Window Dressing among Responsible Investment Funds

Huiqiong Tang[1], Bart Frijns[2], Aaron Gilbert[3], Ayesha Scott[4]

Abstract

Window dressing is a strategy fund managers use to manipulate portfolio holdings for risk level, financial performance, and investment style to allow funds to report more favourable information to the investors. This paper examines the presence of financial and ESG performance-based window dressing in US domestic equity responsible investment funds (RIFs). Our results support the existence of financial and ESG performance-based window dressing in RIFs. We also identify that RIFs with poor past performance, higher tracking error, or those managed by companies with a lower commitment to sustainable investment are more likely to exhibit ESG performance-based window dressing behaviours.

JEL classifications: G10, G11, G12

Keywords: Responsible investment funds, window dressing, ESG performance

[1] (Corresponding Author) Auckland University of Technology, 55 Wellesley Street, Auckland 1010, New Zealand. E-mail: huiqiong.tang@aut.ac.nz.
[2] Open Universiteit, 177 Valkenburgerweg, Heerlen 6401 DL, Netherlands. E-mail: bart.frijns@ou.nl.
[3] Auckland University of Technology, 55 Wellesley Street, Auckland 1010, New Zealand. Tel: +64 9 921 9999 ext 5713. E-mail: aaron.gilbert@aut.ac.nz.
[4] Auckland University of Technology, 55 Wellesley Street, Auckland 1010, New Zealand. Tel: +64 9 921 9999 ext 5617. E-mail: ayesha.scott@aut.ac.nz.

Acknowledgements: We thank Jacquelyn Humphrey, for her insightful comments on an earlier version of the manuscript, and the AUT Finance Department seminar attendees for their feedback.

1. Introduction

Responsible Investing seeks to promote Environmental, Social, and Governance (ESG) factors in its investment objectives (Mallin et al., 1995). Sustainable investment has shown considerable growth in recent years with global assets under management in responsible investments increasing 15% between 2018-2020 to reach US$35.3 trillion (equating to 36% of all professionally managed assets) at the beginning of 2020 (Global Sustainable Investment Alliance, 2021). A considerable fraction of this capital is managed through Responsible Investment Funds (RIFs) – managed funds that consider the ESG attributes of investments when making investment allocations.[5]

As sustainable investment grows, so too does academic interest in RIFs. Extant literature predominantly focuses on the financial performance of RIFs versus conventional funds (van Dijk-de Groot & Nijhof, 2015). Much of this literature finds that RIF investors do not sacrifice returns to achieve their non-financial goals (Friede, Busch & Bassen, 2015; Bialkowski & Starks, 2016), despite RIFs having fewer diversification opportunities. Currently, only a few regulations exist to govern the ESG strategies of RIFs. Given the assumption RIF investors are more likely to consider ESG and the evidence that a smaller investment universe does not sacrifice performance, it is reasonable – arguably essential – to ask whether RIFs make investment allocations that are different from conventional funds. That is, do RIFs live up to their ESG-incorporation strategies or only state they are “RIFs”? The most direct method to verify our question is to consider fund holdings information, as in Dorfleitner et al., (2012), Utz and Wimmer (2014) and Joliet and Titova (2018). These studies assess disclosed companies held by RIFs to determine their ESG

[5]ESG factors are typically addressed using one or more of several possible strategies, include negative/exclusionary screening, positive/best-in-class screening, norms-based screening, ESG integration, sustainability themed investing, impact/community investing, and corporate engagement and shareholder action (Global Sustainable Investment Alliance, 2021).

performance. However, there are two potential problems with such analysis. First, holdings are only available quarterly,[6] which means holdings between two reporting dates are unknown. Second, most studies rely on funds' self-reported information, but relatively little is known about the veracity of such disclosed information. Managers may be less willing to show investors a portfolio that held poorly performing stocks since investors are at least partially evaluating fund managers based on their disclosed portfolio holdings (Solomon et al., 2014). The interest of investors in fund holdings gives fund managers an incentive to manipulate their holdings close to a reporting date to give a more favourable impression, a behaviour commonly known as 'window dressing' (Agarwal et al., 2014).

Extant literature has shown some fund managers window-dress their holdings with respect to risk level (Morey & O'Neal, 2006), financial performance (Agarwal et al., 2014), and investment style (Meier & Schaumburg, 2006) to show a more favourable portfolio to investors, attract new investors and cash inflow, and avoid losing investors to other funds. We argue that RIF managers could adopt the same practice since RIFs are affected by the same pressures as traditional mutual funds. In addition to altering holdings to better performing, less risky firms, RIFs may also adjust holdings with regard to ESG aspects. Specifically, they may strategically sell (buy) companies with low (high) ESG performance prior to a reporting date, giving a more ethical appearance to investors who care about responsible investment. We refer to this as ESG performance-based window dressing. ESG performance window dressing misleads investors, competitors, and rating agencies and can negatively impact a funds' value due to unnecessary rebalancing costs and

[6]Since 2004, mutual funds in the US are required to disclose their complete portfolio holdings quarterly, with a 60-day delay (Securities and Exchange Commission, 2004).

potentially taking more risk than advertised. However, relatively little research has addressed this issue. This paper attempts to fill this gap.

Our study examines if US domestic equity RIFs window-dress along the ESG dimension. We employ fund holdings information and ESG scores of companies held by funds to proxy for funds' ESG performance. We test for two types of window dressing: financial and ESG performance-based window dressing. We compute the "rank gap" and "backward holding return rap (BHRG)" as in Agarwal et al. (2014) to detect financial performance-based window dressing. The first measure detects the rank gap between performance-based and average rankings based on disclosed holding proportions of winner and loser stocks. The second method captures the difference between a fund's actual return and the hypothetical returns imputed from a fund's reported holdings. Both methods demonstrate that RIFs engage in financial performance-based window dressing.

To assess ESG performance-based window dressing, we follow Fama and French (1992) and construct an ESG factor to capture the return premium on a strategy that is long in high ESG stocks and short in low ESG stocks. By regressing the BHRG on the ESG factor, we find that 68 out of 196, or 34.6%, of RIFs alter their holdings to companies with a higher ESG score close to reporting dates. We also use daily return data and conduct an event study considering windows of 5, 10, 15, and 20 days before the reporting date. Depending on the window, we find between 11.76% and 24.26% of RIFs have a significantly higher loading on the ESG factor just prior to the reporting date. This end-of-quarter rebalancing is more likely to be window dressing than routine holding adjustments, which should be uniformly distributed throughout the entire period. To further support the above finding, we compute daily coefficients based on a Kalman filter. The result from a subsample of funds with significant positive coefficients on the ESG factor indicates the sensitivity of returns to sustainable funds persistently increases 12 days before the disclosure date. This finding shows some RIFs adjust their holdings in accordance with ESG performance-based window dressing immediately prior to the reporting date. We further identify that RIFs with poor past performance, higher tracking error, or those managed by companies with lower sustainable investment levels overall are more likely to engage in ESG performance-based window dressing behaviours.

Our paper offers two contributions to the RIFs and window dressing literature. First, we supplement the RIF literature by assessing ESG performance. RIFs are marketing themselves as investments with higher ESG values. They attract investors who, at least in part, are concerned about the social impacts of their investments and want to improve the world they live in. However, extant studies often assume that RIF investments comply with their ethical principles and therefore do not assess RIFs consistency with their investment objectives (Capelle-Blancard and Monjon, 2012; Nitsche and Schröder, 2018). This paper evaluates the sensitivity of RIF returns to an ethical index and thus provides investors with a broader picture of how these funds' ESG performance is. Second, unlike extant papers that investigate window dressing in the context of risk level or performance-based window dressing, this paper provides unique insights into ESG performance-based window dressing. Such "deceptive" behaviour may mislead investors about their actual investment and the social impact they are generating. In addition, widespread window dressing may render RIFs ineffective in promoting social change. Therefore, it is essential to know whether RIFs engage in ESG performance-based window dressing.

2. Literature Review

2.1 Financial and non-financial return of RIFs

The literature on RIFs has grown rapidly over the past two decades. One strand of literature assesses the financial performance of RIFs compared to either all or characteristics-matched conventional funds and aims to detect a potential performance penalty caused by ethical considerations. However, the comparison results are mixed. The majority of these studies demonstrate the risk-adjusted performance of ethical funds does not differ significantly when compared to conventional funds, either before or after fees (Friede, Busch & Bassen, 2015; Bialkowski & Starks, 2016; Thompson et al., 2011). In contrast to previous literature, a few studies highlight that RIFs are outperforming non-RIFs (Gil-Bazo et al., 2009; Lean et al., 2015; Alda, 2020) and RIFs underperforming non-RIFs (El Ghoul & Karoui, 2017; Ibikunle & Steffen, 2017, Azmi et al., 2020) in specific settings.

Another strand of literature studies the non-financial performance of RIFs. These studies focus on portfolio composition and assessing whether RIFs deserve the label "Responsible". For example, Bello (2005) compares assets held by ethical funds and randomly selected conventional funds and finds no significant difference between these two groups. Similar results have also been reported by Chieffe and Lahey (2009). More recently, Utz and Wimmer (2014) rank all US mutual funds from the Asset4 database into quintiles based on their annual ESG scores and find only 11.58% of RIFs are in the highest quintile, compared with 20.13% for conventional funds. In addition, approximately 36.7% of RIFs lie below the average level of ESG performance of all funds. These figures suggest that RIFs may not be achieving the expected ESG performance. However, contrasting results are obtained by other researchers. For example, Kempf and Osthoff (2008) rank

US equity funds based on their holding stocks' KLD rating and find that RIFs have a significantly higher ethical ranking than conventional funds. Bialkowski and Starks (2016) examine the ESG scores of US RIFs and conventional funds holdings companies between 2002 to 2011 and state that RIFs have higher ESG profiles for five of seven categories.[7] More recently, Nitsche and Schröder (2018) employ the ESG information provided by three rating agencies (Oekom research AG, Sustainalytics and ASSET4) and find RIFs have, on average, higher ESG rankings based on the top 10 fund holdings in the European and Global fund universe. Joliet and Titova (2018) find a similar result in the US. They examine 47 US domestic equity funds between September 2009 and November 2015 and find RIFs have significantly higher ESG scores than conventional funds, on average.

Both types of performance-based analyses lead to a number of subsequent studies that compare RIFs with conventional funds in terms of general fund characteristics, such as size, age, turnover ratio and investment style (Alda, 2020; Ivanisevic Hernaus, 2019); fund manager characteristics, such as age, gender, skill, whether they are group managers or not (Agarwal et al., 2014; Muñoz et al., 2015); the intensity of ethical screens applied by RIFs, such as positive screens, negative screens and best in class (Nofsinger & Varma, 2014; Trinks & Scholtens, 2017); investor behaviour (Amel-Zadeh & Serafeim, 2018; Riedl & Smeets, 2017); and market states, crisis and non-crisis periods (Muñoz et al., 2015; Nofsinger & Varma, 2014). However, although these studies have addressed both the financial and non-financial performance of ethical funds, they tend to use announced information by funds and assume reported holdings accurately represent the

[7] MSCI ESG research rating covers seven categories of social responsibility: community, corporate governance, diversity, employee relations, environment, human rights and products. Bialkowski and Starks (2016) find that RIFs offer higher positive exposures to these categories, except community and diversity.

entire investment period, which may not be the case in practice. This paper aims to discuss this point and expand the related existing knowledge of RIFs.

2.2 Window dressing

Window dressing has occupied many researchers over the years. One of the pioneer works is Haugen and Lakonishok (1987), who argue window dressing may be an alternative explanation for the January effect. Since then, considerable empirical research has been conducted to test whether fund managers are “dressing up” their portfolios. According to Agarwal et al. (2014), window dressing behaviour may be caused by potential agency problems between mutual fund companies and their investors. Investors would like the fund company to maximize risk-adjusted fund returns within the constraints of the company's stated investment style and objective. The fund manager is seeking to maximize their fees by maximizing its assets under management. Ideally, a fund manager can satisfy both objectives by earning high returns, achieving investors’ goals, thus attracting new investors and thereby rapidly increasing the assets under management and their manager fees. Attracting new funds is the quickest way to increase the size of the fund and requires a fund to at least appear to outperform its peers. At the very least, managers want to retain their existing funds meaning they need to avoid looking worse than their peers. According to Morey and O'Neal (2006), in the case of poor past performance, investors are more likely to stay if the fund is holding recent top-performing securities. For attracting new fund flows, Elton et al. (2011) state that by changing holdings to better-performing securities, investors may get the impression that the fund is holding winners and infer that this fund held winner stocks that have superior performance. Therefore, window dressing may help managers to retain existing or attract new fund flows and thus generate additional fee income and even consequently increase their standing and job security in the fund family (Cici et al., 2021). Window dressing, therefore, can allow managers with either poor performance or deficient managerial skills to appear better than they are giving them a stronger incentive to engage in window dressing (Hung et al., 2020; Agarwal et al., 2014).

The existing literature has discussed two types of window dressing. The first type occurs due to performance-chasing behaviour. Managers try to sell poorly performing companies and/or buy winner stocks before the holding disclosure date to “make up” their portfolio. For example, O'Neal (2001) investigates 195 US equity mutual funds’ 6-month window rolling return and states that December (just before required disclosure time) exhibits the strongest evidence of window dressing. Similar tests have been done by Meier and Schaumberg (2006). They investigate 4,025 US domestic equity mutual funds between 1997 to 2002 and examine the difference between the actual fund return with a hypothetical buy-and-hold portfolio based on disclosed holdings. They find strong evidence of window dressing during the last days of the quarter. More recently, Ortiz et al. (2015) and Hung et al. (2020) extend the literature by studying Spain and Taiwan, respectively. They both find that equity funds attempt to increase the weights of return-winner and decrease the return-loser stocks in the disclosure months. Ortiz et al. (2015) also find that non-disclosure months show the opposite trend, which supports the existence of window dressing. Another interesting finding discussed by Solomon et al. (2014) is that investors respond positively to funds which include media-covered past winner stocks, although the premium return of winner stocks may not be captured by these funds.

The second type of window dressing happens when the managers try to change the risk characteristics of a fund. Before the disclosure date, managers may purchase less risky stocks and/or decrease high-risk assets to show a less-risky portfolio to attract investors who prefer a safer investment. For instance, Chevalier and Ellison (1997) find that growth and income funds modify their holdings in the last quarter of the year to alter the riskiness of their investment portfolio. Similar results have been found in bond funds. Morey and O'Neal (2006) assess portfolio credit quality holding and daily return of US corporate bond funds between 1998 and 2001. They point out that bond funds invest in significantly more government bonds during reporting periods compared with non-reporting periods, presumably to disclose a less risky portfolio to investors. More recently, Patton and Ramadorai (2013) investigate 14,194 hedge funds and funds of funds between 1994 and 2009, and they find that hedge funds exhibit significant day-of-month seasonality changes on risk exposure, which may be caused by intra-month window dressing.

Both types of window dressing may have negative consequences for fund investors. If fund managers window-dress their holding portfolio, the investors are not only misled about their investment but are also bearing unnecessary transaction costs. O'Neal (2001) estimates that within the American equity funds market, the annual costs attributable to window dress portfolio rebalancing may exceed $1 billion,[8] an economically meaningful sum. Agarwal et al. (2014) also state that due to unnecessary transactions, window dressing may potentially impact fund value adversely.

2.3 Window dressing in RIFs

Although literature documents evidence of window dressing behaviour in mutual funds, limited attention has been given to RIFs, who typically have constraints on their investment universe. In addition to performance and risk exposure-based window dressing, RIFs may exhibit another type of window dressing, ethical performance window dressing. RIF fund managers may change their holdings from socially-conscious investments to some high performance but low-ESG

[8] Funds incur costs with every buy and sell trade conducted as a result of the bid-ask spread.

performance industries between two reporting dates. A few papers mention window dressing in RIFs. For example, Elaut et al. (2015) state that because of potential window dressing, the use of ethical funds' current holdings to assess RIFs' performance may lead to upward biased results[9]. However, to date, limited work has been done in this field.

Kempf and Osthoff (2008) mention ethical rankings of RIFs are higher than conventional funds, and they state that this superior ethical ranking is not caused by window dressing. In their paper, they conduct two tests to rule out the impacts of window dressing. The first measure compares the ethical ranking of mid-year reported holdings and end-year reported holdings. They find the difference is not statistically significant and argue there is no evidence for window dressing. The main drawback of this measure is that it assumes the mid-year reporting is not affected by window-dressing, which may not be the case. The second test detects window dressing by analysing fund performance against conventional and ethical indices. They assume that if window dressing exists, the sensitivity of fund returns is expected to be higher shortly before the reporting date than at other time periods. However, this method is impacted by data availability. They only test this method on 66 RIFs between 2001 and 2004 (the original sample is between 1998 and 2004) based on semi-annual holding data and find no significant outcomes. Additionally, it is worth noting that the sample period of Kempf and Osthoff (2008) represents the early days of responsible investing, resulting in a small number of funds that were likely committed to the goals of ethical investing. Since 2004 there have been significant increases in providers, including more traditional fund families offering responsible products as a portion of their total offered funds. This raises the possibility that window dressing, which has been extensively documented in conventional funds,

[9] This is because the current holdings introduce a look-ahead bias. At the same time, the window dressing may reveal a portfolio that performed well historically, potentially leading to an overestimated performance.

may have become more prevalent in RIFs as a result of the introduction of more conventionally managed funds. Therefore, this paper is going to re-examine their method by employing a more frequent[10], representative[11] and recent RIF sample.

Relatively less work has been done in ESG performance-based window dressing (altering fund holdings for a better ESG score before reporting date) compared to performance-based window dressing. However, several papers examine a similar concept: greenwashing, commonly defined as manipulating and disseminating environmental information to mislead the public (Lyon & Maxwell, 2011). This term is also related to the active management of misleading information and selective disclosure in the mutual fund field. However, compared with ESG performance-based window dressing in RIFs, greenwashing is more related to environmental performance. According to Delmas and Burbano (2011), greenwashing is derived from consumer and investor demand, incentive structures, ethical climate and an optimistic bias. The existence of greenwashing may reduce consumer confidence and undermine the green investment market. Several papers have detected the existence of greenwashing in RIFs. For example, Findlay and Moran (2018)[12] find evidence some funds are presented as impact investments but do not fulfil their announced definition. More recently, Brandon et al. (2021) discovers greenwashing behaviour in US RIFs. Examining greenwashing reflects researchers' growing interest in assessing the integrity of ethical

[10] The Securities and Exchange Commission (SEC) required US mutual funds to disclose their complete portfolio holdings to shareholders on a quarterly basis (with a 60-day delay period) after 2004. The SEC states that this change aims to ask funds to provide more information for investors to assess the fund strategies and managers' skills, as well as to reach a better monitor function on some market manipulations, such as window dressing and portfolio pumping. While the opinion on the value of this change is still mixed (see Parida & Teo, 2018; Parida, 2017; Schwarz & Potter, 2016; Gormley et al. 2019; Frank et al. 2004). We are going to take advantage of this change, employing the quarterly holding date to analyse window dressing in RIFs.

[11] Kempt and Osthoff (2008) identify RIFs based on Morningstar, while our research will identify RIFs base on the USSIF report, which is a comprehensive report specialised in the responsible investment industry in the US

[12] Findlay and Moran (2018) refine the definition of greenwashing to purpose-washing by emphasising intentionality. They defined the purpose-washing occurs when investors are misled about a manager's impact intentions (including measurement) or an investment's potential impact.

funds. This paper will further explore this criticism faced by RIFs and assess if they are living up to their stated ESG goals.

3. Hypotheses

The extant literature suggests that window dressing is a strategy fund managers implement to attract investors via more favourable holding information. Responsible investors seek two sorts of benefits: financial benefits and benefits related to non-financial consider ions (Levitt & List, 2007). Both benefits impact investors' decisions to invest responsibly (Døskeland & Pedersen, 2016). To accomplish investors' dual objectives and show favourable holdings to investors, RIF managers may engage in two types of window dressing. First, for investors who are primarily driven by financial performance, and for whom ESG is a secondary but important consideration, RIFs managers may engage in financial performance window dressing. Specifically, they might attempt to increase (decrease) the proportion of winner (loser) stocks in reported holdings. Second, for pure responsible investors, who prioritize ESG performance over financial returns, fund managers have incentives to distort holdings to companies that have a better ESG score. That is, *ESG* performance-based window dressing.

To test the existence of the two possibilities above, we propose the following two hypotheses:

Hypothesis 1: RIFs exhibit financial performance-based window dressing

Hypothesis 2: RIFs exhibit ESG performance-based window dressing

4. Data

To investigate window dressing by RIFs, we obtain data from three sources. The Forum for Sustainable and Responsible Investment (US SIF) reports provide lists of RIFs. USSIF is a non-profit membership association that focuses on sustainable investment practices in America and is one of the most widely used providers of RIF information in the literature (Benson & Humphrey, 2008; Humphrey et al., 2015; In et al., 2014). Morningstar Direct is used to retrieve information about fund holdings, fund returns, and other characteristics, such as the fund management company, share classes, net assets, and the expense ratio. Information on stock prices and ESG scores of the companies held by RIFs is obtained from Thomson Reuters Eikon[13]. In line with the existing literature, we identify RIFs based on the fund list provided in the USSIF trend reports between 2007-2020.[14] To avoid survivorship bias, RIFs are kept in the sample once they appear in one of these report lists. From reports, we initially identify 1,193 different share classes.[15]

Since this paper focuses on US equity RIFs, we exclude balanced, bond, and global money market funds (Agarwal et al., 2014; In et al., 2014; Kempf & Osthoff, 2008). Based on Morningstar Global Category, funds belonging to US Equity Large Cap Blend, US Equity Large Cap Growth, US Equity Large Cap Value, US Equity Mid Cap and US Equity Small Cap remain in the sample. After filtering for fund category, the sample size is decreased to 747 different share classes. This

[13] The scores are currently published by Refinitiv.
[14] As the 2020 report is not published as we started to collect data. The 2020 list is derived from the US SIF website.
[15] Morningstar provides funds information at the share classes level, which means the different classes of the same funds are treated as separate observations. To avoid double-counting, the share class level information has been aggregated to fund level in the later part of the analysis.

list includes not only active classes but also liquidated and merged classes to avoid survivorship bias (Kempf & Osthoff, 2008).[16]

Some funds in our sample have multiple share classes. The main differences between the share classes are their loads and expense ratios, while portfolio holdings remain the same (Alda et al., 2020; Doshi et al., 2015; Humphrey et al., 2015; Ibikunle & Steffen, 2017; Kurniawan et al., 2016; O'Neal, 2001). Following the existing literature, we aggregate each share class at the fund level based on share-class total net assets to obtain the value-weighted monthly return and annual expense ratio. The 747 different share classes have therefore been aggregated into 216 funds.

Then, consistent with Agarwal et al. (2014), funds with less than 24 monthly returns within the sample period are excluded. We also exclude funds where less than 75% of their holdings can be successfully matched with stock information from TR Eikon[17] (Borgers et al., 2015). This cleaning process results in a final sample of 196 RIFs between January 2005 and December 2019. Fund level descriptive statistics are presented in Table 1. The average fund is about 18 years (215 months) old, has US$3672 million of assets under management, and has a turnover ratio of 55%.

[16] Following the work of Alda et al., (2020), we use information provided by Morningstar under the labels "Obsolete type", "Obsolete date", "Merged into Security", and "Merged into Security ID". The first item indicates if the share class has been liquidated or merged with others. The second item indicates the date of liquidation or merge. The rest two indicate the acquiring fund name and ID, which can tell me if the mergers happened within-family or across families. There are 119 share classes liquidated between 2005-2020. 126 share classes merged with another fund. Within these merged share classes, 55 share classes merged into other share classes of the same. 57 share classes merged into different funds, but the target funds still include in the RIF list. 14 share classes merged into different funds, but the target funds are not included in the RIF list. For liquidated and merged classes, we keep them in sample until their obsolete date.

[17] The portfolio holding downloaded from Morningstar Excel Add-in only includes company name, SECID and proportion held by funds. The identifier SECID is only used in Morningstar. We begin by downloading all the listed US companies' price information and identifier (such as ISIN and Ticker) from Eikon. We then use the identifiers from Eikon to match firms with Morningstar Direct and then use the SECID and companies name provide by Morningstar Direct to match downloaded portfolio holding.

[Insert Table 1 about here]

5. Methodology and results

5.1 Financial performance-based window dressing

Agarwal et al. (2014) develop the "rank gap" method to measure performance inconsistency and use this as a relative measure of window dressing. This method has been widely used in recent papers (Hung et al., 2020; Marques et al., 2020; Cici, et al., 2021). The rank gap is calculated based on the following steps. First, the returns of all listed companies are ranked at the end of each month. Second, we allocate all companies to quintile portfolios. The first (fifth) portfolio includes companies with the highest (lowest) return. Third, we use the list of companies belonging to the first and fifth quintile portfolios. Based on the quarterly reported holdings[18] of each fund, we calculate the proportion of each fund's assets invested in the first and fifth quintile stocks. The two proportions are defined as the winner and loser proportions. Fourth, for each month and each fund that reported holdings, we compute three alternative rankings: 1. we rank all funds based on their past return (the highest return is given the highest rank); 2. we rank all funds based on their winner proportion (the highest proportion is given the highest rank); 3. we rank all funds based on their loser proportion (the lowest proportion is given the highest rank). A higher rank means that a fund has a higher return, a larger proportion of its investments in higher return companies and a smaller proportion in lower return companies. In the spirit of Agarwal et al. (2014), a fund with a low-performance ranking but a high ranking in the winner and loser proportion is more likely to engage

[18] Most funds in sample report their holdings quarterly. We use current quarterly end reported holdings and assume that the holdings keep consistent from the beginning of the current quarter.

in window dressing. The last step is to calculate the *rank gap*, the difference between performance ranking and average proportion ranking. Agarwal et al. (2014) scales the difference by dividing by 200 to produce a theoretical bound around the *rank gap* between -0.495 and +0.495 (as shown in equation 1). The larger the *rank gap* is, the greater performance inconsistency is, and thus the higher the likelihood that window dressing is occurring.

$$rank\ gap_{i,t} = \left(performance\ rank_{i,t} - \frac{winner\ rank_{i,t} + loser\ rank_{i,t}}{2}\right) \div 200 \qquad (1)$$

Table 2 reports summary statistics for the rank gap and the percentage of loser and winner holdings. The average monthly rank gap is 0.0021, and the median is about 0.0025.[19] These figures indicate that financial performance-based window dressing exists in our sample. Although the average is near zero, the high standard deviation of 0.1077[20] and the large range (-0.46, 0.48) indicate that the level of window dressing across RIFs varies. The mean percentage of loser stocks is 4.9%, and the corresponding standard deviation is 4.87. The mean percentage of winner stocks is 15.28%, and the corresponding standard deviation is 9.16. On average, RIFs hold a higher proportion of winner stocks than loser stocks.

[Insert Table 2 about here]

[19] Mean and median in our study are larger than that in Agarwal et al. (2014). Their mean and median are -0.0003 and -0.0025, respectively. In a more recent paper, Bai et al. (2019) state their mean and median rank gap are 0.00 and -0.01. Both existing papers discuss US equity funds, while our sample only includes RIFs. To the best of our knowledge, no other paper assesses rank gap in RIFs.

[20] The standard deviation of our finding is much smaller than that from some other recent papers. For instance, Hung et al. (2020) find the standard deviation of rank gap is 32.16 Taiwan’s mutual funds. This may be caused by the fact that we follow the method of Agarwal et al. (2014) to scale rank gap within a range of (-0.495, +0.495), while Hung et al. (2020) does not.

Following Agarwal et al. (2014), we also employ the "backward holding return gap" (BHRG) to measure the potential performance inconsistency of RIFs between their announced and actual returns. We first collect holding information for each RIF at their quarterly reporting date, and then combine fund holdings with stock information obtained from Thomson Reuter Eikon to calculate a hypothetical portfolio return called the backward holding return (BHR). BHR is calculated based on the buy and hold returns of the disclosed holdings at the end of each quarter. Specifically, we match each fund's holdings with the return of individual companies for each quarter and then, calculate the value-weighted portfolio return for those stocks with available return data, based on the holding weights. We then assess the percentage of holdings that can be matched with stock information. To remain in our sample, a fund needs to have at least 75% of its holdings that can be successfully matched with the stock information. Finally, we scale the value-weighted portfolio return by normalizing portfolio weights to one to get the BHR for each fund. BHRG is defined as the difference between BHR and RIFs' Actual Return (AR).[21]

$$\mathrm{BHRG}_{i,t} = Backward\ holdings\ return\ (BHR)_{\ i,t} - Actual\ return\ (AR)_{i,t} \quad (2)$$

Table 3 describes summary statistics of monthly AR, BHR and BHRG for our sample between 2005 and 2019. The average AR, BHR and BHRG are positive (0.0084, 0.0090, and 0.0007, respectively). According to Agarwal et al. (2014), the higher the BHRG is, the greater the likelihood of performance-based window dressing occurring. As shown in Table 3, monthly

[21] Actual Return is calculated by adding back monthly expense ratios to funds' monthly returns.

BHRGs range from -0.11 to 0.11, which may indicate the existence of financial performance-based window dressing as there are differences between RIFs' announced and actual returns.

[Insert Table 3 about here]

5.2 ESG performance-based window dressing

5.2.1 Full sample analysis

Agarwal et al. (2014) calculates the average value of BHRG to detect the existence of financial performance-based window dressing. The higher the BHRG, the greater the likelihood of window dressing. However, the BHRG cannot explain ESG performance-based window dressing. We cannot directly check holdings information as it is unavailable during the non-reporting period. We overcome this limitation by computing an ESG factor to represent the return premium on an ESG investment strategy (long high-ESG stocks and short low-ESG stocks), in the spirit of Fama and French (1992). By regressing the BHRG on the ESG factor, we can infer whether the BHRG is tilted towards high or low ESG score companies.

To build an ESG factor, all stocks that have ESG scores[22] are ranked according to their previous year's market capitalization. They are then grouped into two portfolios, big and small portfolios,

[22] As sustainable investment has been experiencing notable growth, various rating agencies provide an ESG rating to assess the sustainability performance of companies. We employ the Refinitiv ESG combined score (provided by Thomson Reuters Eikon) to proxy the firms' overall ESG performance. Thomson Reuters ESG score is calculated based on reported ESG-related information, which is one of the most widely acceptable measurements of ESG performance (Dorfleitner et al.,2020; Drempetic et al., 2020). According to Refinitiv (2021), Refinitiv ESG company scores cover 9,000 companies globally and have time-series historical data since 2002. In our sample, 3988 companies have ESG scores.

based on median market value. Within each portfolio, we rank companies based on their ESG scores and assign the highest 30% of companies to the high ESG portfolio and the lowest 30% of companies to the low ESG portfolio. Next, we calculate the value-weighted return for four portfolios (small high ESG, small low ESG, big high ESG and big low ESG). The *ESG* factor is the average return on the two high ESG portfolios, minus the average return on the two low ESG portfolios.[23] ESG factor's portfolios are rebalanced at the beginning of each year.

Following the prior literature, the models can be expressed as follows:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{0i}\left(R_{m,t} - R_{f,t}\right) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{4,i}ESG_t + \varepsilon_{it} \quad (3)$$

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{0i}\left(R_{m,t} - R_{f,t}\right) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{3i}MOM_t + \beta_{4,i}ESG_t + \varepsilon_{it} \quad (4)$$

Where $R_{i,t}$ - $R_{f,t}$ is monthly excess return of RIF i. $R_{m,t}$ - $R_{f,t}$ is monthly excess return of the benchmark.[24] β_{0i} is the slope of regression for RIFs and the corresponding benchmark. $\beta_{1i,}\beta_{2i}$, β_{3i}, β_{4i} are coefficients for the size factor, book-to-market ratio factor, momentum factor and ESG factor, respectively. ε_{it} is error term in month t. $R_{m,t} - R_{f,t}$, *SMB*, *HML* and *MOM* factors data are from Kenneth French Data Library.

We run three regressions (AR, BHR, and BHRG, separately) for each of 196 US equity RIFs based on equations (3) and (4) between 2005M1 and 2019M12. Table 4 summarises the regression results. Panel A and panel B present the 3-factor and 4-factor models, respectively.

[23] ESG factor = 1/2 * (small high ESG + big high ESG) - 1/2 * (small low ESG + big low ESG).
[24] Based on the description of Fama/French factors, $R_{m,t}$ - $R_{f,t}$ includes all firms listed on the NYSE, AMEX, and NASDAQ

[Insert Table 4 about here]

The coefficients for the ESG factor, β_{4i}, the estimate of window dressing, show two interesting results. First, the average β_{4i} for both BHR and AR are negative, indicating that, on average, RIFs' returns, both recorded actual return and backtracked return from disclosed holdings, are tilted towards low ESG score companies. This may not fit the general expectations of RIFs. The logical assumption is that RIFs tilt towards higher ESG score companies. Some investors would even be willing to accept lower expected returns and/or higher management fees to invest in accordance with their social preferences (Riedl & Smeets, 2017). However, the negative average β_{4i} may raise concerns about whether RIFs are investing in what investors expect them to invest in. Second, the average β_{4i} based on the BHRG is positive. That is to say, the return differences between the reporting portfolio and the actual return for these funds are more toward high ESG companies. This may indicate funds are altering holdings to higher ESG score companies, which is consistent with ESG performance-based window dressing behaviour.

[Insert Figure 1 about here]

Table 4 reports the average of the coefficients, while Figure 1 presents the number of funds that have a significant β_{4i} in the sample. The three and four-factor models show similar results. The AR and BHR results show more funds with a significantly negative β_{4i}, indicating that most RIFs with a significant exposure to the ESG factor are investing in lower ESG score companies. However, the BHRG results show that most funds have a significantly positive β_{4i}.

α in Table 4 measures the performance of RIFs compared to the benchmark. A significant positive (negative) alpha indicates that the fund outperforms (underperforms) the market. Both AR and BHR in Table 4 outperform the market between 2005-2009. However, the average α for BHR (0.0009)[25] is larger than that for AR (0.0002). That is to say, disclosed holdings could generate better financial performance than the actual return obtained by investors. This inconsistency may support the existence of financial performance-based window dressing. The percentage reported in the following two columns may also partially support this finding. The percentage of funds with a significantly positive α for the BHR (36.22%) is higher than that for the AR (14.8%), which means that more funds could outperform the market if RIFs were investing in what they disclose to investors. The β_{0i} captures market risk exposure. In both the 3- and 4-factor models, the average β_{0i} generated from the AR is slightly larger than that from the BHR,[26] which may indicate that the actual holding is riskier than the disclosed portfolio. The percentages of funds with β_{0i} greater (less) than one present a more noticeable contrast. Based on the AR, about 48.47% of funds are riskier than the market, whereas the percentage for the BHR is only 39.29%. This suggests that reported holdings may be 'safer', i.e., having a lower market exposure, than the actual portfolios held by RIFs. This idea is in line with Chevalier and Ellison (1997). They find that fund managers modify their holdings in the last quarter of the year to alter the riskiness of their investment portfolio. β_{1i} measures exposure to the size factor. A significantly positive (negative) β_{1i} would provide evidence that funds, on average, are inclined more toward small (big) companies. The average β_{1i} for the AR is smaller than that for the BHR. This may suggest that RIFs' disclosed portfolios are more small-stock-oriented than their actual holdings are. β_{2i} is the coefficient of the

[25] In table 3, the difference between the two means has been tested using paired Two-Sample t-Tests. In 3-factor model, P-value *for* α*, β0i, β1i, β2i and β4i* are 0.0197, 0.0271, 0.0156, 0.9273 and 0.000, respectively. In 4 factor model, P-value *for* α*, β0i, β1i, β2i, β3i and β4i* are 0.0220, 0.0822, 0.00090, 0.3695, 0.2417, and 0.0000, respectively.
[26] Coefficients for AR and BHR from the 3-factor model are not significantly different from 1

HML factor. A significantly positive (negative) β_{2i} indicates that fund's return is more value-oriented (growth-oriented). The average β_{2i} for the AR and the BHR are both negative for the 3-factor model. For the 4-factor model, the average β_{2i} of the AR is negative, but β_{2i} of the BHR is positive. However, the difference in means is statistically insignificant for both models. The coefficient of the momentum factor, β_{3i}, indicates a preference for winner stocks over loser stocks. Both AR and BHR are tilted towards loser stocks, and the difference between the two means is statistically insignificant.

5.2.2 Subsample analysis

The above analysis indicates that a proportion of RIFs (funds' BHRG have significant β_{4i}) may exhibit ESG performance-based window dressing. In this section, we identify common characteristics of these funds.

Based on the value of the funds' coefficient on the ESG factor, we split the sample into three subsamples. Funds with 1) significantly positive, 2) significantly negative, 3) insignificant β_{4i} (ESG). Table 5 presents the mean value of characteristics on these three subsamples.

[Insert Table 5 about here]

Funds that exhibit ESG performance-based window dressing are larger (average size are 8417, 751 and 1151 million, respectively), older (average age are 283, 231 and 176 months, respectively), and have a lower turnover ratio (46%, 49% and 61%, respectively) than funds with significant negative and insignificant β_{4i} (ESG).

5.3 Performance comparison testing

The results generated from the ESG factor models may be interpreted as the effects of fund managers' general trading activity. However, normal portfolio adjustment should be uniformly distributed throughout the entire period, whereas window dressing is more likely to happen close to the reporting date. To rule out the possibility of mistaking ordinary portfolio adjustment for window dressing, we follow Kempf and Osthoff (2008) who use performance comparison to detect if RIFs shift holdings from high ESG to low ESG companies between two report dates. If RIFs apply window dressing, buy (sell) companies with higher (lower) ESG scores before the disclosure dates, a higher exposure is expected to the ethical index just before the reporting dates. Therefore, the daily return of each RIFs will be used to run the following regression:

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}(R_{St} - R_{ft}) + \beta_{2i}(R_{Et} - R_{ft}) + \beta_{3i}D_{it}(R_{St} - R_{ft}) + \beta_{4i}D_{it}(R_{Et} - R_{ft}) + \varepsilon_{it}, \tag{5}$$

where $R_{it} - R_{ft}$ is the daily excess return of RIF i, and $R_{St} - R_{ft}$ is daily excess return of the conventional index. R_{Et} is excess return of the ethical index that has been made orthogonal to the standard index (i.e., the ethical index, MSCI KLD 400 Social Index, is regressed against the conventional index, then the sum of the residual and the intercept forms R_{Et}). D_{it} is a dummy variable that equals one if day t is within the event period (5,10,15, 20 days before the reporting date[27]), and equals 0 for other days. If RIFs apply window dressing in non-ethical stocks, we should find a significantly positive β_{4i}.

We obtain daily fund price and net asset at the share class level from Morningstar. Then, we aggregate the share classes' daily return to the fund level based on each class's daily net asset value. Daily data availability limits our sample size to 136 funds.[28] Following Joliet and Titova (2018), the period starts from the fourth quarter of 2009 to avoid the effects of the financial crisis on investment decisions. Table 6 shows the percentage of funds that have a significant and positive β_{4i} based on Equation (5) between September 2009 and December 2019. Fund numbers are summarized based on different event periods and significance levels. For example, when we examine 15 days before the disclosure date as the event period, 25 out of 136 (18.38%) RIFs show significant positive β_{4i} at the 5% significance level. Table 6 provides further evidence that some RIFs implement a window dressing strategy (shifting towards the orthogonalized ethical index) before the reporting date. This result contrasts with Kempf and Osthoff (2008), who find no indication of window dressing by RIFs between 2001 and 2004. Table 6 also shows that the percentage for the 15-day event period is the highest of all other periods, which may indicate a 'usual' time interval for RIF managers adjusting their holdings before reporting date.

[Insert Table 6 about here]

[27] Similar to Kempf and Osthoff (2008), we do not know when the fund managers adjust their portfolios for window dressing. Therefore, we apply 5, 10, 15 and 20 days before the quarterly report date as the event period.
[28] Only funds that have at least 2 years of data have been retained in the sample.

An interesting finding becomes apparent when we further trim the sample period to January 2015 to December 2019. Reducing the sample period decreases the sample size from 136 to 118 RIFs. Comparing with the results for 2009-2019, the number of funds that show significant positive β_{4i} within the 5- and 10-day event period have increased, while the number of funds that show significant positive β_{4i} within 15- and 20-day event period have decreased between 2015 and 2019. The trend can be seen more directly as percentages (as shown in Figure 2), with a higher proportion of funds showing significantly positive β_{4i} within the 5- and 10-day event period in the last years, and a lower proportion of funds within the 15- and 20-day event period. This indicates that in more recent years, RIFs are more likely to buy (sell) higher (lower) ESG score companies within a period that is closer to the disclosure date.[29] This finding may suggest that RIFs are facing more pressure to obtain attractive financial performance than in the earlier portion of the sample period, and thus they hold better financial performing but low ESG stocks for a longer period to maximize their financial gain.

[Insert Figure 2 about here]

5.4 Kalman filter

The discussion so far has concentrated on standard asset pricing models to estimate coefficients. However, some literature argues that those models may have limitations since they impose stability on the beta parameter, which is usually unlikely in practice (Ortas et al., 2012; Swinkels & Van Der Sluis, 2006). To address this issue, rolling regressions are widely used to estimate the

[29] The reasons behind this phenomenon are beyond the scope of this paper but may relate to increased competition, managerial skills of fund managers or past performance of the RIFs.

coefficients over a certain length of time. By deleting the first observation and adding the next observation, rolling regression generates time-varying exposures. However, this approach still assumes that the exposure stays constant within window (24 or 36 months). Swinkels and Van Der Sluis (2006) propose using a Kalman filter to capture a more accurate time-varying exposure than traditional rolling window regressions. The Kalman filter was introduced initially in Kalman (1960). It is a recursive algorithm for a sequentially updating one-step-ahead estimate of the state mean and variance given new information. This model has been extensively used in autonomous, assisted navigation and other engineering fields. More recently, due to the advantages of estimating dynamic systems, Kalman filters have been used in the economic and finance fields.

In this paper, following Swinkels and Van Der Sluis (2006) method, we use the Kalman filter state-space market model and daily data to capture dynamic betas of RIFs to ethical and conventional indices. Then, we plot the average coefficient to the ethical index of a 51-day-window (-25, +25) and examine if the coefficient exhibits patterns before and after reporting dates. The use of this method may overcome the stated methodological limitation of our previous analysis.

Following the work of Swinkels and Van Der Sluis (2006), the model can be expressed as follow:

$$R_{i,t}^{fund} = \alpha_{i,t} + \beta_{1,i,t} * R_t^{index\ 1} + \beta_{2,i,t} * R_t^{index\ 2} + \varepsilon_{i,t} \quad (6)$$

$$\alpha_{i,t} = \alpha_{i,t-1} \quad (7)$$

$$\beta_{1,i,t} = \beta_{1,i,t-1} + \xi_{1,i,t} \quad (8)$$

$$\beta_{2,i,t} = \beta_{2,i,t-1} + \xi_{2,i,t} \quad (9)$$

Where, $R_{i,t}^{fund}$ is the fund excess return. $R_t^{index\ 1}$is the excess return on the market index, $R_t^{index\ 2}$ is the excess return on orthogonalized ethical index. $\varepsilon_{i,t}$ and $\xi_{i,t}$ are error terms $\varepsilon_{i,t}$ ~ NID $(0,\sigma_\varepsilon^2)$, $\xi_{j,i,t}$ ~ NID $(0, \sigma_{j,i,\xi}^2)$. $\beta_{1,i,t}$ and $\beta_{2,i,t}$ are time-varying exposures to market indices and the orthogonalized ethical index at time *t*. $\beta_{2,i,t}$ is the main indicator in this model, as it represents the sensitivity of RIFs to the ethical index. By looking at the value of $\beta_{2,i,t}$before and after the disclosure date, we can infer if RIFs change their exposure to the ethical index pre-and post-holding information announcement date.

The model is in state-space form. Equation (6) is the measurement equation; it links RIFs' excess returns to conventional and ethical indices. Equations (7), (8) and (9) are transition equations, which gives the state evolution process. We keep manager ability α_t to be constant over time. The exposure coefficients $\beta_{1,i,t}$ and $\beta_{2,i,t}$ are estimated based on exposure at the previous period (*t*-1) plus an associated error term that follows a normal distribution with mean zero and a variance $\xi_{i,t}$.

[Insert Figure 3 about here]

Figure 3 presents the average $\beta_{2,i,t}$ obtained from the Kalman filter (equations (6)-(8)) from 25 days before to 25 days after reporting date (the last day of each quarter).[30] The average coefficient is negative, which means, on average, funds are negatively correlated to the orthogonalized ethical

[30]There are some outliers that appear at the start of the sample period. This is caused by the Kalman filter procedure. The model needs a learning period before it becomes stable. Therefore, the first 7 observation of each fund has been removed.

index. If the coefficient increases when the reporting date approaches, it means that funds are less negatively correlated to the ethical index, and this may indicate the appearance of ESG-performance window dressing. Figure 3 Panel A reports average $\beta_{2,i,t}$ for the entire sample. For the pre-disclosure window (-25, -1), average daily coefficients first gradually decrease and then increase with a small fluctuation. It then enters a fluctuating period, with three rises and peaks within (-13, -12), (-7, -5) and (-1,0). On the disclosure date, t_0, coefficient reaches a relatively higher level (-0.1637) compared to the pre-disclosure window. For the post-disclosure window, the average coefficient immediately declines after the disclosure date from -0.1637 to -0.1677 at *t+5* and then rises to a peak (-0.1606) till it starts decreasing from *t+22*.

The three peaks that happen within (-13, 0) in Panel A might be an indication of potential window dressing behaviour, as they are obvious changes in the coefficient before the reporting date. However, as Panel A displays the entire sample of RIFs, the overall result might be offset by funds within the sample but are unlikely to window dress. Therefore, a sub-sample is created using the funds that are more likely to be engaged in window dressings and the result is shown in Panel B. This subsample includes RIFs with a significantly positive coefficient on the ESG factor in Equation (3). The average daily coefficient for the subsample shows a more obvious increasing trend between (-12, 0), reaching the highest level across the period at the reporting date (t_0). This upward trend indicates potential window dressing, which supports our finding in Table 4 that a segment of funds exhibits ESG performance-based window dress behaviour before the reporting date.

5.5 Determinants of window dressing

The extant literature on window dressing demonstrates that financial performance-based window dressing may be related to poor past performance (Agarwal et al., 2014) and some unobserved influences at the fund management company level (Gil-Bazo et al.,2010). Morey and O'Neal (2006) point out that funds with higher risk exposure are more likely to change their risk characteristics prior to the disclosure date. However, whether these factors also play roles in interpreting the possibility of ESG performance-based window dressing is still unknown. To test the potential determinations of ESG performance-based window dressing, we run a logistic regression:

$$WD_{it} = \gamma_0 + \gamma_1 specialist\ level_{i,t} + \gamma_1 12\ month\ alpha_{i,t} + \gamma_2 tracking\ error_{i,t} + \gamma_3 age_{i,t} + \gamma_4 size_{i,t} + \upsilon_i + \varepsilon_{i,t} \quad (7)$$

where the dependent variable, WD_{it} is a binary variable that equals 1 when ESG performance-based window dressing exists (significant positive β_{4i} in a 24-month rolling window), and 0 otherwise for RIF *I* in month *t*.

$specialist\ level_{i,t}$ is the responsible investing specialist level of fund *i*'s management company in month *t* measured as the fund management company's percentage of assets under management in RIFs[31]. Following the work of Gil-Bazo et al. (2010), we hypothesize that management companies' specialization levels in the management of RIFs is a key in explaining the differences between RIFs and conventional funds.

[31] Our sample reduces to 156 RIFs after considering information at the fund management company level due to data availability.

$12\ month\ alpha_{i,t}$ is the fund's annualized alpha obtained using monthly raw returns and the 4-factor model in a 12-month rolling window, used to control for past performance (Ammann, et al, 2019).

$tracking\ error_{i,t}$ is the past 12-month's cumulative R-squared, estimated from the 4-factor model by using a 12-month rolling window, used to measure return volatility.

We additionally control for $age_{i,t}$ and $size_{i,t}$, which is the number of months since the oldest share class was established, and the logarithm of fund size for RIF i in month t. υ_i is fund-fixed effects, and $\varepsilon_{i,t}$ denotes the error term.

[Insert Table 7 about here]

Table 7 presents the results of our logistic regression in (7). The estimated coefficients of the management company's specialization level, past performance and size are significantly negative. Tracking error and age exhibit significant positive relationships. These results indicate that, RIFs managed by companies with lower specialist levels and poor past performance are more likely to exhibit ESG performance-based window dressing. For volatility, RIFs with higher tracking errors are more likely to engage in ESG performance-based window dressing. The coefficient of size is significantly positive at 1% level, and the coefficient of age is significantly negative at 1 % level. This indicates that the effect of size changed from larger funds are more like to window dress (as shown in Table 5) to smaller funds after we employ the multivariate model (as shown in Table 7).

6. Conclusion

Extensive research has been conducted on the financial and non-financial returns of RIFs based on their self-reported information. However, the accuracy of this information is relatively unknown. This paper attempts to address the issue. In this paper, we examine the presence of financial- and ESG- performance based window dressing of US domestic equity RIF between 2005-2019 using the rank gap, backward holding return gap and factor model methods. The results of those analyses deliver several novel findings. First, by comparing the rank difference between performance-based rank and the rank based on the proportions of winner and loser stocks reported by funds, we find that a proportion of RIFs exhibit financial performance-based window dressing. This finding is supported by the difference between funds' hypothetical holding return calculated using the buy and hold strategy and the actual return reported by funds. Second, RIFs' return differences between the reporting portfolio and actual return show a preference for high ESG companies, which indicates the existence of ESG performance-based window dressing. Third, both recorded actual and backtracked returns from disclosed holdings of funds are tilted towards companies with lower ESG scores. This finding may cause investors' concerns regarding the reliability of RIFs' attitude to ESG consideration.

Our performance comparison test suggests that over a quarter of funds shift towards the ethical index just before reporting date, and the result is robust to the dynamic daily average coefficient on the ethical index. This finding rules out the possibility that our detected ESG performance-based window dressing is caused by normal portfolio adjustment. We also find that funds are leaving it later to window dress their portfolios in more recent periods, moving from 15 days prior to the reporting date to 5 or 10 days in 2015-2019. We argue that this result may reflect that RIFs face more competition and suffer more pressure to obtain better dual performance, forcing managers to adjust holdings regarding better ESG performance in a shorter period prior to the reporting date.

This paper also investigates what drives the ESG performance-based window dressing. We identify that similar to financial performance-based window dressing, RIFs with poor past performance, higher tracking errors and those managed by companies with lower specialist levels, are more likely to exhibit ESG performance-based window dressing.

The results may have interesting implications. There is a possibility that RIFs are not fulfilling their announced ESG criteria, in which case investors need to be very careful in selecting RIFs to achieve their non-financial goals. At the same time, regulators, such as SEC, should encourage RIFs to be more open and transparent in their holdings to build consumer and investor confidence in responsible investment.

References

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

Alda, M. (2020). ESG fund scores in UK SRI and conventional pension funds: Are the ESG concerns of the SRI niche affecting the conventional mainstream? *Finance research letters*, *36*, 101313.

Alda, M., Muñoz, F., & Vargas, M. (2020). Socially responsible mutual fund exit decisions. *Business Ethics: A European Review*, *29*(1), 82-97.

Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, *74*(3), 87-103.

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

Azmi, W., Mohamad, S., & Shah, M. E. (2020). Ethical investments and financial performance: An international evidence. *Pacific-Basin Finance Journal*, *62*, 101147.

Bai, J. J., Ma, L., Mullally, K. A., & Solomon, D. H. (2019). What a difference a (birth) month makes: The relative age effect and fund manager performance. *Journal of financial economics*, *132*(1), 200-221.

Bello, Z. (2005). Socially responsible investing and portfolio diversification. *Journal of Financial Research, 28*(1), 41-57.

Benson, K. L., & Humphrey, J. E. (2008). Socially responsible investment funds: Investor reaction to current and past returns. *Journal of Banking & Finance, 32*(9), 1850-1859.

Bialkowski, J., & Starks, L. T. (2016). SRI funds: Investor demand, exogenous shocks and ESG profiles. Working Paper, University of Texas.

Borgers, A., Derwall, J., Koedijk, K., & ter Horst, J. (2015). Do social factors influence investment behavior and performance? Evidence from mutual fund holdings. *Journal of Banking & Finance, 60*, 112-126.

Brandon, R. G., Glossner, S., Krueger, P., Matos, P., & Steffen, T. (2021). *Do Responsible Investors Invest Responsibly?*. ECGI Working Paper Series in Finance.

Capelle-Blancard, G., & Monjon, S. (2012). Trends in the literature on socially responsible investment: looking for the keys under the lamppost. *Business Ethics: A European Review, 21*(3), 239-250.

Chevalier, J., & Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of political economy, 105*(6), 1167-1200.

Chieffe, N., & Lahey, K. E. (2009). Helping clients select SRI mutual funds and firms. *Journal of Financial Planning*, *22*(2).

Cici, G., Hendriock, M., & Kempf, A. (2021). The impact of labor mobility restrictions on managerial actions: Evidence from the mutual fund industry. *Journal of Banking & Finance*, *122*, 105994.

Delmas, M. A., & Burbano, V. C. (2011). The drivers of greenwashing. *California management review*, *54*(1), 64-87.

Dorfleitner, G., Kreuzer, C., & Sparrer, C. (2020). ESG controversies and controversial ESG: about silent saints and small sinners. *Journal of Asset Management*, *21*(5), 393-412.

Dorfleitner, G., Leidl, M., & Reeder, J. (2012). Theory of social returns in portfolio choice with application to microfinance. *Journal of Asset Management*, *13*(6), 384-400.

Doshi, H., Elkamhi, R., & Simutin, M. (2015). Managerial activeness and mutual fund performance. *Review of Asset Pricing Studies*, *5*(2), 156-184.

Døskeland, T., & Pedersen, L. J. T. (2016). Investing with brain or heart? A field experiment on responsible investment. *Management Science*, *62*(6), 1632-1644.

Drempetic, S., Klein, C., & Zwergel, B. (2020). The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics*, *167*(2), 333-360.

El Ghoul, S., & Karoui, A. (2017). Does corporate social responsibility affect mutual fund performance and flows? *Journal of Banking & Finance, 77*, 53-63.

Elaut, G., Frömmel, M., & Verbeeck, B. (2015). What is the price of a clear conscience? The performance of socially responsible investments in the BRICS countries.

Elton, E. J., Gruber, M. J., Blake, C. R., Krasny, Y., & Ozelge, S. O. (2011). The effect of holdings data frequency on conclusions about mutual fund behavior. In *Investments And Portfolio Performance* (pp. 195-205). World Scientific.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427-465.

Findlay, S., & Moran, M. (2018). Purpose-washing of impact investing funds: motivations, occurrence and prevention. *Social Responsibility Journal*.

Frank, M. M., Poterba, J. M., Shackelford, D. A., & Shoven, J. B. (2004). Copycat funds: Information disclosure regulation and the returns to active management in the mutual fund industry. *The Journal of Law and Economics*, *47*(2), 515-541.

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, *5*(4), 210-233.

Gil-Bazo, J., Ruiz-Verdú, P., & Santos, A. A. P. (2009). The performance of socially responsible mutual funds: The role of fees and management companies. *Journal of Business Ethics, 94*(2), 243-263.

Global Sustainable Investment Alliance. (2021). *2020 Global Sustainable Investment Review*.

Gormley, T., Kaplan, Z., & Verma, A. (2019). *Can disclosure decrease price efficiency? Evidence from mutual fund disclosures*. Working paper.

Haugen, R. A., & Lakonishok, J. (1987). *The incredible January effect: The stock market's unsolved mystery*. Irwin Professional Pub.

Humphrey, J. E., Warren, G. J., & Boon, J. (2015). What is different about socially responsible funds? A holdings-based analysis. *Journal of Business Ethics, 138*(2), 263-277.

Hung, P.-H., Lien, D., & Kuo, M.-S. (2020). Window dressing in equity mutual funds. *The Quarterly Review of Economics and Finance, 78*, 338-354.

Ibikunle, G., & Steffen, T. (2017). European green mutual fund performance: A comparative analysis with their conventional and black peers. *Journal of Business Ethics, 145*(2), 337-355.

In, F., Kim, M., Park, R. J., Kim, S., & Kim, T. S. (2014). Competition of socially responsible and conventional mutual funds and its impact on fund performance. *Journal of Banking & Finance, 44*, 160-176.

Ivanisevic Hernaus, A. (2019). Exploring the strategic variety of socially responsible investment: Financial performance insights about SRI strategy portfolios. *Sustainability Accounting, Management and Policy Journal, 10*(3), 545-569.

Joliet, R., & Titova, Y. (2018). Equity SRI funds vacillate between ethics and money: An analysis of the funds' stock holding decisions. *Journal of Banking & Finance, 97*, 70-86.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems.

Kempf, A., & Osthoff, P. (2008). SRI Funds: Nomen est omen. *Journal of Business Finance & Accounting, 35*(9-10), 1276-1294.

Kurniawan, M., How, J., & Verhoeven, P. (2016). Fund governance and style drift. *Pacific-Basin Finance Journal, 40*, 59-72.

Lean, H. H., Ang, W. R., & Smyth, R. (2015). Performance and performance persistence of socially responsible investment funds in Europe and North America. *The North American Journal of Economics and Finance, 34*, 254-266.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic perspectives, 21*(2), 153-174.

Lyon, T. P., & Maxwell, J. W. (2011). Greenwash: Corporate environmental disclosure under threat of audit. *Journal of Economics & Management Strategy, 20*(1), 3-41.

Mallin, C. A., Saadouni, B., & Briston, R. J. (1995). The financial performance of ethical investment funds. *Journal of Business Finance & Accounting, 22*(4), 483-496.

Marques, M. R., Sampaio, J. O., & Silva, V. A. B. (2020). Window dressing in Brazilian investment funds. *Revista Contabilidade & Finanças, 31*(82), 116-128.

Meier, I., & Schaumberg, E. (2006). Do funds window dress? Evidence for US domestic equity mutual funds, HEC Montreal, Working Paper.

Morey, M. R., & O'Neal, E. S. (2006). Window dressing in bond mutual funds. *Journal of Financial Research, 29*(3), 325-347.

Muñoz, F., Vicente, R., & Ferruz, L. (2015). Stock-picking and style-timing abilities: a comparative analysis of conventional and socially responsible mutual funds in the US market. *Quantitative Finance, 15*(2), 345-358.

Nitsche, C., & Schröder, M. (2018). Are SRI funds conventional funds in disguise or do they live up to their name? In *Research handbook of investing in the triple bottom line*. Edward Elgar Publishing.

Nofsinger, J., & Varma, A. (2014). Socially responsible funds and market crises. *Journal of Banking & Finance, 48*, 180-193.

O'Neal, E. S. (2001). Window dressing and equity mutual funds. *Babcock Graduate School of Management Working Paper*.

Ortas, E., Moneva, J. M., & Salvador, M. (2012). Does socially responsible investment equity indexes in emerging markets pay off? Evidence from Brazil. *Emerging Markets Review*, *13*(4), 581-597.

Ortiz, C., Ramírez, G., & Vicente, L. (2015). Mutual fund trading and portfolio disclosures. *Journal of Financial Services Research, 48*(1), 83-102.

Parida, S. (2017). Impact of reporting delays on profitability of front-running strategies against mutual funds. *Managerial Finance*.

Parida, S., & Teo, T. (2018). The impact of more frequent portfolio disclosure on mutual fund performance. *Journal of Banking & Finance, 87*, 427-445.

Patton, A. J., & Ramadorai, T. (2013). On the high-frequency dynamics of hedge fund risk exposures. *The Journal of Finance, 68*(2), 597-635.

Refinitiv (2021). *Refinitiv ESG Company Scores.* https://www.refinitiv.com/en/sustainable-finance/esg-scores

Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance, 72*(6), 2505-2550.

Schwarz, C. G., & Potter, M. E. (2016). Revisiting mutual fund portfolio disclosure. *The Review of Financial Studies*, *29*(12), 3519-3544.

Solomon, D. H., Soltes, E., & Sosyura, D. (2014). Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics, 113*(1), 53-72.

Swinkels, L., & Van Der Sluis, P. J. (2006). Return-based style analysis with time-varying exposures. *The European Journal of Finance, 12*(6-7), 529-552.

The Forum for Sustainable and Responsible Investment. (2020). *US SIF Trends Report 2020 Executive Summary*.

Thompson, J., Engle Sr, A. D., & Spain, J. W. (2011). " A rose by any other name": Models of social responsibility as predictors of financial performance. *Journal of Financial and Economic Practice, 11*(1), 37.

Trinks, P. J., & Scholtens, B. (2017). The opportunity cost of negative screening in socially responsible investing. *Journal of Business Ethics, 140*(2), 193-208.

U.S. Securities and Exchange Commission. (2004). *Final Rule: Shareholder Reports and Quarterly Portfolio Disclosure of Registered Management Investment Companies*. Retrieved from U.S. Securities and Exchange Commission: https://www.sec.gov/rules/final/33-8393.htm

Utz, S., & Wimmer, M. (2014). Are they any good at all? A financial and ethical analysis of socially responsible mutual funds. *Journal of Asset Management, 15*(1), 72-82.

van Dijk-de Groot, M., & Nijhof, A. H. J. (2015). Socially responsible investment funds: a review of research priorities and strategic options. *Journal of Sustainable Finance & Investment, 5*(3), 178-204.

Table 1. Fund level descriptive statistics

This table reports the summary statistics of 196 US domestic equity RIFs from January 2005 to March 2010. Size is the average fund size for the studied period, reported in millions of dollars. Age is the number of months since the fund's inception date to the end of the study period or the last month that the fund has return information. Turnover ratio % is computed by taking the lesser of purchases or sales (excluding all securities with maturities of less than one year) and dividing by average monthly net assets[32].

	Size (Million)	**Turnover Ratio % (annual)**	**Age (month)**
Mean	3671.69	55.48	214.87
Standard Error	1029.85	2.92	12.53
Median	232.13	45.47	187.00
Standard Deviation	14417.93	40.45	175.41
Kurtosis	61.16	6.21	7.34
Skewness	7.14	1.93	2.34
Minimum	1.94	3.47	28.00
Maximum	150220.29	279.49	1101.00
Count	196	192	196

[32] Four RIFs have no turnover ratio information.

Table 2 Descriptive statistics of rank gap

This table presents the overall summary statistics of the monthly rank gap and the proportions of winner and loser holdings for 196 funds between January 2005 and December 2019. Following Agarwal et al. (2014), the rank gap is calculated based on Equation (1). Percentage of loser (winner) stocks is the percentage of holdings that invest in the stocks that belongs to the quintile portfolio of the lowest (highest) return.

	Rank Gap	**Percentage of loser stocks**	**Percentage of winner stocks**
Mean	0.0021	4.9013	15.2797
Median	0.0025	3.5249	14.0495
Maximum	0.4800	48.8879	69.8622
Minimum	-0.4625	-2.5800	-0.6100
Std. Dev.	0.1077	4.8736	9.1630
Observations	25256	25727	25727

Table 3. Descriptive statistics of AR, BHR, BHRG

This table reports the summary statistics of AR, BHR, BHRG for 196 US domestic equity RIFs. AR is RIF's value-weighted actual return, which is aggregated from the share classes return based on the net asset values of each share class. BHR is the backward holding return calculated based on the disclosed holdings by applying the buy and hold strategy. BHRG is the difference between the BHR and the AR.

	AR	**BHR**	**BHRG**
Mean	0.0084	0.0090	0.0007
Median	0.0129	0.0136	0.0007
Maximum	0.3141	0.2607	0.1122
Minimum	-0.3229	-0.3680	-0.1057
Std. Dev.	0.0453	0.0454	0.0085
Sum	212.7604	231.9007	17.8571
Observations	25347	25656	25237

Table 4 Summary of coefficients from 3,4-factor models

This table reports the regression results of equations (3) & (4) for 196 US domestic equity RIFs between 2005M1 and 2019M12.

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{0i}\left(R_{m,t} - R_{f,t}\right) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{4,i}ESG_t + \varepsilon_{it} \quad (3)$$

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{0i}\left(R_{m,t} - R_{f,t}\right) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{3i}MOM_t + \beta_{4,i}ESG_t + \varepsilon_{it} \quad (4)$$

$R_{i,t}$ - $R_{f,t}$ is monthly excess return of RIF i. $R_{m,t}$ - $R_{f,t}$ is monthly excess return of the benchmark. β_{0i} is the slope of the regression for RIFs and the corresponding benchmark. $\beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i}$ are the coefficients for the size factor, book-to-market ratio factor, momentum factor and ESG factor, respectively. ε_{it} is error term in month t. Panel A and panel B present the 3-factor and 4-factor model, respectively. We also report the percentage of funds that has a significant positive/negative α, and the % of funds has a significant β0i greater (less) than1. *, **, *** denote significant differences from zero (except for $\boldsymbol{\beta_{0i}}$, test is if the average significant different from one) at 10%, 5% and 1%.

	α			β_{0i}			β_{1i}(SMB)	β_{2i}(HML)	β_{3i}(MOM)	β_{4i}(ESG)
	Mean	**% of funds significant positive α**	**% of funds significant negative α**	**Mean**	**% of funds β0i significantly >1**	**% of funds β0i significantly <1**	**Mean**	**Mean**	**Mean**	**Mean**
Panel A 3 factors model										
AR	0.0002	14.80%	6.12%	1.0013	48.47%	51.53%	0.1547***	0.0066		-0.1191***
BHR	0.0009***	36.22%	18.88%	0.9906	39.29%	60.71%	0.1671***	0.0062		-0.0767***
BHRG	-0.0004	22.45%	55.10%	-0.0073***	0.00%	55.10%	0.018***	0.0019		0.0415***
Panel B 4 factors model										
AR	0.0002*	18.88%	3.57%	0.9955	47.96%	52.04%	0.1523***	-0.0034	-0.0127*	-0.1257***
BHR	0.0009***	35.71%	18.37%	0.9868*	39.29%	60.71%	0.1663***	0.0006	-0.0088	-0.0814***
BHRG	-0.0004	22.45%	57.65%	-0.0058***	0.00%	58.16%	0.0189***	0.0047	0.0028	0.0425***

Table 5 Characteristics of the subsamples

This table presents the mean value of characteristics on three subsamples: funds with 1) significant positive, 2) significant negative, 3) insignificant. Significance of differences has been tested using the Welch's t-test and the results are reported in Appendix A. Size is the average of fund size for the studied period, reported in millions of dollars. Age is the number of months since the fund's inception date to the end of the study period or the last month the fund has return information. Turnover ratio % is computed by taking the lesser of purchases or sales (excluding all securities with maturities of less than one year) and dividing by average monthly net assets.

	Size (million)	**Age (month)**	**Turnover % (annually)**	**Observations**
1) Funds with significant positive β4i(ESG)	8417.13	283.13	46.12	68
2) Funds with significant negative β4i(ESG)	750.66	231.29	48.89	7
3) Funds with insignificant β4i(ESG)	1173.81	175.55	61.32	121

Table 6 Performance comparison tests

This table summarises the results for Equation (5).

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}(R_{St} - R_{ft}) + \beta_{2i}(R_{Et} - R_{ft}) + \beta_{3i}D_{it}(R_{St} - R_{ft}) + \beta_{4i}D_{it}(R_{Et} - R_{ft}) + \varepsilon_{it} \tag{5}$$

$R_{it} - R_{ft}$ is daily excess return of RIF i. $R_{St} - R_{ft}$ is daily excess return of the conventional index. R_{Et} is excess return of the ethical index that has been made orthogonal to the standard index (regress ethical index against the conventional index, then use the sum of residual and the Intercept to form R_{Et}). D_{it} is a dummy variable that equals one if the day t is within the event period (5,10,15, 20 days before the reporting date), and equal to 0 for other days. If RIFs apply window dressing in low ESG stocks, we should find a significant positive β_{4i}. We summarise the percentage of funds that have a significant positive β_{4i} before the reporting date for different event periods (5, 10, 15, 20 days before disclosure date) with different significant levels. For example, when applying 5 days before the disclosure date as the event period, 10.29% of RIFs in our sample show significant positive β_{4i} at the 5% significant significance level.

Event periods	**10%**	**5%**	**1%**
5	11.76%	10.29%	8.82%
10	15.44%	13.24%	6.62%
15	24.26%	18.38%	9.56%
20	18.38%	14.71%	4.41%

Table 7 Determinates of ESG-performance-based window dressing.

This table reports the results of logistic regression for the determinates of ESG-performance-based window dressing.

$$WD_{it} = \gamma_0 + \gamma_1 specialist\ level_{i,t} + \gamma_1 12\ month\ alpha_{i,t} + \gamma_2 tracking\ error_{i,t} + \gamma_3 age_{i,t} + \gamma_4 size_{i,t} + \upsilon_i + \varepsilon_{i,t} \quad (7)$$

WD_{it}, is the binary variable that equals 1 when ESG performance-based window dressing exists, and 0 otherwise for RIF i in month t. $specialist\ level_{i,t}$ is responsible investment specialist level of fund i's management company in month t. $12\ month\ alpha_{i,t}$ is fund annualized alpha that obtained using monthly raw returns and the 4-factor model in a 12-month rolling window. $tracking\ error_{i,t}$ is the past 12-month's cumulative R-square that estimated from 4-factor model by using12-month rolling window. $age_{i,t}$ is the number of months since the oldest share class was established for RIF *i* in month *t*. $size_{i,t}$ is the logarithm of the fund size for RIF *i* in month *t*. υ_i is the fund-fixed effect, $\varepsilon_{i,t}$ denotes the error term.

Dependent variable: binary variable of ESG performance-based window dressing

Specialist level	**12 month alpha**	**tracking error**	**age**	**size**
-0.9547*	-5.7364***	0.3745***	0.0075***	-0.7177***

Figure 1 Funds with significant β_{4i}.

This figure presents the number of funds that has significant β4i based on equations (3) and (4).

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{0i}(R_{m,t} - R_{f,t}) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{4,i}ESG_t + \varepsilon_{it} \quad (3)$$

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_{0i}(R_{m,t} - R_{f,t}) + \beta_{1i}SMB_t + \beta_{2i}HML_t + \beta_{3i}MOM_t + \beta_{4,i}ESG_t + \varepsilon_{it}$$

(4)

$R_{i,t}$ - $R_{f,t}$ is monthly excess return of RIF i. $R_{m,t}$ - $R_{f,t}$ is monthly excess return of the benchmark. β_{0i} is the slope of the regression for RIFs and the corresponding benchmark. β_{1i}, β_{2i}, β_{3i}, β_{4i} are the coefficients for the size factor, book-to-market ratio factor, momentum factor and ESG factor, respectively. ε_{it} is error term in month t.

The upper part summarises the result from Equation (4), and the lower part shows the result from Equation (3)

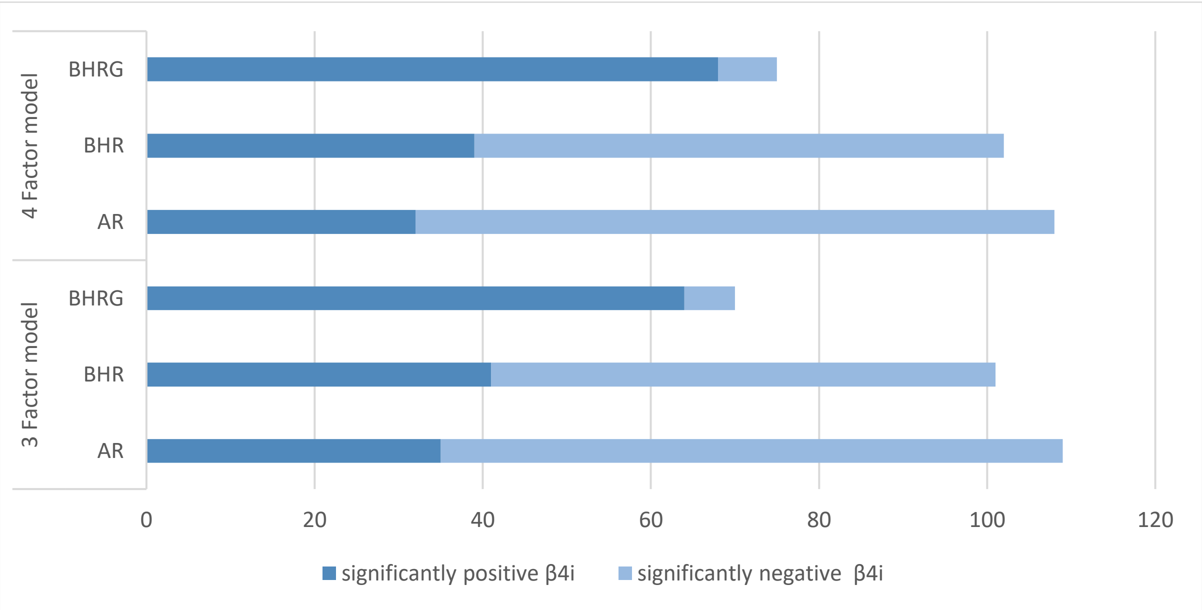


Figure 2 Performance comparison test (Subsamples)

This Figure reports results for Equation (5) for different sample periods.

$$R_{it} - R_{ft} = \alpha_i + \beta_{1i}(R_{St} - R_{ft}) + \beta_{2i}(R_{Et} - R_{ft}) + \beta_{3i}D_{it}(R_{St} - R_{ft}) + \beta_{4i}D_{it}(R_{Et} - R_{ft}) + \varepsilon_{it} \tag{5}$$

$R_{it} - R_{ft}$ is daily excess return of RIF i. $R_{St} - R_{ft}$ is daily excess return of the conventional index. R_{Et} is excess return of the ethical index that has been made orthogonal to the standard index (regress ethical index against the conventional index, then use the sum of residual and the Intercept to form R_{Et}). D_{it} is a dummy variable that equals one if the day t is within the event period (5,10,15, 20 days before the reporting date), and equal to 0 for other days. The Figure presents the percentage of funds with significant positive β_{4i} for different event periods at different significant levels. The left part summarises the result of 136 RIFs between January 2009 and December 2019. The right part summarises the result for a sub-period (between January 2015 and December 2019, and there are 118 RIFs within the sub-period.

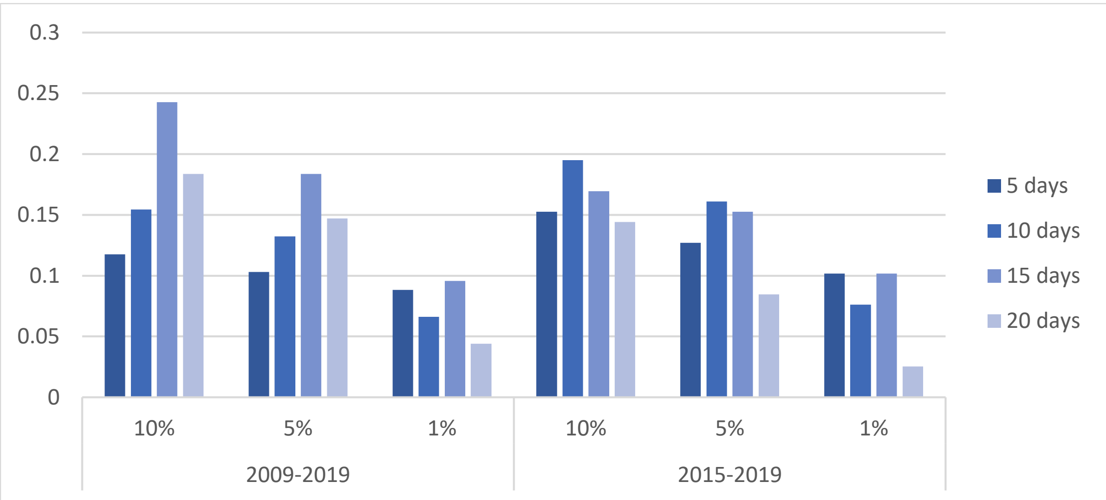


Figure 3 Average exposure obtained from Kalman filter for the 51-day-window.

This figure reports the average $\beta_{2,i,t}$ of Equation (6) for the 51-day-window.

$$R_{i,t}^{fund} = \alpha_{i,t} + \beta_{1,i,t} * R_t^{index\ 1} + \beta_{2,i,t} * R_t^{index\ 2} + \varepsilon_{i,t} \quad (6)$$

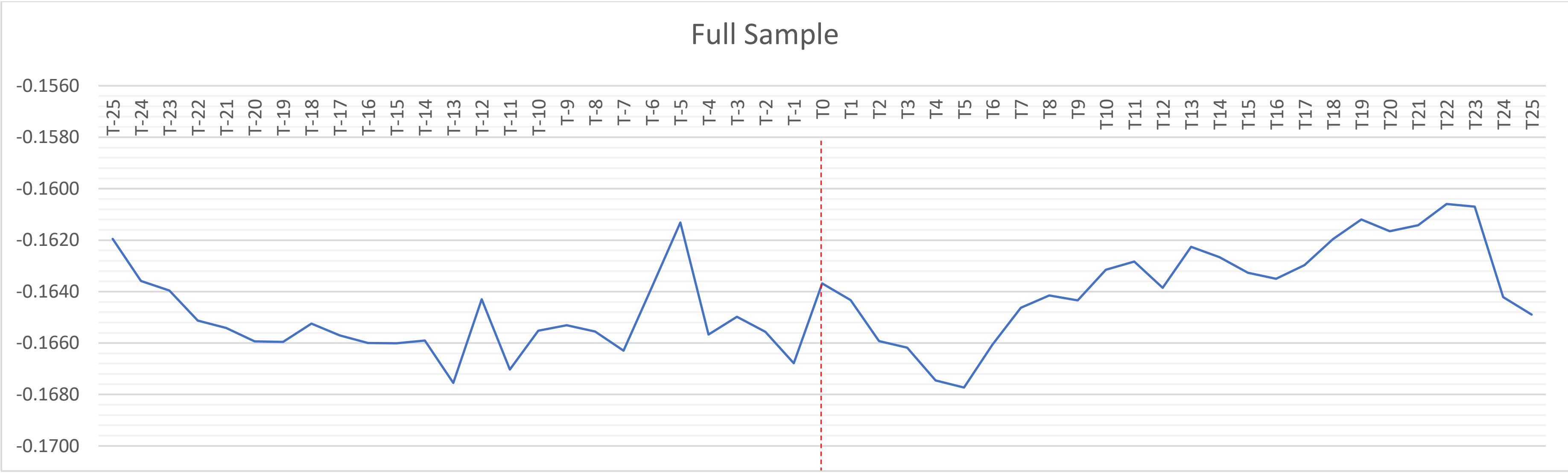
$$\alpha_{i,t} = \alpha_{i,t-1} \quad (7)$$

$$\beta_{1,i,t} = \beta_{1,i,t-1} + \xi_{1,i,t} \quad (8)$$

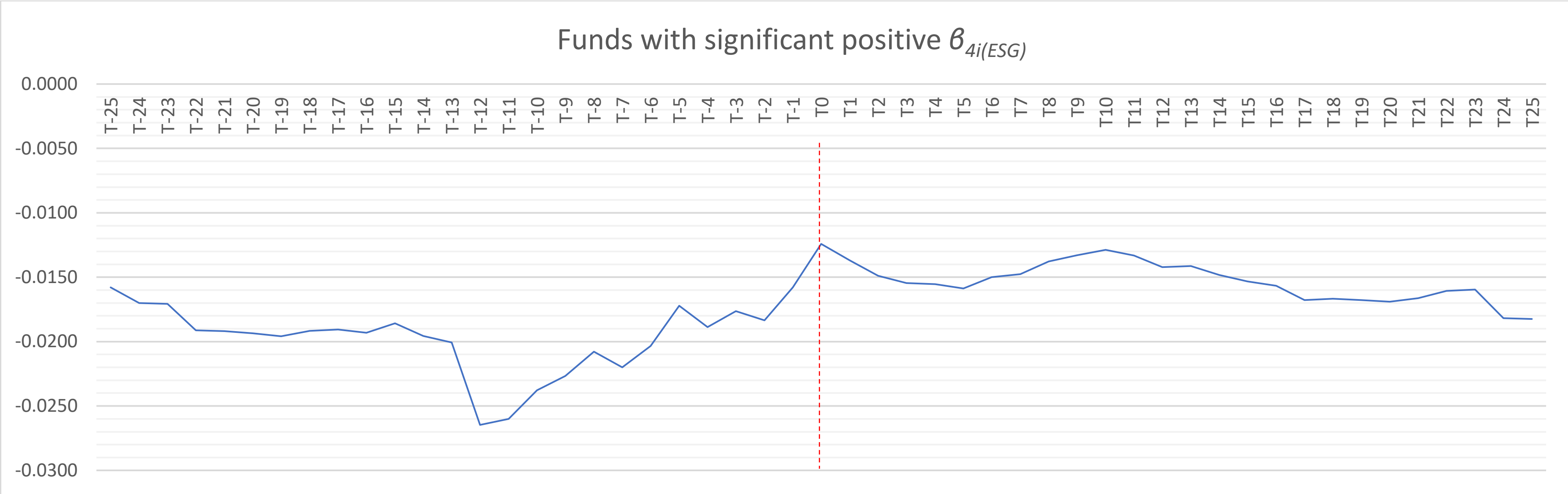
$$\beta_{2,i,t} = \beta_{2,i,t-1} + \xi_{2,i,t} \quad (9)$$

$R_{i,t}^{fund}$ is the fund excess return. $R_t^{index\ 1}$is the excess return on market index, $R_t^{index\ 2}$ is the excess return on orthogonalized ethical index. $\varepsilon_{i,t}$ and $\xi_{i,t}$ are the error terms $\varepsilon_{i,t}$ ~ NID (0,σ_ε^2), $\xi_{j,i,t}$ ~ NID (0,$\sigma_{j,i,\xi}^2$). $\beta_{1,i,t}$and $\beta_{2,i,t}$ are the time-varying exposures to market indices and the orthogonalized ethical index at time t. $\beta_{2,i,t}$ is the main indicator in this model, as it represents the sensitivity of RIFs to the ethical index. Panel A shows the results for 136 funds that have daily data. Panel B shows the results for the subsample, which includes the funds that have significant positive $\beta_{4i(ESG)}$ based on Equation (3). The red dash line indicates T0.

Panel A


Full Sample
-0.1560
-0.1580
-0.1600
-0.1620
-0.1640
-0.1660
-0.1680
-0.1700
T-25 T-24 T-23 T-22 T-21 T-20 T-19 T-18 T-17 T-16 T-15 T-14 T-13 T-12 T-11 T-10 T-9 T-8 T-7 T-6 T-5 T-4 T-3 T-2 T-1 T0 T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15 T16 T17 T18 T19 T20 T21 T22 T23 T24 T25

Panel B


Funds with significant positive $\beta_{4i(ESG)}$
0.0000
-0.0050
-0.0100
-0.0150
-0.0200
-0.0250
-0.0300
T-25 T-24 T-23 T-22 T-21 T-20 T-19 T-18 T-17 T-16 T-15 T-14 T-13 T-12 T-11 T-10 T-9 T-8 T-7 T-6 T-5 T-4 T-3 T-2 T-1 T0 T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12 T13 T14 T15 T16 T17 T18 T19 T20 T21 T22 T23 T24 T25

Appendix A

t-Test: Two-Sample Assuming Unequal Variances for the comparison between subsample 1 and 3

	Size		**Age**		**Average turnover %**	
	subsample 1	***subsample 3***	***subsample 1***	***subsample 3***	***subsample 1***	***subsample 3***
Mean	8417.13	1173.81	283.13	175.55	46.12	61.32
Variance	542607286.82	15244207.86	50059.43	16805.85	1116.89	1902.01
Observations	68.00	121.00	68.00	121.00	68.00	117.00
Hypothesized Mean Difference	0.00		0.00		0.00	
df	69.00		93.00		169.00	
t Stat	2.54		3.64		-2.66	
P(T<=t) two-tail	0.0132		0.0005		0.0086	
t Critical two-tail	1.99		1.99		1.97	