

Secondary Publication



Muck, Matthias; Schmidl, Thomas

Relative to whom? : The impact of peer groups on ESG ratings and financial performance

Date of secondary publication: 05.05.2026

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-114951x

Primary publication

Muck, Matthias; Schmidl, Thomas (2026): Relative to whom? : The impact of peer groups on ESG ratings and financial performance, in: Journal of Asset Management, Basingstoke: Palgrave Macmillan, Vol. 27, No. 2, 11, pp. 1–26, doi: 10.1057/s41260-026-00444-4.

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>



Relative to whom? The impact of peer groups on ESG ratings and financial performance

Matthias Muck¹ · Thomas Schmidl¹

Revised: 15 December 2025 / Accepted: 19 January 2026
© The Author(s) 2026

Abstract

This paper investigates how peer groups affect the financial informativeness of ESG ratings. Many ESG rating agencies evaluate firms relative to their industry or sector peers, but the consequences of this methodological choice remain largely anecdotal and have received little systematic empirical attention. Using the Refinitiv ESG framework (which applies a best-in-industry approach) we construct an alternative best-in-sector rating and assess their respective abilities to reflect sustainability-related financial risks. Our findings show that firms with low ESG ratings consistently earn higher abnormal returns than those with high ratings, consistent with the existence of a sustainability risk premium. While this pattern holds across both methodologies, the return spread is significantly larger under the best-in-sector approach. This suggests that broader peer groups are more effective in distinguishing firms with financially material sustainability risks.

Keywords ESG ratings · ESG category scores · Recalculation

Mathematics Subject Classification M14 · G24

Introduction

According to the Global Sustainable Investment Alliance review, the demand for sustainable investment opportunities has increased by 20% in the last two years and reached a total amount of 30.3 trillion assets under management in 2022 (GSIA 2022). This displays the fact that fund managers align their flows and investment strategies with environment, social, and governance (ESG) objectives (Hartzmark and Sussman 2019). This growing demand has driven the importance of ESG ratings, which are widely used to measure firms' sustainability performance and financial risks.

A key feature of many ESG rating systems is their reliance on relative benchmarking. Firms are commonly evaluated against their industry or sector peers. While this approach controls for structural differences across industries, it may

also lead to counter-intuitive outcomes. For example, superior ESG ratings for firms in environmentally harmful industries and low ratings for firms in sustainability-focused industries. Such divergences could affect investment decisions, capital allocation, and ultimately the cost of capital.

Although such inconsistencies are frequently cited in public discourse, they remain largely anecdotal and have received little systematic empirical attention. It remains unclear to what extent the peer group influences ESG ratings and whether this affects their informativeness for investors. This paper aims to close this gap by empirically investigating how peer groups shape the financial informativeness of ESG ratings. Specifically, we study Refinitiv's ESG rating methodology, which employs a best-in-industry approach. This means, that firms are evaluated relative to others in the same industry group. We then construct an alternative best-in-sector rating, in which firms are benchmarked against a broader set of peers within the same economic sector.

This peer group change allows us to examine whether ESG ratings based on broader peer groups offer more relevant signals to financial markets. In doing so, we compare the performance of U.S. stocks sorted by both rating types and analyse the return spread (Δ -Alpha) between firms with low and high ESG scores. If ESG ratings are informative,

✉ Thomas Schmidl
thomas.schmidl@uni-bamberg.de

Matthias Muck
matthias.muck@uni-bamberg.de

¹ Chair of Banking and Financial Control, University of Bamberg, 96045, Bamberg, Germany



we expect stocks with poor sustainability engagement to earn a return premium that compensates for higher sustainability-related risk. The magnitude of this spread provides a measure of how well the rating distinguishes between financially relevant sustainability information.

Our results contribute to the existing literature in several ways. First, we examine the mechanisms through which peer groups affect ESG ratings under the Refinitiv methodology, in which final ratings are calculated as weighted averages of category-specific scores. Both the category scores and the associated weights are influenced by the composition of the peer group. On average, firms tend to receive more favourable scores when benchmarked against less sustainable peer groups. However, the effect on category scores is not necessarily uniform across all firms within an industry. Moreover, the impact of changing weights is theoretically ambiguous. As a result, the overall effect of the peer group on ESG ratings is *ex ante* unclear.

Second, we construct best-in-industry and best-in-sector ESG ratings following the Refinitiv (2023) methodology. Our recalculated best-in-industry scores closely match the original Refinitiv ratings, validating our approach. Nevertheless, small deviations arise due to data limitations and differences in peer group composition. In particular, we do not observe the full global universe of firms used by Refinitiv, nor do we have access to historical industry assignments. Additional discrepancies may stem from firm turnover (e.g., delistings, mergers) and potential refinements in the Refinitiv (2023) methodology over time. Despite these limitations, our analysis of U.S. firms shows that while some ratings remain unchanged across both methodologies, a substantial number of firms receive different ratings if the peer group is changed.

Third and most importantly, we examine the financial informativeness of ESG ratings by sorting firms into portfolios based on their best-in-industry and best-in-sector scores. We find that firms with low ESG ratings consistently earn higher abnormal returns than those with high ratings. This finding holds across both peer group definitions, indicating that both best-in-class measures contain financially relevant information. However, the return spread between the lowest and highest rated portfolios is significantly larger under the best-in-sector approach. This implies that broader peer groups are more effective in identifying financially material sustainability differences between firms. The results are robust across various model specifications, portfolio rebalancing strategies, and sample periods.

Our study is related to two strands of sustainable finance literature. The first strand focuses on the construction and evaluation of ESG measures. Previous research compares ESG rating providers e.g., Billio et al. (2021) and documents significant disagreement among them (Chatterji et al. 2016),

which is often based on the relative importance of different ESG attributes (Berg et al. 2022). Moreover, ESG ratings are subject to construction biases, as shown by Drempetic et al. (2020) and Dobrick et al. (2023) in the context of Refinitiv. While a growing number of studies aim to improve ESG measurement frameworks e.g., Dumrose et al. (2022), little attention has been paid to the role of peer groups in score construction. To the best of our knowledge, we are the first to explicitly analyse the best-in-industry approach as implemented in the Refinitiv (2023) methodology.

The second related strand of literature addresses the relevance of sustainability in asset pricing. Prior research suggests that sustainable firms exhibit lower downside risk and greater resilience in times of market stress e.g., Albuquerque et al. (2020), implying that ESG risks may be priced by financial market participants. Empirical evidence shows that sustainability-related information, including ESG scores, is reflected in stock returns e.g., Hübel and Scholz (2020). Our study contributes to this by examining how the construction of ESG measures, and specifically the peer groups, affects the financial informativeness and usefulness for investment decision-making.

The remainder of this study is structured as follows: The “[Related literature](#)” section gives an overview of related literature. “[The Refinitiv rating](#)” section addresses the construction of the Refinitiv rating and the difference between best-in-sector and best-in-industry scores. The “[Methodology and data](#)” section shows the used data and methodology. The “[Results](#)” section provides our empirical results and a discussion of implications, while the “[Robustness](#)” section presents the robustness checks. The last section concludes.

Related literature

ESG considerations have become a central topic in academic and practical discussions on corporate finance and investment strategies due to a growing emphasis on sustainability from regulatory initiatives and investor changing preferences. In this context, ESG ratings play a crucial role in assessing corporate sustainability performance and informing investment decisions. However, the methodologies underlying these ratings, as well as their financial implications, remain widely debated. This section reviews the existing literature on sustainability and financial returns (“[Sustainability and financial returns](#)” section), the ESG rating market (“[The market of ESG ratings](#)” section), and the methodological challenges associated with ESG ratings (“[Methodological challenges in ESG ratings](#)” section).



Sustainability and financial returns

The significance of sustainability information has grown considerably in recent years. From a societal point of view, sustainability is widely acknowledged as a critical driver of long-term development, enhancing corporate resilience while mitigating adverse environmental and social impacts. In the context of investment decisions, the importance of sustainability-related data has also gained substantial attention. There is a growing body of literature, that examines the connection between firms' increasing sustainability engagement and their financial performance (Friede et al. 2015; Whelan et al. 2021). Specifically, sustainable firms are often perceived as less risky (e.g. Albuquerque et al. (2019); and Hübel and Scholz (2020)). Moreover, being rated as sustainable also has a positive effect on fund inflows (Hartzmark and Sussman 2019). But, motives for investors are diverse. For example, retail investors may deviate their decisions from positive emotions connected with the term sustainability (Heeb et al. 2023), independently of a financial risk-return consideration. Others view stocks with superior ESG performance as financial hedges, due to their enhanced resiliency during times of the financial crisis (Lins et al. 2017) or the COVID-19 pandemic (Albuquerque et al. 2020; Broadstock et al. 2021). As a result, sustainable firms may benefit from lower financing costs (El Ghoul et al. 2011).

The market of ESG ratings

In response to the growing demand for sustainability information, regulatory frameworks have introduced reporting obligations designed to enhance transparency and ensuring that firms provide comprehensive insights into their ESG practices. These regulations have led to an increasing emphasis on sustainability reporting across various countries, with regulatory bodies recognizing the necessity of clear standards to ensure the availability of reliable and comparable ESG data (ESMA 2021). The primary objective of these reporting requirements is to foster transparency and provide investors with essential information to facilitate informed decision-making. ESG ratings, as combined indicators of environmental (e.g., Bolton and Kacperczyk (2021)) and social (e.g., Hong and Kacperczyk (2009); and Edmans (2011)) performance, play a crucial role in this context by consolidating complex and detailed sustainability information into a single, easily interpretable metric (Berg et al. 2022). This simplification allows financial market participants to efficiently assess a firm's sustainability performance.

The ESG rating market is highly diverse (Billio et al. 2021). Specialized agencies, such as the Carbon Disclosure Project (CDP) or RepRisk, focus on particular aspects

of ESG spectrum. In contrast, larger providers like MSCI, Bloomberg, and Refinitiv combine their ESG data with investment tools to extend their services (Douglas et al. 2017). These providers generally rely on public sources to collect sustainability data, which they supplement with proprietary information, such as internal company questionnaires (e.g., RobecoSAM) or disclosures from non-profit organizations. Despite the fact that their data basis is very similar to a large extent, their approaches of processing and interpretation of this is in turn very different. Hence, rating agencies substantially disagree in their sustainable assessment (Chatterji et al. 2016; Berg et al. 2022). Notably, this divergence tends to increase with the volume of disclosed ESG information (Christensen et al. 2022), highlighting the inherent difficulties in objectively assessing ESG engagement of firms.

Methodological challenges in ESG ratings

ESG ratings have been critically examined in the literature. For example, Drempevic et al. (2020) and Dobrick et al. (2023) identify a size bias in Refinitiv's ESG ratings, where larger firms tend to receive higher ratings. Furthermore, Muck and Schmidl (2024) analyse the aggregation mechanism of the Refinitiv ESG rating methodology and find that the weighting of different ESG categories plays a crucial role in determining the financial informativeness of the final rating.

However, a fundamental concern in ESG rating methodologies is the widespread adoption of the best-in-industry approach. Under this framework, a firm's ESG performance is assessed by comparing it only to firms within the same industry, rather than using an objective benchmark to define sustainability. This method reflects the assumption that each industry may face unique sustainability risks, which could vary across industries (Eccles et al. 2012). As a result, firms in different industries disclose different types of ESG information, and failure to disclose certain data may negatively impact their ratings. Moreover, the best-in-industry methodology is believed to incentivize firms in lower ESG-performing industries to enhance sustainability efforts, as targeted investments can substantially improve their ratings.

Despite its theoretical justifications, the best-in-industry approach presents several limitations. A key drawback is that companies operating in industries with lower sustainability standards may receive higher ESG ratings than firms in inherently more sustainable industries. This undermines the comparability of ESG ratings across sectors, making direct comparisons of sustainability performance challenging. Additionally, this approach can lead to situations where firms obtain high ESG ratings, even though their industries as a whole might be excluded from sustainable investment



portfolios due to misalignment with broader sustainability principles.

While there are anecdotal concerns regarding the best-in-industry approach, empirical evidence on its financial implications remains scarce. In particular, little is known about how this methodology affects the financial informativeness of ESG ratings. An ESG rating is considered financially informative if it enables investors to identify and assess the risks associated with investment opportunities. Despite theoretical concerns regarding the best-in-industry approach, its actual influence on investment decision-making has not yet been thoroughly analysed. This study empirically examines whether and to what extent this methodology introduces biases in financial decision-making, with a particular focus on the ESG rating model employed by Refinitiv.

The Refinitiv rating

This section outlines the methodological foundations of the Refinitiv (2023) ESG rating system, which follows a best-in-industry approach. Accordingly, our focus lies on understanding the role of the peer group in shaping ESG scores and ratings. In particular, we examine how the exclusion of peers influences individual scores and the overall rating outcome.

The section is structured as follows: The “[Rating categories](#)” section provides an overview of the ESG categories and associated letter grades. The “[Calculation](#)” section describes the calculation methodology. The “[Impact of the peer group](#)” section analyses the impact of peer groups on individual scores, while the “[Differences in ESG ratings](#)” section investigates how these peer effects translate into differences in overall ESG ratings.

Rating categories

The Refinitiv (2023) ESG Rating is a numerical value ranging from 0 to 100, with higher scores indicating stronger ESG performance. These scores are subsequently translated into letter grades as shown in Table 1. The ESG rating is derived from ten underlying categories, grouped into three pillars: environmental, social, and governance.

The environmental pillar assesses a firm’s resource use, emissions, and environmental innovation. The resource use (*RESU*) category evaluates factors such as energy efficiency, water consumption, and reliance on renewable energy sources. The emissions (*EMIS*) category captures sustainability metrics, including total CO₂ emissions relative to revenue, NOx emissions, and corporate emission reduction policies. Environmental innovation (*ENVI*) reflects efforts to develop environmentally friendly products and investments in substantiality research and development.

The social pillar consists of four categories. Workforce (*WORK*) reflects diversity metrics, trade union representation, employee turnover rates, training programs, and workplace safety. Human rights (*HUMR*) considers corporate policies on fundamental labour rights, including freedom of association and the prohibition of forced or child labour. Community (*COMM*) focuses on a firm’s business ethics, fair competition policies, anti-corruption measures, and community engagement initiatives. Product responsibility (*PROD*) assesses whether a company upholds high standards in customer health and safety, data privacy, and product accountability.

The governance pillar encompasses three categories. Management evaluates board composition, leadership skills, audit committee independence, and executive compensation relative to revenue. Shareholders examines shareholder rights, equal voting power, and corporate governance policies. Corporate social responsibility (CSR) strategy assesses whether a firm discloses its sustainability initiatives in

Table 1 ESG score range and grade description

Score range	Grade	Description
$0.0000 \leq \text{score} \leq 8.3333$	D–	‘D’ score indicates poor relative ESG performance and insufficient degree of transparency in reporting material ESG data publicly.
$8.3333 < \text{score} \leq 16.6666$	D	
$16.6666 < \text{score} \leq 25.0000$	D+	
$25.0000 < \text{score} \leq 33.3333$	C–	‘C’ score indicates satisfactory relative ESG performance and moderate degree of transparency in reporting material ESG data publicly.
$33.3333 < \text{score} \leq 41.6666$	C	
$41.6666 < \text{score} \leq 50.0000$	C+	
$50.0000 < \text{score} \leq 58.3333$	B–	‘B’ score indicates good relative ESG performance and above-average degree of transparency in reporting material ESG data publicly.
$58.3333 < \text{score} \leq 66.6666$	B	
$66.6666 < \text{score} \leq 75.0000$	B+	
$75.0000 < \text{score} \leq 83.3333$	A–	‘A’ score indicates excellent relative ESG performance and high degree of transparency in reporting material ESG data publicly.
$83.3333 < \text{score} \leq 91.6666$	A	
$91.6666 < \text{score} \leq 100.0000$	A+	

This table presents the translation from the Refinitiv ESG score (displayed as numbers between 0 and 100) to the Refinitiv ESG Rating (displayed as grades between A+ and D–).

Source Refinitiv (2023).



alignment with established frameworks such as the Global Reporting Initiative.

Each category score is derived from multiple underlying data points, which can be either numerical (e.g., emissions, water consumption, waste production) or Boolean (e.g., presence of an emission reduction policy, diversity policies). The ESG scores are calculated based on a firm's relative performance within its peer group, as outlined in the next section.

Calculation

The Refinitiv (2023) ESG score is a weighted average of all ten category scores:

$$ESG_i = \sum_{c=1}^{10} w_{c,j} C_{c,i} \quad (3.1)$$

where ESG_i is the ESG score of firm i , and $C_{c,i}$ and $w_{c,j}$ denotes the category score and category weight of category c for firm i within peer group j .

Category scores are determined relative to the peer group. In the case of Refinitiv, peer groups are defined as all firms within the same Thomson Reuters Business Classification (TRBC) industry group (best-in-industry approach) for the categories within the environmental and social pillars. For the governance pillar, the peer group comprises all firms headquartered in the same country. For each data point of a category, all firms within a peer group are ranked according to sustainability. For a numerical data point, such as total emissions, firms are ranked in ascending order, while for a Boolean data point, firms that meet a given sustainability criterion (score = 1) are ranked above those that do not meet the criterion or do not report the metric (score = 0). The score $Score_{d,i}$ for a data point d of firm i is computed as:

$$Score_{d,i} = \frac{no. \text{ worse} + \frac{no. \text{ same}}{2}}{no. \text{ value}} \cdot 1_{eligible} \quad (3.2)$$

where $Score_{d,i}$ is firm i 's percentile score for data point d . The variable $no. \text{ worse}$ represents the number of firms with a worse sustainability performance for the data point, $no. \text{ same}$ denotes the number of firms with identical performance, and $no. \text{ value}$ is the total number of firms within the peer group reporting the given data point. Some firms are assigned a score of zero by construction. This applies, for instance, in Boolean scoring frameworks when a firm fails to meet a predefined sustainability criterion or when no information is available. These firms are treated as non-eligible ("non-eligible firm") and are represented by an indicator variable $1_{eligible}$ equal to zero. Firms meeting the criterion receive

a value of one ("eligible firms"). The individual data point scores within a category are then aggregated into a single category score using a similar scoring mechanism.

Category weights capture the relative importance of each ESG category for a given industry and are specific to the respective peer group. Refinitiv employs a data-driven methodology that assigns higher weights to categories considered more relevant based on an industry's exposure to negative sustainability screening criteria. For example, industries with high emission levels are assigned greater weight in the emissions category. This approach is operationalized by comparing the median values of sustainability indicators across industries. Industries are then sorted into deciles based on these medians. Categories linked to industries in less sustainable deciles receive higher weights, reflecting their presumed materiality. As a result, category weights vary across industries and aim to align ESG assessments with the specific sustainability challenges most relevant to each sector. A modified procedure is applied to the categories within the governance pillar, as detailed below in the "Differences in ESG ratings" section.

Impact of the peer group

In this section, we examine the impact of peer groups on scores. Consider a sector that is divided into two industries, A and B , consisting of M and N firms, respectively. The effect of the peer group is isolated by comparing ESG scores based on best-in-industry versus best-in-sector benchmarking. The analysis generalizes to sectors with more than two industries, as each industry can be subdivided into smaller groups without loss of generality.

Definitions: Let $s_A = \frac{M}{M+N}$ and $s_B = 1 - s_A = \frac{N}{M+N}$ denote the relative sizes of industries A and B . In industry A (B), all x_0 (y_0) firms are non-eligible, where higher scores reflect better sustainability performance. Let $z_S = \frac{x_0 + y_0}{M+N}$ be the share of non-eligible firms in the sector, and $z_A = \frac{x_0}{M}$ and $z_B = \frac{y_0}{N}$ the corresponding shares within industries A and B .

For a given firm i , define $v_{A,i} = \frac{x_{H,i}}{M}$ and $w_{A,i} = \frac{x_{L,i}}{M}$, where $x_{H,i}$ and $x_{L,i}$ represent the number of firms in industry A with strictly higher and lower scores, respectively. Similarly, define $v_{B,i} = \frac{y_{H,i}}{N}$ and $w_{B,i} = \frac{y_{L,i}}{N}$ for industry B . Note that $x_{L,i} + x_{H,i} \leq 1$ and $y_{L,i} + y_{H,i} \leq 1$, as some firms may have identical scores to firm i .

Average Scores: The score of a firm is defined in Eq. (3.2). For calculating average scores, we assume without loss of generality that all eligible firms have distinct,



non-zero scores.¹ The average best-in-sector score μ_S is given by:

$$\mu_S = \frac{\sum_i Score_i}{M+N} = \frac{\sum_{i=x_0+y_0}^{M+N-1} \frac{x+\frac{1}{2}}{M+N}}{M+N} = \frac{1}{2} (1 - z_S^2)$$

Similarly, the average best-in-industry scores $\mu_{A,B}^{ind}$ are:

$$\mu_{A,B}^{ind} = \frac{1}{2} (1 - z_{A,B}^2)$$

Average scores equal $\frac{1}{2}$ when all firms are eligible, implying that peer group definitions do not influence average outcomes in such cases.

If some firms are non-eligible, the average scores may deviate from the benchmark value of $\frac{1}{2}$ and can be lower, but still positive. The magnitude of this deviation increases with the share of non-eligible firms in the respective peer group. Define $\bar{\Delta}$ as the difference between average best-in-industry and best-in-sector scores:

$$\begin{aligned} \bar{\Delta} &= (s_A \mu_A^{ind} + (1 - s_A) \mu_B^{ind}) - \mu_S \\ &= -\frac{1}{2} (1 - s_A) s_A (z_A - z_B)^2 \end{aligned}$$

Thus, the average best-in-industry score across all firms can never exceed the average best-in-sector score. The difference is non-positive and increases in absolute value with the squared difference $(z_A - z_B)^2$.

Remark: In the case of Boolean scores, where only firms meeting a sustainability criterion are eligible and receive a score of 1, it holds that $w_{A,i} = \bar{w}_A = z_A$ and $w_{B,i} = \bar{w}_B = z_B$. Hence, $(z_A - z_B)^2 = (\bar{w}_A - \bar{w}_B)^2$. This implies that the greater the heterogeneity in sustainability performance between industries that is, the larger the difference between \bar{w}_A and \bar{w}_B - the more pronounced the divergence between best-in-industry and best-in-sector scores becomes.

Scores Using Eq. (3.2), the scores of eligible firms satisfy:

$$\begin{aligned} S_{S,i} &= \frac{1}{2} + \frac{1}{2} [s_A (w_{A,i} - v_{A,i}) + (1 - s_A) (w_{B,i} - v_{B,i})] \\ S_{D,i}^{ind} &= \frac{1}{2} + \frac{1}{2} (w_{D,i} - v_{D,i}), \quad D \in \{A, B\} \end{aligned} \quad (3.3)$$

where $S_{S,i}$ denotes the best-in-sector score. Depending on the industry, $S_{A,i}^{ind}$ or $S_{B,i}^{ind}$ is the corresponding best-in-industry score. The scores of non-eligible firms are always zero.

Intuitively, $r_{A,i} = w_{A,i} - v_{A,i}$ and $r_{B,i} = w_{B,i} - v_{B,i}$ represent firm i 's peer-relative scores within industries A

and B , respectively. Recall that w denotes the share of less sustainable peers, while v denotes the share of more sustainable ones. A firm receives a best-in-sector score above $\frac{1}{2}$ if and only if its peer-relative score is positive. To obtain a positive score, the firm must therefore outperform the median firm in its industry. The overall best-in-sector score is then computed as a weighted average of the firm's peer-relative scores across industries, with weights corresponding to the relative sizes of the respective industries.

Remark: For Boolean scores, if a firm has a score of 1, then $v_{A,i} = v_{B,i} = 0$. The corresponding scores are:

$$\begin{aligned} S_{S,i} &= \frac{1}{2} + \frac{1}{2} [s_A \bar{w}_A + (1 - s_A) \bar{w}_B] \\ S_{D,i}^{ind} &= \frac{1}{2} + \frac{1}{2} w_{D,i} \quad D \in \{A, B\} \end{aligned}$$

The best-in-industry score is positively related to the number of non-eligible peers within the same industry. Hence, firms that meet a sustainability criterion benefit more from industry-specific benchmarking when a large share of their industry peers is non-eligible.

Changes in Scores of a Firm: Let $\Delta_{i,A}$ and $\Delta_{i,B}$ denote the difference in scores between best-in-industry and best-in-sector benchmarks. Then, for an eligible firm, it holds that:

$$\begin{aligned} \Delta_{A,i} &= S_{A,i}^{ind} - S_{S,i} = -\frac{1}{2} (1 - s_A) [(w_{B,i} - v_{B,i}) - (w_{A,i} - v_{A,i})] \\ \Delta_{B,i} &= S_{B,i}^{ind} - S_{S,i} = -\frac{1}{2} (1 - s_B) [(w_{A,i} - v_{A,i}) - (w_{B,i} - v_{B,i})]. \end{aligned} \quad (3.4)$$

This is the central result of the section. It shows that firms with identical sustainability characteristics may receive different ESG scores under the best-in-industry approach, depending on their peer-relative scores within the industries that make up the sector. A firm benefits from best-in-industry benchmarking - specifically, when it receives a higher score than under sector-based evaluation - if its peer-relative score within its own industry exceeds that in the other industry. This mechanism tends to favour firms in weaker-performing industries and penalize those in stronger ones, although the effect may not be uniform (as illustrated in the example below). Moreover, the effect is amplified when the firm belongs to a relatively small industry. Firms that are non-eligible receive a score of zero under both the best-in-industry and best-in-sector approaches.

Example (Non-Uniform Effect within an Industry): Consider a highly skewed distribution within industry A , where a single firm I achieves the highest score in the entire sector, while all other firms in A score below all firms in B . All firms are eligible, and the sustainability criterion is numerical (e.g., tons of waste). In this case, we have

¹ See Appendix A.1 for details.



$0 < w_{A,I} < 1$, $w_{B,I} = 1$, and $v_{A,I} = v_{B,I} = 0$. Thus, firm I 's peer-relative score is less than one within industry A , but exactly one within industry B .

For all other firms $i \neq I$ in industry A , it holds that $0 \leq w_{A,i} < 1$, $w_{B,i} = 0$, $0 \leq v_{A,i} < 1$, and $v_{B,i} = 1$. Their peer-relative score in industry A is greater or equal than in industry B , where it equals -1 . From Eq. (3.4) it follows that:

$$\begin{aligned}\Delta_{A,I} &< 0 \\ \Delta_{A,i} &\geq 0 \quad \forall i \neq I\end{aligned}$$

This example illustrates that the direction of score changes within a single industry may not be uniform. Although such cases may be rare, even a high-performing firm can be penalized under industry-specific benchmarking if it belongs to an otherwise weak industry.

Remark: In the Boolean case, Eq. (3.4) simplifies to:

$$\begin{aligned}\Delta_{A,i} &= -\frac{1}{2}(1 - s_A)(\bar{w}_B - \bar{w}_A) \\ \Delta_{B,i} &= -\frac{1}{2}(1 - s_B)(\bar{w}_A - \bar{w}_B)\end{aligned}$$

In this case, it is sufficient to consider the share of firms that do not meet the sustainability criterion and are therefore non-eligible. Firms operating in industries with a higher proportion of non-eligible peers receive higher best-in-industry scores, and vice versa.

Differences in ESG ratings

Differences in ESG ratings between best-in-industry and best-in-sector benchmarking, denoted by ΔESG_i , arise from differences in both category scores and the associated category weights. According to the Refinitiv methodology, category-level scores are first computed as described in the [Impact of the peer group](#) section and then aggregated using the scoring mechanism defined in Eq. (3.2), applied to the sum of individual data point scores (SIDPS) within each category. Consequently, differences in category scores are expected to be positively related to differences in individual scores. These category scores are subsequently weighted to reflect their relative importance in the overall ESG assessment.

The weights are based on negative sustainability screening criteria (e.g., tons of emissions or the absence of specific social policies) considered to be key for each category (key indicator). For each industry, the medians of these key indicators are computed and ranked into deciles.² Industries

² If a category contains multiple indicators, the value of the screening criterion is the average of them.

with the worst median value for a given sustainability criterion are assigned a ‘‘magnitude weight’’ of 10, whereas those with the best median receive a weight of 1. An exception applies to the Governance pillar, which uses fixed magnitude weights.³ The final category weight is calculated as the ratio of a category’s magnitude weight to the sum of all magnitude weights. Thus, magnitude weights and category weights are positively related. In line with the recalculation of the scores, we change the peer group. In doing so, we use the Refinitiv (2023) approach with the same key indicators to derive the sector weightings.

Definitions: The analytical setup follows the ‘‘[Impact of the peer group](#)’’ section. The sector consists of two industries, A and B , comprising M and N firms, respectively. For a given ESG category, let the difference between best-in-industry and best-in-sector scores be denoted by $\Delta_{D,i} = S_{D,i}^{ind} - S_{S,i}$, where $D \in \{A, B\}$, and $\Delta_{D,i}$ is defined in Eq. (3.4).

Let $\bar{S}_{D,i}^{ind}$ and $\bar{S}_{D,i}^S$ denote the weighted average of category scores across all other categories (excluding the one under consideration), under best-in-industry and best-in-sector benchmarking, respectively. To isolate the effect of the category in question, we assume $\bar{S}_{D,i}^{ind} = \bar{S}_{D,i}^S$. The firm’s weight for the respective category is denoted by $\omega_{D,i}^{ind}$ (best-in-industry) and $\omega_{D,i}^S$ (best-in-sector), with both weights bounded between 0 and 1.

ESG Ratings Difference: The overall difference in ESG scores is given by:

$$\begin{aligned}\Delta ESG_i &= \omega_{D,i}^{ind} S_{D,i}^{ind} + (1 - \omega_{D,i}^{ind}) \bar{S}_{D,i}^{ind} - [\omega_{D,i}^S S_{S,i}^S + (1 - \omega_{D,i}^S) \bar{S}_{D,i}^S] \\ &= \omega_{D,i}^{ind} \Delta_{D,i} + (\omega_{D,i}^{ind} - \omega_{D,i}^S) (S_{S,i} - \bar{S}_{D,i}).\end{aligned}\quad (3.5)$$

The first term shows that ESG rating differences are directly related to category score differences. The second term reflects the impact of weight differences between the two benchmarking approaches. This effect is positive when the category score exceeds the average of the other categories, and negative otherwise. It follows that the overall impact on ESG ratings depends on both the magnitude and the direction of category score and weight differences.

Weights: As a key indicator, consider a numerical variable such as emissions or waste, where higher values indicate lower sustainability performance. Let M_A and M_B denote the medians of this indicator in industries A and B , respectively, and assume that $M_A > M_B$. Then the sector-wide median M_S lies within the interval $[M_B, M_A]$.

Because medians are positively related to firm rankings—and these rankings, in turn, determine category

³ In the Governance pillar, the magnitude weights for the categories Management, Shareholders, and CSR Strategy are fixed at 10, 3, and 2, respectively.



weights—firms from industries with higher medians (e.g., A) tend to receive higher weights under sector-based benchmarking. Conversely, firms from industries with lower medians (e.g., B) are more likely to receive lower weights.

This effect is further amplified by differences in industry size. The number of firms in each industry influences the position of the overall sector median. When the share of firms from industry A (s_A) is small, the sector median M_S will lie closer to M_B . In this case, removing industry B through best-in-industry benchmarking induces a more pronounced shift in the effective median, thereby intensifying the resulting change in category weights. A symmetric argument applies to industry B .

Total Effect: The firm's category score difference $\Delta_{D,i}$ influences the direction of the ESG rating change positively through the first term in Eq. (3.5). Each category score is derived from a positively defined sustainability metric (SIDPS). Let m_D denote the median of the SDPS within industry D . According to Eq. (3.4), the category score tends to be positively related to m_A and negatively related to m_B . As previously noted, this effect is particularly pronounced in smaller industries.

However, the overall impact on an individual firm's ESG rating remains ambiguous. As motivated by the example in the “Differences in ESG ratings” section, a lower median $m_{A,i}$ compared to $m_{B,i}$ does not necessarily imply that $\Delta_{A,i} > 0$ (and vice versa), since the influence of medians on score differences may not be uniform across firms.

Even when abstracting from this non-uniformity and assuming a plausible empirical scenario - namely, that the key indicators and SDPS underlying category weights and scores, and thus M_D and m_D , are perfectly aligned in how they reflect sustainability performance across industries - the total effect on the ESG rating remains unclear. In such a setting, changes in category scores and weights would tend to move in the same direction. However, the final impact on the ESG rating also depends on how a firm's best-in-sector score in a given category compares to its average score across all categories. Since this deviation can be either positive or negative, the two components in Eq. (3.5) - score and weight effects - may either reinforce or offset one another.

Numerical example

To illustrate the impact of the recalculation on firms' ratings, we consider the energy sector, which consists of five industries (coal, oil & gas, oil & gas related equipment and services, renewable energy, and uranium). For the whole sample and examination period, the sector-based ESG scores in the coal industry have a mean value of 35.04 and exhibit

a minimum of 2.12 and a maximum of 81.99.⁴ In contrast, the industry-based ESG score and the associated category scores show mostly higher means. This demonstrates that the changed peer group has a significant influence on the sustainability assessment of a company. By comparison, the companies in the renewable energy industry increase on average their resource use and environmental innovation category scores. The aggregated measure slightly decreases since some social category scores diminish through the recalculation.

To explore the economic impact of the calculation approach, we illustrate the changes in the category scores and the aggregated ESG score using a numerical example. Therefore, we consider two firms from the coal and renewable energy industries, respectively. Table 2 shows that the overall ESG scores of both coal firms decrease after the recalculation, while the two renewable energy firms show an increase in the overall ESG score. The firms in the renewable energy industry show substantial improvements across the environmental and social categories after the sector-based recalculation. For example, Renewable Energy Firm I exhibits a growth of 11.16 in the resource use score, which reflects a rating change from C+ to B-. In the environmental innovation category, both firms record an increase of more than two rating levels. Additionally, the two firms show a similar pattern in the social categories with a sharp increase in the product responsibility score of Renewable Energy Firm I.

In contrast, the majority of the environmental and social categories of the two coal firms decrease. Coal Firm I exhibits a decrease of 12.08 in the emissions category. This illustrates that when coal firms are not compared with themselves, they perform worse than comparable firms of other industries in terms of emissions. However, the resource use and workforce category scores of Coal Firm II increase by 0.66 and 5.26, respectively.

In summary, this provides anecdotal evidence suggesting that some industries benefit from the Refinitiv best-in-industry approach, since they are evaluated only against their peers. These observations demonstrate that best-in-sector comparisons could result in ESG scores that are more in line with the economic expectations or intuition reflected in sustainability ratings.

Methodology and data

This section outlines the empirical methodology to evaluate the impact of peer groups on the informativeness of ESG ratings. We begin by introducing the methodological

⁴ For further explanation, see Appendix Table 11.



Table 2 Recalculation example

Panel A	Coal Firm I			Coal Firm II			Industry-Means	
	<i>Diff</i>	Sec-ESG	Ind-ESG	<i>Diff</i>	Sec-ESG	Ind-ESG	Sec-ESG	Ind-ESG
ESG	24.79	38.08	62.87	4.79	56.35	61.15	35.04	43.36
RESU	1.14	90.11	91.25	- 0.66	66.91	66.25	38.52	44.77
EMIS	12.08	48.63	60.71	7.13	63.11	70.24	40.31	44.77
ENVI	0.00	0.00	0.00	0.00	0.00	0.00	5.66	6.33
WORK	13.66	51.12	64.77	- 5.26	74.58	69.32	47.64	50.94
HUMR	4.89	9.63	14.52	12.40	48.89	61.29	16.51	21.55
COMM	25.35	16.70	42.05	7.25	80.25	87.50	33.39	47.06
PROD	8.07	59.43	67.50	1.23	22.52	23.75	33.38	32.86
Panel B	Renewable Energy Firm I			Renewable Energy Firm II			Industry-Means	
	<i>Diff</i>	Sec-ESG	Ind-ESG	<i>Diff</i>	Sec-ESG	Ind-ESG	Sec-ESG	Ind-ESG
ESG	- 17.10	62.85	45.75	- 12.89	70.70	57.81	40.04	42.80
RESU	- 11.16	57.77	46.61	- 10.48	97.77	87.29	42.16	37.87
EMIS	3.59	34.93	38.52	5.56	85.42	90.98	34.23	38.69
ENVI	- 17.46	23.02	5.56	- 17.46	73.02	55.56	63.15	45.33
WORK	- 11.81	63.22	51.41	2.07	73.28	75.35	37.47	47.26
HUMR	- 2.10	92.10	90.00	- 11.58	83.58	72.00	25.10	26.09
COMM	- 15.54	92.46	76.92	13.68	40.94	54.62	32.42	44.40
PROD	- 65.85	73.23	7.38	- 17.63	59.43	41.80	53.67	41.66

This table illustrates a numerical example of how the category and aggregated ESG scores change after the recalculation for the year 2021. We present two firms from the coal industry and two firms from the renewable energy sector. *Ind-ESG* is the recalculated industry-based score, *Sec-ESG* is the recalculated sector-based score, and *Diff* is the difference between them. A positive difference indicates that the industry-based scores are higher than the sector-based scores, while a negative difference indicates the opposite. The means of the ESG and category scores are the mean values across the industry peers over the examination period.

framework, including the construction of best-in-industry and best-in-sector ESG scores and the portfolio sorting approach used to assess financial performance, in the “[Methodology](#)” section. We then describe the underlying dataset, which combines ESG information from Refinitiv with financial market data to estimate return differentials across rating-based portfolios, in the “[Data](#)” section.

Methodology

Our main objective is to assess the impact of the peer group on the financial informativeness of ESG ratings. Specifically, we examine how the choice of the peer group affects the ability of ESG scores to reflect the inherent sustainability risk exposure of a firm. Therefore, we compare two types of ESG ratings: best-in-industry and best-in-sector ratings. Both are based on the Refinitiv methodology.

In practice, only best-in-industry ratings are available at the Refinitiv database. However, (historical) ESG scores could include small discrepancies in terms of the underlying methodology and benchmark, which may introduce sample selection biases. To address this issue, we reconstruct best-in-industry ratings ourselves to ensure that the benchmark is consistent with our available sample. This approach also guarantees that average category ratings do not exceed 50, as discussed in the “[Impact of the peer group](#)” section.

For industry group and economic sector classification, we rely on Refinitiv’s TRBC scheme. TRBC is a five-level system consisting of 13 economic sectors, 33 business sectors, 62 industry groups, 154 industries, and 898 activities. The Refinitiv (2023) ESG methodology utilizes the industry group as the relevant benchmark. Accordingly, we use this level to recalculate the best-in-industry scores. In contrast, best-in-sector scores are based on the broader economic sector. This is the highest level at which peer-specific weights are still applied under the Refinitiv approach. This distinction affects the peer groups used for calculating scores within the environmental and social pillars. In contrast, scores for categories in the governance pillar remain unaffected, as firms are benchmarked against firms in the same country.

We sort firms into decile and quintile portfolios based on their best-in-industry or best-in-sector ESG score. Firms with the lowest (highest) scores are assigned to the bottom (top) portfolios. The Refinitiv scores are updated at fiscal year’s end. Hence, we rebalance our portfolios at the end of December in year t , and holding them from the beginning of January until the end of December in the following year $t + 1$.

To analyse the informativeness of the ESG scores, we use the five-factor model by Fama and French (2015):



$$R_t = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \epsilon_t, \quad (4.1)$$

where R_t is the return reduced by the risk-free rate at time t , and α is the unexplained abnormal return (i.e., the sustainability premium) by the model. The Fama and French (2015) five-factor model captures the return of the market (MKT), the return of a size factor (SMB), the return of a value factor (HML), the return of a profitability factor (RMW), and the return of an investment factor (CMA).

To evaluate the financial informativeness of the ESG scores, we rely on the semi-strong form of the efficient market hypothesis, that all publicly available information is priced by financial market participants (Fama 1970). Specifically, we assume that all sustainability information in the form of ESG scores is reflected in the prices. If ESG scores contain priced information, we expect α to be lower for high-ESG firms and, consequently, higher for low-ESG firms. Following Khan et al. (2016), we focus on the Δ -Alpha, defined as the difference in abnormal returns between the top and bottom portfolios. Consequently, a higher Δ -Alpha signals greater informativeness of ESG scores. From an investment perspective, this corresponds to the return of an investment strategy that has long positions in stocks with low ESG scores and short positions in stocks with high ESG scores, based on the respective peer-group-specific ESG score.

Data

The focus of the study is the U.S. stock market. However, the recalculation of the best-in-industry and best-in-sector ratings requires all worldwide rated firms by Refinitiv as reference. Hence, we use an initial universe of 10,510 companies that are included in the Refinitiv ESG rating procedure in the year 2021. For each of these firms, we obtained the required data to recalculate both best-in-industry and best-in-sector scores. Since the examination period starts in the year 2010 and ends in 2022, our final sample consists of 59,978 firm-year observations.

In the following examination, we use a subsample consisting of 3,098 firms having their headquarters in the U.S., resulting in 18,078 firm-year observations. With 29.24% of the firms, U.S.-based firms represent the largest share in our dataset. For each of these U.S. firms, price, accounting, and ESG data are taken from Refinitiv database. Factor data for the Fama and French (2015) five-factor model is obtained from Kenneth R. French's data library.⁵

Table 3 presents the best-in-industry (Panel A) and best-in-sector (Panel B) environmental and social category scores. In the whole sample, the means of the best-in-sector category scores are slightly higher than those of the best-in-industry category scores, while the standard deviations (reflecting the extent to which ESG ratings could differentiate between firms based on sustainability performance) stay on a similar level across both peer group approaches.

Table 3 Descriptive statistics

	U.S. Sample			Whole sample		
	Obs	Mean	S.D	Obs	Mean	SD
<i>Panel A: Industry-based score</i>						
RESU	18,078	25.25	31.65	59,978	36.80	33.10
EMIS	18,078	24.18	30.22	59,978	37.48	33.03
ENVI	18,078	16.06	26.71	59,978	23.27	30.48
WORK	18,078	37.62	25.82	59,978	49.39	29.16
HUMR	18,078	20.02	29.86	59,978	26.71	32.56
COMM	18,078	56.43	25.14	59,978	47.24	30.13
PROD	18,078	39.12	25.77	59,978	43.12	30.43
ESG	18,078	38.38	18.93	59,978	42.70	20.69
<i>Panel B: Sector-based score</i>						
RESU	18,078	25.31	31.46	59,978	37.15	33.20
EMIS	18,078	24.09	29.90	59,978	37.88	33.11
ENVI	18,078	16.17	27.36	59,978	23.42	31.37
WORK	18,078	37.14	25.05	59,978	49.68	29.22
HUMR	18,078	20.28	30.09	59,978	26.83	32.71
COMM	18,078	56.51	24.27	59,978	47.31	30.26
PROD	18,078	38.21	26.29	59,978	43.53	30.89
ESG	18,078	32.67	21.20	59,978	39.91	22.63

This table presents the number, the mean, and the standard deviation of the recalculated ESG scores. The U.S. sample is the examination sample and includes firms having their headquarters in the U.S., while the whole sample contains every recalculated firm-year-observation.

⁵ Web-page: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.



Turning to the U.S. sample, we observe that companies located in the U.S. exhibit substantially lower average ESG category scores. The largest differences of around 12 points are observed in the resource use (*RESU*), emissions (*EMIS*), and workforce (*WORK*) categories. These differences are the equivalent of one and a half letter grade. The only exception is the Community (*COMM*) category, where the average score increases from ca. 47 to 56. Overall, the U.S. results are consistent to the global patterns: the means and standard deviations of scores remain broadly stable across benchmarking approaches.

Table 4 illustrates the rating transition matrix between best-in-industry and best-in-sector ESG ratings for U.S. firms. After the change of the benchmark, around 29% (5,287 of 18,078 ratings) of the companies remain at the identical rating level, while 56% (10,195 of 18,078 ratings) of the companies receive a downgrade. For example, 240 of the 841 firm-year observations rated between A+ and A- (grey marked horizontal area in Table 4), according to the industry-based approach, receive a downgrade that results in a B+ rating or worse. However, there are also upgraded companies. 305 companies that receive a rating in the B rating range (grey marked area at the right in Table 4) could enhance their rating to at least an A- due to the sector-based recalculation. The largest decreases are observed in one firm-year observation from A- to C+ and in three firm-year observations from A to B.

Results

This section presents the empirical results of our analysis. We begin by validating the plausibility of our recalculated best-in-industry ESG scores by comparing them to the original Refinitiv scores (“Fit of recalculated best-in-industry ratings” section). We then examine the relationship between ESG scores and stock returns using the Fama and French (2015) five-factor model, comparing the informativeness of best-in-industry and best-in-sector scores (“ESG ratings and financial returns” section). Finally, we explore several alternative model specifications to disentangle the roles of category weights, firm characteristics, and sustainability-related risk factors (“Alternative model specifications” section).

Fit of recalculated best-in-industry ratings

As discussed in the “Methodology” section, we use the original Refinitiv (2023) methodology to recalculate the industry-based category scores. To validate the appropriateness of this approach, we compare them with the original Refinitiv scores. Table 5 shows the results by category and economic sector. We find that the correlations between the original and

Table 4 Change in ratings

	Best-in-sector										Total		
	D-	D	D+	C-	C	C+	B-	B	B+	A-		A	A+
D-	193	49	3	0	0	0	0	0	0	0	0	0	245
D	455	902	251	20	0	0	0	0	0	0	0	0	1,628
D+	267	1,716	826	311	19	1	0	0	0	0	0	0	3,140
C-	12	1,126	1,304	737	265	53	4	1	0	1	0	0	3,503
C	0	246	1,002	702	533	248	56	6	0	2	1	0	2,796
C+	0	23	350	438	499	461	256	56	7	2	0	0	2,092
B-	0	0	46	119	206	386	412	256	69	2	1	0	1,497
B	0	0	0	18	48	161	321	429	265	47	1	0	1,290
B+	0	0	0	0	2	27	108	278	377	234	20	0	1,046
A-	0	0	0	0	0	1	7	40	177	298	77	1	601
A	0	0	0	0	0	0	0	3	12	80	117	11	223
A+	0	0	0	0	0	0	0	0	0	1	14	2	17
Total	927	4,062	3,782	2,345	1,572	1,338	1,164	1,069	907	667	231	14	18,078

This table displays a matrix of the recalculated industry-based ESG ratings (vertical axis) and the sector-based ESG ratings (horizontal axis) of U.S. firms. Firms on the diagonal line (in bold) have no rating change, companies above the diagonal line receive a rating upgrade through the peer group change, and companies below the diagonal line are downgraded.



Table 5 Validation of recalculated ESG category scores

	RESU	EMIS	ENVI	WORK	HUMR	COMM	PROD
<i>Academic & Educational Serv.</i>							
Correlation	0.9572	0.9608	1.0000	0.8989	0.9529	0.8562	0.9550
Regression	0.9358	0.9428	1.0000	0.8799	0.9282	0.8378	0.9467
Mean	-1.6487	0.0638	0.0000	-1.0980	0.0658	1.2755	0.0233
<i>Basic Materials</i>							
Correlation	0.9820	0.9802	0.9681	0.9699	0.9944	0.9807	0.9890
Regression	0.9826	0.9780	0.9852	0.9797	0.9983	1.0081	0.9839
Mean	-1.1605	-1.5082	1.0249	-2.0510	-0.1521	-1.8768	-1.0591
<i>Consumer Cyclical</i>							
Correlation	0.9488	0.9607	0.9743	0.8964	0.9944	0.9320	0.9701
Regression	0.9409	0.9488	0.9612	0.9253	0.9813	0.9833	0.9628
Mean	-1.8804	-1.9238	-0.9066	-2.9315	-0.6165	-3.6699	-1.3697
<i>Consumer Non-Cyclical</i>							
Correlation	0.9517	0.9542	0.9648	0.9064	0.9890	0.9362	0.9623
Regression	0.9449	0.9508	0.9550	0.9330	0.9757	0.9831	0.9564
Mean	-2.0927	-2.0023	-0.4856	-3.7461	-0.6779	-3.2996	-2.5723
<i>Energy</i>							
Correlation	0.9760	0.9776	0.9881	0.9704	0.9852	0.9596	0.9809
Regression	0.9540	0.9588	0.9921	0.9721	0.9597	0.9742	0.9580
Mean	-2.7765	-2.8702	0.0353	-3.6522	-0.9463	-3.4714	-1.4846
<i>Financials</i>							
Correlation	0.9936	0.9779	0.9936	0.9804	0.9949	0.9609	0.9897
Regression	0.9809	0.9606	0.9840	0.9812	0.9919	1.0060	0.9832
Mean	-0.5576	-0.9613	-0.4607	-2.1838	-0.1304	-4.1361	-1.1422
<i>Healthcare</i>							
Correlation	0.9812	0.9806	0.9687	0.9608	0.9928	0.9795	0.9758
Regression	0.9605	0.9701	0.9711	0.9530	0.9707	1.0037	0.9682
Mean	-0.7510	-0.4230	-0.0862	-2.4773	-0.5461	-3.1609	-1.9311
<i>Industrials</i>							
Correlation	0.9774	0.9775	0.9755	0.9450	0.9932	0.9700	0.9652
Regression	0.9664	0.9628	0.9636	0.9539	0.9838	0.9932	0.9699
Mean	-0.7656	-1.1395	-0.6131	-2.3416	-0.3928	-2.7793	-0.8122
<i>Real Estate</i>							
Correlation	0.9871	0.9856	0.9805	0.9730	0.9932	0.9727	0.9850
Regression	0.9761	0.9628	0.9830	0.9610	0.9845	1.0012	0.9698
Mean	-1.3641	-1.3661	-0.3310	-2.0765	-0.1137	-2.7694	-1.2508
<i>Technology</i>							
Correlation	0.9687	0.9693	0.9671	0.9337	0.9939	0.9633	0.9706
Regression	0.9517	0.9511	0.9541	0.9446	0.9740	0.9943	0.9645
Mean	-1.7319	-1.7350	-0.8158	-3.8414	-0.6489	-3.4790	-2.4715
<i>Utilities</i>							
Correlation	0.9768	0.9857	0.9122	0.9682	0.9958	0.9822	0.9921
Regression	0.9686	0.9796	0.8835	0.9795	0.9965	0.9770	0.9854
Mean	-1.4386	-1.9181	-2.8188	-2.4839	0.0377	-1.2876	-1.4711
<i>Average</i>							
Correlation	0.9749	0.9745	0.9743	0.9470	0.9931	0.9612	0.9767
Regression	0.9620	0.9595	0.9648	0.9557	0.9816	0.9935	0.9693
Mean	-1.3322	-1.4737	-0.4683	-2.7297	-0.4163	-3.1553	-1.4977

This table presents the correlation coefficients, regression coefficients, and differences in means between the original Refinitiv category scores and the recalculated industry-based category scores. We regress the recalculated category scores as a dependent variable with the Refinitiv original scores as an independent variable and year-fixed effects. All correlation and regression coefficients are highly significant (p -value < 0.01). The mean is the difference in means between the recalculated and the original Refinitiv category scores



recalculated category scores are consistently high and close to 1. On average, the correlation coefficients range between 0.95 to 1.00, with one matching the original score (*ENVI* in *Academic & Educational Services*). To further examine the fit of the self-constructed scores, we regress the recalculated scores (dependent variable) with the original Refinitiv scores (independent variable) and year-fixed effects. The estimated coefficients are close to 1 and highly significant, indicating a nearly perfect linear relationship, and consequently, show that the recalculated scores closely replicate the original ratings. In addition, we find a minor systematic deviation in means between the category scores. These are within the range of 0.5 to 3 points in a 100-point scale.

There are three main reasons why the fit between original and recalculated scores may not be perfect. First, ESG ratings depend on a specific peer group used for the benchmarking. Differences in peer group composition will lead to differences in scores. In our recalculation procedure, we take 10,510 companies for 2021 as the basis and assume that these are also rated in the previous years. As a result, our list does not include companies that are part of the rating calculation process in the past but no longer appear in the base year. This may be the case if a respective company was involved, for example, in past insolvency issues, mergers, or acquisitions. We must therefore assume that our list is partially incomplete, and we do not match the exact peer groups for the recalculation.

Second, it is possible for firms to be assigned to a different industry over time. According to the TRBC classification of Refinitiv, the assignment to an industry is based on sales generated in an individual business segment. Since the rated companies usually generate sales in several industry segments and some of them are thematically very close to each other (e.g., *Aerospace & Defense* and *Government Activity*), volatility in sales leads to a different industry allocation over time. This leads to changes in the peer group and, consequently, in the recalculated scores. However, due to the lack of historical industry classification of the individual firms, these changes cannot be taken into account.

Third, the scores are fixed after 5 years (Refinitiv 2023). This means, if 2021 is the current fiscal year in which Refinitiv discloses its category scores, all scores prior to fiscal year 2017 are no longer updated, even if the underlying data points are corrected or the calculation method has changed. As a result, scores in our sample dating further back in time may reflect outdated information and are particularly affected by this inaccuracy.

Despite these limitations, we find a strong overall fit between the recalculated and original Refinitiv scores - both statistically and economically.

ESG ratings and financial returns

Financial informativeness

Table 6 presents the regression coefficients of the Fama and French (2015) five-factor model for the top and bottom decile and quintile portfolios. Panel A shows the results of the industry-based ESG scores and Panel B of the sector-based ESG scores. We find significant abnormal returns (Alphas) across all portfolios. For the industry-based and the sector-based scores, the top portfolios exhibit negative Alphas, while the bottom portfolios display positive Alphas. In line with prior literature, these findings point to the negative relationship between stock returns and sustainability engagement (e.g., Bolton and Kacperczyk (2021); Hübel and Scholz (2020)). In addition, this suggests the existence of a sustainability premium, and that both peer-group-specific ESG scores are capable of identifying it.

However, differences emerge when comparing best-in-industry and best-in-sector results. For industry-based scores, we find an annualized Alpha difference (Δ -Alpha) of 9.64% for the decile portfolios and 7.42% for the quintile portfolios. In comparison, when sorting according to the sector-based scores, the Δ -Alpha increases to 13.62% for decile portfolios and 9.74% for quintile portfolios. The bottom portfolios primarily drive those differences. This means, firms with the weakest ESG scores. Differences in unexplained returns (Alphas) for the top portfolios are much smaller.

In summary, the results suggest that the peer group has a meaningful impact on the financial informativeness of ESG ratings. In particular, the best-in-sector approach appears to more effectively detect firms with weak sustainability performance. This, improves the identification of a return premium required by investors for firms with greater sustainability risk exposures.

Firms with downgraded ratings

The previous section highlights that the choice of the peer group primarily affects firms in the bottom ESG portfolios. But what does this mean in practice for portfolio managers of sustainable funds that invest in firms with high sustainability commitments? Does it matter if they use best-in-industry or a best-in-sector ratings to assess a firm's ESG risk? To answer this question, we examine three sustainable investment strategies based on minimum ESG thresholds: "Strategy A-" targets firms rated at least A-, "Strategy B" includes all firms rated B+ or higher, and "Strategy B" applies a B+ cutoff. These thresholds are based on best-in-industry ratings, reflecting how such strategies might be



Table 6 Regression results of different ESG scores

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Industry-based score (N = 155)								
<i>MKT</i>	0.9694***	(20.6642)	1.0153***	(43.1018)	0.9975***	(25.7156)	1.0027***	(50.2973)
<i>SMB</i>	1.0009***	(10.9887)	0.1573***	(3.4388)	0.8646***	(11.4800)	0.2485***	(6.4214)
<i>HML</i>	0.0790	(0.9711)	0.2178***	(5.3325)	0.0853	(1.2678)	0.2181***	(6.3088)
<i>RMW</i>	-0.2934***	(-2.7036)	0.1570***	(2.8805)	-0.2628***	(-2.9280)	0.2087***	(4.5249)
<i>CMA</i>	-0.2044	(-1.6311)	-0.0563	(-0.8951)	-0.0864	(-0.8336)	-0.0395	(-0.7421)
<i>Intercept</i>	0.5001**	(2.5024)	-0.2938***	(-2.9282)	0.4553***	(2.7553)	-0.1795**	(-2.1138)
R^2	0.8700		0.9446		0.9021		0.9605	
Annualized alpha	6.17**		-3.47***		5.60***		-1.82**	
Δ Alpha				9.64***				7.42***
Panel B: Sector-based score (N = 155)								
<i>MKT</i>	0.9024***	(17.2796)	0.9985***	(51.9741)	0.9448***	(21.8361)	1.0312***	(51.4980)
<i>SMB</i>	0.8729***	(8.6096)	0.1067***	(2.8599)	0.8506***	(10.1267)	0.2286***	(5.8794)
<i>HML</i>	-0.2283**	(-2.5210)	0.2553***	(7.6611)	-0.1582**	(-2.1088)	0.2559***	(7.3688)
<i>RMW</i>	-0.8128***	(-6.7279)	0.1300***	(2.9240)	-0.6790***	(-6.7834)	0.1628***	(3.5151)
<i>CMA</i>	-0.2866**	(-2.0548)	0.0496	(0.9659)	-0.2003*	(-1.7334)	0.0289	(0.5402)
<i>Intercept</i>	0.8499***	(3.8206)	-0.2475***	(-3.0244)	0.6248***	(3.3902)	-0.1661*	(-1.9473)
R^2	0.8380		0.9609		0.8812		0.9621	
Annualized Alpha	10.69***		-2.93***		7.76***		-1.98*	
Δ Alpha				13.62***				9.74***

This table presents coefficient estimates and t-statistics of the recalculated industry-based and sector-based ESG score decile and quintile regressions. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: *MKT*, *SMB*, *HML*, *RMW*, *CMA*. The results are based on the monthly average returns between the years 2010 and 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks.

***, ** and * reflect significance at the 1%, 5%, and 10% level.

implemented in practice using the standard Refinitiv ESG data.

However, under the best-in-sector approach, some firms meeting these thresholds while others would fall short. We focus on these “downgraded” firms by forming a portfolio and computing their Alphas. Table 7 presents the Alpha estimates of the three possible investment strategies. Panel A displays portfolios consisting of A+ to A- firms according to the recalculated industry-based ESG ratings. Following this, the one-level downgrade portfolio contains all superior rated firms (according to the best-in-industry approach) with at least one rating decrease through the sector-based recalculation, which are now below the threshold. The two- and three-level downgrade portfolios include firms with at least two or three rating downgrades, respectively. If sustainable funds are just allowed to invest in at least A- rated firms (Panel A), we would expect, that the downgraded firms should exhibit higher positive excess returns since market participants price the risk according to the actual sustainability risk and not to the third-party rating.

According to the hypothesis, in Panel A, the two-level downgrade portfolio exhibits significantly positive returns.

Moreover, the annualized Alpha estimates are significant for the one- and two-level downgrade portfolios for the B+ (Panel B) and B (Panel C) strategy. The results show that especially the downgraded firms (according to the best-in-sector approach) are responsible for the positive financial returns within a sustainable portfolio. As a consequence, the best-in-industry ratings allow portfolio managers to invest in superior rated firms without having to forgo positive returns. We assume, that those high yields originate to a large extent from firms in environmentally harmful industries. In summary, our results show that the best-in-industry approach favours green-washing suspicions for sustainable investment strategies of mutual funds.

Alternative model specifications

This section explores alternative model specifications to better understand the observed differences in the financial informativeness of ESG scores across peer groups. Specifically, we aim to unbundle the influence of weighting schemes, address potential endogeneity concerns, and examine the explanatory power of sustainability-related



Table 7 Regression regression results of rating strategy

	Best-in-class		1 level downgrade		2 level downgrade		3 level downgrade	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Strategy								
<i>Intercept</i>	-0.0280	(-0.1989)	0.4795	(1.2345)	1.2325**	(2.1998)	0.0285	(0.0252)
Annualized alpha	-0.34		5.91		15.84**		0.34	
Panel B: Strategy B								
<i>Intercept</i>	0.2387	(1.1313)	1.1467**	(2.4399)	2.7781*	(1.7583)	0.9487*	(1.7031)
Annualized alpha	2.90		14.66**		38.93*		12.00*	
Panel C: Strategy B								
<i>Intercept</i>	0.2430	(0.9725)	0.8555*	(1.6706)	2.2776	(1.4601)	0.3469	(0.8234)
Annualized alpha	2.96		10.76*		31.03		4.24	

This table shows coefficient estimates and t-statistics of four investment strategies. Best-in-Class represents a strategy that invests in all companies that have a better industry-based rating than a defined lower bound. For Panel A, this lower bound is an A- ESG rating, for Panel B, a B+ ESG rating, and for Panel C, a B ESG rating. 1 level downgrade represents an investment portfolio that includes the part of companies of the former strategy that have at least a one-level downgrade of their ESG rating through the peer group change. 2 level downgrade and 3 level downgrade represent investment portfolios of companies with at least two level or three level downgrades in the ESG rating. All values are presented in percent. Those investment portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are omitted, and according Fama and French (2015): *MKT*, *SMB*, *HML*, *RMW*, *CMA*.

***, ** and * reflect significance at the 1%, 5%, and 10% level.

risk factors. Therefore, we introduce three complementary approaches: a constant-weighting method to isolate the impact of the peer group (“**Weights**” section), a two-step procedure that accounts for firm-level fundamentals (“**Two step procedure**” section), and an extended factor model that includes an ESG-specific risk factor (“**ESG adjusted model**” section).

Weights

As shown in “**The Refinitiv rating**” section, the difference between best-in-sector and best-in-industry ESG scores depends not only on differences in category scores but also on deviations in category weights. The latter effect is ambiguous and may relativize the impact of score differences. As noted by Muck and Schmidl (2024), the weighting scheme has a significant empirical influence on the financial informativeness of ESG ratings. To isolate the effect of weighting, we fix category weights while allowing scores to vary. Specifically, for each firm i , we set the best-in-sector weight equal to the firm i 's best-in-industry weight, i.e.,

$\omega_{S,i} = \omega_{D,i}^{ind}$, where D denotes the industry and S denotes the sector to which firm i belongs.

Table 8 presents the regression estimates of the decile and quintile portfolios. Panel A displays the results of the industry-based ESG scores. To facilitate the comparison, we replicate these results from Table 6. Panel B reports the coefficients for the sector-based ESG scores with constant weighting. For both decile and quintile portfolios, the estimated Δ -Alpha is higher under the best-in-sector approach than under the best-in-industry benchmark, even though

category weights are held constant. This suggests that differences in category scores alone contribute meaningfully to the financial informativeness of ESG scores. However, the Δ -Alpha estimates remain lower than those reported in Table 6.

In particular, we find that the differences in abnormal returns between best-in-sector and best-in-industry scores in the bottom portfolios are substantially reduced when best-in-industry weights are applied instead of best-in-sector weights. For example, using best-in-sector weights, the abnormal return in the bottom decile portfolio is 4.52 percentage higher than under the best-in-industry benchmark. Isolating the benchmark effect, this difference decreases to 1.84 percentage. These observations suggest that differences in scores and weights tend to operate in the same direction. More importantly, they provide empirical evidence that variation in category weights is a relevant driver of the observed differences in financial informativeness between best-in-sector and best-in-industry ESG scores.

Two step procedure

As discussed in the “**Related literature**” section, ESG ratings and financial performance may be affected by each other. To address potential endogeneity concerns, we follow a two-step procedure. First, we estimate an ESG score with respect to firm size, market-to-book ratio, profitability, and leverage, while controlling for year and sector fixed effects.⁶ In the second step, we sort firms into decile and quintile portfolios according to their estimated best-in-industry and

⁶ For further explanation, see Appendix A.2.



Table 8 Additional tests

	Deciles			Quintiles		
	Annualized alpha			Annualized alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Industry-based score						
Constant weighting						
<i>Intercept</i>	0.5001**	-0.2938***		0.4553***	-0.1795**	
Annualized alpha	6.17**	-3.47***	9.64***	5.60***	-1.82**	7.42***
Two step procedure						
<i>Intercept</i>	0.7811***	-0.1145		0.5813***	-0.1117	
Annualized alpha	9.79***	-1.37	11.16***	7.20***	-1.33	8.53***
ESG adjusted model						
<i>Intercept</i>	-0.0532	-0.1702		-0.0577	-0.0845	
Annualized alpha	-0.64	-2.02	1.38	-0.69	-1.01	0.32
Panel B: Sector-based score						
Constant weighting						
<i>Intercept</i>	0.6440***	-0.1812**		0.5784***	-0.1332	
Annualized alpha	8.01***	-2.15**	10.16***	7.17***	-1.59	8.76***
Two step procedure						
<i>Intercept</i>	0.7936***	-0.1932**		0.6751***	-0.1786*	
Annualized alpha	9.95***	-2.29**	12.24***	8.41***	2.12*	10.53***
ESG adjusted model						
<i>Intercept</i>	0.2800	-0.1037		0.0540	-0.0135	
Annualized alpha	3.41	-1.24	4.65**	0.65	-0.16	0.81

This table shows annualized alpha estimates and differences of the additional tests. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: *MKT*, *SMB*, *HML*, *RMW*, *CMA*. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. Under Constant weighting, we report the Alpha estimates and differences of industry-weighted ESG scores. Under Two step procedure, we report the Alpha estimates and differences for estimated ESG scores with respect to firm size, market-to-book ratio, profitability, year fixed effects, and sector fixed effects. Under ESG adjusted model, we report Alpha estimates and differences for a Fama and French (2015) five-factor model with an additional *ESG*-factor.

***, ** and * reflect significance at the 1%, 5%, and 10% level.

best-in-sector ESG scores and repeat the regression analysis. This procedure mitigates concerns that ESG scores might merely reflect firm characteristics rather than sustainability-related information.

Table 6 presents the results of this two-step procedure. Panel A shows the results of the industry-based ESG scores. We find statistically significant abnormal returns for the bottom portfolios, while the estimates for the top decile and quintile portfolios are not statistically significant. Nevertheless, the difference in abnormal returns (Δ -Alpha) between the bottom and top portfolios is significant, suggesting that best-in-industry ESG scores reflect financial informativeness after controlling for specific firm characteristics.

Panel B reports the results of the sector-based scores. In contrast to the industry-based case, all estimated abnormal returns in this setting are statistically significant across decile and quintile portfolios. The estimated differences in abnormal returns (Δ -Alpha) are also larger than those obtained under the best-in-industry approach. In line with previous findings, these results confirm the conclusion that

broader peer groups enhance the financial informativeness of ESG ratings.

ESG adjusted model

To further investigate the source of the return difference between bottom and top portfolios and to ensure that our results are driven by the portfolio sorting procedure, we include an additional ESG factor into the standard Fama and French (2015) five-factor model. Therefore, we construct the factor in order to isolate a sustainability premium and repeat the empirical analysis.⁷

Table 8 presents the results of the ESG adjusted model. Across all regression models, the abnormal returns (Alphas) for the industry-based and sector-based ESG score portfolios are no longer statistically significant. Hence, the former excess returns are largely absorbed by the additional ESG factor. This suggests that the informativeness in form of

⁷ For further explanation, see Appendix A.3.



abnormal returns can be traced back to the related ESG risk and our sorting methodology.

Robustness

We conduct further robustness checks to validate our previous findings. Table 9 provides a battery of additional tests for the recalculated industry-based (Panel A) and the sector-based ESG score (Panel B). First, we need to take into account that extraordinary events such as the aftermath of the Financial Crisis or the outbreak of COVID-19 have a significant impact on financial markets and could bias our results. Consequently, we analyse different sub-samples of

the original sample. For the period January 2013 to December 2022 (excluding the aftermath of the Financial Crisis), we find similar results. The sector-based score points higher Alpha differences than the industry-based score. This is in line with the results between February 2010 and February 2020 (exclusion of the COVID-19 observations). Moreover, we exclude both events, which results in an analysis period of January 2013 to February 2020. Decile and quintile bottom portfolios exhibit significant risk premiums, and the top portfolios show significant negative returns, while the sector-based ESG score points higher Alpha differences. These results indicate that if we neglect the effects of extraordinary events, our constructed sector-based ESG score leads to

Table 9 Robustness tests

	Deciles			Quintiles		
	Annualized alpha			Annualized alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Industry-based score						
<i>Alternative sub-periods</i>						
Excluding fin. crisis	6.71**	- 3.78**	10.49***	5.92**	- 2.40**	8.32***
Excluding COVID-19	5.30**	- 3.29***	8.59***	5.04***	- 2.34***	7.38***
Excluding both	6.15**	- 3.33***	9.48***	5.43**	- 2.42**	7.85***
<i>Alternative factor models</i>						
Three-factor alpha	4.54*	- 3.21***	7.75***	4.49**	- 1.66	6.15***
Six-factor alpha	6.76***	- 2.86**	9.62***	6.38***	- 1.55	7.93***
<i>Alternative rebalancing</i>						
Monthly rebalancing	6.25**	- 3.46***	9.71***	5.60***	- 2.13**	7.73***
<i>Pillar models</i>						
Environmental pillar	4.45**	- 2.64**	7.09***	4.99***	- 2.50**	7.49***
Social pillar	5.99***	- 2.25**	8.24***	4.21**	- 1.22	5.43***
Panel B: Sector-based score						
<i>Alternative sub-periods</i>						
Excluding fin. crisis	12.84***	- 2.89***	15.73***	8.80***	- 2.01*	10.81***
Excluding COVID-19	9.43***	- 3.76***	13.19***	7.09***	- 2.78***	9.87***
Excluding both	11.39***	- 3.67***	15.06***	8.19***	- 2.77***	10.96***
<i>Alternative factor models</i>						
Three-factor alpha	7.06**	- 2.39**	9.45***	4.94*	- 1.41	6.35**
Six-factor alpha	11.44***	- 2.13**	13.57***	8.38***	- 1.21	9.59***
<i>Alternative rebalancing</i>						
Monthly rebalancing	10.69***	- 2.93***	13.62***	7.76***	- 1.98*	9.74**
<i>Pillar models</i>						
Environmental pillar	5.68***	- 2.36**	8.04***	5.68***	- 2.25**	7.93***
Social pillar	7.38***	- 2.20**	9.58***	6.23**	- 1.31	7.54***

This table shows annualized Alpha estimates and differences of the robustness tests. All values are presented in percent. The portfolios are sorted by ESG scores. The analyses for the pillar (environmental and social) models are conducted in the same way as for the ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: *MKT*, *SMB*, *HML*, *RMW*, *CMA*. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks. Under Alternative sub-periods, we divide the sample into three sub-periods. Under Alternative factor models, we report the Alpha estimates and differences of the Fama and French (1993) three and six-factor regression models. Under Alternative rebalancing, we sort the portfolios on a monthly basis. Under Pillar models, the portfolios are sorted by the environmental and the social pillar score.

***, ** and *reflect significance at the 1%, 5%, and 10% level.



better differentiation and could therefore enhance the measurement of sustainability risk.

The next series of robustness tests includes a subset of different factor models. We re-estimate the Alphas with two alternative factor models to assess the robustness of our findings. Using the Fama and French (1993) three-factor and the Fama and French (2015) five-factor model with a momentum factor (Carhart 1997), we re-run our analysis. While the Alpha difference of the Fama and French (1993) three-factor model is 1.7% per year (7.75% and 9.45%) for decile portfolios, we find an Alpha difference of around 4% per year (9.62% and 13.57%) for the six-factor model. For both ESG scores, the returns of the top quintile portfolios remain negative, but are no longer statistically significant. In general, the results of the alternative factor models confirm the previous out-performance of the sector-based recalculated ESG scores.

The third robustness check assesses the reliability of the findings if we use an alternative rebalancing methodology for the portfolios. To capture the dynamic nature of sustainability and its impact on stock returns, we use a more granular approach. Hence, we rebalance the ESG score portfolios on a monthly basis and re-run the analysis. We find significantly higher Alpha differences for the recalculated sector-based ESG scores (13.62%) as for the recalculated industry-based ESG score (9.71%) in the decile portfolios. In addition, the Alpha difference slightly decreases for the quintile portfolios (7.73% and 9.74%). Regarding our main findings, the Alpha estimates of this robustness check are close to the previous regression results. This is because Refinitiv calculates its ESG scores on the basis of companies' annual reports. As the majority of companies in our study disclose sustainability information at fiscal year's end, the underlying data of the ESG scores is also updated on a yearly basis.

The fourth series of robustness tests the recalculated environmental and social pillars. To rule out the possibility that our results are mainly driven by a single ESG pillar, we analyse the effects at a more granular level. Therefore, we sort the portfolios according to the environmental and social pillar scores. Regarding the Alpha differences in the decile and quintile portfolios for the environmental pillar, the coefficients are significant and similar for both. The change in the calculation method has also an effect in this case. The differentiation, especially in the environmental portfolios, is as pronounced as in the previous results. In contrast, the recalculated social score estimates exhibit higher Alpha differences for the decile (9.58% and 8.24%) and quintile (7.54% and 5.43%) portfolios. This implies that

the recalculation procedure strongly enhances the explanatory power of the social component of the ESG rating, while slightly affecting the environmental pillar.

In summary, we find across all specifications similar results. In line with our main findings, the bottom portfolios, including companies with worse ESG scores, exhibit a risk premium, while financial market participants forgo returns for companies in the top portfolios. The recalculated sector-based ESG scores point to larger differences in abnormal returns, suggesting that those scores better reflect the inherent sustainability risks of a firm.

Conclusion

In this paper, we investigate the role of peer groups for the financial informativeness of ESG ratings. Due to the methodology, differences in ESG scores arise from both category scores and category weights, which reflect the sustainability performance of the industry and the firm, respectively. Therefore, we construct best-in-industry and best-in-sector ESG ratings and analyse the financial performance of U.S. stocks under each approach. We find that both ratings are financially informative, but best-in-sector scores yield a larger differentiation in returns between firms with strong and weak sustainability profiles. In particular, the return spread between bottom and top ESG portfolios is significantly larger under the best-in-sector methodology. This suggests that broader peer groups could more effectively capture sustainability-related financial risks.

These findings have important implications for retail and institutional investors who rely on ESG ratings for portfolio construction and risk assessments. Best-in-industry ratings may be less suitable than best-in-sector ratings for identifying sustainability risks and estimating the associated risk premium. Moreover, this methodological approach could potentially favour green-washing suspicions or problems for sustainable funds that align their investment strategies with those ratings. As a consequence, those findings highlight the limitations of relying solely on relative, best-in-class ESG scores. Instead, financial market participants should complement such ratings with their own absolute assessments of firms' environmental and social exposures. For asset managers, this adds another layer of complexity to the integration of ESG criteria into investment decisions.

Our analysis focuses on the Refinitiv ESG framework. A promising avenue for future research is to examine whether similar effects arise under alternative ESG methodologies, which may apply different peer-group definitions or



weighting schemes. Further research could also explore whether peer group effects are similarly relevant in global or emerging markets, or in sectors with particularly salient sustainability risks, such as those related to climate change.

Another important direction for future research concerns the governance pillar. Since the calculation and weighting methodology of the management, shareholders, and CSR strategy category differs from that of the environmental and social dimensions, these were kept constant in our study. However, the weighting scheme of the governance categories is fixed, and each industry and country is assigned the same weighting. Prior evidence suggests that the weighting of all categories, including those in the governance pillar, may increase the financial informativeness of ESG ratings (Muck and Schmidl 2024). One might therefore expect that a change in the weighting in the governance pillar could further influence our results. Hence, examining how alternative weighting schemes affect the aggregated ESG scores would clarify the extent to which methodological choices in the governance pillar influence the informational content of ESG assessments for investment decisions and capital allocation.

Appendix A

A.1: Numerical scores: firms with identical ratings and the average score

Assume that there are k firms with the same rating and that there are $\xi > 0$ firms with a worse score than that. The total number of firms in a peer group is N . It holds that:

$$\sum_{j=1}^k \frac{\xi + \frac{k}{2}}{N} = \frac{k(\xi + \frac{k}{2})}{N} = \frac{\sum_{j=0}^{k-1} \xi + j + \frac{1}{2}}{N}.$$

The right-hand side shows the corresponding sum for k firms with different numerical scores, where the first is better than ξ firms, the second better than $\xi + 1$ firms, the third better than $\xi + 2$ firms, and so on. Hence, the average is not affected by the number of firms with identical numerical scores.

A.2: ESG score estimate

For the estimation of the ESG scores based on firm characteristics, we employ the following cross-sectional regression model:

$$\widehat{ESG}_{i,t} = \alpha_{i,t} + \beta_1 SIZE_{i,t} + \beta_2 MTB_{i,t} + \beta_3 ROA_{i,t} + \beta_4 LEV_{i,t} + \gamma_t + \delta_j + \epsilon_{i,t}$$

where $SIZE_{i,t}$ is the natural logarithm of firm i 's book value of total assets, $MTB_{i,t}$ is the ratio of firm i 's market capitalization and book value of equity, $ROA_{i,t}$ is the ratio of firm i 's EBIT (earnings before interest and taxes) and the book value of total assets, $LEV_{i,t}$ is the ratio of firm i 's total debt and the book value of total assets, γ_t are year fixed effects, and δ_j are sector fixed effects.

A.3: Additional ESG factor

Following the Fama and French (1993) methodology, we construct an ESG factor. We sort firms into six portfolios based on their market capitalization and ESG scores. As breakpoints for market capitalization, we utilize the median. For the ESG score, we use the bottom 30%, the middle 40%, and the top 30%. We only apply four portfolios and omit the two middle portfolios: *SmallLowESG* stands for a portfolio with small market capitalization and low ESG scores, *BigLowESG* for big market capitalization and low ESG scores, *SmallHighESG* for small market capitalization and high ESG scores, and *BigHighESG* for big market capitalization and high ESG scores. As all Refinitiv ESG scores and pillars are updated on a yearly basis, we adjust the sorting of portfolios on a yearly basis. Finally, we receive the monthly return of our ESG factor (ESG_t):

$$ESG_t = 0.5(SmallLowESG_t + BigLowESG_t) - 0.5(SmallHighESG_t + BigHighESG_t) \quad (A.1)$$

According to Hübel and Scholz (2020), we interpret this sustainability factor as the return of a long-short investment strategy, that has long positions in low ESG firms and short positions in firms with high ESG scores. To measure the excess stock returns in terms of Alpha, we use the Fama and French (2015) five-factor model plus the constructed ESG_t factor, which captures the factors market (MKT_t), size (SMB_t), book-to-market (HML_t), profitability (RMW_t), investment (CMA_t), and sustainability (ESG_t):

$$R_t = \alpha + \beta_1 MKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \beta_6 ESG_t + \epsilon_t \quad (A.2)$$

where R_t is the return on the diversified decile or quintile portfolio in month t reduced by the risk-free return. Alpha (α) is the excess return, that is not explained by the other firm characteristic factors, and ϵ_t is the error term.



Table 10 TRBC sector and industry classification

Academic	Basic Materials
Miscellaneous Educational Service Providers	Chemicals
Professional & Business Education	Construction Materials
School, College & University & Business Education	Containers & Packaging
Consumer Cyclical	Metals & Mining
Automobiles & Auto Parts	Paper & Forest Products
Diversified Retail	Consumer Non-Cyclicals
Homebuilding & Construction Supplies	Beverages
Hotels & Entertainment Services	Consumer Goods Conglomerates
Household Goods	Food & Drug Retailing
Leisure Products	Food & Tobacco
Media & Publishing	Personal & Household Products & Services
Specialty Retailers	Financials
Textiles & Apparel	Banking Services
Energy	Collective Investments
Coal	Insurance
Oil & Gas	Investment Banking & Investment Services
Oil & Gas Rel.	Investment Holding Companies
Renewable Energy	Healthcare
Uranium	Biotechnology & Medical Research
Industrials	Healthcare Equipment & Supplies
Aerospace & Defense	Healthcare Providers & Services
Construction & Engineering	Pharmaceuticals
Diversified Industrial	Technology
Goods Wholesale	Communications & Networking
Freight & Logistics Serv.	Computers, Phones & Household Electronics
Machinery, Tools, Heavy	Electronic Equipment & Parts
Vehicles, Trains & Ships	Financial Technology (Fintech) & Infrastructure
Passenger Transportation Serv.	Integrated Hardware & Software
Professional & Commercial Serv.	Office Equipment
Transport Infrastructure	Semiconductors & Semiconductor Equipment
Real Estate	Software & IT Services
Real Estate Operations	Telecommunications Services
Residential & Commercial REITs	
Utilities	
Electrical Utilities & IPPs	
Multiline Utilities	
Natural Gas Utilities	
Water & Related Utilities	

This table shows sector and industry classifications according to Thomson Reuters Business Classification (TRBC) system. The sectors Government Activity and Institutions are not displayed, since they are equal to their subordinate industry groups.



Table 11 Descriptive statistics for industries (whole sample)

	Obs	Mean	SD	Min.	P25	Median	P75	Max.
<i>Academic & Educational Services</i>	174	31.08	18.91	6.19	14.99	23.41	48.20	76.79
Miscellaneous Educational Service Providers	104	30.17	20.35	6.19	14.46	20.69	48.95	76.79
Professional & Business Education	31	28.42	15.43	9.02	15.76	23.39	38.66	58.77
Schools, Colleges and Universities	39	35.63	16.95	7.94	19.12	33.52	51.63	62.38
<i>Basic Materials</i>	6,049	44.26	20.96	0.44	26.82	44.50	60.69	92.06
Chemicals	1,904	48.81	21.34	1.29	31.54	51.03	66.00	92.06
Construction Materials	533	45.69	20.20	7.99	29.59	46.47	60.69	87.04
Containers & Packaging	460	49.90	19.67	4.75	34.91	49.59	66.51	86.24
Metals & Mining	2,887	39.61	20.19	0.44	22.53	37.74	55.13	90.23
Paper & Forest Products	265	49.60	19.04	5.54	37.39	48.25	64.32	87.31
<i>Consumer Cyclical</i>	8,256	39.20	22.05	0.09	19.80	36.29	57.17	94.81
Automobiles & Auto Parts	1,438	51.22	23.07	1.00	32.73	53.44	70.46	94.81
Diversified Retail	601	35.07	20.43	1.48	18.68	30.88	50.03	85.14
Homebuilding & Construction Supplies	896	43.89	20.64	0.98	24.98	44.46	59.96	91.91
Hotels & Entertainment Services	1,487	35.96	20.17	1.37	18.52	33.46	51.40	83.85
Household Goods	348	40.69	23.92	1.53	19.05	38.01	62.72	89.58
Leisure Products	282	36.05	20.88	0.78	20.37	32.19	51.84	87.06
Media & Publishing	1,059	31.09	20.55	0.09	14.06	24.31	47.20	82.88
Specialty Retailers	1,436	33.60	19.29	0.81	18.19	28.61	46.62	86.02
Textiles & Apparel	709	43.17	21.52	1.51	22.71	43.23	60.47	87.54
<i>Consumer Non-Cyclical</i>	4,318	43.83	22.42	0.29	24.88	42.95	62.78	96.32
Beverages	684	46.20	22.44	0.29	27.94	45.79	65.66	92.84
Consumer Goods Conglomerates	472	44.69	21.76	4.27	26.16	44.61	60.92	92.38
Food & Drug Retailing	743	41.60	21.92	0.74	23.25	40.98	60.47	92.03
Food & Tobacco	1,895	41.62	21.44	0.30	23.45	40.55	58.88	90.70
Personal & Household Products & Services	524	51.13	25.12	7.23	26.02	53.47	74.29	96.32
<i>Energy</i>	3,400	42.13	20.46	1.09	24.55	40.42	58.54	94.42
Coal	324	35.04	17.35	2.12	22.07	33.37	47.05	81.99
Oil & Gas	1,822	46.47	21.39	1.09	27.27	47.36	63.80	94.42
Oil & Gas Related Equipment and Services	1,012	37.10	18.25	1.78	22.81	33.99	50.08	90.37
Renewable Energy	242	40.04	18.15	9.94	24.27	38.63	52.14	84.85
<i>Financials</i>	9,556	37.39	24.34	0.41	16.80	30.84	57.19	96.35
Banking Services	5,292	38.93	24.71	0.49	17.88	31.75	59.83	96.35
Collective Investments	404	14.18	8.91	0.41	8.07	12.16	17.55	49.07
Insurance	1,771	42.18	24.17	1.65	20.23	40.08	62.60	95.66
Investment Banking & Investment Services	1,884	34.20	22.43	0.52	14.80	30.20	51.87	92.21
Investment Holding Companies	205	31.34	24.32	0.45	11.66	23.47	49.15	92.04
<i>Healthcare</i>	5,312	30.38	23.67	0.15	11.31	20.40	47.54	95.76
Biotechnology & Medical Research	1,535	15.54	12.38	0.15	8.42	11.91	17.14	77.71
Healthcare Equipment & Supplies	1,333	35.01	24.04	0.83	13.92	28.64	54.67	89.36
Healthcare Providers & Services	697	30.42	19.70	0.19	14.74	24.85	44.60	86.38
Pharmaceuticals	1,747	39.86	25.97	0.27	15.52	36.37	61.07	95.76
<i>Industrials</i>	8,622	43.40	20.12	0.62	27.01	42.48	59.52	94.41
Aerospace & Defense	567	44.61	20.51	3.41	26.16	43.62	60.43	85.30
Construction & Engineering	1,169	43.80	20.05	0.62	27.53	41.57	60.49	89.50
Diversified Industrial Goods Wholesale	154	51.77	21.34	8.00	37.05	57.21	68.37	92.82
Freight & Logistics Services	910	40.42	18.01	1.29	26.18	40.11	54.86	83.03
Machinery, Tools, Heavy Vehicles, Trains & Ships	2,978	46.71	20.80	1.82	30.16	46.22	62.82	94.41
Passenger Transportation Services	724	44.40	20.04	2.34	28.73	46.75	60.80	85.09
Professional & Commercial Services	1,567	38.53	18.26	0.77	23.55	36.61	52.12	86.11
Transport Infrastructure	553	38.52	19.59	1.88	21.06	37.66	55.04	81.24
<i>Real Estate</i>	4,164	36.63	21.79	0.47	18.28	32.68	53.31	92.22
Real Estate Operations	1,998	37.17	22.21	0.47	17.51	35.14	53.71	92.22
Residential & Commercial REITs	2,166	36.12	21.39	0.52	18.54	30.42	53.00	92.14
<i>Technology</i>	7,656	40.02	23.73	0.54	18.65	36.20	60.17	94.38



Table 11 (continued)

	Obs	Mean	SD	Min.	P25	Median	P75	Max.
Communications & Networking	458	40.53	24.73	3.54	18.23	36.34	60.70	94.38
Computers, Phones & Household Electronics	656	51.64	26.11	0.59	27.55	55.69	74.96	93.75
Electronic Equipment & Parts	728	40.99	22.21	1.07	21.33	39.45	59.08	88.12
Financial Technology (Fintech) & Infrastructure	136	24.36	18.52	1.19	12.17	17.68	32.22	75.89
Integrated Hardware & Software	34	34.81	17.18	8.34	18.30	32.13	50.70	69.66
Office Equipment	137	62.12	23.04	1.00	48.82	68.12	81.52	90.79
Semiconductors & Semiconductor Equipment	1,191	45.23	21.09	4.66	27.65	44.31	62.07	92.42
Software & IT Services	2,931	30.58	20.42	0.54	14.33	24.56	43.37	90.72
Telecommunications Services	1,385	48.81	23.39	1.37	27.17	52.72	69.30	88.22
<i>Utilities</i>	2,471	45.54	20.05	2.49	29.95	45.24	62.08	89.41
Electric Utilities & IPPs	1,519	45.91	19.87	3.56	30.37	45.61	62.61	89.41
Multiline Utilities	392	52.04	20.68	4.19	38.38	55.31	68.52	86.14
Natural Gas Utilities	337	40.56	18.87	3.79	25.64	38.75	55.63	83.99
Water & Related Utilities	223	39.13	18.01	2.49	24.89	35.68	53.46	78.91

This table shows summary statistics of the recalculated sector-based ESG scores across TRBC (The Refinitiv Business Classification) economic sectors and industries. It reports the number of observations, the mean, the standard deviation, the minimum, the 25%-quartile, the median, the 75%-quartile, and the maximum of the sector-based ESG scores. The whole sample contains all recalculated firm-year observations.



Table 12 Descriptive statistics for industries (U.S. Sample)

	Obs	Mean	S.D.	Min.	P25	Median	P75	Max.
<i>Academic & Educational Services</i>	81	23.26	15.21	6.19	14.91	18.27	23.42	74.15
Miscellaneous Educational Service Providers	60	24.44	17.37	6.19	14.38	16.66	26.02	74.15
Professional & Business Education	10	19.28	3.32	14.59	16.72	18.54	22.13	24.62
Schools, Colleges & Universities	11	20.45	5.43	14.56	17.00	18.73	22.51	34.46
<i>Basic Material</i>	942	45.27	19.88	8.38	28.37	43.96	59.87	91.36
Chemicals	388	50.63	21.74	8.82	28.86	51.52	70.26	91.36
Construction Materials	52	30.55	14.43	11.02	15.76	31.55	41.20	53.48
Containers & Packaging	162	48.37	15.56	12.07	36.04	48.54	60.01	73.39
Metals & Mining	308	39.52	18.30	8.38	26.29	36.67	49.21	85.29
Paper & Forest Products	32	43.85	8.94	27.91	37.59	43.36	51.52	64.56
<i>Consumer Cyclicals</i>	2,711	32.55	19.68	0.09	17.58	26.04	45.71	93.40
Automobiles & Auto Parts	261	44.50	24.14	4.24	23.71	38.71	65.04	93.40
Diversified Retail	204	36.33	19.05	6.42	20.14	33.48	50.94	80.05
Homebuilding & Construction Supplies	328	32.59	17.18	7.01	19.99	26.93	40.32	91.91
Hotels & Entertainment Services	549	32.58	19.77	5.08	15.68	25.78	48.08	79.54
Household Goods	139	37.61	20.86	7.19	19.63	32.94	55.52	82.22
Leisure Products	155	34.92	20.95	4.53	18.64	29.06	47.54	87.06
Media & Publishing	328	21.06	14.81	0.09	10.90	17.10	24.61	69.23
Specialty Retailers	566	29.03	15.86	1.19	17.97	24.22	35.28	81.70
Textiles & Apparel	181	36.79	20.22	7.00	19.42	33.74	53.09	80.36
<i>Consumer Non-Cyclicals</i>	866	45.75	24.00	1.81	24.37	44.04	67.68	93.67
Beverages	113	48.47	24.89	4.49	28.51	56.38	68.49	87.31
Consumer Goods Conglomerates	60	59.55	25.76	9.18	45.68	64.92	82.04	92.38
Food & Drug Retailing	127	39.46	22.16	1.81	23.54	38.22	56.64	92.03
Food & Tobacco	382	45.53	22.60	6.69	24.35	46.66	65.41	89.21
Personal & Household Products & Services	184	44.38	25.16	8.33	22.79	36.23	71.83	93.67
<i>Energy</i>	1,037	37.36	19.96	4.02	21.36	32.52	51.10	90.37
Coal	42	34.38	19.96	4.97	13.31	35.99	52.79	63.96
Oil & Gas	510	41.90	20.76	4.02	24.23	40.49	56.87	87.55
Oil & Gas Related Equipment and Services	399	32.12	18.21	4.54	18.53	26.67	41.45	90.37
Renewable Energy	86	36.16	16.42	12.67	24.22	31.71	42.94	79.11
<i>Financials</i>	3,000	26.49	18.34	1.23	14.65	19.99	31.43	92.21
Banking Services	1,882	23.77	15.96	1.23	14.61	19.03	25.26	88.77
Collective Investments	17	8.73	7.36	1.93	2.54	7.53	8.04	26.57
Insurance	613	30.33	18.98	4.29	16.26	22.87	40.91	84.48
Investment Banking & Investment Services	468	33.51	23.28	2.37	14.15	25.07	49.69	92.21
Investment Holding Companies	20	15.08	7.18	7.77	10.92	11.89	19.25	32.40
<i>Healthcare</i>	2,729	22.50	20.62	0.15	9.68	14.08	25.53	88.32
Biotechnology & Medical Research	1,151	13.13	8.67	0.15	8.32	11.28	15.30	71.56
Healthcare Equipment & Supplies	753	29.31	22.74	1.71	12.33	19.87	42.27	87.54



Table 12 (continued)

	Obs	Mean	S.D.	Min.	P25	Median	P75	Max.
Healthcare Providers & Services	324	29.32	21.02	0.19	13.31	21.12	43.71	86.38
Pharmaceuticals	501	29.39	27.11	0.27	10.08	15.43	48.02	88.32
<i>Industrials</i>	2,278	37.23	18.03	0.62	23.27	33.31	50.72	86.23
Aerospace & Defense	242	39.59	19.63	10.67	25.17	31.40	55.87	83.87
Construction & Engineering	168	30.82	16.58	0.62	18.44	26.16	38.48	72.95
Diversified Industrial Goods Wholesale	6	19.80	3.56	15.08	17.98	19.65	20.62	25.81
Freight & Logistics Services	241	39.07	18.64	9.31	23.66	33.57	56.96	76.71
Machinery, Tools, Heavy Vehicles, Trains & Ships	864	38.41	17.22	3.86	25.20	35.62	51.68	86.23
Passenger Transportation Services	105	47.84	18.98	10.32	32.19	52.65	64.28	78.52
Professional & Commercial Services	652	34.23	17.29	0.77	21.16	31.07	45.04	86.11
<i>Real Estate</i>	1,331	33.75	20.75	0.52	17.47	26.43	48.37	92.22
Real Estate Operations	145	31.47	24.25	7.99	13.49	19.72	44.35	92.22
Residential & Commercial REITs	1,186	34.03	20.28	0.52	18.14	26.78	49.56	92.14
<i>Technology</i>	2,530	31.70	21.18	0.54	15.08	25.14	45.78	92.42
Communications & Networking	213	37.05	23.78	5.76	17.31	30.11	54.41	90.58
Computers, Phones & Household Electronics	157	39.57	23.06	0.59	20.93	32.82	55.19	87.35
Electronic Equipment & Parts	198	30.85	15.86	7.16	18.18	26.86	40.02	76.26
Financial Technology (Fintech) & Infrastructure	60	26.35	20.14	2.97	12.84	18.24	43.60	70.38
Integrated Hardware & Software	16	24.21	15.14	8.34	15.23	17.51	28.78	58.16
Office Equipment	41	60.54	21.68	7.93	51.65	62.15	79.56	86.15
Semiconductors & Semiconductor Equipment	473	40.41	20.41	6.55	25.02	35.91	55.70	92.42
Software & IT Services	1,173	26.25	19.04	0.54	12.96	19.03	34.13	90.72
Telecommunications Services	199	28.26	21.16	1.37	12.88	21.17	35.89	82.01
<i>Utilities</i>	573	50.10	19.05	2.49	34.82	51.76	66.16	88.13
Electric Utilities & IPPs	308	55.72	17.80	11.01	42.15	59.86	68.83	88.13
Multiline Utilities	118	55.17	13.33	22.17	45.93	55.31	64.39	86.14
Natural Gas Utilities	80	33.09	12.49	13.32	23.58	30.61	40.44	70.29
Water & Related Utilities	67	35.62	20.34	2.49	18.96	31.94	50.80	78.03

This table shows summary statistics of the recalculated sector-based ESG scores across TRBC (The Refinitiv Business Classification) economic sectors and industries. It reports the number of observations, the mean, the standard deviation, the minimum, the 25%-quartile, the median, the 75%-quartile, and the maximum of the sector-based ESG scores. The U.S. sample contains firms having their headquarters in the U.S.



Table 13 Regression results for the original Refinitiv score

	Deciles			Quintiles		
	Annualized alpha			Annualized alpha		
	Bottom	Top	Difference	Bottom	Top	Difference
Panel A: Main results						
Main results	8.01***	- 2.81***	10.82***	5.61***	- 1.82*	7.43***
Panel B: Additional tests						
Alternative weighting	6.40***	- 2.44***	8.84***	6.03***	1.93**	7.96***
Endogeneity concern	10.32***	- 1.34	11.46***	7.20***	- 1.33	8.53***
Noise concern	0.78	- 1.78*	2.56	- 0.65	- 0.85	0.20
Panel C: Robustness tests						
<i>Alternative sub-periods</i>						
Excluding fin. crisis	9.01***	- 2.76**	11.77***	5.78**	- 1.74	7.52***
Excluding COVID-19	6.44***	- 3.44***	9.88***	5.11***	- 2.77***	7.88***
Excluding both	7.49***	- 3.33***	10.82***	5.47**	- 2.67***	8.14***
<i>Alternative factor models</i>						
Three-factor alpha	6.34**	- 2.11**	8.45***	4.06**	- 1.16	5.22**
Six-factor alpha	8.77***	- 2.34**	11.11***	6.59***	- 1.25	7.84***
<i>Alternative rebalancing</i>						
Monthly rebalancing	7.79***	- 2.77***	10.56***	5.55***	- 1.82*	7.37***
<i>Pillar Models</i>						
Environmental pillar	6.55***	- 2.23**	8.78***	6.00***	- 1.93**	7.93***
Social pillar	4.94**	- 1.69*	6.63***	4.07**	- 0.60	4.67***

This table shows annualized Alpha estimates and differences of the original Refinitiv ESG scores. All values are presented in percent. The portfolios are sorted by ESG scores. Companies with the worst scores are in the bottom portfolio, and companies characterized by the best scores are assigned to the top portfolio, respectively. Those portfolios are rebalanced on a yearly basis. The dependent variable is the equal-weighted monthly return of the ESG portfolios minus the risk-free rate. Independent variables are the following Fama and French (2015) factors: *MKT*, *SMB*, *HML*, *RMW*, *CMA*. The results are based on the monthly average returns between February 2010 and December 2022. The difference in annualized Alphas reflects the returns of a long-short portfolio, with long positions in stocks with low ESG ratings and short positions in high ESG stocks.

***, ** and * reflect significance at the 1%, 5%, and 10% level.

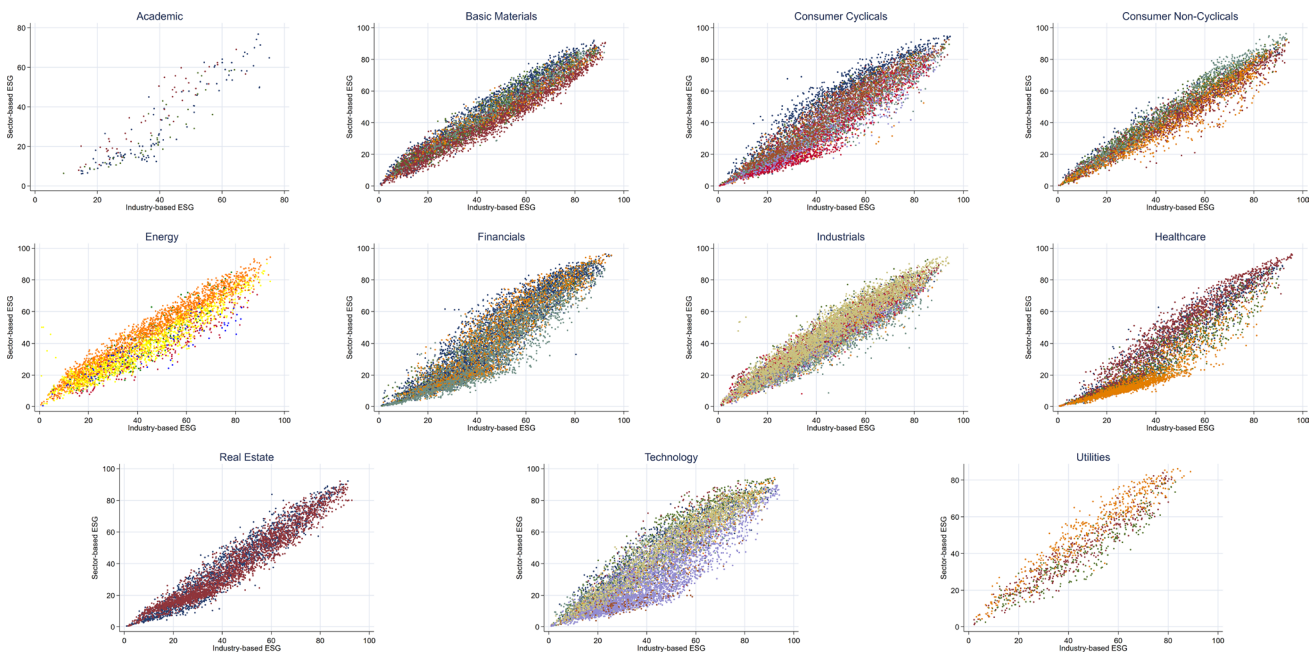


Fig. 1 Scatter plot ESG scores. This figure shows the distribution of the recalculated and original industry-based score for the sectors academic, basic materials, consumer cyclical, consumer non-cyclical,

financials, industrials, healthcare, utilities. The sample consists of data points between the years 2010 and 2022



Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Generative AI and AI-assisted technologies in the writing process During the preparation of this work the authors used ChatGPT and DeepL in order to copy edit the original text. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albuquerque, R., Y. Koskinen, S. Yang, and C. Zhang. 2020. Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. *The Review of Corporate Finance Studies* 9: 593–621.
- Albuquerque, R., Y. Koskinen, and C. Zhang. 2019. Corporate social responsibility and firm risk: Theory and empirical evidence. *Management Science* 65: 4451–4469.
- Berg, F., J.F. Kölbel, and R. Rigobon. 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance* 26: 1315–1344.
- Billio, M., M. Costola, I. Hristova, C. Latino, and L. Pelizzon. 2021. Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management* 28: 1426–1445.
- Bolton, P., and M. Kacperczyk. 2021. Do investors care about carbon risk? *Journal of Financial Economics* 142: 517–549.
- Broadstock, D.C., K. Chan, L.T.W. Cheng, and X. Wang. 2021. The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. *Finance Research Letters* 38: 101716.
- Carhart, M.M. 1997. On persistence in mutual fund performance. *The Journal of Finance* 52: 57–82.
- Chatterji, A.K., R. Durand, D.I. Levine, and S. Touboul. 2016. Do ratings of firms converge? Implications for managers, investors and strategy researchers. *Strategic Management Journal* 37: 1597–1614.
- Christensen, D.M., G. Serafeim, and A. Sikochi. 2022. Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review* 97: 147–175.
- Dobrick, J., C. Klein, and B. Zwergel. 2023. Size bias in Refinitiv ESG data. *Finance Research Letters* 55: 104014.
- Douglas, E., T. van Holt, and T. Whelan. 2017. Responsible investing: Guide to ESG data providers and relevant trends. *Journal of Environmental Investing* 8: 92–114.
- Drempetic, S., C. Klein, and B. Zwergel. 2020. The influence of firm size on the ESG score: Corporate sustainability ratings under review. *Journal of Business Ethics* 167: 333–360.
- Dumrose, M., S. Rink, and J. Eckert. 2022. Disaggregating confusion? The EU taxonomy and its relation to ESG rating. *Finance Research Letters* 48: 102928.
- Eccles, R.G., M.P. Krzus, J. Rogers, and G. Serafeim. 2012. The need for sector-specific materiality and sustainability reporting standards. *Journal of Applied Corporate Finance* 24: 65–71.
- Edmans, A. 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. *Journal of Financial Economics* 101: 621–640.
- El Ghoul, S., O. Guedhami, C. C. Y. Kwok, and D. R. Mishra. 2011. Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance* 35: 2388–2406.
- ESMA. 2021. *ESMA letter to EC on ESG ratings*. Paris: ESMA.
- Fama, E.F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25: 383–417.
- Fama, E.F., and K.R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.
- Fama, E.F., and K.R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22.
- Friede, G., T. Busch, and A. Bassen. 2015. ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment* 5:210–233.
- GSIA. 2022. Global Sustainable Investment Review 2022. <https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf>
- 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: 2789–2837.
- Heeb, F., J.F. Kölbel, F. Paetzold, and S. Zeisberger. 2023. Do investors care about impact? *The Review of Financial Studies* 36: 1737–1787.
- Hong, H., and M. Kacperczyk. 2009. The price of sin: The effects of social norms on markets. *Journal of Financial Economics* 93 (1): 15–36.
- Hübel, B., and H. Scholz. 2020. Integrating sustainability risks in asset management: The role of ESG exposures and ESG ratings. *Journal of Asset Management* 21: 52–69.
- Khan, M., G. Serafeim, and A. Yoon. 2016. Corporate sustainability: First evidence on materiality. *The Accounting Review* 91: 1697–1724.
- Lins, K.V., H. Servaes, and A. Tamayo. 2017. Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance* 72: 1785–1824.
- Muck, M., and T. Schmidl. 2024. Comparing ESG score weighting approaches and stock performance differentiation. *Finance Research Letters* 67 : 105924.
- Refinitiv. 2023. *Environmental, social and governance scores from LSEG*. London: LSEG.
- Whelan, T., U. Atz, T. van Holt, and C. Clark. 2021. *ESG and Financial Performance: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020*. New York: NYU Stern Center for Sustainable Business and Rockefeller Asset Management.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

