estimation in Table 5 Column "Piecewise-linear Function". The y-axis is abnormal ending-minus-beginning net ESG trading in percentage, and the x-axis is sustainability percentage ranking from Morningstar. A higher (lower) sustainability rating percentage score indicates better (worse) ESG performance. Morningstar grades funds into 5 categories: 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%. According to the rating cutoff, we estimate a piecewise linear function via the following regression: $y_{i,t,l} = b_0 + f_E(p) \times \mathbb{I}_{E1} + f_B(p) \times \mathbb{I}_{B1} + g(p) \times buy\_ratio_{i,t,l} + \alpha_{i,t} + \epsilon_{i,t,l}$, where $f_E(p)$, $f_B(p)$, and $g(p)$ are piecewise linear functions partitioned on rating categories, and p is percentage ranking from 0% to 100%. We plot the estimated function of $f_E(p) - f_B(p)$. The solid line is the point estimation, and the shadow area is the one-standard-deviation error band. Standard errors are calculated via bootstrap with 500 replications.

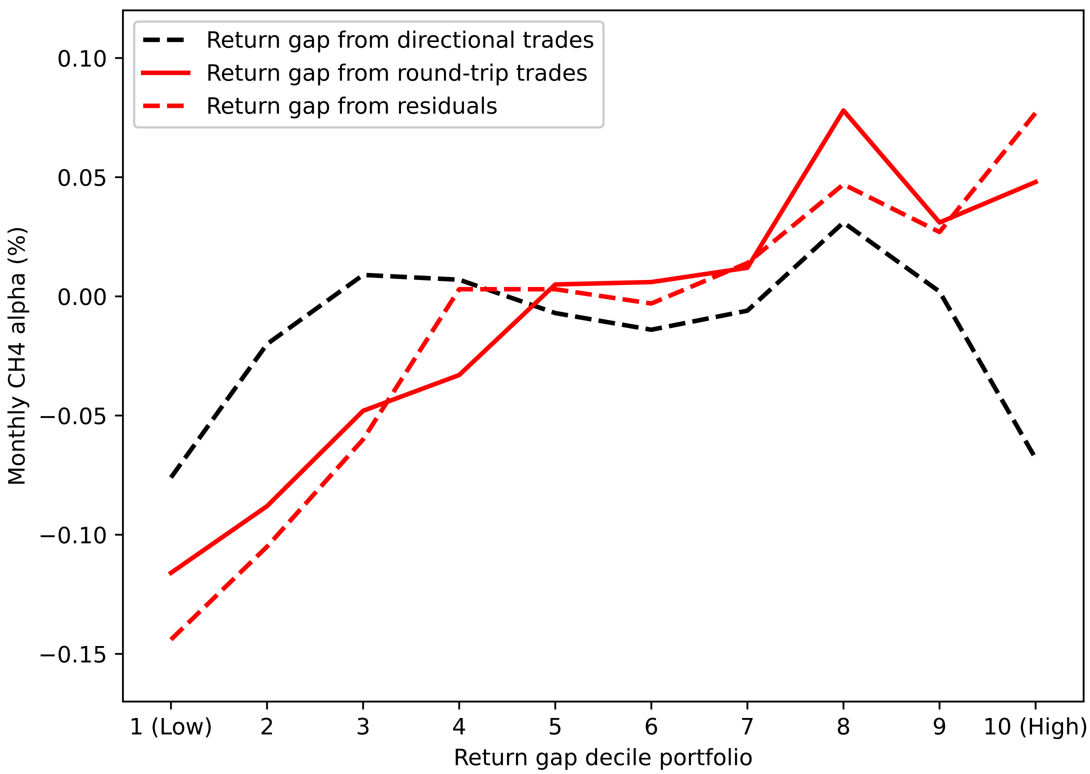

Figure 4: Decomposition of return gap predictability. This figure shows the mutual fund monthly risk-adjusted returns from the model in Carhart (1997) after sorting by each component of the return gap over the period between 2001 and 2022Q2. Based on our trading detection method, we decompose the return gap in Kacperczyk, Sialm, and Zheng (2008) into three components related to directional trades, round-trip trades, and residuals, respectively. We sort funds into 10 groups at the end of each quarter using each component of the return gap with at least a 3-month gap to ensure information is publicly available.