Muck, Matthias; Schmidl, Thomas

Comparing ESG score weighting approaches and stock performance differentiation

Date of secondary publication: 17.03.2025

Version of Record (Published Version), Article

Persistent identifier: urn:nbn:de:bvb:473-irb-1070452

Primary publication

Muck, Matthias; Schmidl, Thomas (2024): Comparing ESG score weighting approaches and stock performance differentiation, in: Finance Research Letters, New York: Elsevier Science, Vol. 67, Part B, Nr. 105924, pp. 1–9, doi: 10.1016/j.frl.2024.105924.

Finance Research Letters 67 2024) 105924

Contents lists available at ScienceDirect

Finance Research Letters

journal homepage: www.elsevier.com/locate/frl

Comparing ESG score weighting approaches and stock performance differentiation

Matthias Muck, Thomas Schmidl *

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

ARTICLE INFO

JEL classification:
M14
G24

Keywords:
ESG ratings
Category weightings

ABSTRACT

This study examines how weighting methodologies in ESG ratings for sustainability categories align with their financial relevance. We analyse Refinitiv's data-driven weights against equal weights and SASB's expert-based weights. Our findings show all methods differentiate firms by returns, but equal and SASB weights lead to a stronger effect, particularly after the Paris Agreement. This suggests alternative weighting approaches might better capture the financial significance of ESG categories.

1. Introduction

Sustainability information on environmental (E), social (S), and governance (G) factors has become increasingly important for investors in recent years. As such, firms are increasingly mandated to provide ESG reporting, satisfying the growing demand from investors (Ilhan et al., 2023). ESG ratings aggregate this sustainability information into a single number, but the market for these ratings remains unregulated (ESMA, 2021). This lack of regulation can lead to disagreement between rating agencies (e.g. Chatterji et al., 2016; Berg et al., 2022). ESG ratings combine information from various categories, such as CO2 emissions, human rights policies, or corporate social responsibility (CSR) strategies, with an assessment of their relevance to investors. This materiality concept ensures the information used significantly impacts a company's financial return. Material information increases the price informativeness of the rating, while immaterial information has no financial impact (Grewal et al., 2021). The relevance of each category is reflected in its weighting, which signifies its relative importance for the final ESG score. Typically, category weighting is industry-specific. For example, a category highly relevant in the energy sector (e.g., carbon emissions) may be less important in the financial services sector.

In this paper, we explore to what extent ESG weighting systems can be used to distinguish between different types of investments and assess their suitability for investors. Specifically, we focus on Refinitiv's ESG score[1] and weighting methodology, which is referenced in over 1,200 academic articles by the end of 2020 (Berg et al., 2020). According to an OECD report about ESG investing practices, Refinitiv is one of the three major rating providers (Boffo and Patalano, 2020) and is one of the considered ratings in the white paper "Seeking Return on ESG" as an intermediary for sustainable information during the World Economic Forum (WEF, 2019). Standard methods for assessing the materiality of ESG categories involve analysing economic channels within an industry, which can be prone to errors. Refinitiv's approach aims to circumvent this issue through a "data-driven methodology". Refinitiv assigns weights to categories, based on quantifiable data, such as a company's CO2 emissions (Refinitiv, 2023). While this method simplifies the calculation and promotes transparency, it does not analyse the direct impact of these categories (e.g., CO2 emissions)

* Corresponding author.
E-mail addresses: matthias.muck@uni-bamberg.de (M. Muck), thomas.schmidl@uni-bamberg.de (T. Schmidl).

[1] For consistency within this paper, we will refer to Refinitiv, even though they became LSEG (London Stock Exchange Group) data and analytics in 2023.

https://doi.org/10.1016/j.frl.2024.105924
Received 15 March 2024; Received in revised form 20 June 2024; Accepted 31 July 2024
Available online 3 August 2024

on a company's performance as reflected in the profit and loss statements or cash flows. We raise the question whether such a weighting system aligns with the financial relevance of categories.

To assess the relationship between financial performance and ESG weighting, we compare Refinitiv's ESG score to two alternative indices: an informed rating and a uniformed rating. All ESG ratings use the same category scores but different weightings. The informed benchmark rating enriches the Refinitiv rating by incorporating the Sustainability Accounting Standards Board's (SASB) materiality assessment for different industries. This materiality assessment is based on expert group judgement and aims to identify ESG material sustainability categories. The uniformed benchmark rating, in contrast, assigns equal weight to all ESG categories, offering a baseline comparison. Following Khan et al. (2016), we sort the stocks into decile and quintile portfolios based on their ESG rating. We then estimate the Alpha (the abnormal return) using the Fama and French (2015) five-factor model. We measure the performance difference between stocks with the highest and lowest ESG ratings as the differences in abnormal returns (Δ Alpha).

Our findings contribute to previous research in two important ways. First, we demonstrate that the weighting of ESG ratings could enhance the differentiation of firms based on stock returns. Stocks with the highest sustainability ratings exhibit lower returns compared to those with lower ratings. This effect is more pronounced for weights that incorporate industry-specific materiality through SASB information and for the naive equal-weighted approach. While Refinitiv's weights also differentiate between financially under- and over-performing stocks, the performance differences are smaller.

Second, to ensure the robustness of our results, we divided the sample into pre- and post-Paris Agreement (2015) sub-periods. Interestingly, significant return differences are primarily observed in the post-Agreement period. This suggests that the Agreement potentially leads investors to re-evaluate the importance of sustainability. Furthermore, when comparing the ESG indices, we find that the performance differences were larger for the naive and SASB ratings compared to Refinitiv's weighting.

In summary, our findings suggest that both the uninformed, equally weighted score and the SASB-based score may be more effective in differentiating stocks based on their future financial performance compared to the existing Refinitiv score. This enhanced effectiveness is particularly evident after the Paris Agreement, likely driven by an increased recognition of ESG factors by investors.

The remainder of this study is structured as follows: Section 2 addresses the related literature and details the research question. Section 3 describes the data and the methodology used. Section 4 provides the results. Section 5 checks for robustness. Section 6 concludes.

2. Related literature and research question

Sustainability ratings have become increasingly important for asset managers. They factor into best-in-class investment decisions (Drempetic et al., 2020) and influence how mutual funds align their strategies with ESG ratings (Hartzmark and Sussman, 2019). Investors' willingness to pay a premium for sustainable investments further strengthens this trend (Heeb et al., 2023). Additionally, successful investor engagement in sustainability can lead to increased institutional ownership (Dimson et al., 2015). Recent findings highlight that superior sustainable firms have lower systematic risk (Albuquerque et al., 2019), which translates to a lower cost of capital[2] (El Ghoul et al., 2011; Apergis et al., 2022; Wong et al., 2021) due to their perceived resiliency during economic downturns. This makes them function as a financial hedge for investors (Albuquerque et al., 2020; Broadstock et al., 2021). Consequently, a link between sustainability information and financial performance is evident (e.g., Whelan et al., 2021; Friede et al., 2015). Interestingly, studies by Drempetic et al. (2020) and Dobrick et al. (2023) identify a potential size bias in Refinitiv ESG scores, where larger companies tend to have better ratings.

The relationships described above highlight the importance of ESG ratings. These ratings make sustainability measurable by combining various social (e.g. Hong and Kacperczyk, 2009; Edmans, 2011) and environmental proxies (e.g. Bolton and Kacperczyk, 2021). In addition to traditional credit rating agencies, several financial data providers offer their own sustainability assessments. Refinitiv, MSCI, and Bloomberg are listed as the three major rating agencies, according to the OECD.[3] Alongside these, specialized providers exist, such as the Carbon Disclosure Project (CDP) or the Institutional Shareholder Services (ISS) (Douglas et al., 2017).

ESG ratings combine category scores and weights. Category scores are firm-specific characteristics that measure various aspects of sustainability. Weights measure the importance of the category and are industry-specific.[4] This acknowledges that a category important in one industry may be less relevant in another. Specifically, consider firm i from industry j. Its ESG score is given by

$$ESG_i = \sum_{c=1}^{N} C_{c,i} \cdot w_{c,j}, \tag{2.1}$$

where $C_{c,i}$ is the firm's score for category $c \in \{1, \dots, N\}$ and $w_{c,j}$ is the weight for category c in industry j.

ESG ratings differ with regard to categories and weighting methods, despite similar data collection. For example, Refinitiv includes the most indicators, while MSCI only contains 37 issues in their calculation (Billio et al., 2021). Notably, MSCI and Refinitiv are similar concerning their methodology. Both are industry-based measures with weights reflecting the relevance of different themes (Eccles and Stroehle, 2018; Refinitiv, 2023).

The literature on ESG rating construction mainly addresses the choice of categories. For example, Chatterji et al. (2016) and Berg et al. (2022) analyse rating disagreements between providers and the relative importance of different ESG categories for ratings.

[2] However, the probability of default is not normally distributed but depends on the respective industry. See, for example, Palmieri et al. (2023).

[3] See the report "ESG Investing: Practices, Progress and Challenges" (Boffo and Patalano, 2020).

[4] This aligns with, for example, Eccles et al. (2012) who state that sustainability risks vary significantly across industries.

These differences in ratings are considered as financial risk, and Gibson et al. (2021) find that financial market participants price that divergence with a risk premium. To reduce such a divergence, regulatory frameworks like the EU Taxonomy could be used. For example, it defines a set of environmental properties, and Dumrose et al. (2022) analyse the consistency of rating categories with the EU Taxonomy for several providers. Their findings show that ESG measures from three out of four agencies are significantly related to the EU Taxonomy.

While significant research focuses on selecting relevant categories for ESG ratings, less attention has been paid to how those categories are weighted. The literature suggests that the materiality of categories drives their financial importance for investors, but this materiality may vary across industries (Eccles et al., 2012). Therefore, the Sustainability Accounting Standards Board (SASB) also defines the materiality of categories with respect to the industry. Using categories from the MSCI sustainability ratings, Khan et al. (2016) find that mainly material sustainability categories matter to discriminate stocks according to their performance. Moreover, (Schiehll and Kolahgar, 2021) and (Consolandi et al., 2022) consider financial disclosure and show that material information is most relevant for stock returns.

In this paper, we investigate whether the weighting of categories in Refinitiv's ESG ratings aligns with their financial importance for investors. Refinitiv is considered one of the most relevant ESG ratings and employs a data-driven approach similar to MSCI ratings. We explore whether the weights assigned to categories in Refinitiv's ESG ratings better reflect their financial importance compared to a simple, equal weighting scheme (detailed in the next section) and further benchmark these weights against those obtained from SASB's materiality matrix, a comprehensive framework that defines industry-specific materiality of ESG categories based on expert assessments.

3. Data and methodology

Our sample consists of 3,098 firms having their headquarters in the U.S. The examination period starts in the year 2010 and ends in 2022, resulting in 18,352 firm-year observations. ESG and price data are taken from the Refinitiv database. We sort the stocks into equally weighted decile and quintile portfolios based on their ESG score, as detailed below. Companies with the lowest (highest) scores are assigned to the bottom (top) portfolios. In line with Refinitiv's updating of sustainability measures, we rebalance portfolios annually at the beginning of the fiscal year.

We consider three different sets of weights, including "Refinitiv weights", "SASB weights", and "naive weights" leading to different ESG scores. Refinitiv weights reflect Refinitiv's assessment of a category for an industry. They apply a data-driven approach: Each industry group and ESG category is assigned a magnitude weight. The proxy for the magnitude weight can be quantitative (measured with units, e.g., tons of carbon emissions) or binary (yes/no answer, e.g., availability of a firm's human rights policy). Industry medians are determined for quantitative categories. For binary categories, the share of companies with a positive answer within an industry is determined. Industry results are then ranked and mapped on scales from 1 (low exposure) to 10 (high exposure).[5] Category weights are calculated as

$$w_{c,j} = \frac{MagnitudeWeight_{c,j}}{\sum_{c=1}^{N} MagnitudeWeight_{c,j}}. \tag{3.1}$$

Governance categories use a different approach, which is not elaborated on here.[6]

The SASB introduced investor-focused guidance for sustainability disclosure in February 2014. This guidance identifies specific sustainability risks with different financial impacts in industries. SASB distinguishes between material and immaterial sustainability risks for each industry based on expert group judgement (Khan et al., 2016). We compute SASB weights to address the economic channels of Refinitiv's categories by enriching them with data from the SASB's materiality matrix. For coherence, SASB categories and industries are mapped to Refinitiv's categories and industries (see Table A2 in the Online Appendix). Magnitude weights are then assigned as 10 for material categories and 5 for immaterial categories, reflecting the best possible score and Refinitiv's default value. This weighting scheme essentially doubles the influence of material categories compared to immaterial ones. Weights are calculated using the same formula as Refinitiv's weights (Eq. (2.1)). By using SASB weights, we make a strong assumption: the identification of material and immaterial categories was independent of stock returns in our sample. This means past stock performance did not influence the classification of sustainability factors. We address this potential issue in greater detail in the robustness Section 5. Finally, naive weights[7] serve as a benchmark, assuming no information about the importance of categories is available. In this case, we assume that all weights are equal, resulting in an equally weighted sum of the magnitude weights.

We explore the financial informativeness of ESG scores by comparing the excess returns of the top and bottom portfolios.[8] We follow the most common approach in the literature and compute excess stock returns Alpha according to the Fama and French (2015) five-factor model, which captures the factors market (MKT), size (SMB), book-to-market (HML), profitability (RMW),

[5] See Online Appendix for details.

[6] The weightings of the governance categories deviate from the number of underlying indicators. In other words, the more data points are included in the calculation, the higher the weighting. For a more detailed explanation, see Table A1 (Online Appendix). For the next steps, we will use the same weighting approaches as for the other categories.

[7] Equal weighting could be used, if the evidence about different themes and their contributions are not clear. Nearly 45 percent of sustainability indices, such as the Human Development Index, use an equal-weight approach (Gan et al., 2017).

[8] See for similar methodical approaches, for example, Hübel and Scholz (2020), Avramov et al. (2022), and Pedersen et al. (2021).

Table 1
Descriptive statistics: ESG scores and ESG pillars.

	Obs.	Mean	Std. Dev.	Minimum	P25	Median	P75	Maximum
Original score								
ESG	18,352	40.90	19.34	0.63	25.76	37.76	54.71	95.16
ENV	18,352	24.08	27.12	0.00	0.00	13.96	43.02	98.55
SOC	18,352	41.56	20.73	0.44	25.76	38.03	55.55	98.01
GOV	18,352	48.10	22.35	0.16	30.15	48.33	66.00	99.46
Naive unweighted score								
ESG^{Equal}	18,352	35.25	19.59	0.34	20.36	28.85	47.78	95.51
ENV^{Equal}	18,352	22.34	26.23	0.00	0.00	10.61	40.51	98.48
SOC^{Equal}	18,352	39.99	20.13	0.45	25.00	35.91	52.18	98.07
GOV^{Equal}	18,352	41.84	20.40	0.08	26.55	39.88	55.85	99.48
SASB risk-weighted score								
ESG^{SASB}	18,352	36.20	20.02	0.50	21.00	29.93	49.04	96.39
ENV^{SASB}	18,352	23.20	27.25	0.00	0.00	10.65	42.96	98.69
SOC^{SASB}	18,352	40.99	20.27	0.41	25.83	37.26	53.76	98.31
GOV^{SASB}	18,352	41.59	20.50	0.08	26.11	39.49	55.52	99.48

Notes: This table shows summary statistics of the original ESG score (ESG), its environmental pillar (ENV), social pillar (SOC), governance pillar (GOV). The naive-unweighted scores are ESG^{Equal}, ENV^{Equal}, SOC^{Equal}, GOV^{Equal}. The SASB risk-weighted scores are ESG^{SASB}, ENV^{SASB}, SOC^{SASB}, GOV^{SASB}. The mean, standard deviation, minimum, 25% quantile, median, 75% quantile, and maximum are reported. The sample includes data points between February 2010 and December 2022.

and investment (CMA). As a market benchmark, we follow the standard procedure. This benchmark is computed from all stocks listed on the NYSE, AMEX, and NASDAQ and represents approximately the whole U.S. market capitalization.[9]

Note that this procedure implicitly assumes the efficient market hypothesis in the semi-strong form, that is, all publicly available ESG information is included in prices (Fama, 1970). The advantage of our standard approach is that results are comparable to the existing literature on ESG ratings (in particular Khan et al. (2016)). An alternative performance measure would be to use forward-looking information like analyst-based implied cost of capital (ICC) instead of Alpha. This has been done, for example, by Pástor et al. (2022). However, this procedure implicitly assumes that analysts' financial forecasts are accurate and unbiased.[10]

4. Results

Table 1 presents descriptive statistics of ESG scores. The naive-unweighted and SASB scores have similar means and standard deviations. While the Refinitiv score also matches the other scores in terms of standard deviations, it has substantially higher means. These observations also apply to the environmental, social, and governance pillars considered. However, the discrimination of firms is indicated by the dispersion of scores. Across all weighting methods considered, the environmental categories have the highest dispersion in terms of standard deviations of scores. Specifically, the standard deviations of the environmental scores are around 27 compared to standard deviations around 20 for the social and the governance scores.

Table 2 details the distributions of ESG scores across industries. Consistent with the overall data, naive-unweighted and SASB scores again show similar means across industries. Refinitiv scores, however, exhibit higher means compared to the others. This difference is most pronounced in the financial, real estate, and academic & educational services industries. Conversely, utilities and consumer non-cyclicals are less affected. Importantly, the standard deviations of all scores remain similar across industries.

Table 3 shows the regression results of a Fama and French (2015) five-factor model applied to top and bottom portfolios formed by ESG scores calculated with Refinitiv weights (Panel A), SASB weights (Panel B), and naive weights (Panel C). Concerning Fama and French (2015) factors, the estimated models show no major differences across the ESG metrics. Moreover, for all scores, the top portfolio has significantly negative Alphas. This suggests that sustainability is negatively associated with stock returns, which aligns with existing literature.

The performance difference (Δ Alpha) is the key metric of our analysis. It is printed in bold and measures the difference in unexplained returns (Alpha) between top and bottom portfolios. These differences are statistically significant. Decile portfolios, compared to quintile portfolios, exhibit higher levels of significance. This suggests that all ESG scores can differentiate firms based on their financial performance, with decile portfolios providing a clearer distinction. The largest difference is observed for SASB weights, followed by naive weights. However, these two metrics show very similar abilities to distinguish. In contrast, Refinitiv weights lead to substantially smaller performance differences (Δ Alpha) with lower levels of statistical significance, especially for quintile portfolios. In summary, our results imply that while all metrics use the same information, SASB weights seem to marginally improve the identification of financially relevant categories compared to naive weights. Refinitiv weights, however, appear less effective in this regard. These findings are relevant for investment approaches that select their stocks based on Refinitiv ESG ratings. Their data-driven approach is not as effective in measuring financial risk as an uninformed equal-weighted rating. As a consequence, investors do not receive the expected risk-return relation, which is reflected by the rating. However, they could address this shortcoming by integrating, for example, a SASB weighting to enhance the informativeness (Khan et al., 2016) of Refinitiv ESG ratings.

9 The market benchmark and the other factors are obtained through Kenneth R. French's web-page: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

10 For a discussion, see, for example, Gebhardt et al. (2001), Easton and Monahan (2005), and Guay et al. (2011).

Table 2
Descriptive Statistics: ESG scores by industry.

	Obs.	ESG	SD	ESG^{Equal}	SD^{Equal}	Diff.	ESG^{SASB}	SD^{SASB}	Diff.
Academic & Educational Services	77	34.52	15.25	27.29	13.23	**7.22**	27.63	12.84	**6.89**
Basic Materials	975	45.39	20.48	43.29	19.10	**2.10**	43.08	19.95	**2.31**
Consumer Cyclicals	2,615	41.37	18.91	35.89	18.95	**5.48**	36.05	19.82	**5.32**
Consumer Non-Cyclicals	901	48.07	24.48	46.25	24.00	1.82	45.32	25.25	**2.75**
Energy	1,050	39.04	20.74	36.51	19.70	**2.53**	38.69	20.04	0.35
Financials	2,983	38.59	15.90	29.50	16.11	**9.08**	31.90	16.37	**6.68**
Healthcare	2,966	35.81	18.29	28.66	17.53	**7.15**	30.69	18.04	**5.12**
Industrials	2,265	41.48	18.81	37.35	18.83	**4.12**	37.24	19.77	**4.24**
Real Estate	1,345	42.73	20.11	35.09	18.43	**7.64**	34.57	19.10	**8.16**
Technology	2,589	42.05	19.51	36.85	20.83	**5.20**	38.30	21.25	**3.74**
Utilities	586	50.39	18.76	48.65	18.95	1.74	48.32	20.45	**2.07**
Total	18,352	35.25	19.34	40.90	19.59	**5.65**	36.20	20.02	**4.69**

Notes: This table presents the differences between the original Refinitiv, the naive-unweighted, and the SASB risk-weighted ESG score across industries. The number of observations, the mean, the standard deviation of all scores, and the average differences are reported. Bold coefficients are significant at the 10% level. The sample includes data points between February 2010 and December 2022.

Table 3
Regression results: performance differences.

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Refinitiv ESG score ($N = 155$)								
MKT	0.9701***	22.8602	1.0101***	53.8631	1.0212***	27.0196	1.0228***	52.0335
SMB	0.7699***	9.3446	0.0959***	2.6344	0.7177***	9.7803	0.1817***	4.7623
HML	0.0091	0.1238	0.2053***	6.3134	0.0430	0.6558	0.2049***	6.0099
RMW	−0.4079***	−4.1548	0.1258***	2.8999	−0.3521***	−4.0266	0.1402***	3.0838
CMA	−0.0308	−0.2717	0.0812	1.6203	−0.1276	−1.2640	0.0630	1.2008
$Intercept$	0.1546	0.8552	−0.2278***	−2.8514	0.0958	0.5950	−0.1616*	−1.9295
R^2	0.8757		0.9626		0.9029		0.9615	
Annualized Alpha	1.87		−2.70***		1.16		−1.92*	
Δ Alpha				**4.57****				**3.08***
Panel B: SASB risk-weighted ESG score ($N = 155$)								
MKT	0.9581***	22.1052	1.0112***	55.1359	0.9928***	27.6461	1.0383***	52.2352
SMB	0.7252***	8.6189	0.0510	1.4335	0.7204***	10.3317	0.1903***	4.9316
HML	−0.0965	−1.2844	0.2180***	6.8543	−0.0448	−0.7194	0.2243***	6.5057
RMW	−0.3786***	−3.7764	0.1236***	2.9128	−0.3188***	−3.8369	0.1336***	2.9052
CMA	−0.0663	−0.5724	0.0753	1.5372	−0.0467	−0.4866	0.0425	0.7998
$Intercept$	0.3743**	2.0272	−0.1696**	−2.1707	0.1608	1.0510	−0.1595*	−1.8838
R^2	0.8638		0.9637		0.9058		0.9621	
Annualized Alpha	4.59**		−2.02**		1.95		−1.90*	
Δ Alpha				**6.61*****				**3.85****
Panel C: Naive-unweighted ESG score ($N = 155$)								
MKT	0.9649***	22.6174	1.0125***	55.0927	0.9917***	27.5742	1.0300***	51.4703
SMB	0.7039***	8.4984	0.0361	1.0120	0.7254***	10.3884	0.1766***	4.5466
HML	−0.0501	−0.6772	0.2388***	7.4904	−0.0293	−0.4698	0.2237***	6.4441
RMW	−0.4283***	−4.3395	0.1132***	2.6619	−0.4239***	−5.0948	0.1567***	3.3850
CMA	−0.1144	−1.0039	0.0515	1.0484	−0.0856	−0.8905	0.0315	0.5888
$Intercept$	0.3442*	1.8938	−0.1634**	−2.0877	0.1832	1.1956	−0.1527*	−1.7918
R^2	0.8697		0.9637		0.9081		0.9608	
Annualized Alpha	4.21*		−1.94**		2.22		−1.82*	
Δ Alpha				**6.15*****				**4.04****

Notes: This table shows coefficient estimates and t-statistics of the Refinitiv, SASB risk-weighted, and naive-unweighted ESG 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 Fama and French (2015): 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% levels.

Table 4
Regression results: Paris Agreement.

	Deciles				Quintiles			
	Bottom		Top		Bottom		Top	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Panel A: Before Paris Agreement (N = 70)								
Refinitiv								
Intercept	−0.0320	−0.1590	−0.2684***	−2.7136	−0.0137	−0.0903	−0.2255***	−2.9717
Annualized Alpha	−0.38		−3.17***		−0.16		−2.67***	
Δ Alpha				**2.79**				**2.51**
SASB risk-weighted								
Intercept	0.1597	1.0608	−0.2959***	−3.6265	−0.0168	−0.1336	−0.2065***	−2.8106
Annualized Alpha	1.93		−3.49***		−0.20		−2.45***	
Δ Alpha				**5.42*****				**2.25**
Naive-unweighted								
Intercept	0.0134	0.0797	−0.2633***	−3.1756	−0.0515	−0.4176	−0.1516**	−2.0564
Annualized Alpha	0.16		−3.11***		−0.62		−1.80**	
Δ Alpha				**3.27**				**1.18**
Panel B: After Paris Agreement (N = 50)								
Refinitiv								
Intercept	0.1998	0.7695	−0.2997**	−2.6326	0.2754	1.3196	−0.2693**	−2.1724
Annualized Alpha	2.42		−3.54**		3.36		−3.18**	
Δ Alpha				**5.96*****				**6.54*****
SASB risk-weighted								
Intercept	0.4638**	2.0617	−0.1329	−1.0100	0.3508*	1.8848	−0.3076**	−2.2145
Annualized Alpha	5.71**		−1.58		4.29*		−3.63**	
Δ Alpha				**7.29*****				**7.92*****
Naive-unweighted								
Intercept	0.5198**	2.2734	−0.1376	−1.1142	0.4364**	2.1810	−0.3128**	−2.2910
Annualized Alpha	6.42**		−1.64		5.36**		−3.69**	
Δ Alpha				**8.06*****				**9.05*****

Notes: This table shows coefficient estimates of the intercepts and t-statistics of the Refinitiv, naive-unweighted, and SASB risk-weighted ESG regressions before and after the Paris Agreement in December 2015. 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 of the Fama and French (2015) model are hidden. The period before the Paris Agreement starts in February 2010 and ends in November 2015, and the period after the Paris Agreement is between January 2016 and February 2020. 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% levels.

5. Robustness

As a robustness test, we follow Bolton and Kacperczyk (2021) and consider the Paris Agreement (December 2015) as an exogenous shock. According to Duong et al. (2022) "the Agreement is considered as the most significant event in climate finance history". The Agreement's focus on climate change makes it particularly relevant in our context for two reasons. First and as indicated above, environmental categories have the highest dispersion of scores (in terms of standard deviations), implying the most pronounced differentiation between firms. Second, it is reasonable to assume the Agreement served as an important turning point, prompting investors to recognize the broader relevance of sustainability issues (including climate risks) to asset pricing. Therefore, we divide our sample into two subsamples: the first from February 2010 to November 2015 (before Paris Agreement) and the second from January 2016 to February 2020 (after Paris Agreement). We exclude data from March 2020 onwards to eliminate potential influence from the COVID-19 crisis on our findings.

Results in Table 4 reveal that after the Paris Agreement, all ESG metrics display a statistically significant ability to differentiate investments, but with varying degrees of effectiveness as measured by the performance difference (Δ Alpha). Naive-unweighted and SASB risk-weighted ESG scores exhibit the strongest distinction, with significantly higher returns (positive Alpha) for the bottom quintile portfolios. This effect is also observed for decile portfolios. Refinitiv ESG scores, on the other hand, show the weakest differentiation, with significant Alphas only being observed for the top quintile (and decile) portfolios. Hence, naive-unweighted and SASB risk-weighted ESG metrics may be more sensitive to identifying stocks with differing financial performance in the post-Agreement era.

In the pre-Agreement period, none of the performance differences (Δ Alpha) are statistically significant for quintile portfolios. For decile portfolios, only the top decile with SASB weights exhibited a significant performance difference. However, since the SASB guidance was not established before February 2014, we cannot rule out that this finding might be due to endogeneity. Performance differences for other ESG weightings were not significant during this period. Interestingly, we also find that top portfolios across all ESG metrics were associated with potentially lower future returns before the Paris Agreement.

Table 5
Regression results: robustness tests.

	Decile Δ Alpha			Quintile Δ Alpha		
	Refinitiv	SASB	Unweighted	Refinitiv	SASB	Unweighted
Alternative Regions						
Developed Countries	6.80***	7.63***	7.52***	4.84***	6.29***	6.34***
Alternative Sample Periods						
Excluding COVID	3.80**	6.08***	5.23***	3.96***	4.49***	4.28***
Excluding Fin. Crisis	4.48*	6.53***	6.65***	2.35	3.48*	3.98*
Excluding Both Events	3.49*	5.93***	5.57***	3.60**	4.42***	4.51**
Alternative Rebalancing						
Monthly Rebalancing	4.79**	6.03***	5.68***	3.04*	3.82**	4.05**
Alternative Portfolio Sorting						
Residual ESG	5.02**	5.35**	5.67**	2.47	3.73**	3.09*

Notes: This table presents the annualized Alpha differences of the robustness tests. All values are presented in percent. 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 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. The portfolios are sorted by ESG scores. Under "Alternative Regions", we examine a different sample and extend it to 6,900 firms with their headquarters in the developed countries. Under "Alternative Sample Periods", we report differences in Alpha for sub-periods. The "Excluding COVID" period starts in February 2010 and ends in February 2020. The "Excluding Fin. Crisis" period does not include the aftermath of the financial crisis and starts in January 2013 and ends in December 2022. The "Excluding Both Events" period includes observations between January 2013 and February 2020. Under "Alternative Rebalancing", we rebalance the portfolios on a monthly basis. Under "Alternative Portfolio Sorting", we estimate an ESG score with respect to firm size, market-to-book ratio, return-on-assets, and leverage and conduct the sorting based on those estimates. ***, **, and * reflect significance at the 1%, 5%, and 10% levels.

In summary, our results suggest a shift in investor behaviour after the Paris Agreement. The Paris Agreement might have influenced how investors perceive ESG factors when making investment decisions, leading to performance differences captured by ESG scores to varying degrees.

Table 5 displays a series of additional robustness tests. To generalize our results, we first re-run the analysis with firms headquartered in the U.S., Europe, Asia, Canada, and Australia. The Alpha differences increase for all weightings, but the SASB and naive approaches better reflect the differentiation. In the next series of robustness tests, we exclude observations from the aftermath of the Financial Crisis and the COVID-19 period. All alternative sub-samples confirm our main results. To capture the dynamic nature of ESG scores, we rebalance the portfolios on a monthly basis. Notably, the quintile portfolio results diverge only minimally from the yearly rebalancing. This is because most disclosed ESG data is updated once a year (Refinitiv, 2023). In the final robustness test, we address possible endogeneity concerns. We estimate an ESG score based on firm characteristics (size, market-to-book ratio, return-on-assets, and leverage) and sort the decile and quintile portfolios based on the new estimates. The Alpha differences are more similar, but the general trend remains evident. Across all additional tests, we confirm our main findings: The largest differences are displayed for SASB weights and naive weights.

6. Conclusion

Sustainability and financial performance are closely linked. Consequently, ESG ratings are valuable tools for managers and academics, providing information about the sustainability of investments and enhancing transparency in financial markets. From a policymaker's point of view, understanding how weighting methodologies impact the differentiation of firms based on ESG factors can inform policymakers seeking to establish robust ESG rating frameworks.

This study investigates the weighting of categories in Refinitiv's ESG rating, recognized as one of the most important metrics in the field. Refinitiv employs a data-driven approach similar to the methodology of MSCI ratings. We focus on whether the weighting of categories in Refinitiv's ESG ratings aligns with their financial importance for investors. We compare Refinitiv's weights with a simple, equal weighting scheme and with SASB weights calculated based on expert assessments.

We find that all weighting methods can differentiate firms based on stock returns. However, equal weights and SASB weights lead to a more pronounced differentiation of firm performance compared to Refinitiv's data-driven weighted score, especially following the Paris Agreement. This suggests that alternative weighting approaches might better capture the financial relevance of ESG categories.

Specifically, one future research direction is to identify procedures that lead to weights enabling a clearer differentiation of investment opportunities based on ESG criteria. From a regulatory perspective, it is important to identify weights that better align with the financial relevance of categories. This would further enhance market transparency and inform investment decisions. Our work could be extended by analysing the weighting methodology of other providers in detail to see if our results hold true for these other major rating agencies. Finally, while the Fama and French (2015) five-factor model was used to test return differences, another interesting avenue for future research could involve incorporating more forward-looking information, such as analyst-based implied cost of capital.

CRediT authorship contribution statement

Matthias Muck: Writing – review & editing, Supervision, Methodology, Conceptualization. **Thomas Schmidl:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization.

Data availability

The authors do not have permission to share data.

Declaration of Generative AI and AI-assisted technologies in the writing process

Statement: During the preparation of this work the authors used Google Gemini and DeepL in order to copy editing 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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.frl.2024.105924.

References

Albuquerque, R., Koskinen, Y., Yang, S., Zhang, C., 2020. Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. Rev. Corp. Financ. Stud. 9 (3), 593–621.
Albuquerque, R., Koskinen, Y., Zhang, C., 2019. Corporate social responsibility and firm risk: Theory and empirical evidence. Manage. Sci. 65 (10), 4451–4469.
Apergis, N., Poufinas, T., Antonopoulos, A., 2022. ESG scores and cost of debt. Energy Econ. 112, 106186.
Avramov, D., Cheng, S., Lioui, A., Tarelli, A., 2022. Sustainable investing with ESG rating uncertainty. J. Financ. Econ. 145 (2, Part B), 642–664.
Berg, F., Fabisik, K., Sautner, Z., 2020. Rewriting history II: The (Un)predictable past of ESG ratings. Working Paper.
Berg, F., Kölbel, J.F., Rigobon, R., 2022. Aggregate confusion: The divergence of ESG ratings. Rev. Financ. 26 (6), 1315–1344.
Billio, M., Costola, M., Hristova, I., Latino, C., Pelizzon, L., 2021. Inside the ESG ratings: (Dis)agreement and performance. Corp. Soc. Responsib. Environ. Manag. 28 (5), 1426–1445.
Boffo, R., Patalano, R., 2020. ESG investing: Practices, progress and challenges. https://www.oecd.org/finance/ESG-Investing-Practices-Progress-Challenges.pdf.
Bolton, P., Kacperczyk, M., 2021. Do investors care about carbon risk? J. Financ. Econ. 142 (2), 517–549.
Broadstock, D.C., Chan, K., Cheng, L.T.W., Wang, X., 2021. The role of ESG performance during times of financial crisis: evidence from COVID-19 in China. Financ. Res. Lett. 38, 101716.
Chatterji, A.K., Durand, R., Levine, D.I., Touboul, S., 2016. Do ratings of firms converge? Implications for managers, investors and strategy researchers. Strateg. Manag. J. 37 (8), 1597–1614.
Consolandi, C., Eccles, R.G., Gabbi, G., 2022. How material is a material issue? Stock returns and the financial relevance and financial intensity of ESG materiality. J. Sustain. Financ. Invest. 12 (4), 1045–1068.
Dimson, E., Karakaş, O., Li, X., 2015. Active ownership. Rev. Financ. Stud. 28 (12), 3225–3268.
Dobrick, J., Klein, C., Zwergel, B., 2023. Size bias in refinitiv ESG data. Finance Res. Lett. 55, 104014.
Douglas, E., van Holt, T., Whelan, T., 2017. Responsible investing: Guide to ESG data providers and relevant trends. J. Environ. Invest. 8 (1), 92–114.
Drempetic, S., Klein, C., Zwergel, B., 2020. The influence of firm size on the ESG score: Corporate sustainability ratings under review. J. Bus. Ethics 167 (2), 333–360.
Dumrose, M., Rink, S., Eckert, J., 2022. Disaggregating confusion? The EU taxonomy and its relation to ESG rating. Finance Res. Lett. 48, 102928.
Duong, H.N., Kalev, P.S., Kalimipalli, M., Trivedi, S., 2022. Do firms benefit from carbon risk management? Evidence from the credit default swaps market. Working Paper.
Easton, P.D., Monahan, S.J., 2005. An evaluation of accounting-based measures of expected returns. Account. Rev. 80 (2), 501–538.
Eccles, R.G., Krzus, M.P., Rogers, J., Serafeim, G., 2012. The need for sector-specific materiality and sustainability reporting standards. J. Appl. Corp. Finance 24 (2), 65–71.
Eccles, R.G., Stroehle, J.C., 2018. Exploring Social Origins in the Construction of ESG Measures, (no. 3212685), Rochester, NY.
Edmans, A., 2011. Does the stock market fully value intangibles? Employee satisfaction and equity prices. J. Financ. Econ. 101 (3), 621–640.
El Ghoul, S., Guedhami, O., Kwok, C.C.Y., Mishra, D.R., 2011. Does corporate social responsibility affect the cost of capital? J. Bank. Financ. 35 (9), 2388–2406.
ESMA, 2021. In: European Securities and Markets Authority (Ed.), ESMA Letter to EC on ESG Ratings. pp. 1–4.
Fama, E.F., 1970. Efficient capital markets: A review of theory and empirical work. J. Finance 25 (2), 383–417.
Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. J. Financ. Econ. 116 (1), 1–22.
Friede, G., Busch, T., Bassen, A., 2015. ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. J. Sustain. Financ. Invest. 5 (4), 210–233.
Gan, X., Fernandez, I.C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., Wu, J., 2017. When to use what: Methods for weighting and aggregating sustainability indicators. Ecol. Indic. 81, 491–502.
Gebhardt, W.R., Lee, C.M.C., Swaminathan, B., 2001. Toward an implied cost of capital. J. Account. Res. 39 (1), 135–176.
Gibson, B.R., Krueger, P., Schmidt, P.S., 2021. ESG rating disagreement and stock returns. Financ. Anal. J. 77 (4), 104–127.
Grewal, J., Hauptmann, C., Serafeim, G., 2021. Material sustainability information and stock price informativeness. J. Bus. Ethics 171 (3), 513–544.
Guay, W., Kothari, S.P., Shu, S., 2011. Properties of implied cost of capital using analysts' forecasts. Aust. J. Manag. 36 (2), 125–149.
Hartzmark, S.M., Sussman, A.B., 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. J. Finance 74 (6), 2789–2837.
Heeb, F., Kölbel, J.F., Paetzold, F., Zeisberger, S., 2023. Do investors care about impact? Rev. Financ. Stud. 36 (5), 1737–1787.
Hong, H., Kacperczyk, M., 2009. The price of sin the effects of social norms on markets. J. Financ. Econ.
Hübel, B., Scholz, H., 2020. Integrating sustainability risks in asset management: The role of ESG exposures and ESG ratings. J. Asset Manag. 21 (1), 52–69.
Ilhan, E., Krueger, P., Sautner, Z., Starks, L.T., 2023. Climate risk disclosure and institutional investors. Rev. Financ. Stud. 36 (7), 2617–2650.
Khan, M., Serafeim, G., Yoon, A., 2016. Corporate sustainability: First evidence on materiality. Account. Rev. 91 (6), 1697–1724.
Palmieri, E., Ferilli, G.B., Stefanelli, V., Geretto, E.F., Polato, M., 2023. Assessing the influence of ESG score, industry, and stock index on firm default risk: A sustainable bank lending perspective. Finance Res. Lett. 57, 104274.

Pástor, Ľ., Stambaugh, R.F., Taylor, L.A., 2022. Dissecting green returns. J. Financ. Econ. 146 (2), 403–424.
Pedersen, L.H., Fitzgibbons, S., Pomorski, L., 2021. Responsible investing: The ESG-efficient frontier. J. Financ. Econ. 142 (2), 572–597.
Refinitiv, 2023. Environmental, social and governacne scores from LSEG. LSEG.
Schiehll, E., Kolahgar, S., 2021. Financial materiality in the informativeness of sustainability reporting. Bus. Strategy Environ. 30 (2), 840–855.
WEF, 2019. Seeking return on ESG: Advancing the reporting ecosystem to unlock impact for business and society (produced in collaboration with Allianz SE and Boston Consulting Group). https://www3.weforum.org/docs/WEF_ESG_Report_digital_pages.pdf.
Whelan, T., Atz, U., van Holt, T., Clark, C., 2021. In: NYU Stern Center for Sustainable Business and Rockefeller Asset Management (Ed.), ESG and Financial Performance: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015 - 2020.
Wong, W.C., Batten, J.A., Ahmad, A.H., Mohamed-Arshad, S.B., Nordin, S., Adzis, A.A., 2021. Does ESG certification add firm value? Finance Res. Lett. 39, 101593.