ESG Ratings Management*

Jess Cornaggia[a]
Pennsylvania State University

Kimberly Cornaggia[b]
Pennsylvania State University

June 4, 2024

Abstract

This paper examines how corporations respond to changes in ESG ratings criteria. Using data from a leading ESG rater, we find that firms' reported performance on certain criteria improves within the same month of the rater changing its model to place more emphasis on the criteria. This effect is stronger among firms with ESG-focused institutional investors and customers. We find no evidence that reported performance predicts real changes in firms' ESG behavior. Rather, the improvements appear cosmetic, suggesting the ratings management appeases investors and consumers who rely on ESG ratings. Overall, the results show how firms influence their ESG ratings when they are allowed to engage with ESG raters during the rating process.

Keywords: ESG, ESG Ratings, Sustainability, Greenwashing, CSR

JEL Classifications: G34, G24, D22

* We thank Yuchen Li for research assistance. For helpful comments, we thank Florian Berg, Janet Gao, Stefan Lewellen, seminar participants at Penn State University, University of Illinois at Urbana-Champaign, University of Bielefeld and University of Dortmund Joint Finance Seminar, Korea University Business School and Korea Advanced Institute of Science and Technology, and audience members at the Pitt/CMU/PSU Finance Conference, Bretton Woods Accounting and Finance Ski Conference, and the University of Connecticut Finance Conference. All errors belong to the authors.
[a] Corresponding author. Telephone: 814.863.2390, E-mail address: jcornaggia@psu.edu.
[b] Telephone: 814.865.2243, E-mail address: kcornaggia@psu.edu.

1. Introduction

Environmental, social, and governance (ESG) ratings provide information on corporate performance with respect to non-investor stakeholder interests (Larcker, Pomorski, Tayan, and Watts, 2022). Whether firms care about their ESG ratings is an open question. On one hand, ESG activities may not align with the traditional goal of maximizing shareholder wealth, and ESG ratings often disagree (Berg, Kölbel, and Rigobon, 2022; Berg, Kölbel, Pavlova, and Rigobon, 2023). On the other hand, many investors rely on ESG ratings to guide investment decisions (Avramov, Cheng, Lioui, and Tarelli, 2022). The purpose of this paper is to examine the sensitivity of firms' behavior to criteria underlying ESG ratings. When an ESG rater adjusts the methodology it uses to produce ratings, do firms adjust their behavior in response to manage their ESG ratings?

Data from Sustainalytics, a leading ESG rater, provide an opportunity to address this question. Sustainalytics' ratings are weighted sums, meaning they reflect a combination of raw scores and weights. A firm's overall ESG rating reflects a weighted sum of E, S, and G pillar scores. Each of the E, S, and G pillar scores, in turn, reflects a weighted sum of dozens of underlying criteria scores. (Details on Sustainalytics' rating methodology are available in Section 2.) The basic approach in this paper is to test whether, and over what time horizon, firms improve their raw criteria scores in response when Sustainalytics places greater weight on a given underlying criteria. The sample runs from 2009 through 2019, a period when the rater provides monthly updates on companies' ESG ratings.

The main analysis tests the sensitivity of criteria raw scores to criteria weights. We find that a 1 percentage point increase in weight for a given criteria is associated with an increase in raw score of about 14% of a typical change. This response occurs in the *same month* as the weight change. Our approach forces estimation to come from variation at the firm-criteria-month level.

Firm-criteria fixed effects absorb variation in raw scores associated with, for example, firms' average performance on certain criteria. Firm-month fixed effects absorb variation associated with, for example, ESG ratings "drift", whereby firms exhibit improvements in ratings over time (Crespi and Migliavacca, 2020; D.E. Shaw & Co., 2022). Criteria-month fixed effects absorb variation associated with, for example, social justice trends that motivate firms to improve performance on related issues.

We test the robustness of our results and find similar effects when we use changes in criteria weights rather than levels. The results are also robust in the time series: Although they vary in statistical significance, we observe positive, contemporaneous relations between rater criteria weights and firms' raw performance scores in ten out of 11 sample years. We also find the results are robust in the cross section and are not driven by any particular industry. However, results tend to be strongest for criteria associated with the environmental pillar. Results for criteria associated with the social and governance pillars are less robust.

We consider several explanations for how firms change their raw ESG scores so quickly in response to changes in criteria weights. First, firms may legitimately and nimbly adjust their ESG behavior. To test this possibility, we incorporate data from RepRisk, a news aggregator that tracks firm involvement in reputation-harming ESG incidents. We begin by decomposing firms' raw criteria scores into two components, one related to the criteria weights applied by the rater and the other a residual component. We focus on the first component. If the portion of raw scores associated with criteria weights reflects real firm behavior, then this component should predict a decrease in the likelihood that firms receive negative press for their involvement in ESG incidents. We find no evidence of this effect. The weight-driven component of raw scores does not predict the likelihood that firms engage in reputation-harming ESG incidents, either in the same month or over the next year. This is true whether we examine equally weighted incidents or if the incidents are weighted by severity, novelty, or reach.

As an alternative approach, we incorporate data from the U.S. Environmental Protection Agency's Toxic Releases Inventory (TRI) Program. This analysis focuses on firms' environmental criteria, as social and governance scores are unrelated to firms' release of toxic chemicals. As before, we decompose firms' raw environmental criteria scores into weight-driven and residual components. We find scant evidence that the weight-drive component has predictive power for the toxic chemicals firms release, recycle, recover, treat or transfer over time horizons up to two years in the future. Together with the analysis based on RepRisk incidents, these results cast doubt on the possibility that real firm behavior drives the relation between criteria weights and ESG scores.

Another explanation is that Sustainalytics changes how it generates raw scores at the same time it changes criteria weights. That is, criteria updates could manifest through the way scores are measured, as well as weighted. This explanation predicts heightened volatility in raw scores when criteria weights change, and this is indeed what we observe. However, this explanation does not predict which direction raw scores should change. We find increases in criteria weights are associated with increases in raw scores, but not decreases. Further, volatility in raw scores is smaller and more symmetric around decreases in criteria weights.

A related possibility is that Sustainalytics caters to firms, not the other way around. We test this possibility by studying the extensive margin of criteria weight changes. Specifically, we focus on instances where Sustainalytics stops using certain criteria altogether in the creation of its ESG ratings. If Sustainalytics deemphasizes criteria when firms struggle to perform well under the criteria, then raw scores should decline leading up to criteria terminations. We find no evidence of this pattern. If anything, raw scores associated with terminated criteria exhibit a slight upward trend through the month of termination. We also find that raw scores during the month of termination are significantly *higher* than raw scores among criteria that are newly introduced.

A fourth explanation is that firms manage their ESG ratings by exerting influence on ESG raters. This opportunity could present itself when ESG raters incorporate feedback from firms during the rating process. For example, in a 2017 methodology brief, Sustainalytics indicates that the last step of its rating process is to solicit feedback from the rated company (Sustainalytics, 2017).[1] Simon MacMahon, head of ESG research at Sustainalytics, describes this step as follows:

> *"Once we have completed our ratings process, we send the profile to the company for feedback. During those conversations, we're looking for any additional information or clarification that can enhance our analysis. New information doesn't always lead to a change in our rating, but we do listen. As ESG rating outcomes become more important, we certainly hear from people inside firms who forcefully argue for their point of view."*[2]

The results so far are consistent with this explanation. We examine this explanation more thoroughly using several approaches. First, we examine the sensitivity of our results among subsamples based on criteria characteristics. Sustainalytics intends its criteria to measure preparedness, disclosure, or performance. We find the contemporaneous relation between criteria weights and raw scores is greatest among criteria designed to measure preparedness. Many of these criteria involve the drafting of policies on topics such as money laundering, conflict minerals, data privacy, and so forth. Firms could create simple versions of such policies on short notice upon learning Sustainalytics places greater emphasis on the criteria. It would be more difficult to credibly adjust performance-related criteria such as establishing a corporate foundation or increasing racial diversity among a board of directors. In this sense, our results provide perspective on a recent investigation by Bloomberg of MSCI, another leading ESG rater. The investigators

---

[1] The language is as follows: *"A draft report is sent to every company that we research for feedback. In our company contact process our goal is to gather feedback as well as additional and updated information from the company."* This practice remains in place as of 2021 (Sustainalytics, 2021).
[2] "The Challenge of Rating ESG Performance" Harvard Business Review. URL accessed July 11, 2023: https://hbr.org/2020/09/the-challenge-of-rating-esg-performance

concluded that MSCI "rewards the most rudimentary business practices," and that many companies receive upgrades for "doing nothing but surfing the wave of methodology changes, reweightings, or similar tweaks."[3]

A consequence of the ratings management hypothesis is that firms' incentive to manage ESG ratings should vary in the cross section with monitoring by ESG-focused stakeholders. For example, firms with more ESG-focused institutional investors should have greater incentive to manage their ESG ratings. We examine 13f filings and sort firms based on equity holdings by ESG-focused institutional investors. We indeed find the relation between criteria weights and raw scores is more pronounced for firms with high ESG investor ownership. We conduct a similar analysis based on ESG-focused customers. European investors and consumers have long exhibited greater demand for corporate social responsibility than investors and consumers in the U.S. and elsewhere.[4] We collect data from FactSet on the geographic distribution of firms' revenue generation. We find that the relation between criteria weights and raw scores is more pronounced among firms that derive more revenue from Europe. These findings are surprising given that ESG ratings often diverge and, as Azarmsa and Shapiro (2023) argue, this divergence may lead to the under-provision of effort by firms to improve their ESG performance. Our findings indicate firms manage ESG ratings to appeal to ESG-focused stakeholders even though Sustainalytics is only one rater in an unconcentrated market.

2. Background

2.1. Corporate ESG Ratings

[3] "The ESG Mirage" Bloomberg. URL accessed July 10, 2023: https://www.onepak.com/the-esg-mirage/
[4] "ESG Headwinds: Embraced In Europe, Under Fire in America" Forbes. URL accessed May 9, 2023: https://www.forbes.com/sites/forbesfinancecouncil/2023/04/11/esg-headwinds-embraced-in-europe-under-fire-in-america/?sh=2916c0071f21

ESG ratings provide information on corporate performance with respect to non-investor stakeholder interests.[5] The number of companies providing ESG ratings has expanded rapidly in recent years to at least 160 worldwide.[6] Despite the growth in this industry and widespread reliance on ESG ratings, many observers question the reliability of these ratings. One criticism relates to the independence of ESG raters, as several prominent ESG raters interact with rated firms during the rating process. For example, Institutional Shareholder Services (ISS) produces ESG ratings while selling advising services that help firms "stay ahead of emerging shifts in environmental and social norms."[7] Sustainalytics, a company owned by Morningstar and the focus of this paper, explicitly incorporates feedback from rated firms prior to releasing rating updates. See Figure 1 excerpted from Sustainalytics (2017) and Sustainalytics (2021). According to Sustainalytics, it has over 1,000 clients, serves 18 of the top 20 asset managers in the world, and covers over 20,000 companies in 172 countries.[8] It also features prominently in academic research. Its ratings have been used or referenced in nearly 2,000 academic articles since 2009.[9]

[Insert Figure 1 here.]

What should ESG ratings measure, and how should they measure it? These are difficult questions, but they require answers because ESG ratings play an increasingly important role in the

---

[5] See Larcker, Pomorski, Tayan, and Watts (2022) for a primer on the role and function of ESG ratings in the economy. Whether these goals are aligned with the traditional corporate objective of shareholder wealth maximization remains an area of active debate, dating back at least to Milton Friedman's essay, "The Social Responsibility of Business Is to Increase Its Profits" (Friedman, 1970). Recent review articles on the role of ESG (or precursors to ESG, such as CSR) objectives in corporate decision making include Cornell and Damodaran (2020) and Edmans (2023).

[6] "The Signal and the Noise. Measurement of ESG data needs a big overhaul" The Economist. URL accessed August 20, 2022: https://www.economist.com/special-report/2022/07/21/the-signal-and-the-noise. Avetisyan and Hockerts (2016) discuss consolidation in the ESG ratings industry.

[7] "Improve ESG Ratings" ISS Corporate Solutions. URL accessed August 20, 2022: https://www.isscorporatesolutions.com/improve-esg-ratings/

[8] "Who We Are" Morningstar Sustainalytics. URL accessed August 20, 2022: https://www.sustainalytics.com/about-us

[9] A search on https://app.dimensions.ai/discover/publication for the keyword "Sustainalytics" returns 1,955 publications as of March 10, 2023.

economy (Starks, Venkat, and Zhu, 2023). Existing research weighs the roles of ESG data disclosure (Landi and Sciarelli, 2019; Dimson, Marsh, and Staunton, 2020; Christensen, Serafeim, and Sikochi, 2022; Liu, 2022; Ilhan, Krueger, Sautner, and Starks, 2022; and Krueger, Sautner, Tang, and Zhong, 2022), preferences for integrating new criteria (Escrig-Olmedo, Muñoz-Torres, and Fernández-Izquierdo, 2010; Escrig-Olmedo, Fernández-Izquierdo, Ferrero-Ferrero, Rivera-Lirio; and Muñoz-Torres, 2019), coordination on taxonomies (Dumrose, Rink, and Eckert, 2022), and company preferences (Clementino and Perkins, 2021). Some commentators suggest E, S, and G should not be integrated into a single metric, as the three categories are separable.[10]

A growing literature documents that ESG raters often disagree and examines the determinants of this disagreement. Berg, Kölbel, and Rigobon (2022) is a seminal paper on the topic. Others include Dorfleitner, Halbritter, and Nguyen (2015); Chatterji, Durand, Levine, and Touboul (2016); Zumente and Lāce (2021); and Charlin, Cifuentes, and Alfaro (2022), who note that ESG ratings exhibit even less agreement than wine ratings. Berg, Kölbel, and Rigobon (2022) examine six prominent ESG raters and provide a detailed analysis deconstructing the sources of this disagreement. They deconstruct ratings along three dimensions and show that 56% of the divergence stems from measurement differences, 38% stems from differences in scope, and 6% stems from weight differences. Billio, Costola, Hristova, Latino, and Pelizzon (2021) likewise document differences among raters in terms of ESG characteristics, attributes, and standards.

2.2. ESG Investment and Investor Reliance on ESG Ratings

ESG funds invest in corporations with favorable ESG profiles (Krueger, Sautner, and Starks, 2020). This investment practice has grown significantly in recent years. For example, assets in European sustainable funds have grown over tenfold since 2010, from EUR 112 billion to EUR

[10] "Investors should separate out the E, S, and G" Financial Times. URL accessed July 12, 2023: https://www.ftadviser.com/investments/2022/03/18/investors-should-separate-out-the-e-s-and-g/

1,202 billion in 2020.[11] Despite growing interest in ESG investing, whether ESG ratings explain stock returns remains and open question. Rzeźnik, Hanley, and Pelizzon (2022) show that less-sophisticated retail investors are more likely than mutual funds to rely on ESG ratings produced by Sustainalytics. Some evidence shows that stock returns (Shanaev and Ghimire, 2022) and crash risk (Feng, Goodell, and Shen, 2022) respond to ESG ratings. However, Atz, Holt, Liu, and Bruno (2022) provide a meta-analysis of 1,141 peer-reviewed articles and 27 meta-reviews published between 2015 and 2020. These authors conclude that financial performance of ESG investing has on average been indistinguishable from conventional investing.

Part of the challenge to answering this question is the aforementioned divergence in ESG ratings. This divergence muddies any relationship between ESG ratings, investor demand, and performance. Berg, Kölbel, Pavlova, and Rigobon (2022) propose noise-correction procedures to more cleanly estimate the relation between firms' ESG ratings and stock performance. Brandon, Krueger, and Schmidt (2021) show that firms with higher ESG rating disagreement earn a risk premium. However, a complicating factor to understanding the relation between ESG ratings and stock performance is the practice of ESG ratings backdating. Berg, Fabisik, and Sautner (2021) show that ratings produced by Refinitiv ESG (formerly ASSET4) are sometimes rewritten and this re-writing generates the appearance of a positive link between ESG scores and firms' stock market performance. We contribute to this literature with evidence that the reliability of ESG rating may further be affected by "ratings management" on the part of rated firms.

2.3. Parallels between ESG Ratings and Credit Ratings

[11] Source: Morningstar. URL accessed November 22, 2022: https://assets.contentstack.io/v3/assets/blt4eb669caa7dc65b2/blt7a208fcfc3d719a8/61ade16b7de7d945b9c4b8cd/European_ESG_Fund_Landscape_2020.pdf

In several ways, ESG raters are like credit rating agencies such as Moody's, Standard & Poor's, and Fitch. Aslan, Poppe, and Posch (2021) study links between the two types of raters. Both types of raters produce model-dependent metrics. Our paper is like those in the credit ratings space that study the real and financial implications of raters' modeling choices.[12] Both types of raters aggregate information on dimensions of corporate performance and provide public metrics around which market participants may coordinate, as in Boot, Milbourn, and Schmeits (2006). Both also subject to conflicts of interest driven by ownership structure, as shown by Tang, Yan, and Yao (2022). These authors find that firms with the same major shareholders as ESG raters, known as "sister firms", receive higher ESG ratings. This result echoes a rich literature on credit rating agencies that studies the effects of rater ownership and compensation structure on ratings quality.[13] Gyönyörová, Stachoň, and Stašek (2021) emphasize the importance of company sectors and country of domicile when explaining ESG ratings divergence, similar to how Cornaggia, Cornaggia, and Hund (2017) show differences in credit ratings practices across asset classes. One contrast is that the real effects that credit ratings have on the economy appear more limited for ESG ratings (Berg, Heeb, and Kölbel, 2023).[14]

3. Data and Sample

We obtain ESG ratings data from Sustainalytics. We restrict the sample to firms that are rated by Sustainalytics every month from August 2009 through September 2019.[15] This filter

[12] Examples include Griffin and Tang (2012), Griffin, Nickerson, and Tang (2013), Griffin and Nickerson (2017), and Cornaggia, Cornaggia, and Israelsen (2018 and 2023).
[13] Examples include Jiang, Stanford, and Xie (2012), Cornaggia and Cornaggia (2013), Kedia, Rajgopal, and Zhou (2014), Xia (2014), Bonsall (2014), and Bruno, Cornaggia, and Cornaggia (2016).
[14] Examples include Adelino, Cunha, Ferriera (2017), Almeida, Cunha, Ferreira, and Restrepo (2017), Chen, Lookman, Schürhoff, and Seppi (2014), Cornaggia, Cornaggia, and Israelsen (2018), and Kisgen (2019).
[15] Sustainalytics made a major change to how it creates ESG ratings in October 2019, after the conclusion of our sample. Its new methodology is called "ESG Risk Ratings 2.0" (Sustainalytics, 2021). The new method incorporates the extent to which a company is exposed to ESG risks and how well the company manages these exposures. The new method continues to incorporate feedback from rated firms at the final step of the rating process. Rzeźnik, Hanley, and Pelizzon (2022) show that retail investors, in particular, reacted to Sustainalytics' methodology change.

prevents the results from being affected by firms entering or exiting the sample. The data structure we use for most analysis is a firm-criteria-month panel. Each firm-criteria-month observation includes raw scores and weights for criteria underlying environmental, social, and governance pillars. We observe in the data several months where Sustainalytics reports criteria weights as percentages rather than decimals. These months include July and August of 2017 and March through November of 2018, and the practice appears to apply across firms. We exclude from baseline tests observations associated with these months to avoid making judgement calls as to whether we should scale the weights to make them consistent with the rest of the sample. (For example, it is unclear whether a criteria weight of "20" means 20 percent or 20 basis points.) Figure 2 displays mean weights across all firms rated with a given criteria in each sample month.

[Insert Figure 2 here.]

Criteria weights change through time. Because we restrict the sample to firms that are rated by Sustainalytics every month from August 2009 through September 2019, the changing mean sample weights cannot be driven by new firms being rated or other firms exiting the sample. Instead, variation in mean weights derives from at least one of two sources: i) Sustainalytics could increase or decrease the set of criteria it deploys for a given firm-month. ii) Sustainalytics could adjust the criteria weights it applies to the set of existing criteria. Our main analysis controls for the influence of i) by including firm-month fixed effects. The results we document cannot be driven by a growing or shrinking number of criteria being applied to a firm over time. Instead, the results derive from criteria weight adjustments.

Figure 3 displays mean criteria raw scores through time. Raw scores measure how well companies proactively manage the environmental, social, and governance issues that are most

Mutual funds did not react. Firms whose stock prices declined because of the methodology change responded by repurchasing shares.

material to their business (Sustainalytics, 2017). Raw scores exhibit significant variation in magnitude. For example, in Panel A, firms rated under environmental criteria e_1_4, "Environmental Fines and Non-monetary Sanctions" have mean raw scores exceeding 80 points over the sample. In contrast, firms rated under environmental criteria e_1_8, "Programmes and Targets to Increase Renewable Energy Use" have mean raw scores below 20 points over the sample. Figure A.1 in the Internet Appendix displays the distribution of the raw scores in the full sample of firm-criteria-month observations. We observe bunching at certain scores. Although raw scores potentially take any value from zero to 100, we find they only take one of 17 unique values. The most common values are 100 points and zero points, reflecting the binary nature of many criteria. Figure A.2 in the Internet Appendix reveals time series changes in raw scores over the sample. For each month, we compute means of raw scores for environmental, social, and governance firm-criteria. We plot median splines of these values. Criteria performance scores for all three pillars exhibit upward drift suggesting either a loosening of rater standards or changes in corporate ESG performance over time. These patterns are consistent with Crespi and Migliavacca (2020) and D.E. Shaw & Co. (2022) showing that firms' ESG scores improve over time.

[Insert Figure 3 here.]

Figure 4 shows the number of firms rated according to each criteria and month of the sample. Figures 2 through 4 also show the number of criteria Sustainalytics uses for each pillar over the sample. That is, Sustainalytics deploys 59 (61, 43) criteria when measuring all firms' environmental (social, governance) scores over the entire sample. However, individual firm-months are rated with a subset of these criteria. The mean number of environmental (social, governance) criteria used by Sustainalytics for a firm-month is 19.9 (21.1, 21.6). Table A.I in the Internet Appendix shows an example of the data. It provides firm-criteria-month observations for two consecutive months (September and October, 2011) for an anonymized firm. This example includes two instances where criteria weights and raw score increase contemporaneously from one month to the next. Not all increases in criteria weights are accompanied by increases in raw scores. However, there are no instances where decreases in criteria weights are accompanied by decreases in raw scores. Figure A.3 in the Internet Appendix shows how Sustainalytics distributes weights across criteria through time. Criteria weights sum to 100 percent for firm-months. For an average firm-month, Sustainalytics assigns approximately 36 percent (39 percent, 25 percent) of weight to environmental (social, governance) criteria. These allocations are stable through time.

[Insert Figure 4 here.]

Figure 5 shows correlation matrices of criteria weights. We display them by pillar to ease visualization. It is difficult to draw meaningful inferences from correlation matrices with dozens of rows and columns. Therefore, we present correlation coefficients with combinations of colors and sizes. Large, light-shaded areas indicate correlations approaching 1.00. Large, dark-shaded areas indicate correlations approaching -1.00. Across all three panels, we count a roughly equal number of large, light-shaded areas and large, dark-shaded areas. This comparison indicates that criteria weights tend to have positive correlations as often as they have negative. This feature of the data is noteworthy because criteria weights sum to one hundred percent across environmental, social, and governance pillars for firm-months. Figure A.4 in the Internet Appendix provides correlation matrices of month-over-month changes of criteria weights. The interpretation from all three panels of this figure echoes that of Figure 5. When criteria weights change, they tend to be accompanied by changes in the same and opposite direction among other criteria in the pillar with similar frequency.

[Insert Figure 5 here.]

Table I displays summary statistics for the sample used in the main analysis. It includes nearly 1.8 million firm-criteria-month observations. In addition to statistics related to raw scores and weights, we characterize criteria by whether they reflect preparedness-, disclosure-, or performance-related criteria. *Preparedness* (*Disclosure*, *Performance*) is an indicator variable taking a value of one if the criteria relates to ESG preparedness (disclosure, performance). We merge balance sheet and income statement information from Compustat by firm name. We use data from IBES to count the number of equity analysts following the firm. We measure Compustat and IBES data at the firm-year level. We impute these firm-year measures to firm-criteria-month observations in the same calendar year. We exclude firms with sector-specific regulatory constraints, including banks, insurance companies, and utilities. (Results are not sensitive to this filter.) Our final sample includes 300 unique firms across 38 industries, or "peer groups", as Sustainalytics terms them. Our main multivariate tests include firm-month fixed effects, which negate the need for firm-level controls. However, we report summary statistics on a variety of characteristics to characterize the firms in the sample. We report *Firm size*, the natural log of the firm's total assets; *Dividends*, the firm's total dividends divided by total assets; *Cash*, the firm's cash divided by total assets; *Leverage*, the firm's total debt including current divided by total assets; *ROA*, the firm's net income divided by total assets; *BTM*, the firm's book value per share divided by its price close at the end of the calendar year; *CAPX*, the firm's capital expenditures divided by total revenue; *SG&A*, the firm's selling, general, and administrative expense divided by total revenue; and *R&D*, the firm's research and development expense divided by total revenue. *R&D* takes a value of zero if it is missing. We also include *Analysts*, the natural logarithm of one plus the number of equity analysts covering the firm. Table A.II in the Internet Appendix provides summary statistics for subsamples of firm-criteria-months associated with the environmental, social, and governance pillars.

Panel B of Table I provides summary statistics for measures from additional databases we use later in the paper. These measures include counts of ESG incidents from RepRisk. RepRisk is a news aggregator that provides daily observations of each firm's involvement (or lack thereof) in reputation-harming ESG incidents. Such incidents can relate to environmental, social, or governance pillars. RepRisk refers to incidents that involve more than one pillar as "cross-cutting" incidents. In addition to the raw incident counts, RepRisk provides weighting factors ranging in value from 1 to 3 that increase with the severity, novelty, and reach of incidents. We also include information from the Toxic Releases Inventory (TRI) database provided by the U.S. Environmental Protection Agency (EPA). We measure the amount (in pounds) of toxic chemicals released, recycled, recovered, treated, and transferred in a firm-year. We associate these variables with observations related to the environmental pillar. We use firms' 13f filings to measure the percentage of firm-years' shares that are held by ESG-focused institutional investors. We classify a fund as an ESG investor if its name contains any of a list of key terms.[16] Finally, we merge hand-collected data from FactSet to measure the percentage of a firm's revenue that derives from Europe. We collect the data with a snapshot as of December 2022. We impute these measures onto earlier observations, recognizing the values will increase in staleness for older observations.

[Insert Table I here.]

Although Sustainalytics publishes fresh ESG ratings each month, these observations may not be independent. We examine serial correlation in Table II. Panel A reports correlation

[16] We do not classify funds with the term "equity" in the fund name as "ESG funds" because "equity" in this context refers to common stocks rather than social justice. The key terms include "carbon", "clean", "climate", "CSR", "ESG", "environment", "ethic", "ethical", "green", "justice", "planet", "social", "socially", "sustain", "sustainable", "sustainability", "responsible", "responsibility", "water", "women", and "values".

coefficients for *Raw score* and six of its lags. The first column presents the full sample and subsequent columns partition the sample by ESG pillar. Panel B repeats the analysis for *Weight* and six of its lags. All measures for both panels exhibit significant serial correlation, as all correlation coefficients are economically large and statistically significant at the 1% level.

[Insert Table II here.]

4. Analysis and Results

4.1. Univariate Analysis

We start by investigating the univariate relationship between weights and raw scores for ESG criteria. We identify 13,396 (10,949) firm-criteria-months where weights associated with environmental criteria increase (decrease). Additional increases (decreases) occur in 7,888 (5,827) and 6,425 (6,911) firm-criteria-months associated with social and governance criteria, respectively. If firms care about their ESG ratings, then we should observe improvements (increases) in raw ESG scores after Sustainalytics increases criteria weights. However, the time horizon over which such improvements may obtain is unclear. We display the weights and raw scores in the six months following each criteria weight change event This approach allows time for reactions of raw scores to manifest while being sufficiently narrow to avoid the influence of the raw score drift observed in Figure A.2 in the Internet Appendix. We also display the weights and raw scores in the six months preceding each event to allow for the possibility that increases in raw scores anticipate increases in weights. Panel A of Figure 6 plots mean criteria weights and raw scores around weight-increase events for criteria associated with all three ESG pillars.

[Insert Figure 6 here.]

The results indicate no lag between criteria weight increases and raw score increases, as the two change together in the same month. Figure A.4 in the Internet Appendix repeats this analysis after splitting the data into subsets by pillar. The general result in Figure 5 is strongest among environmental criteria. Taken at face value, these results indicate that firms indeed care about their ESG ratings, in that they improve their raw ESG score on a given criteria as soon as Sustainalytics applies more weight to the criteria. The sensitivity is economically significant. A typical change in criteria weight is associated with an increase in raw score equivalent to about 15% of a typical change.[17]

We repeat this analysis for weight-decrease events in Panel B. We do not observe a sharp change in raw scores around criteria weight decreases. If anything, we observe a gradual increase in raw scores from month -6 to month +6. This gradual increase is not surprising given the positive drift in raw scores shown in Figure A.2 in the Internet Appendix. Comparing Panels A and B in Figure 6 reveals an asymmetry in how raw scores respond to changes in criteria weights. Raw scores are more sensitive to increases in criteria weights than they are to decreases in criteria weights. This asymmetry indicates firms more actively respond to criteria weight increases than weight decreases.

Figure 7 extends the analysis in Figure 6. Instead of examining levels of raw scores in the months around weight changes, this figure plots the probabilities that raw scores increase and decrease each month relative to the month of the criteria weight change. Panel A shows that raw scores exhibit a stable probability of increase and decrease leading up to, and following, weight increases. However, we observe an increase in the probability that raw scores increase in the month prior to weight increases, and this probability peaks during the same month the criteria weight increases. Panel B shows this pattern is not present for probabilities of raw score decreases. The

[17] The figure shows that the average increase in criteria weight is 44 basis points (1.54 percentage points to 1.98 percentage points). This change is accompanied by an increase in raw score of 1.0 points (46.2 to 47.2), which is about 15% of 6.56 points, the average change in raw score (see Table I).

probability that raw scores decrease is more stable through the time series, and we observe no peak in the probability that raw scores increase or decrease in the same month that criteria weights decrease. Together, Figures 6 and 7 show that raw scores have a particular propensity to increase in the same month that criteria weights increase.

[Insert Figure 7 here.]

4.2. Multivariate Analysis

The patterns in Figures 6 and 7 are clear: Criteria raw scores increase in the same month that criteria weights increase, but scores do not fall with offsetting decreases in criteria weights. These effects could be influenced by a variety of confounds. We more rigorously examine these patterns with OLS regressions, using a rich set of fixed effects. Specifically, we estimate the following OLS regression:

$$\text{Raw score}_{i,c(p),t} = \alpha + \beta \text{ Weight}_{i,c(p),t} + \text{Lead and lag of Weight}_{i,c(p),t} +$$
$$\text{Firm}_i \times \text{Criteria}_{c(p)} \text{ FE} + \text{Firm}_i \times \text{Month}_t \text{ FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_t + \varepsilon_{i,c(p),t} \qquad (1)$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month. Table II and Figures 6 and 7 indicate significant serial correlation in criteria weights. We therefore include lead and lag values of weights to control for the possibility that raw scores are sensitive to recent weights in addition to contemporaneous weights. This approach "horse races" the explanatory power of the weights. (For robustness, Table A.III in the Internet Appendix reports results with leads and lags through three and six months. The results are not sensitive to the number of leads and lags.)

The fixed effects force estimation of β to come from variation at the firm-criteria-month level. Firm-criteria fixed effects absorb variation in raw scores associated with firms' or industries' average performance on certain criteria or pillars. Firm-month fixed effects absorb variation associated with certain firms' or industries' time series patterns in overall ESG performance. Criteria-month fixed effects absorb variation associated with certain criteria or pillars over time. For example, to the extent that firms increase board diversity (Sustainalytics code g_2_7) or decrease employee turnover (Sustainalytics code s_1_5) over time, this effect will not influence the estimate of β. We cluster standard errors by peer group, month, and criteria levels. This approach is conservative in that it allows for the possibility that the error term is correlated at each of these levels. Alternative approaches to clustering errors (e.g., at the peer group, firm, month, or criteria levels, individually, or at any interaction of these dimensions) produce smaller standard errors for the estimate of β. Table III reports the results.

[Insert Table III here.]

The results echo the univariate analysis. Column (1) reports results for the full sample. We find that a one-standard deviation increase in *Weight* (1.33%, per Table I) is associated with a 1.20-point increase in *Raw score* (1.33% × 0.90 = 1.20 points) The result is statistically significant and economically significant: a 1.20-point increase is approximately 18.3% of a typical change in *Raw score* (6.56 points, per Table I). Columns (2) through (4) repeat the analysis on subsets of the data after splitting by pillar. The results indicate the overall effect is driven by the environmental and social pillars. The coefficient on β for observations associated with the governance pillar is positive but it is not statistically significant.

Panel B provides and alternative approach by using *Δ Weight* in place of *Weight* as the primary independent variable. By using month-over-month changes in firm-criteria weights, this approach mitigates the need to control for leads and lags of *Weight*. Although marginally weaker, we observe results that are similar to those in Panel A. Panel C takes this approach a step further and uses *Δ Raw* score in place of *Raw score* as the dependent variable. This approach mitigates the need to include multi-dimensional fixed effects as control variables. Column (1) of Panel C shows there is no relation between *Δ Raw score* and *Δ Weight*. This result is not surprising given the evidence in Figures 6 and 7 showing that raw scores exhibit an asymmetric response to changes in criteria weights. Specifically, raw scores increase when criteria weights increase, but they are insensitive when criteria weights decrease. We allow for this characteristic in column (2). Here we include separate variables that capture positive and negative changes in criteria weights. The results echo the univariate evidence in Figures 6 and 7. Raw scores are sensitive to positive changes in criteria weights but show no reaction when criteria weights decrease. As in Panels A and B, results in columns (3) through (5) reveal this overall effect is driven by criteria associated with the environmental pillar.

4.3. Robustness

We examine the robustness of the relation between contemporaneous criteria weights and raw scores along several dimensions. We begin with the time series. We divide the full sample into subsamples based on calendar year and replicate the regression specification in equation (1) on each subsample. Figure 8 plots the coefficient of interest for each subsample. We observe coefficient estimates larger than zero in ten out of 11 sample years. Five of the years' coefficients are statistically significant. This figure shows that the overall effect is not driven by any particular year.

[Insert Figure 8 here.]

Next, we examine robustness in the cross section. We divide the full sample into subsamples based on peer groups and replicate the regression specification in equation (1) on each subsample. Figure 9 plots a histogram of the coefficients of interest. We require peer groups to have at least 25,000 observations to prevent peer groups with minimal presence in the data from exerting disproportionate influence on the shape of the distribution. Coefficient estimates around zero are the most common. However, the distribution is right-skewed, and its mean is 1.82. This analysis shows that the coefficients are generally positive across sectors and the overall effect is not driven by any particular industry.

[Insert Figure 9 here.]

5. Mechanism

The contemporaneous relationship between criteria weights and raw scores is surprising. A priori, one would expect a lag between when Sustainalytics places greater emphasis on certain criteria and when firms respond by improving performance along the criteria – assuming firms care to respond at all. In this section, we examine possible explanations for the main result.

5.1. Does Firm Behavior Change?

We test whether firms' ESG behavior legitimately and nimbly adjusts within the same month that criteria weights change. We approach the analysis by incorporating data from RepRisk. RepRisk is a news aggregator that records firms' reputation-harming ESG incidents. We return to the set of events in Figures 6 and 7 where Sustainalytics increases and decreases weights associated with criteria. We compute the probability that firms experience reputation-harming incidents associated with environmental, social, and governance behavior. We plot these values around weight change events in Figure 10. We also combine all criteria weight change events and plot the probability that firms experience "cross-cutting" incidents in the months surrounding them. "Cross-cutting" incidents are those RepRisk classifies as being related to firm behavior in at least two of the environmental, social, and governance categories.

[Insert Figure 10 here.]

If firms legitimately adjust their ESG behavior in the same month that criteria weights change, then these adjustments should be reflected in a contemporaneous decrease in the likelihood that firms find themselves embroiled in reputation-harming incidents. However, the results reveal no discernable change in incidence rates around criteria weights, as probabilities remain stable through time. This lack of evidence casts doubt on the possibility that firms change their behavior.

We more rigorously examine the relation between criteria weights and firms' ESG behavior by extending the analysis to a multivariate setting. We proceed in two stages. In the first stage, we model raw scores with criteria weights, alone. Specifically, we estimate the following OLS regression:

$$\text{Raw score}_{i,c(p),t(y)} = \alpha + \beta\ \text{Weight}_{i,c(p),t(y)} + \varepsilon_{i,c(p),t(y)} \tag{2}$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month (and indexes year, *y*, for control variables in the second-stage regression). Following this regression, we capture the values of *Raw score* predicted by criteria weights, as well as the residual values of *Raw score*. The predicted values estimate the portion of ESG performance scores that are related to ESG rater focus (criteria weights). The residual values represent everything else. We then estimate the following second-stage OLS regression:

$$\begin{aligned}\text{ESG incidents}_{i,c(p),t(y)} &= \alpha + \beta_1\ \text{Predicted Raw score}_{i,c(p),t(y)} + \beta_2\ \text{Residual Raw score}_{i,c(p),t(y)} \\ &+ \text{Firm-year controls}_{i,y} + \text{Firm}_i \times \text{Criteria}_{c(p)}\ \text{FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_{t(y)} + \varepsilon_{i,c(p),t(y)}\end{aligned} \tag{3}$$

The dependent variable in Panels A (B, C) is *Environmental* (*Social*, *Governance*) *ESG incidents*, the number of times firms experience ESG incidents over the next 12 months related to environmental (social, governance) criteria. The dependent variable in Panel D is *Cross-cutting ESG incidents*, the number of times firms experience incidents over the next 12 months related to at least two pillars. The key independent variable, *Predicted Raw score* is the predicted value of

Raw score from the first stage model and allows us to test whether weight-driven raw scores are associated with firms' ESG behavior. The other dependent variable, *Residual Raw score* is the residual value of *Raw score* from the first stage model and allows us to test whether the variation in raw scores that is not driven by criteria is associated with firms' ESG behavior.

We report the results in Table IV. Dependent variables in columns (2), (3), and (4) weight (i.e., multiply) ESG incidents by the severity, novelty, and reach of the incidents, respectively. RepRisk provides the measures of severity, novelty, and reach and they range in value from 1 to 3. We standardize ESG incident measures, *Predicted Raw score*, and *Residual Raw score* to follow mean-zero, unit-variance distributions. This approach improves the ease of interpretation of regression coefficients. We include firm-year controls and suppress their coefficients to conserve space.

[Insert Table IV here.]

The results reveal no relation between weight-driven changes in raw scores and firms' ESG behavior. This non-result obtains across all three pillars, and among incidents that cut across pillars. Further, we observe no results after weighting incidents by their severity, novelty, or reach. We infer that firms do not materially change their operations in response to Sustainalytics' changes in priorities (evidenced by changes in criteria weights).

The only coefficients in Table IV that attain significance are those on *Residual Raw score* for cross-cutting incidents. However, because these coefficients are positive, they are opposite in sign to what one would expect if Sustainalytics' measures capture firm behavior. (Higher raw scores should predict less participation in reputation-harming incidents, not more.) By separating ESG ratings into weight-driven and residual components, these tests complement the approach of Serafeim and Yoon (2022). These authors find that consensus ratings across ESG raters predict

ESG news, but individual raters are not necessarily predictive, particularly when there is significant disagreement among raters.

Next, we take an alternative approach to testing whether firms make real changes to operations in response to changes in ESG rater priorities. We incorporate the Toxic Releases Inventory (TRI) database from the U.S. Environmental Protection Agency (EPA). We focus on criteria associated with the environmental pillar for this analysis because the release of toxic chemicals pertains to environmental behavior and is it not obviously related to social or governance concerns. The TRI data include a variety of measures that capture the propensity for firms to release (both onsite and offsite), recycle, recover, treat, or transfer toxic chemicals. These data are common in climate finance research. See, for example, Duchin, Gao, and Xu (2023), who study whether firms' divestment of polluting plants meaningfully affects pollution levels. These authors find that divesting polluting plants allows firms to "greenwash" without affecting change in pollution levels. As in the analysis with RepRisk data, we proceed in two stages. The first stage estimates the following OLS specification:

$$\text{Raw score}_{i,c,t(y)} = \alpha + \beta \text{ Weight}_{i,c,t(y)} + \varepsilon_{i,c,t(y)} \tag{4}$$

where *i* denotes firm, *c* denotes environmental criteria, and *t* denotes month (and indexes year, *y*, for control variables in the second-stage regression). We capture the values of *Raw score* predicted by criteria weights, as well as the residual values of *Raw score*. The predicted values estimate the portion of environmental criteria raw scores that are related to criteria weights. The residual values represent everything else. We then estimate the following second-stage OLS regression:

$$\begin{aligned}\text{Toxic releases}_{i,c,t(y)} &= \alpha + \beta_1 \text{ Predicted Raw score}_{i,c,t(y)} + \beta_2 \text{ Residual Raw score}_{i,c,t(y)} \\ &+ \text{Firm-year controls}_{i,y} + \text{Firm}_i \times \text{Criteria}_c \text{ FE} + \text{Criteria}_c \times \text{Month FE}_{t(y)} + \varepsilon_{i,c,t(y)} \quad (5)\end{aligned}$$

The dependent variables are firm-year measures of the quantity of toxic chemicals firms release, recycle, recover, treat, or transfer during the same year, next year, and two years later. We measure these variables in the future to allow time for firms to change their operations resulting in lower levels of toxic releases. The key dependent variable, *Predicted Raw score* is the predicted value of *Raw score* from the first stage model and allows us to test whether criteria-driven changes in raw scores are associated with changes in firms' toxic releases. The other dependent variable, *Residual Raw score* is the residual value of *Raw score* from the first stage model and allows us to test whether the variation in raw scores that is not driven by criteria changes is associated with changes in firms' toxic releases.

We standardize toxic release measures, *Predicted Raw score*, and *Residual Raw score* to follow mean-zero, unit-variance distributions. This approach improves the ease of interpretation of regression coefficients. We use firm-year controls instead of firm-month fixed effects because the dependent variables are measured at the firm-year level. We suppress coefficients on firm-year controls to conserve space. We report the results in Table V.

[Insert Table V here.]

Panel A reports results for all firm-criteria-month observations associated with the environmental pillar. Panel B repeats this analysis after restricting the sample to performance-related criteria (discarding preparedness- and disclosure-related criteria). We implement this restriction because toxic chemical releases are directly related to environmental performance. Panel A shows no meaningful relationship between either predicted or residual raw scores for environmental criteria and the amount of toxic chemicals firms release, recycle, recover, treat, or transfer, over a time horizon of up to two years. These results cast doubt on the possibility that firms materially change practices in response to changes in ESG rating standards.

We observe similar non-results in Panel B, with the exceptions of offsite chemical treatments and transfers two years later. These coefficients on *Predicted Raw score* are negative and statistically significant. These results suggest that changes in criteria weights drive firms to materially change offsite chemical handling. However, the magnitude of these coefficients is not economically large. A one-standard deviation increase in *Predicted Raw score* is associated with a 0.0030- and 0.0051-standard deviation reduction in offsite chemical treatments and transfers, respectively. Overall, Figure 10, Table IV, and Table V provide scant evidence that changes in ESG criteria weights indicate meaningful changes in firms' real ESG behavior, either contemporaneously or in the future.

5.2. Which Way Does Causality Run?

An assumption underlying the analysis so far is that any results reflect choices made by firms. That is, when Sustainalytics changes the weights it uses for various criteria, firms may choose to adjust their ESG behavior. However, it is possible that causality runs the other direction. Sustainalytics may choose to adjust criteria weights in response to firms' performance along certain ESG criteria. Under this paradigm, Sustainalytics caters to firms, not the other way around, similar to how credit rating agencies cater to rated firms.[18] However, the results in Table III do not suggest such catering. Such a paradigm predicts increases in raw scores will *lead* increases in criteria weights. The results in Table III indicate the relation between criteria weights and raw scores is strongest in the contemporaneous month. Still, we address this possibility with additional analysis.

We examine instances in the data when Sustainalytics ceases to deploy certain criteria. If Sustainalytics caters its rating standards to firms' ESG behavior, then it should stop using criteria

[18] See, for example, Griffin and Tang (2012), Griffin, Nickerson, and Tang (2013), Bruno, Cornaggia, and Cornaggia (2016), and Cornaggia, Cornaggia, and Israelsen (2023).

under which firms generally perform poorly. We identify 6,054 firm-criteria-months, exclusive of the final month of the sample, when Sustainalytics no longer uses a particular criterion for any firm(s). We compute the average raw score for the particular criterion in the six months leading up to criteria terminations and plot the results in Panel A of Figure 11. We find no downward trend in raw scores leading up to criteria terminations. If anything, raw scores exhibit a mild increase prior to terminations. This result casts doubt on the possibility that Sustainalytics typically caters to firms. For completeness, we perform a similar analysis in Panel B with instances where Sustainalytics introduces criteria. We identify 7,570 firm-criteria-months, exclusive of the first month of the sample, when new criteria appear in the data. Comparing Panels A and B, the mean score among terminated criteria in the month of termination (61.5 points) is significantly larger than the mean score among new criteria in the month the criteria is first deployed (47.6 points). This comparison again indicates that Sustainalytics does not generally cater to firms by initiating criteria under which firms perform well, nor removing criteria under which firms perform poorly.

[Insert Figure 11 here.]

5.3. Do Firms Manage ESG Ratings?

We examine next the possibility that firms influence Sustainalytics by claiming good news about their ESG performance during the rating process. Sustainalytics (2017) indicates that the final step in its rating process is to solicit feedback from rated firms prior to publishing ratings. Figure 1 shows this step in an excerpt from a report produced by Sustainalytics that details its ratings process. Through this interaction, firms may have the opportunity to claim favorable performance on criteria for which Sustainalytics boosts weights. We refer to this behavior as "ratings management" hereafter. A recent article by Simon MacMahon, head of ESG research at Sustainalytics, suggests firms try to influence Sustainalytics. He writes:

"*Once we have completed our ratings process, we send the profile to the company for feedback. During those conversations, we're looking for any additional information or clarification that can enhance our analysis. New information doesn't always lead to a change in our rating, but we do listen. As ESG rating outcomes become more important, we certainly hear from people inside firms who forcefully argue for their point of view.*" (MacMahon, 2022)

If the ratings management hypothesis is correct – if firms signal to Sustainalytics that performance under certain criteria has improved when the criteria weight increases – then this signal should be more credible for criteria that firms can easily adjust. We develop a proxy for ease of adjustment. First, we sort the main sample according to criteria type. Sustainalytics intends its criteria to measure "preparedness", "disclosure", or "performance". Table A.IV in the Internet Appendix shows how we classify all 163 criteria. For example, criteria e_1_1, "Formal Environmental Policy", belongs to the "preparedness" category. Criteria e_1_5, "Participation in Carbon Disclosure Project", belongs to the "disclosure" category. Criteria e_1_9, "Carbon Intensity" belongs to the "performance" category. We make similar categorizations for all criteria, including those underlying social and governance pillars.

We conjecture that the preparedness category is easiest to adjust. Many criteria involve the creation of a policy that could be crafted on short notice. For example, criteria s_4_2_1 is "Human Rights Policy". Criteria g_1_4_1 is "Policy on Money Laundering". Presumably, a firm could react quickly and draft such policies after learning that Sustainalytics has applied greater weight to the criteria. Criteria related to disclosure could also be adjusted easily, although the creation of data reporting systems could be onerous. For example, criteria s_2_2_2 is "Reporting on Supply Chain Monitoring and Enforcement". Improving performance along this dimension likely requires building information technology infrastructure. Criteria related to performance, e.g., criteria s_1_5, "Employee Turnover Rate", seem difficult to adjust over a short time horizon. We replicate our main regression specification in equation (1) on subsets of firm-criteria-month observations split by criteria “preparedness”, “disclosure”, or “performance”. Figure 11 displays the coefficient on *Weight* for each subsample. Consistent with the ESG ratings management hypothesis, the “preparedness” category has the largest and most significant coefficient.

[Insert Figure 12 here.]

5.4. The Role of ESG Investors and Customers

If firms manage their ESG ratings, then they likely do so to satisfy monitoring stakeholders. We test this possibility in the cross section after sorting firms by the extent to which their shares are held by ESG-focused investors. We collect 13f filings for firms to learn the identities of funds that hold firms’ equities each firm-year. We classify a fund as an ESG investor if its name contains any of a list of the key terms described in the data section. We then sort all firm-criteria-month observations into quintiles according to the share of the firm-month’s equity that is held by ESG funds. We replicate our baseline specification in equation (1) on each subsample. Figure 13 reports the coefficient on *Weight* for each subsample. We observe generally larger coefficients for subsamples of firm-years with above-median ESG fund holdings, as the largest point estimate obtains in quintile four. This pattern, while not stark, supports the idea that firms manage their ESG ratings for the benefit of ESG investors.

[Insert Figure 13 here.]

In addition to investors, firms rationally manage their ESG ratings to attract customers. For example, a recent survey by PwC found that 83% of consumers believe companies should actively shape ESG best practices.[19] Interest in firms’ ESG behavior is especially prevalent in Europe, where regulators, investors, and consumers have long preceded the United States in pursuing ESG

[19] “How much does the public care about ESG?” PwC. URL accessed July 12, 2023: https://www.pwc.com/gx/en/services/sustainability/publications/cop26/how-much-does-the-public-care-about-esg-pwc-cop26.html

objectives. Based on these patterns, we conjecture that firms with more customers in Europe (as opposed to the United States or other continents) will face greater pressure to manage ESG ratings. We test this conjecture by collecting information on the geographic distribution of firms' revenue generation. The data are from FactSet and we collect them with a snapshot as of December 2022. We sort all firm-criteria-month observations into quintiles according to the share of the firm's revenue derived from Europe. We replicate our baseline specification in equation (1) on each subsample. If firms manage their ESG ratings to appease ESG-conscious customers, then the relationship between contemporaneous criteria weights and raw scores should be more pronounced among firms that derive more revenue from European customers. Figure 14 reports the coefficient on *Weight* for each subsample. We observe generally larger coefficients for subsamples of firms with above-median revenue generation from Europe, as the largest point estimates obtain in quintiles four and five. This pattern supports the idea that firms manage their ESG ratings to appeal to ESG-conscious customers. We cannot rule out that higher European revenue generation results in additional European ESG regulation. We note only that such regulatory pressure presents an equally rational catalyst for firms to manage their ESG ratings.

[Insert Figure 14 here.]

6. Conclusion

This paper tests whether, and over what time horizon, firms respond to changes in ESG ratings criteria. We use data from an ESG rater that incorporates feedback from firms during the rating process and produces ratings at a monthly frequency. We indeed find that when the rater changes the weight it applies to certain criteria in the creation of its ESG ratings, firms respond by adjusting their reported ESG behavior in the same month. These adjustments appear to be the result of cosmetic, or paperwork-driven changes rather than material changes to firms' operations.

Bloomberg Businessweek recently reviewed a sample of 155 ESG ratings upgrades by MSCI, another leading ESG rater. It concluded: "As many as half of the companies Businessweek analyzed got upgrades for doing nothing but surfing the wave of methodology changes, reweightings, or similar tweaks."[20] Likewise, we do not observe real changes in the likelihood that firms are embroiled in ESG controversies, or that they reduce their release of toxic chemicals because of these adjustments. Rather, it appears firms "manage" their ESG ratings for the benefit of ESG-conscious investors and customers.

[20] "The ESG Mirage" Bloomberg. URL accessed July 10, 2023: https://www.onepak.com/the-esg-mirage/

References

Adelino, M., Cunha, I., and Ferriera, M.A., 2017, "The economic effects of public financing: Evidence from municipal bond ratings recalibration", *Review of Financial Studies* 30 (9), 3223-3268.

Almeida, H., Cunha, I., Ferreira, M.A., and Restrepo, F., 2017, "The real effects of credit ratings: Using sovereign downgrades as a natural experiment", *Journal of Finance* 72, 249-290.

Aslan, A., Poppe, L., and Posch, P., 2021, "Are Sustainable Companies More Likely to Default? Evidence from the Dynamics between Credit and ESG Ratings", *Sustainability* 13 (15), 8568.

Atz, U., Holt, T., Liu, Z., and Bruno, C., 2022, "Does Sustainability Generate Better Financial Performance? Review, Meta-analysis, and Propositions", *Journal of Sustainable Finance and Investment*, forthcoming.

Avetisyan, E., and Hockerts, K., 2016, "The Consolidation of the ESG Rating Industry as an Enactment of Institutional Retrogression", *Business Strategy and the Environment* 26 (3), 316-330.

Avramov, D., Cheng, S., Lioui, A., and Tarelli, A., 2022, "Sustainable investing with ESG rating uncertainty", *Journal of Financial Economics* 145, 642-664.

Azarmsa, E., and Shapiro, J. D., 2023, "The Market for ESG Ratings", working paper.

Berg, F., Heeb, F., and Kölbel, J. F., 2023, "The Economic Impact of ESG Ratings", working paper.

Berg, F., Kölbel, J., Pavlova, A., and Rigobon, R., 2022, "ESG Confusion and Stock Returns: Tackling the Problem of Noise", working paper.

Berg, F., Kölbel, J., and Rigobon, R., 2022, "Aggregate Confusion: The Divergence of ESG Ratings", *Review of Finance* 26, 1315-1344.

Berg, F., Fabisik, K., and Sautner, Z., 2021, "Is History Repeating Itself? The (Un)Predictable Past of ESG Ratings", working paper.

Billio, M., Costola, M., Hristova, I., Latino, C., and Pelizzon, L., 2021, "Inside the ESG ratings: (Dis)agreement and performance", *Corporate Social Responsibility and Environmental Management* 28 (5), 1426-1445.

Bonsall, S., 2014, "The impact of issuer-pay on corporate bond rating properties: Evidence from Moody's and S&P's initial adoptions", *Journal of Accounting and Economics* 57, 89-109.

Boot, A., Milbourn, T., and Schmeits, A., 2006, "Credit Ratings as Coordination Mechanisms", *Review of Financial Studies* 19, 81-118.

Brandon, R. G., Krueger, P., Schmidt, P. S., 2021, "ESG Rating Disagreement and Stock Returns", *Financial Analysts Journal* 77, 104-127.

Bruno, V., Cornaggia, J., and Cornaggia, K., 2016, "Does regulatory certification affect the information content of credit ratings?", *Management Science* 62, 1578-1597.

Charlin, V., Cifuentes, A., and Alfaro, J., 2022, "ESG ratings: an industry in need of a major overhaul", *Journal of Sustainable Finance & Investment*, forthcoming.

Chatterji, A., Durand, R., Levine, D., and Touboul, S., 2016, "Do ratings of firms converge? Implications for managers, investors and strategy researchers", *Strategic Management Journal* 37, 1597-1614.

Chen, Z., Lookman, A.A., Schurhoff, N., and Seppi, D.J., 2014, "Rating-Based Investment Practices and Bond Market Segmentation", *Review of Asset Pricing Studies* 4 (2), 162-205.

Christensen, D., Serafeim, G., and Sikochi, A., 2022, "Why Is Corporate Virtue in the Eye of the Beholder? The Case of ESG Ratings", *The Accounting Review* 97, 147-175.

Clementino, E., and Perkins, R., 2021, "How Do Companies Respond to Environmental, Social, and Governance (ESG) ratings? Evidence from Italy", *Journal of Business Ethics* 171, 379-397.

Cornell, Bradford, and Damodaran, A., 2020, "Valuing ESG: Doing Good or Sounding Good?" working paper.

Cornaggia, J., and Cornaggia, K., 2013, "Estimating the costs of issuer-paid credit ratings", *Review of Financial Studies* 26, 2229-2269.

Cornaggia, J., Cornaggia, K., and Hund, J., 2017, "Credit ratings across asset classes: A long-term perspective", *Review of Finance* 21 (2), 465-509.

Cornaggia, J., Cornaggia, K., and Israelsen, R., 2018, "Credit ratings and the cost of municipal financing", *Review of Financial Studies* 31, 2038-2079.

Cornaggia, J., Cornaggia, K., and Israelsen, R., 2023, "Rating Agency Fees: Pay to Play in Public Finance?", *Review of Financial Studies* 36 (5), 2004-2045.

Crespi, F., and Migliavacca, M., 2020, "The Determinants of ESG Rating in the Financial Industry: The Same Old Story or a Different Tale?", *Sustainability* 12 (16), 6398.

Shaw, D.E. & Co., 2022, "Keep the Change: Analyzing the Increase in ESG Ratings for U.S. Equities", April 2022.

Dimson, E., Marsh, P., and Staunton, M., 2020, "Divergent ESG Ratings", *The Journal of Portfolio Management* 47 (1), 75-87.

Dorfleitner, G., Halbritter, G., and Nguyen, M., 2015, "Measuring the level and risk of corporate responsibility – An empirical comparison of different ESG rating approaches", *Journal of Asset Management* 16, 450-466.

Duchin, R., Gao, J., and Xu, Q., 2023, "Sustainability or Greenwashing: Evidence from the Asset Market for Industrial Pollution", working paper.

Dumrose, M., Rink, S., and Eckert, J., 2022, "Disaggregating confusion? The EU Taxonomy and its relation to ESG rating", *Finance Research Letters* 48, 102928.

Edmans, A., 2023, "The end of ESG", *Financial Management* 52 (1), 3-17

Escrig-Olmedo, E., Fernández-Izquierdo, M. A., Ferrero-Ferrero, I., Rivera-Lirio, J. M., and Muñoz-Torres, M. J., 2019, "Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles", *Sustainability* 11 (3), 915.

Escrig-Olmedo, E., Muñoz-Torres, M., and Fernández-Izquierdo, M. A., 2010, "Socially responsible investing: sustainability indices, ESG rating and information provider agencies", *International Journal of Sustainable Economy* 2 (4), 442-461.

Feng, J., Goodell, J. W., and Shen, D., 2022, "ESG rating and stock price crash risk: Evidence from China", *Finance Research Letters* 46 Part B, 102476.

Friedman, T, 1970, "The Social Responsibility of Business Is to Increase Its Profits" Published in: The New York Times Magazine, September 13, 1970.

Griffin, J., and Nickerson, J., 2017, "Debt correlations in the wake of the financial crisis: What are appropriate default correlations for structured products?" *Journal of Financial Economics* 125, 454-474.

Griffin, J., Nickerson, J., and Tang, D., 2013, "Rating shopping or catering? An examination of the response to competitive pressure for CDO credit ratings", *Review of Financial Studies* 26, 2270-2310.

Griffin, J., and Tang, D., 2012, "Did subjectivity play a role in CDO credit ratings?", *Journal of Finance* 67, 1293-1328.

Gyönyörová, L., Stachoň, M., and Stašek, D., 2021, "ESG ratings: relevant information or misleading clue? Evidence from S&P Global 1200", *Journal of Sustainable Finance & Investment* 13 (2), 1075-1109.

Ilhan, E., Krueger, P., Sautner, Z., and Starks, L., 2023, "Climate Risk Disclosure and Institutional Investors", *Review of Financial Studies* 36, 2617-2650.

Jiang, J., Stanford, M., and Xie, Y., 2012, "Does it matter who pays for bond ratings? Historical evidence", *Journal of Financial Economics* 105, 607-621.

Kedia, S., Rajgopal, S., and Zhou, X., 2014, "Did going public impair Moody's credit ratings?", *Journal of Financial Economics* 114, 293-315.

Kisgen, D., 2019, "The Impact of Credit Ratings on Corporate Behavior: Evidence from Moody's Adjustments", *Journal of Corporate Finance* 58 (C), 567-582.

Krueger, P., Sautner, Z., and Starks, L., 2020, "The importance of climate risks for institutional investors", *Review of Financial Studies* 33 (3), 1067-1111.

Krueger, P., Sautner, Z., Tang, D., and Zhong, R., 2022, "The Effects of Mandatory ESG Disclosure around the World", working paper.

Landi, G., and Sciarelli, M., 2019, "Towards a more ethical market: the impact of ESG rating on corporate financial performance", *Social Responsibility Journal* 15 (1), 1747-1117.

Larcker, D. F., Pomorski, L., Tayan, B., and Watts, E. M., 2022, "ESG Ratings: A Compass Without Direction", *Stanford Closer Look Series*. Corporate Governance Research Initiative, August 2, 2022.

Liu, Min, 2022, "Quantitative ESG disclosure and divergence of ESG ratings", *Frontiers in Psychology* 13, 10.3389/fpsyg.2022.936798.

MacMahon, S., 2020, "The Challenge of Rating ESG Performance", *Harvard Business Review* September-October 2020.

Rzeźnik, A., Hanley, K. W., and Pelizzon, L., 2022, "Investor Reliance on ESG Ratings and Stock Price Performance" working paper.

Serafeim, G., and Yoon, A., 2022, "Stock price reactions to ESG news: the role of ESG ratings disagreement", *Review of Accounting Studies* forthcoming.

Shanaev, S., and Ghimire, B., 2022, "When ESG meets AAA: The effect of ESG rating changes on stock returns", *Finance Research Letters* 46 Part A, 102302.

Starks, L.T., Venkat, P., and Zhu, Q., 2023, "Corporate ESG Profiles and Investor Horizons", working paper.

Sustainalytics, 2017, "Sustainalytics' ESG Rating Research Methodology".

Sustainalytics, 2021, "ESG Risk Ratings – Methodology Abstract".

Tang, D., Yan, J., and Yao, C., 2022, "The Determinants of ESG Ratings: Rater Ownership Matters", working paper.

Xia, H., 2014, "Can investor-paid credit rating agencies improve the information quality of issuer-paid rating agencies?" *Journal of Financial Economics* 111, 450-468.

Zumente, I., and Lāce, N., 2021, "ESG Rating—Necessity for the Investor or the Company?", *Sustainability* 13 (16), 8940.

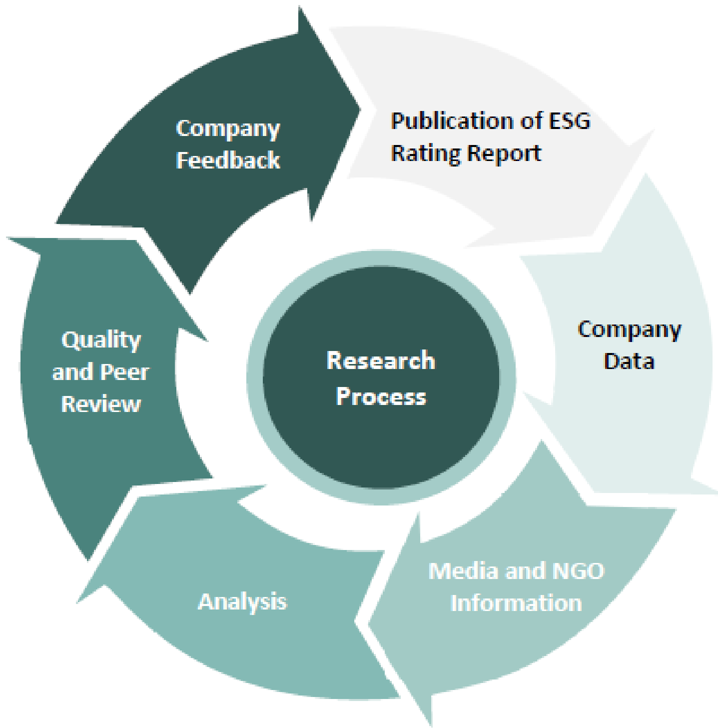
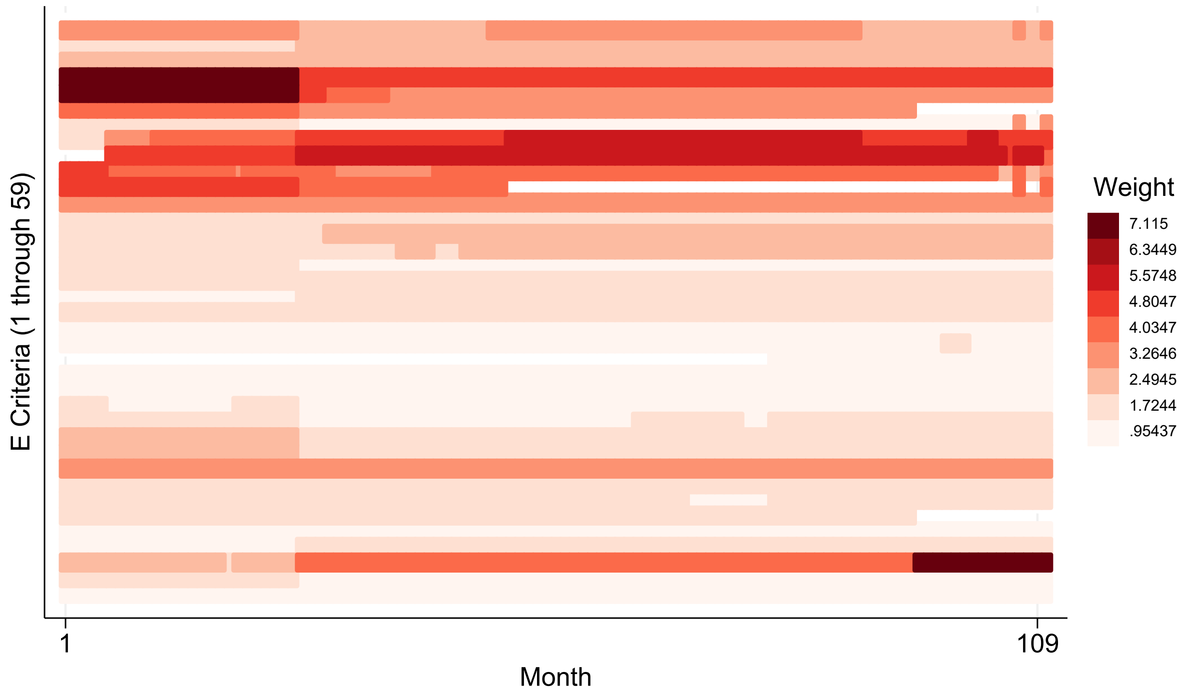
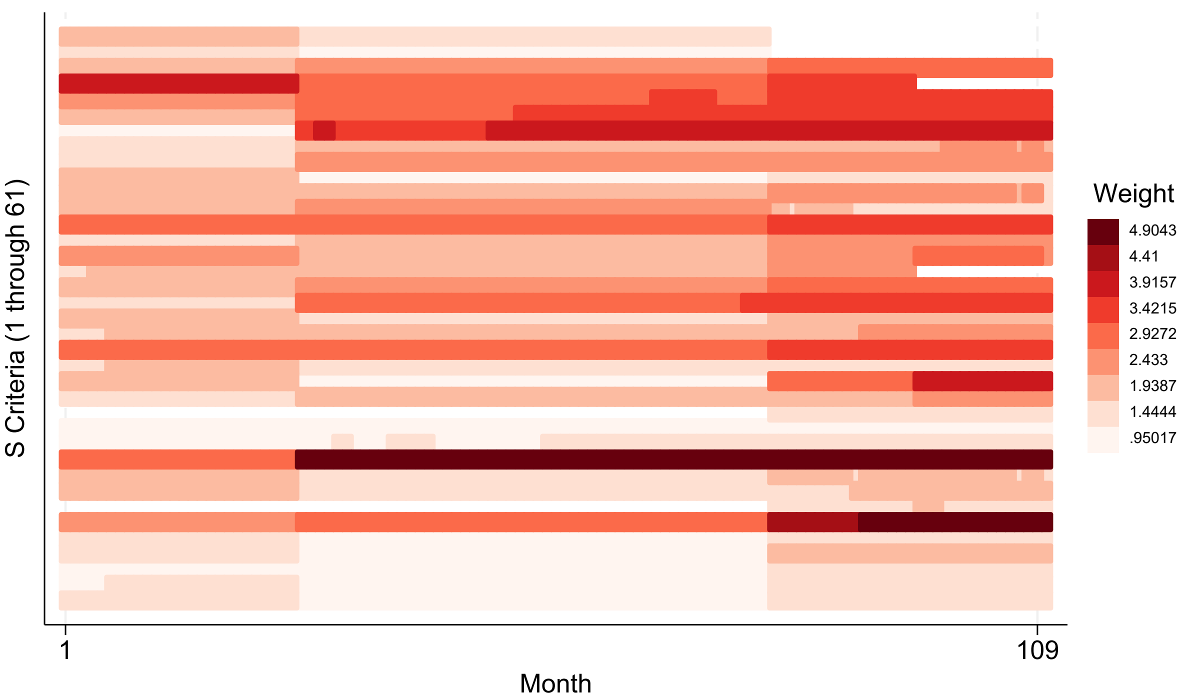


Figure 4. Sustainalytics Research Process

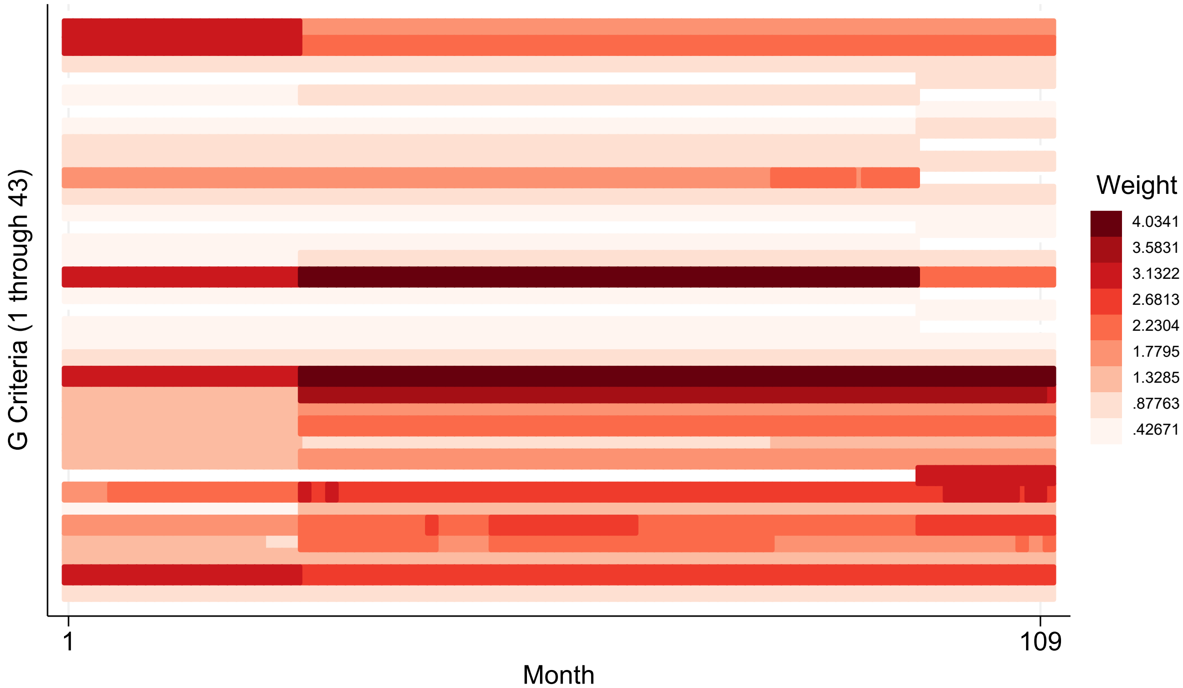
Figure 1. Sustainalytics' ESG rating process. This figure is reproduced from "Sustainalytics' ESG Rating Research Methodology", Sustainalytics (2017). It shows the steps Sustainalytics uses when producing ESG ratings.



Panel A – Environmental criteria weights

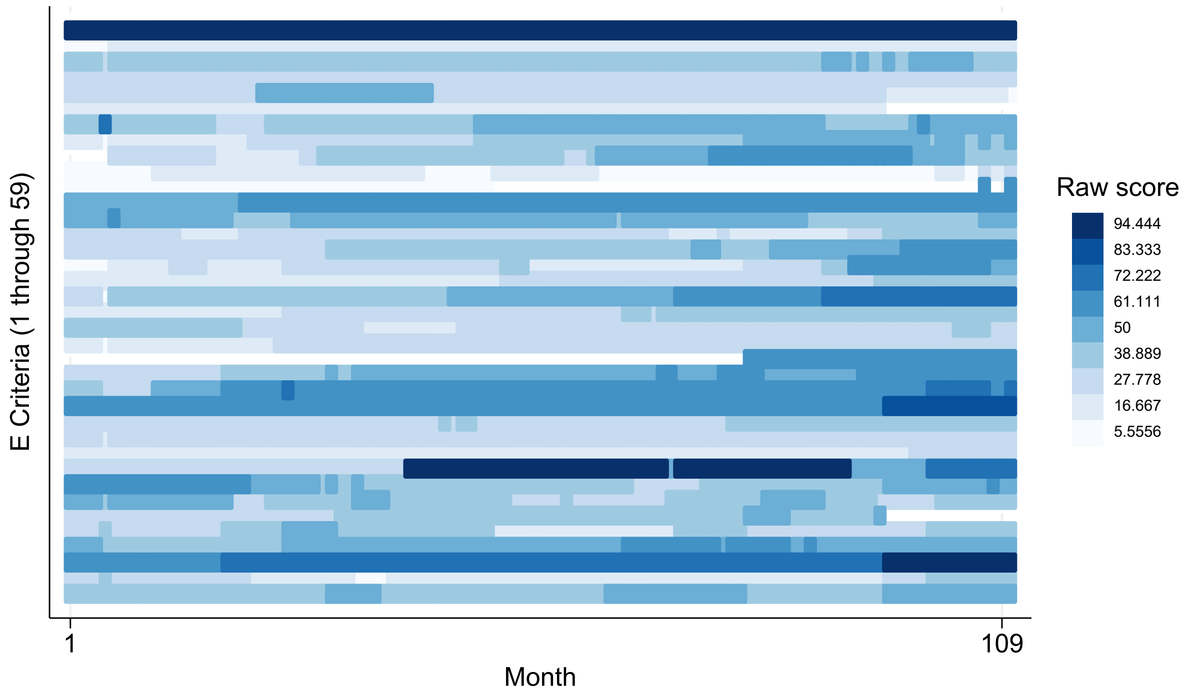


Panel B – Social criteria weights

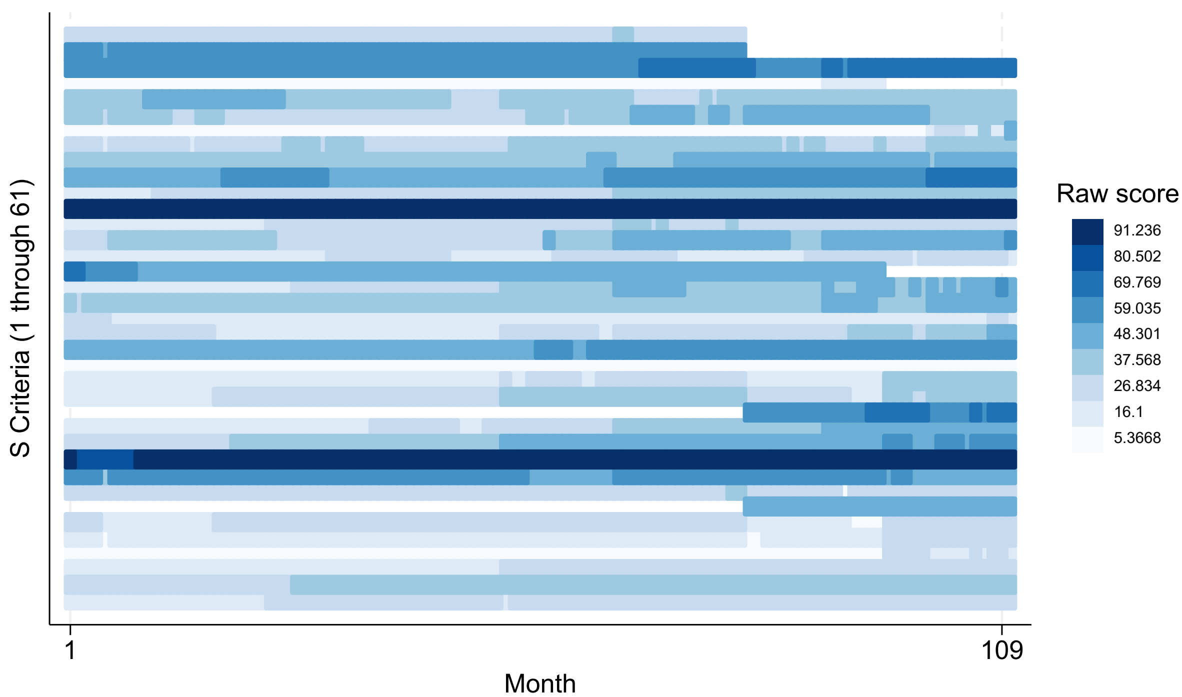


Panel C – Governance criteria weights

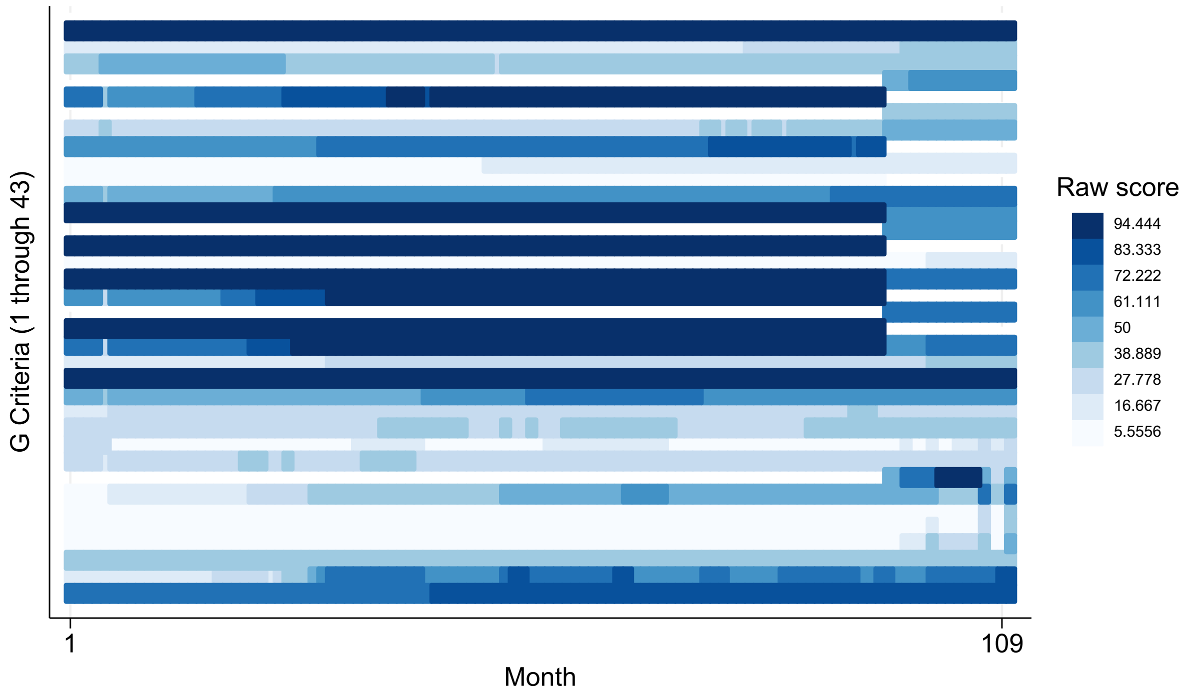
Figure 2. Mean criteria weights through time. This figure displays mean weights for each criteria and month of the sample. The sample runs from August 2009 through September 2019. Panel A (B, C) displays mean weights for criteria underlying the environmental (social, governance) pillar. Data are from Sustainalytics.



Panel A – Environmental criteria raw score



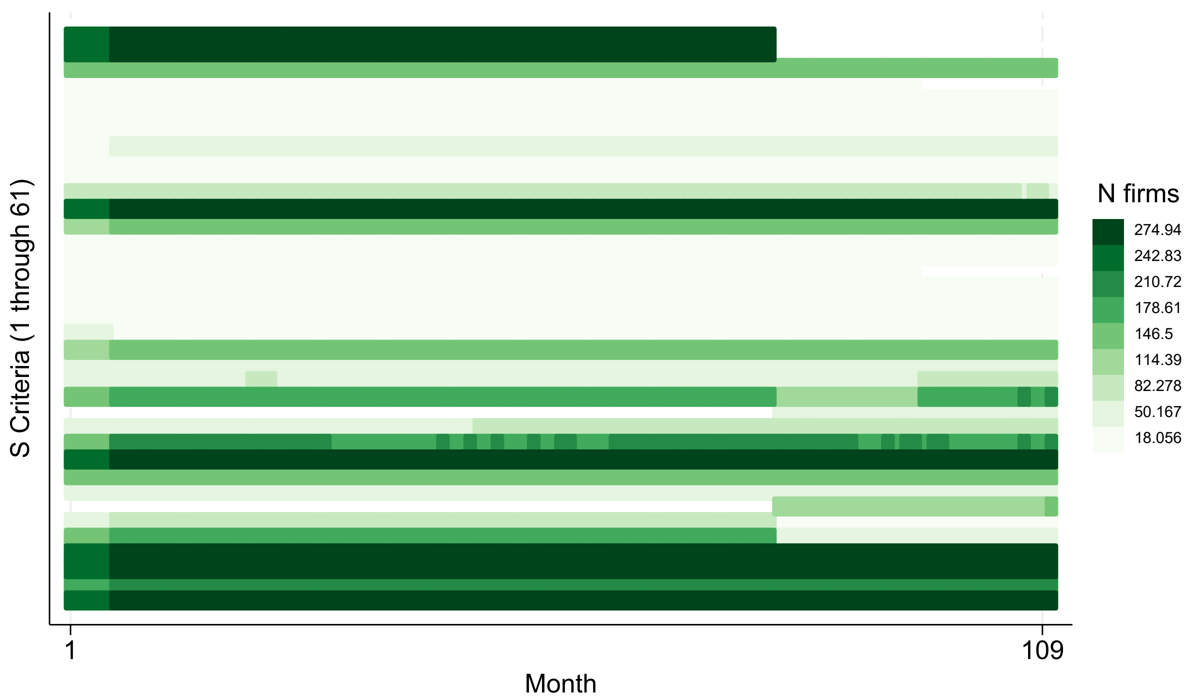
Panel B – Social criteria raw score



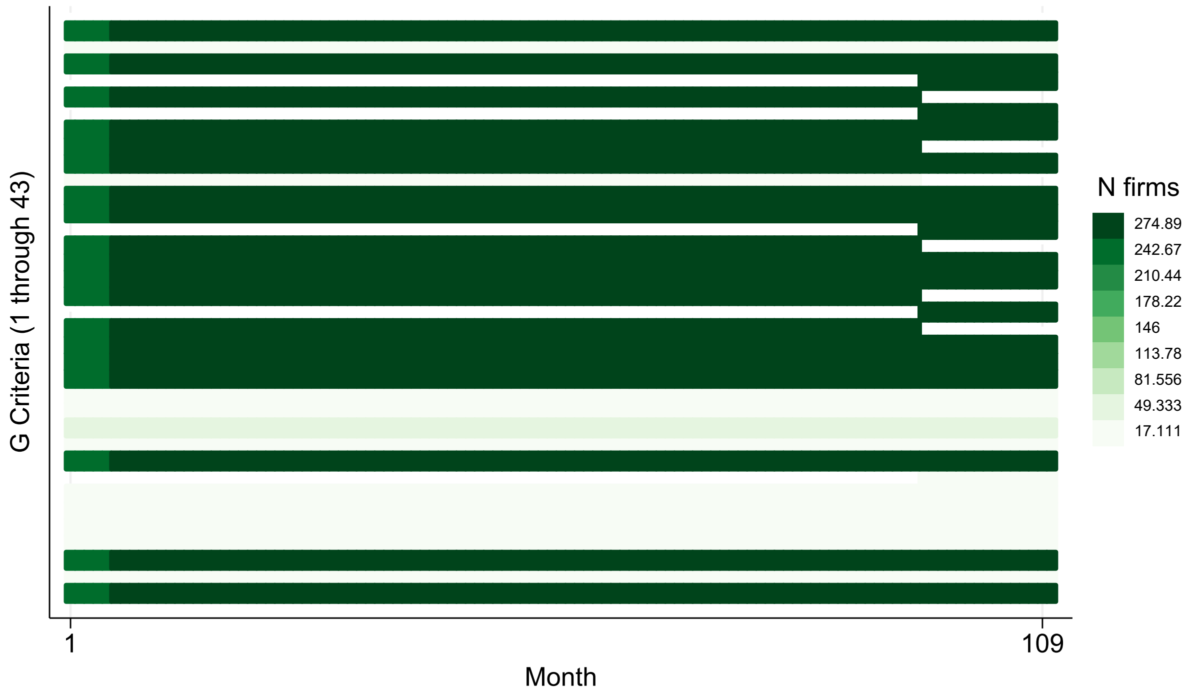
Panel C – Governance criteria raw score

Figure 3. Mean criteria raw scores through time. This figure displays the mean raw score for each criteria and month of the sample. The sample runs from August 2009 through September 2019. Panel A (B, C) displays mean criteria scores underlying the environmental (social, governance) pillar. Data are from Sustainalytics.

Panel A – Environmental criteria N rated firms

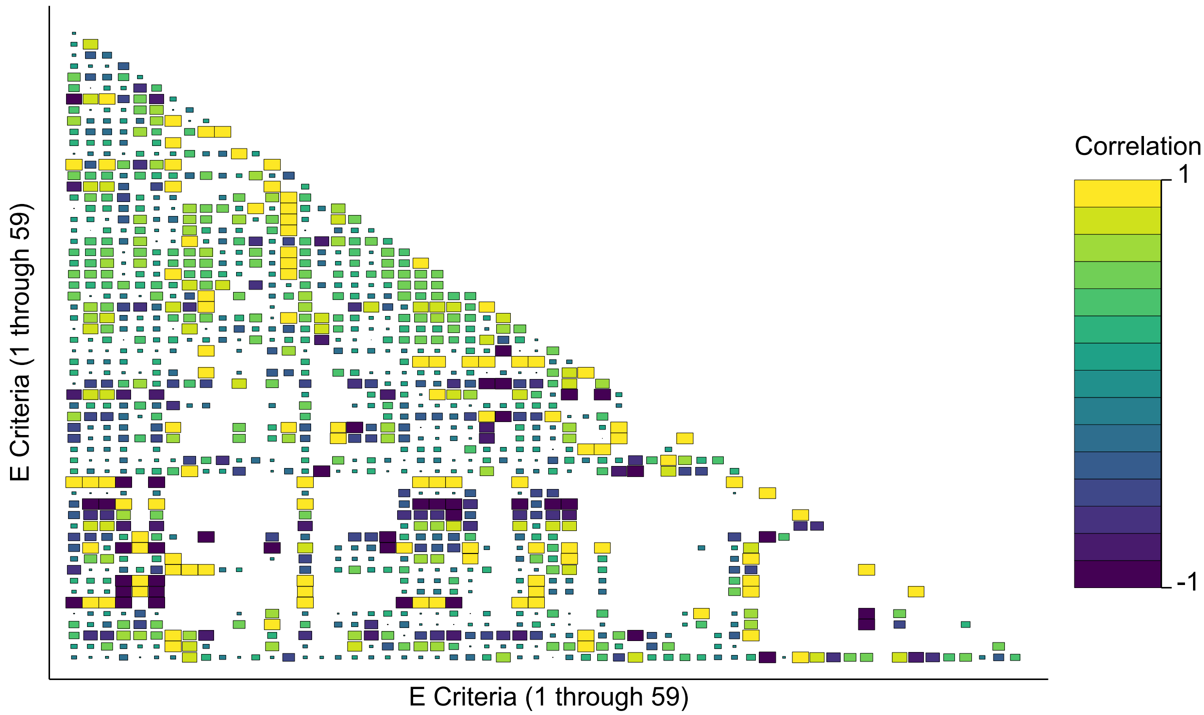


Panel B – Social criteria N rated firms

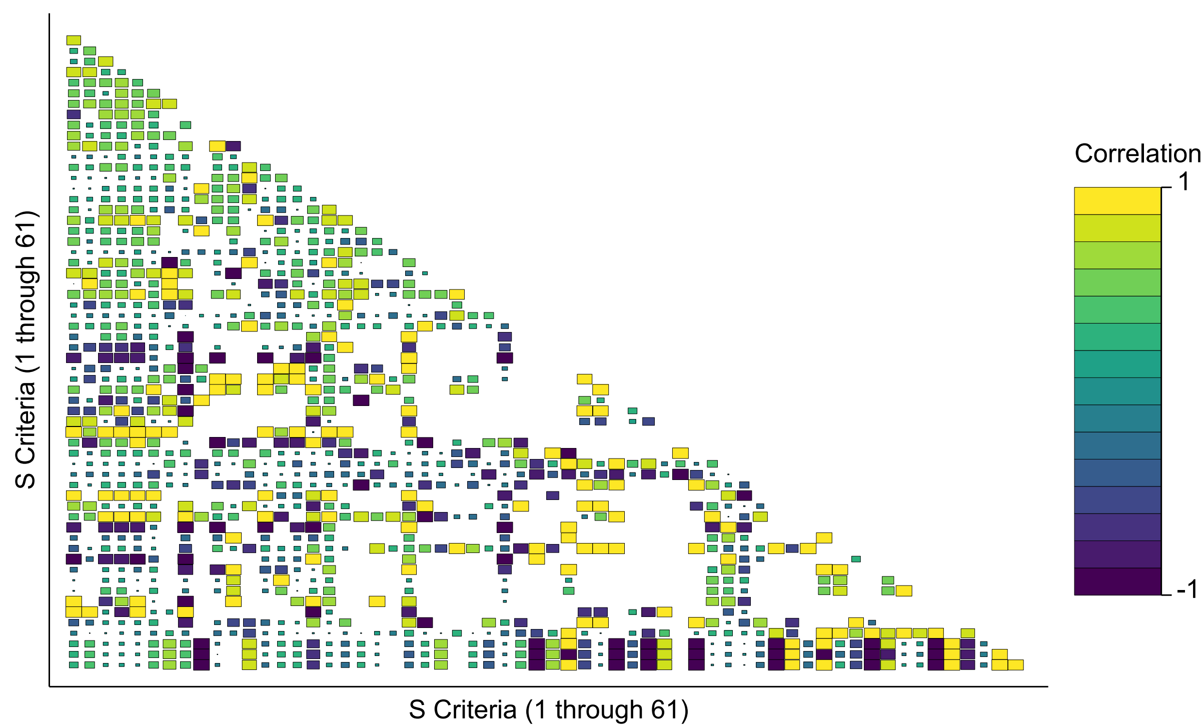


Panel C – Governance criteria N rated firms

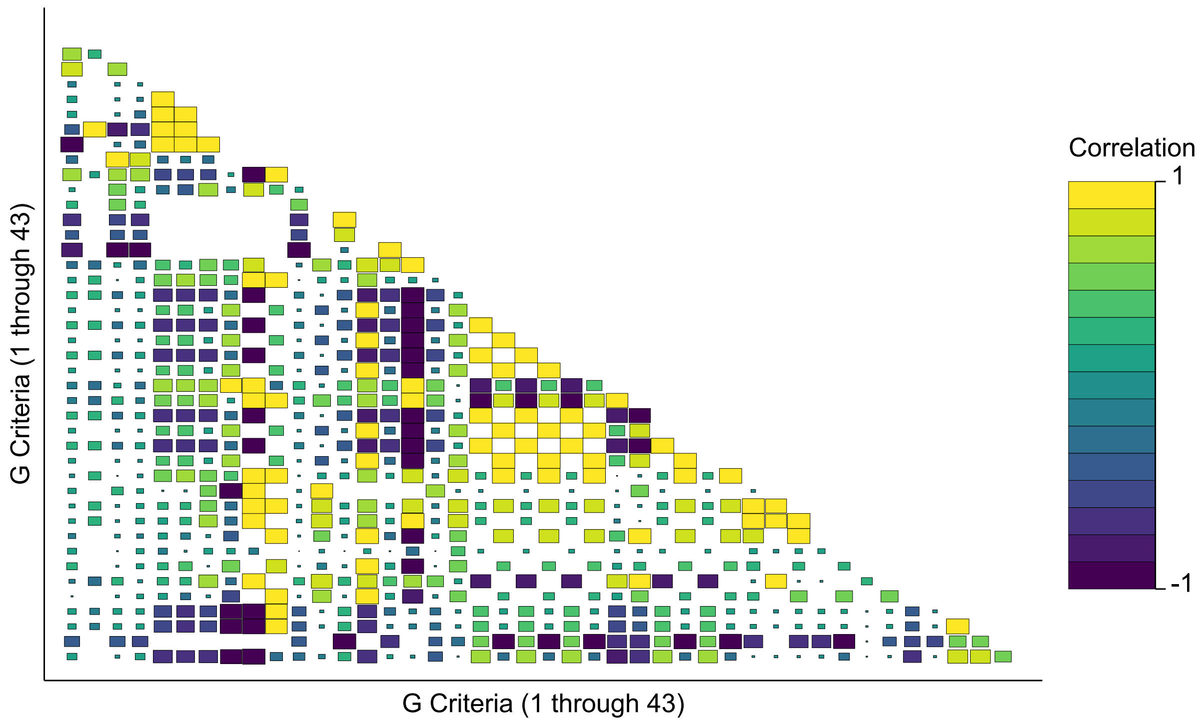
Figure 4. Number of rated firms by criteria through time. This figure displays the number of rated firms under each criteria and month of the sample. The sample runs from August 2009 through September 2019. Panel A (B, C) displays the number of rated firms under each criteria for the environmental (social, governance) pillar. Data are from Sustainalytics.



Panel A – Environmental criteria correlation matrix

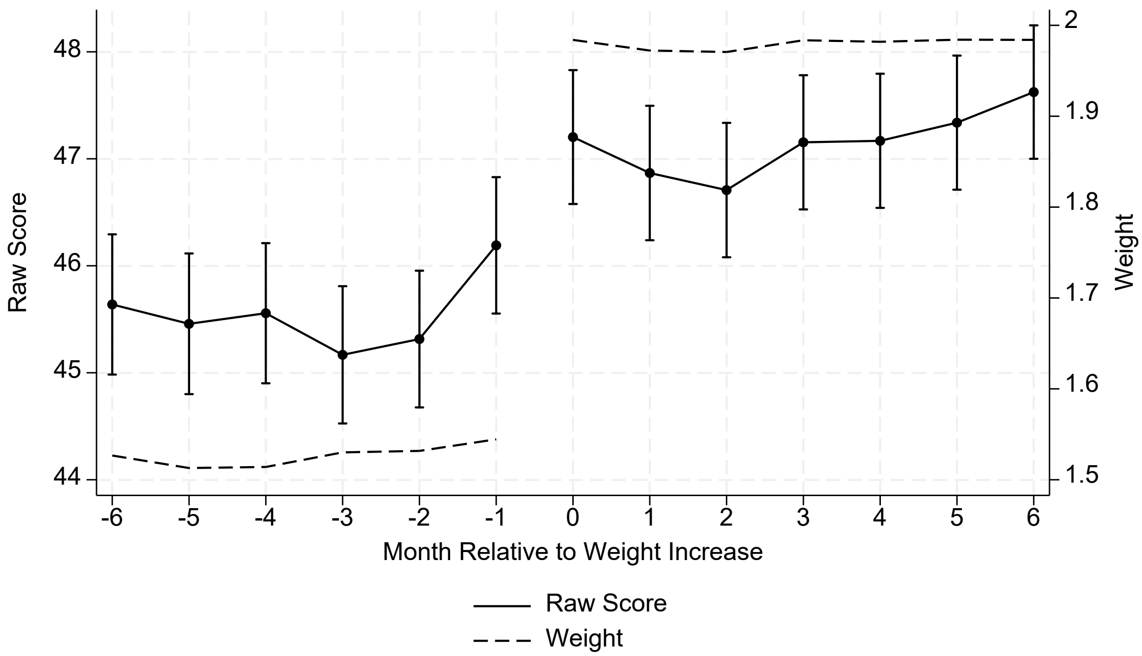


Panel B – Social criteria correlation matrix

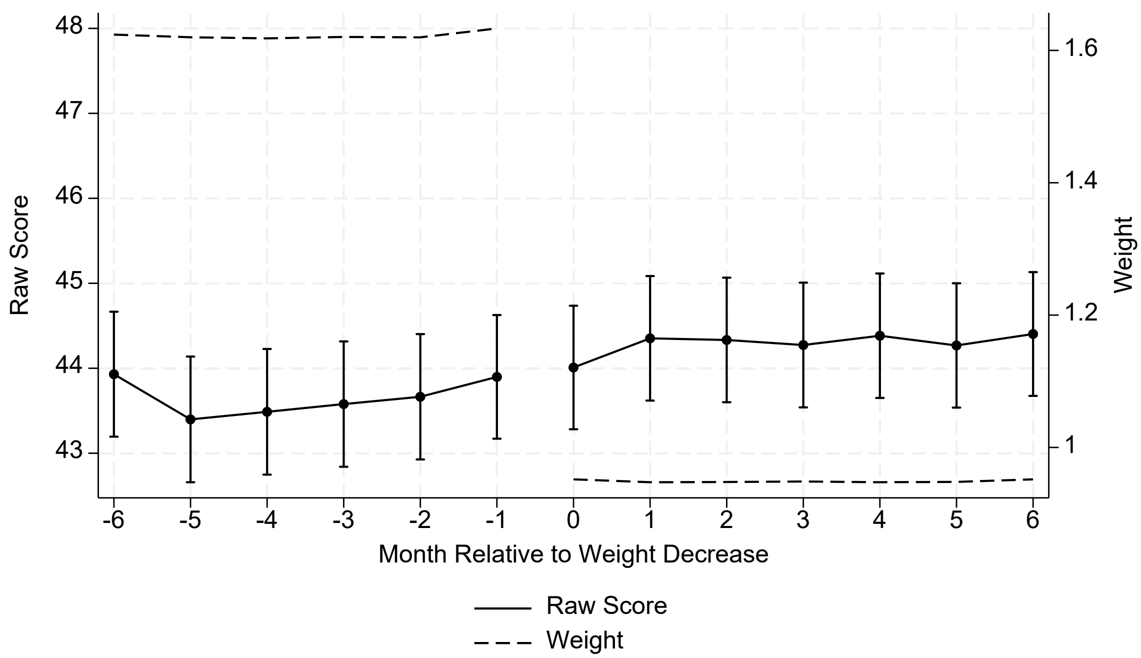


Panel C – Governance criteria correlation matrix

Figure 5. Correlation matrices of criteria weights. This figure displays correlations of criteria weights for firm-month observations. The sample runs from August 2009 through September 2019. Panel A (B, C) displays correlations among criteria for the environmental (social, governance) pillar. Data are from Sustainalytics.

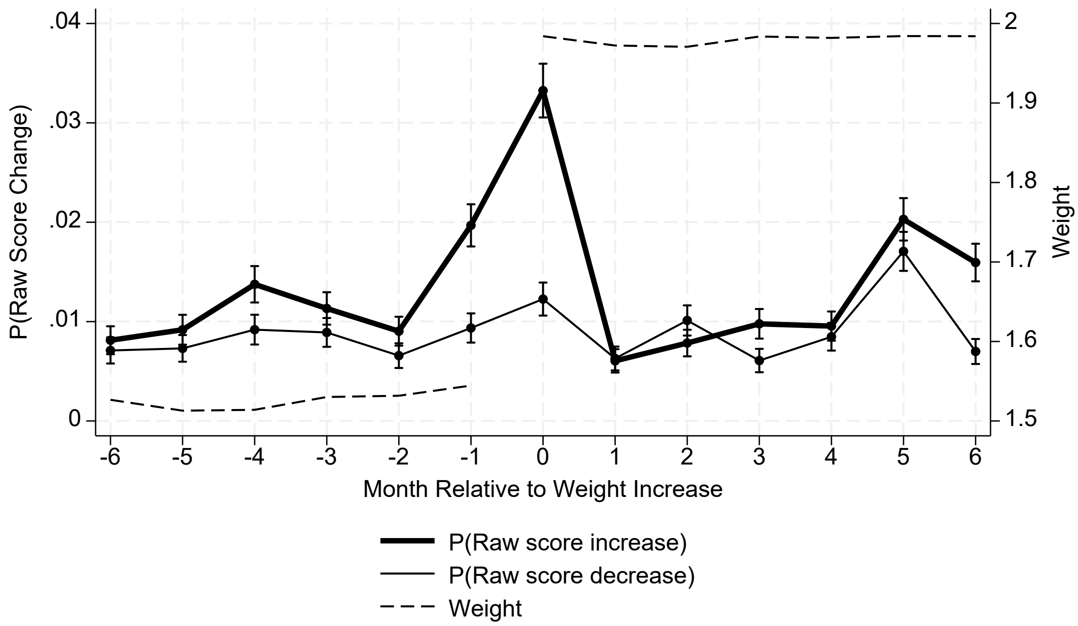


Panel A – Increases in criteria weights

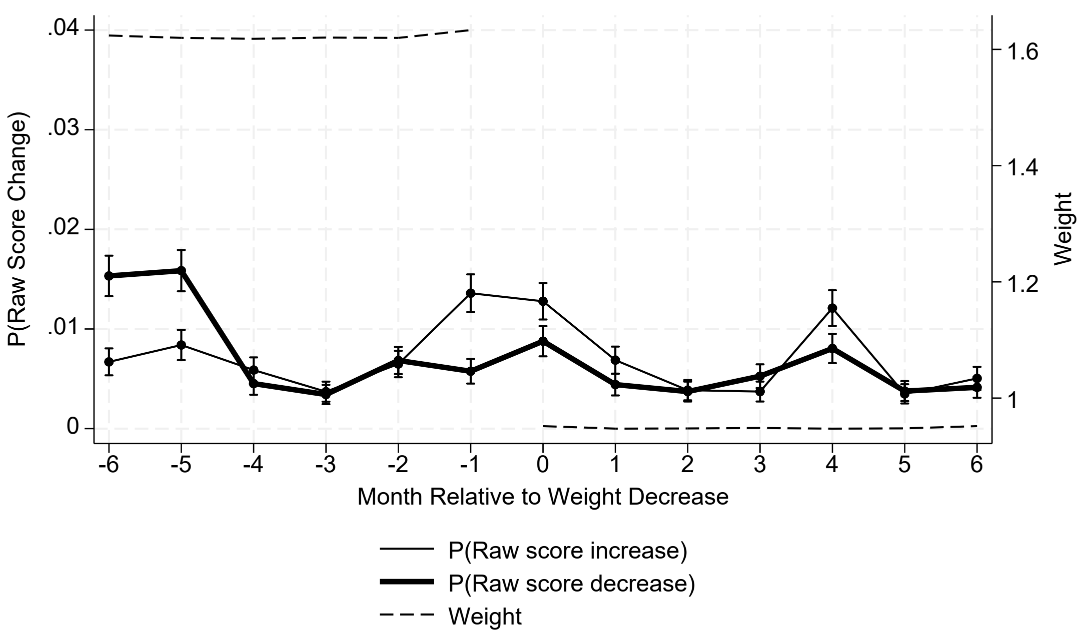


Panel B – Decreases in criteria weights

Figure 6. Raw scores around criteria weight changes. This figure displays mean raw scores around firm-criteria-months when criteria weights increase (Panel A) and decrease (Panel B). Range caps represent 90% confidence intervals. Data are from Sustainalytics.



Panel A – Increases in criteria weights



Panel B – Decreases in criteria weights

Figure 7. Probability of raw score changes around criteria weight changes. This figure displays mean probabilities that raw scores change around firm-criteria-months when criteria weights increase (Panel A) and decrease (Panel B). Range caps represent 90% confidence intervals. Data are from Sustainalytics.

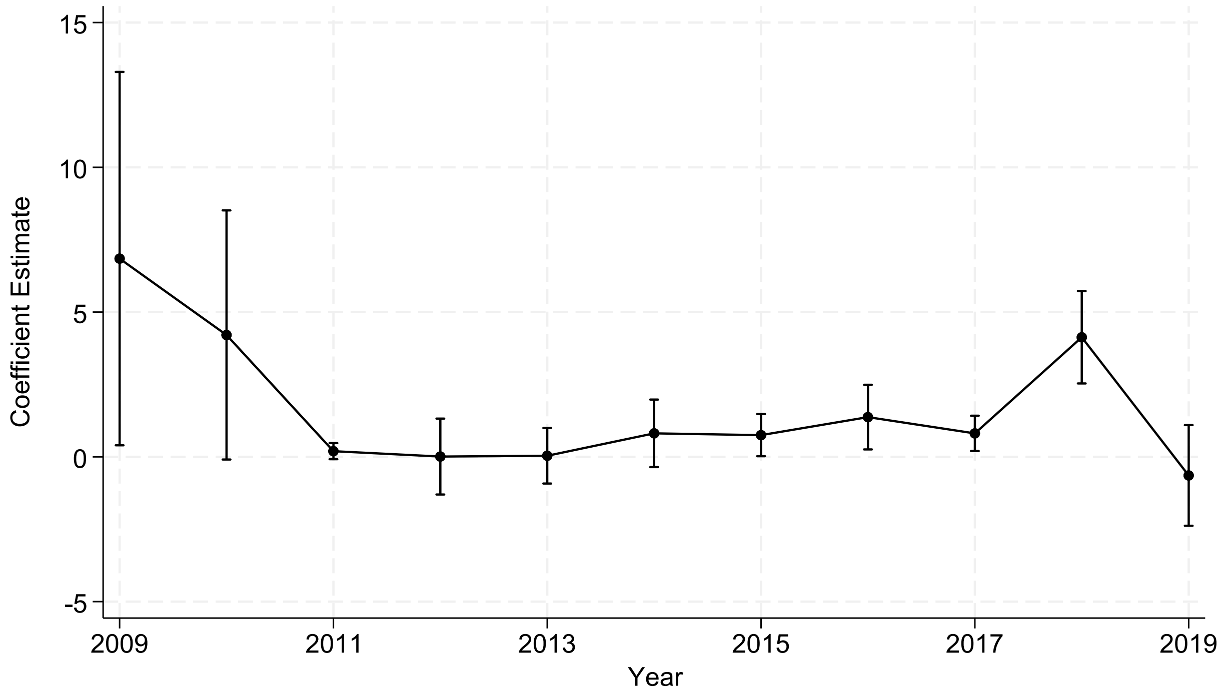


Figure 8. The relation between raw scores and weights in the time series. This figure displays estimates of β from the following OLS regression:

$$\text{Raw score}_{i,c(p),t} = \alpha + \beta\ \text{Weight}_{i,c(p),t} + \text{Lead and lag of Weight}_{i,c(p),t} +$$
$$\text{Firm}_i \times \text{Criteria}_{c(p)}\ \text{FE} + \text{Firm}_i \times \text{Month}_t\ \text{FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_t + \varepsilon_{i,c(p),t}$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month. We repeat this regression for each year of the sample. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Data for raw scores and criteria weights are from Sustainalytics. The sample includes firm-month-criteria observations from August 2009 through September 2019. Standard errors are clustered at the peer group, month, and criteria levels. Range caps represent 90% confidence intervals.

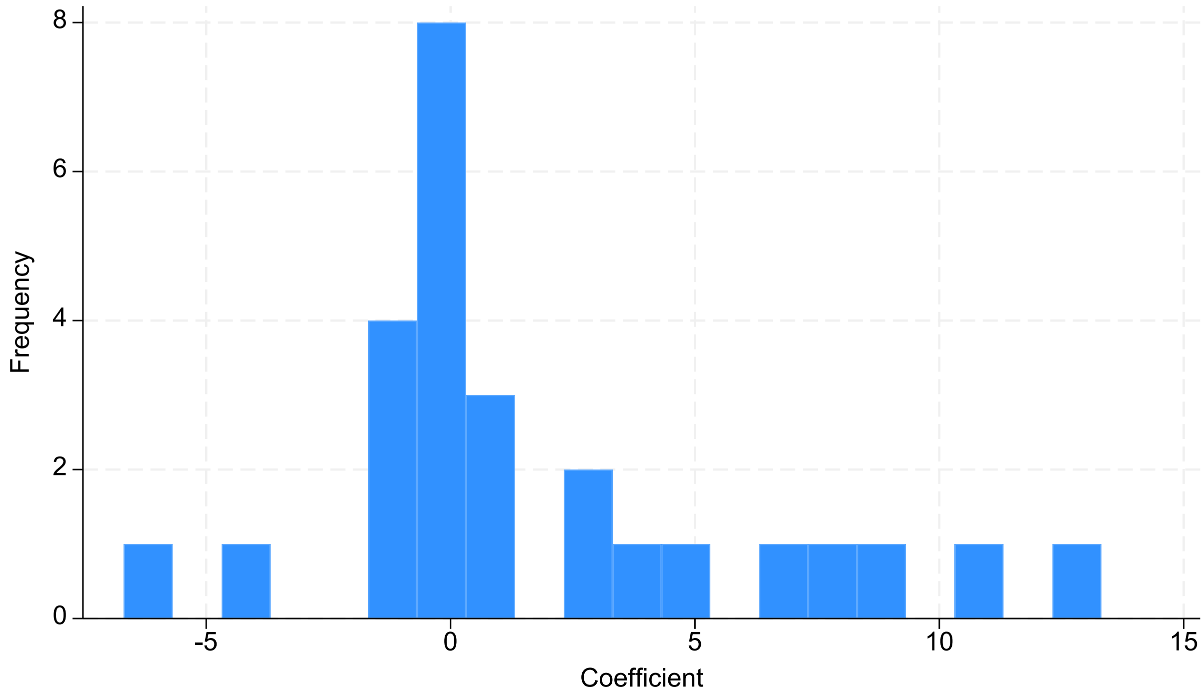


Figure 9. The relation between raw scores and weights by peer group. This figure displays a histogram of estimates of β from the following OLS regression:

$$\text{Raw score}_{i,c(p),t} = \alpha + \beta\ \text{Weight}_{i,c(p),t} + \text{Lead and lag of Weight}_{i,c(p),t} +$$
$$\text{Firm}_i \times \text{Criteria}_{c(p)}\ \text{FE} + \text{Firm}_i \times \text{Month}_t\ \text{FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_t + \varepsilon_{i,c(p),t}$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month. We repeat this regression for each industry peer group root of the sample. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. The sample includes firm-month-criteria observations from August 2009 through September 2019. We restrict the analysis to peer group roots with at least 25,000 firm-month-criteria observations.

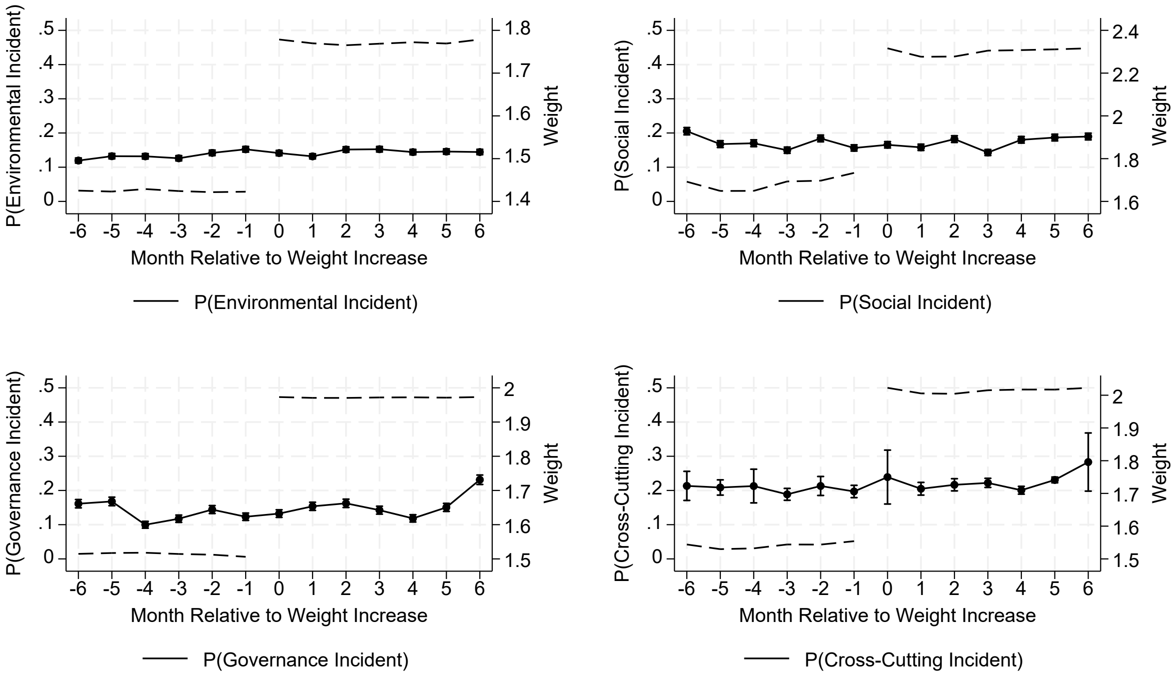
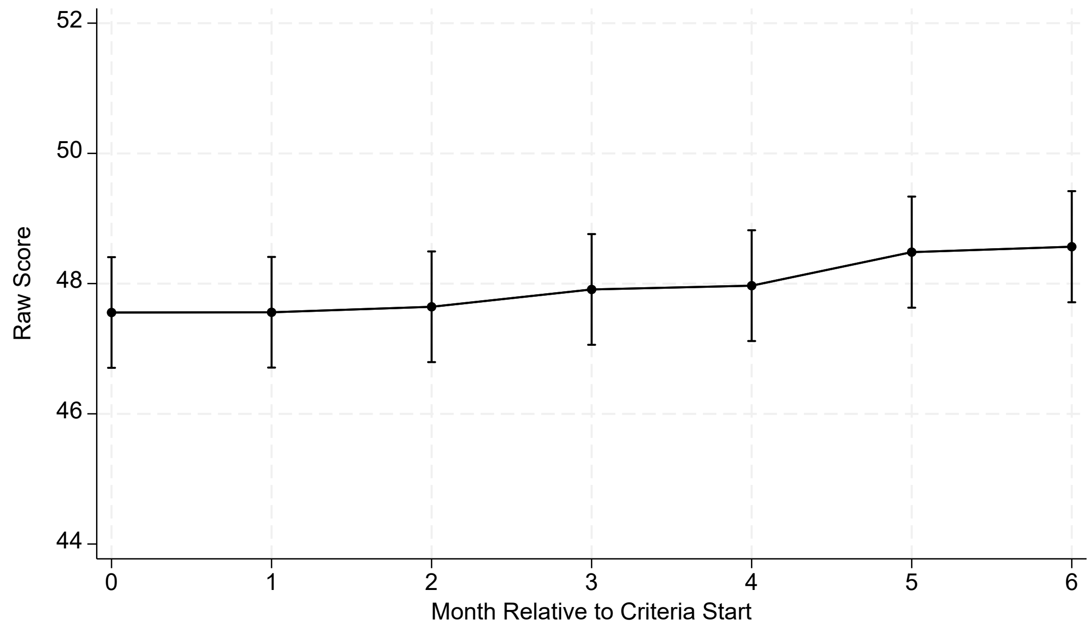
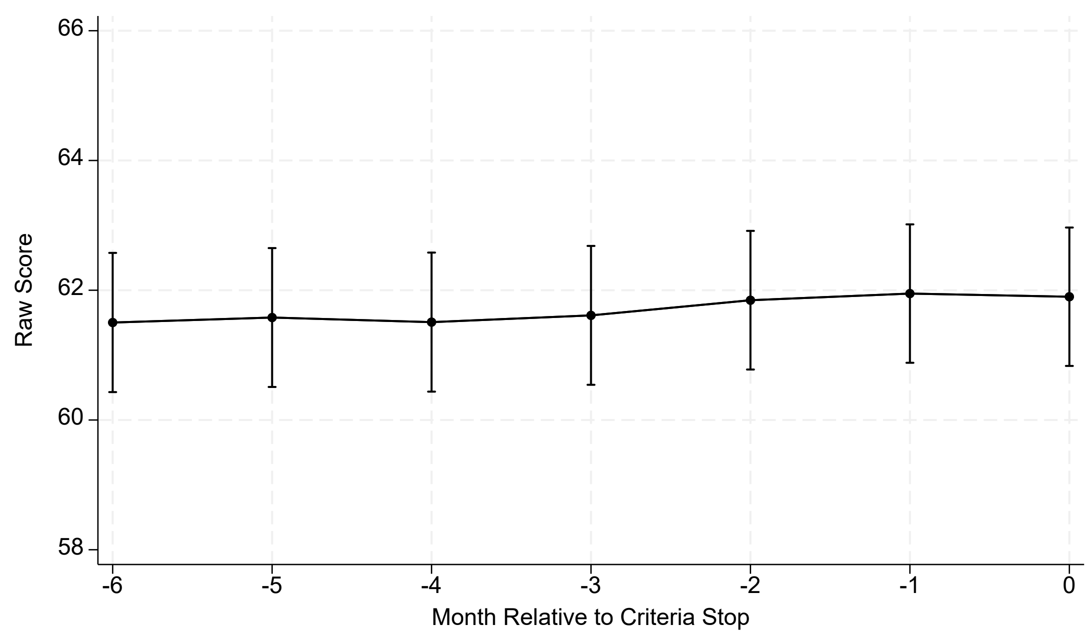


Figure 10. Probability of ESG incidents around criteria weight changes. This figure displays monthly probabilities of firms experiencing ESG-related incidents, as reported by RepRisk, around changes in weights applied to criteria by Sustainalytics in its ESG ratings generating process. We produce separate plots for environmental, social, and governance incident probabilities and criteria weight changes. We also produce a separate plot for incidents that cut across environmental, social, and governance pillars around weight changes for criteria in any of the three pillars. Range caps represent 90% confidence intervals.



Panel A – Criteria starts



Panel B – Criteria stops

Figure 11. Raw scores around criteria starts and stops. Panel A displays mean raw scores in the initial and following months when Sustainalytics introduces a new ESG rating criteria for a firm. Panel B displays mean raw scores in the final and preceding months when an ESG criteria is no longer used for a firm. Range caps represent 90% confidence intervals. Data are from Sustainalytics.

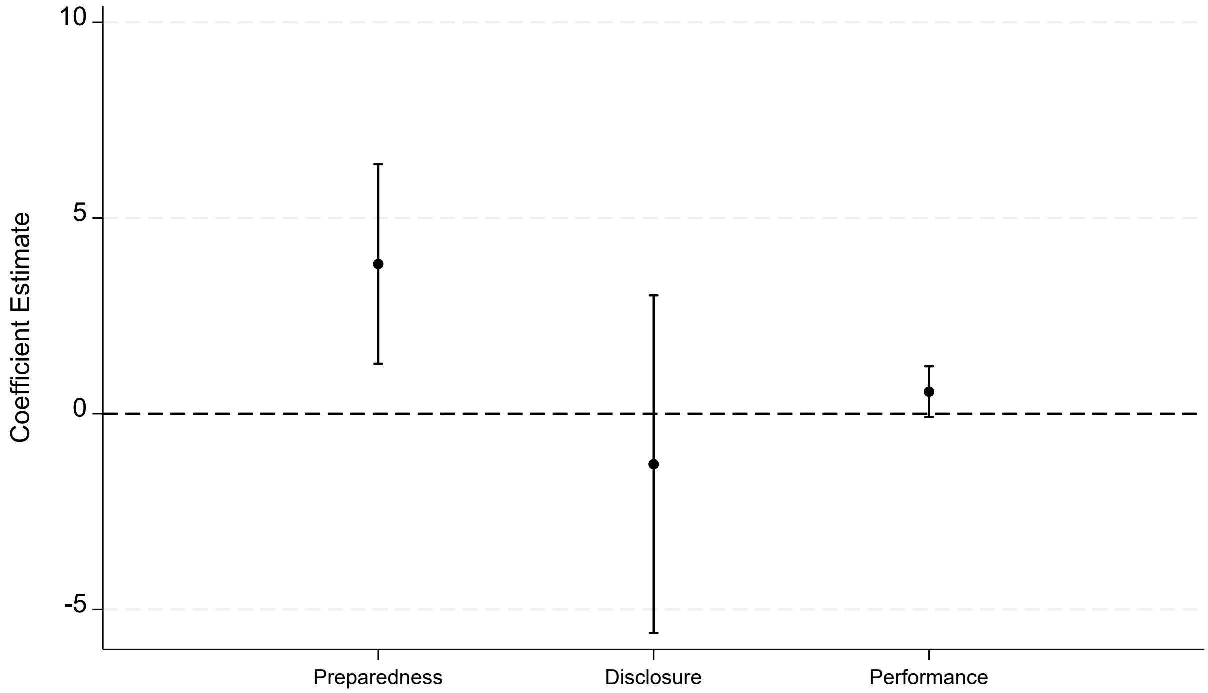


Figure 12. Ratings management as a function of criteria type. This figure displays estimates of β from the following OLS regression:

$$\text{Raw score}_{i,c(p),t} = \alpha + \beta\ \text{Weight}_{i,c(p),t} + \text{Lead and lag of Weight}_{i,c(p),t} + \text{Firm}_i \times \text{Criteria}_{c(p)}\ \text{FE} + \text{Firm}_i \times \text{Month}_t\ \text{FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_t + \varepsilon_{i,c(p),t}$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month. We repeat this regression after partitioning the sample by criteria type. Specifically, we sort criteria according to whether they are based on firms' preparedness, disclosure, or performance. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Data for raw scores and criteria weights are from Sustainalytics. The sample includes firm-month-criteria observations from August 2009 through September 2019. Standard errors are clustered at the peer group, month, and criteria levels. Range caps represent 90% confidence intervals.

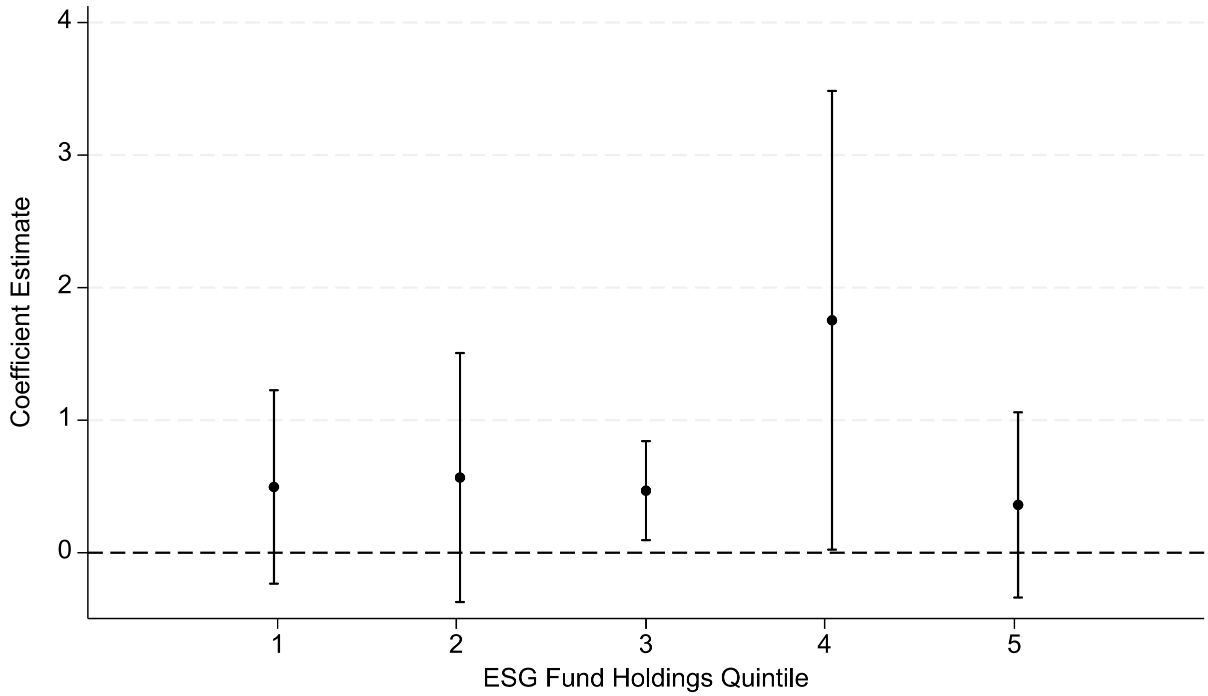


Figure 13. Ratings management and ESG fund holdings. This figure displays estimates of β from the following OLS regression:

$$\text{Raw score}_{i,c(p),t} = \alpha + \beta\ \text{Weight}_{i,c(p),t} + \text{Lead and lag of Weight}_{i,c(p),t} + \\ \text{Firm}_i \times \text{Criteria}_{c(p)}\ \text{FE} + \text{Firm}_i \times \text{Month}_t\ \text{FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_t + \varepsilon_{i,c(p),t}$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month. We repeat this regression after partitioning the sample by ESG fund holdings quintile. Specifically, we use 13f filings to identify ESG funds and amounts of firms' equities they hold. We compute the percentage of a firm-year's equity that is held by an ESG fund and then divide the full sample into quintiles based on this measure. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Data for raw scores and criteria weights are from Sustainalytics. The sample includes firm-month-criteria observations from August 2009 through September 2019. Standard errors are clustered at the peer group, month, and criteria levels. Range caps represent 90% confidence intervals.

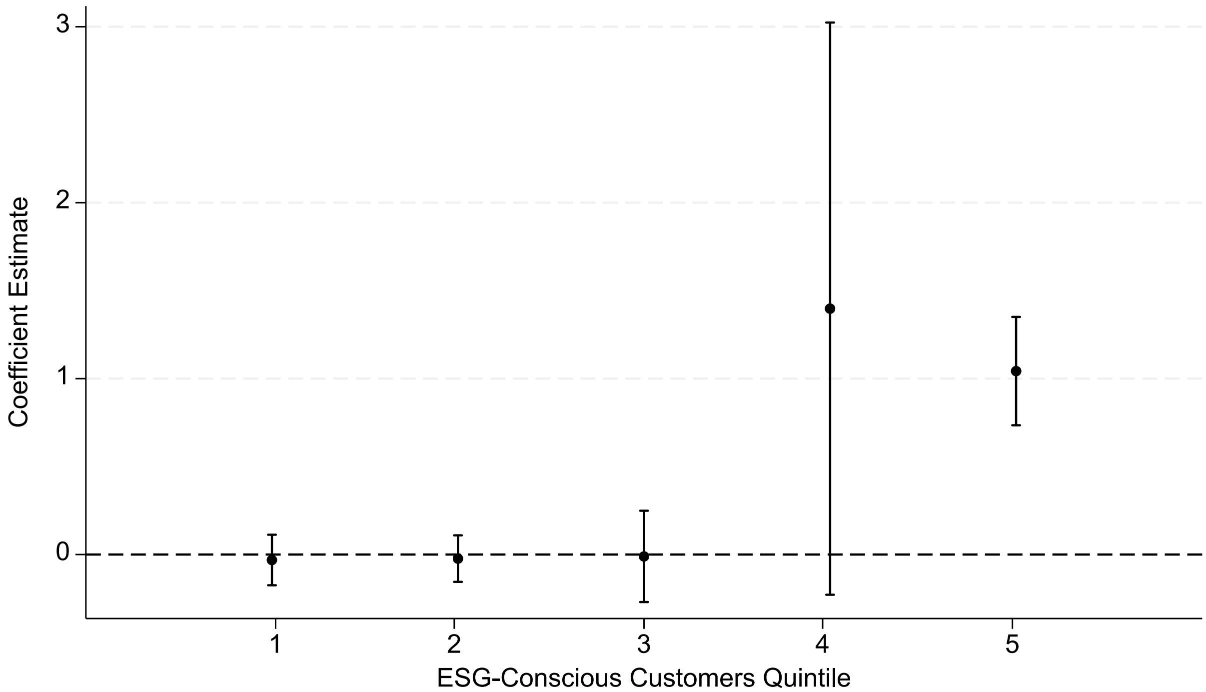


Figure 14. Ratings management and ESG-conscious customers. This figure displays estimates of β from the following OLS regression:

$$\text{Raw score}_{i,c(p),t} = \alpha + \beta \text{ Weight}_{i,c(p),t} + \text{Lead and lag of Weight}_{i,c(p),t} + \\ \text{Firm}_i \times \text{Criteria}_{c(p)} \text{ FE} + \text{Firm}_i \times \text{Month}_t \text{ FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_t + \varepsilon_{i,c(p),t}$$

where *i* denotes firm, *c* denotes criteria (and indexes pillar, *p*), and *t* denotes month. We repeat this regression after partitioning the sample by quintile of revenue the firm receives from Europe. Specifically, we collect regional revenue data from FactSet as of December 2022. We compute the percentage of a firm-year's revenue that comes from Europe and then divide the full sample into quintiles based on this measure. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Data for raw scores and criteria weights are from Sustainalytics. The sample includes firm-month-criteria observations from August 2009 through September 2019. Standard errors are clustered at the peer group, month, and criteria levels. Range caps represent 90% confidence intervals.

Table I
Summary Statistics

Panel A of this table displays summary statistics for firm-month-criteria observations from August 2009 through September 2019. Panel B displays summary statistics for supplemental analysis. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Δ *Raw score* (Δ *Weight*) is the change in *Raw score* (*Weight*) from the preceding firm-month-criteria. Data for raw scores and weights are from Sustainalytics. *Preparedness* (*Disclosure*, *Performance*) is an indicator variable taking a value of one if the criteria relates to ESG preparedness (disclosure, performance). *Firm size* is the natural log of the firm's total assets. *Dividends* is the firm's total dividends divided by total assets. *Cash* is the firm's cash divided by total assets. *Leverage* is the firm's total debt including current divided by total assets. *ROA* is the firm's net income divided by total assets. *BTM* is the firm's book value per share divided by its price close at the end of the calendar year. *CAPX* is the firm's capital expenditures divided by total revenue. *SG&A* is the firm's selling, general, and administrative expense divided by total revenue. *R&D* is the firm's research and development expense divided by total revenue. *R&D* takes a value of zero if it is missing. Balance sheet and income statement data are from Compustat. *Analysts* is the natural logarithm of one plus the number of equity analysts covering the firm. Analyst data are from IBES. *Environmental* (*Social*, *Governance*) *incidents* is the number of times firms experience ESG incidents over the next 12 months related to environmental (social, governance) criteria. *Cross-cutting incidents* captures ESG incidents that involve at least two pillars. We weight (i.e., multiply) ESG incidents by the severity, novelty, and reach of the incidents, respectively. Severity, novelty, and reach range in value from 1 to 3. Data on ESG incidents are from RepRisk. *Toxic releases* (*Onsite releases*, *Offsite releases*, *Offsite recycled*, *Offsite recovered*, *Offsite treated*, *Total transfers*) is the number of pounds of toxic chemicals released (released onsite, released offsite, recycled offsite, recovered offsite, treated offsite, transferred) in a firm-year. We associate these variables with observations related to the environmental pillar. Toxic release data come from the Toxic Releases Inventory (TRI) Program from the U.S. Environmental Protection Agency. *ESG fund holdings* is the percentage of a firm-year's equity that is held by ESG-focused institutional investors. We use firms' 13f filings to measure ESG-focused institutional investor holdings. *ESG customers* is the percentage of a firm-year's revenue that it generates from Europe. We collect data on the geographic origin of firm's revenues from FactSet. We impute firm-year data to firm-month-criteria observations in the same calendar year.

Panel A: Baseline Sample

	N	Mean	Median	SD	10^{th} pctl	90^{th} pctl
Raw score	1,787,228	51.47	50	42.96	0	100
Δ Raw score	1,787,228	0.10	0.00	6.56	0.00	0.00
Weight	1,787,228	1.59	1.00	1.33	0.50	3.00
Δ Weight	1,787,228	0.00	0.00	0.14	0.00	0.00
Preparedness	1,787,228	0.32	0.00	0.47	0.00	1.00
Disclosure	1,787,228	0.08	0.00	0.27	0.00	0.00
Performance	1,787,228	0.60	1.00	0.49	0.00	1.00
Firm size	1,787,228	9.51	9.42	1.23	8.21	11.13
Dividends	1,787,228	0.03	0.02	0.03	0.00	0.06
Cash	1,787,228	0.10	0.08	0.08	0.01	0.21
Leverage	1,787,228	0.27	0.25	0.17	0.06	0.49
ROA	1,787,228	0.06	0.06	0.08	0.00	0.14
BTM	1,787,228	0.39	0.32	0.43	0.10	0.75
CAPX	1,787,228	0.11	0.04	0.59	0.01	0.18
SG&A	1,787,228	0.22	0.19	0.23	0.04	0.40
R&D	1,787,228	0.04	0.00	0.11	0.00	0.14
Analysts	1,787,228	1.37	1.79	1.15	0.00	2.71

Panel B: Supplement

	N	Mean	Median	SD	10^{th} pctl	90^{th} pctl
Environmental incidents	379,917	1.5	0.0	2.7	0.0	5.0
× Severity	379,917	2.4	0.0	4.7	0.0	8.0
× Novelty	379,917	2.4	0.0	4.3	0.0	8.0
× Reach	379,917	2.7	0.0	5.4	0.0	9.0
Social incidents	407,327	2.0	0.0	3.2	0.0	7.0
× Severity	407,327	3.2	0.0	5.6	0.0	10.0
× Novelty	407,327	3.3	0.0	5.2	0.0	11.0
× Reach	407,327	3.8	0.0	6.6	0.0	12.0
Governance incidents	418,489	1.9	1.0	2.9	0.0	6.0
× Severity	418,489	2.8	1.0	4.7	0.0	9.0
× Novelty	418,489	3.1	2.0	4.7	0.0	9.0
× Reach	418,489	3.7	1.0	6.6	0.0	12.0
Cross-cutting incidents	1,205,733	2.6	1.0	3.5	0.0	8.0
× Severity	1,205,733	3.9	1.0	5.9	0.0	12.0
× Novelty	1,205,733	4.1	2.0	5.6	0.0	13.0
× Reach	1,205,733	5.3	2.0	8.2	0.0	17.0
Total releases	639,296	951,235	3,401	5,313,933	3,033	666,717
Onsite releases	639,296	816,126	849	5,079,459	653	346,602
Offsite releases	639,296	135,109	2,748	937,721	1	118,006
Offsite recycled	639,296	119,676	12,130	641,443	0	92,597
Offsite recovery	639,296	1,244,720	43,392	12,100,000	0	459,675
Offsite treated	639,296	165,251	0	1,518,310	0	12,315
Total transfers	639,296	240,249	12,809	1,396,893	0	198,515
ESG fund holdings	1,619,161	0.0024	0.0016	0.0035	0.0002	0.0053
ESG customers	1,787,228	12.5539	10.8000	13.4644	0.0000	28.2000

Table II
Time Series Correlations

Panel A (B, C) displays correlation coefficients for *Raw score* (*Weight*, Δ *Weight*) and its first six lags. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Data for raw scores and criteria weights are from Sustainalytics. This sample includes firm-month-criteria observations from August 2009 through September 2019. Column 1 of each panel uses the full sample of firm-criteria-month observations described in Table I. Columns 2, 3, and 4 partition the sample by the ESG pillar with which each firm-month-criteria observation is associated.

Panel A: Raw score

		Full sample split by pillar		
	Full sample	Environmental	Social	Governance
	(1)	(2)	(3)	(4)
1 month lag	0.99	0.99	0.99	0.99
2 months lag	0.98	0.98	0.98	0.98
3 months lag	0.97	0.97	0.98	0.97
4 months lag	0.97	0.96	0.97	0.97
5 months lag	0.96	0.95	0.97	0.96
6 months lag	0.95	0.94	0.96	0.95

Panel B: Weight

		Full sample split by pillar		
	Full sample	Environmental	Social	Governance
	(1)	(2)	(3)	(4)
1 month lag	0.99	0.99	0.99	1.00
2 months lag	0.99	0.99	0.99	1.00
3 months lag	0.99	0.98	0.99	0.99
4 months lag	0.98	0.98	0.98	0.99
5 months lag	0.98	0.98	0.98	0.99
6 months lag	0.98	0.97	0.97	0.99

Table III
The Relation between Raw Scores and Weights

This table displays OLS regression results with *Raw score* (Panels A and B) and Δ *Raw score* (Panel C) as dependent variables. The independent variables of interest are *Weight* (Panel A) and Δ *Weight* (Panels B and C). *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-criteria-month. Δ *Raw score* (Δ *Weight*) is the change in *Raw score* (*Weight*) from the preceding firm-criteria-month. Data for criteria raw scores and weights are from Sustainalytics. The sample includes firm-criteria-month observations from August 2009 through September 2019. Column 1 of Panel A and B uses the full sample of firm-criteria-month observations. Columns 2, 3, and 4 partition the sample by the ESG pillar with which each firm-month-criteria observation is associated. Columns 1 and 2 of Panel C use the full sample of firm-criteria-month observations and columns 3, 4, and 5 partition the sample by the ESG pillar with which each firm-month-criteria observation is associated. Δ $Weight^{+}$ (Δ $Weight^{-}$) is the change in *Weight* when it takes a value greater (less) than zero. Standard errors are clustered at the peer group, criteria, and month levels. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: *Raw score* is Dependent Variable

		Full sample split by pillar		
	Full sample	Environmental	Social	Governance
	(1)	(2)	(3)	(4)
$Weight_{-1}$	-0.38	-0.53	-0.35	-0.32
	(0.25)	(0.36)	(0.37)	(0.74)
Weight	0.90	1.08	0.90	0.75
	(0.43)**	(0.62)*	(0.36)**	(1.46)
$Weight_{+1}$	-0.44	-0.30	-0.35	-0.79
	(0.26)	(0.44)	(0.25)	(0.75)
Firm × Criteria FE	Yes	Yes	Yes	Yes
Firm × Month FE	Yes	Yes	Yes	Yes
Criteria × Month FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.85	0.83	0.86	0.84
N	1,787,228	568,174	602,371	616,683

Panel B: *Raw score* is Dependent Variable

		Full sample split by pillar		
	Full sample	Environmental	Social	Governance
	(1)	(2)	(3)	(4)
Δ Weight	0.54	0.69	0.56	0.42
	(0.29)*	(0.40)*	(0.34)	(0.62)
Firm × Criteria FE	Yes	Yes	Yes	Yes
Firm × Month FE	Yes	Yes	Yes	Yes
Criteria × Month FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.85	0.83	0.86	0.84
N	1,805,109	615,321	631,690	644,980

Panel C: Δ *Raw score* is Dependent Variable

			Full sample split by pillar		
	Full sample	Full sample	Environment	Social	Governance
	(1)	(2)	(3)	(4)	(5)
Δ Weight	0.03				
	(0.09)				
Δ Weight $^{+}$		0.33	0.59	0.07	0.03
		(0.19)*	(0.31)*	(0.11)	(0.16)
Δ Weight $^{-}$		-0.06	-0.30	0.01	0.27
		(0.22)	(0.42)	(0.16)	(0.19)
Constant	0.10	0.10	0.10	0.10	0.09
	(0.04)**	(0.04)**	(0.04)**	(0.03)***	(0.06)
Adj. R^2	-0.00	0.00	0.00	-0.00	0.00
N	1,805,109	1,805,109	597,125	631,718	645,008

Table IV
Do Weight-Driven Ratings Predict ESG Incidents?

Each panel of this table displays results from second-stage regressions. In the first stage, we generate predicted and residual values from the following regression.

$$\text{Raw score}_{i,c(p),t(y)} = \alpha + \beta \text{ Weight}_{i,c(p),t(y)} + \varepsilon_{i,c(p),t(y)}$$

where *i* denotes firm, *c* denotes environmental criteria (and indexes pillar, *p*), and *t* denotes month (and indexes year, *y*, for control variables in the second-stage regression). Panel A (B, C) uses firm-month-criteria observations associated with the environmental (social, governance) pillar. Panel D uses firm-month-criteria observations associated with all three pillars. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. In the second stage, we regress measures of ESG-related incidents on the predicted and residual values of *Raw score* as follows:

$$\text{ESG incidents}_{i,c(p),t(y)} = \alpha + \beta_1 \text{ Predicted Raw score}_{i,c(p),t(y)} + \beta_2 \text{ Residual Raw score}_{i,c(p),t(y)} + \text{Firm-year controls}_{i,y} + \text{Firm}_i \times \text{Criteria}_{c(p)} \text{ FE} + \text{Criteria}_{c(p)} \times \text{Month FE}_{t(y)} + \varepsilon_{i,c(p),t(y)}$$

Environmental (*Social*, *Governance*) *incidents* is the number of times firms experience ESG incidents over the next 12 months related to environmental (social, governance) criteria. *Cross-cutting incidents* captures ESG incidents that involve at least two pillars. We weight (i.e., multiply) ESG incidents by the severity, novelty, and reach of the incidents, respectively. Severity, novelty, and reach range in value from 1 to 3. Data on ESG incidents are from RepRisk. We standardize ESG incident measures, *Predicted Raw score*, and *Residual Raw score* to follow mean-zero, unit-variance distributions. We suppress firm-year controls to conserve space. Firm-year controls include *Firm size* (the natural log of the firm's total assets), *Dividends* (the firm's total dividends divided by total assets), *Cash* (the firm's cash divided by total assets), *Leverage* (the firm's total debt including current divided by total assets), *ROA* (the firm's net income divided by total assets), *BTM* (the firm's book value per share divided by its price close at the end of the calendar year), *CAPX* (the firm's capital expenditures divided by total revenue), *SG&A* (the firm's selling, general, and administrative expense divided by total revenue), *R&D* (the firm's research and development expense divided by total revenue, which takes a value of zero if it is missing), and *Analysts* (the natural logarithm of one plus the number of equity analysts covering the firm). We impute firm-year balance sheet, income statement, and analyst data to firm-month-criteria observations in the same calendar year. Data for raw scores and criteria weights are from Sustainalytics. Data for ESG incidents are from RepRisk. Balance sheet and income statement data are from Compustat. Analyst data are from IBES. The sample includes firm-month-criteria observations from August 2009 through September 2019. Standard errors are clustered at the peer group, month, and criteria levels. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Environmental ESG incidents

	Dependent variable:			
	Incidents	× Severity	× Novelty	× Reach
	(1)	(2)	(3)	(4)
Predicted Raw score	0.00	-0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.02)
Residual Raw score	0.00	0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.85	0.86	0.86	0.83
N	379,903	379,903	379,903	379,903

Panel B: Social ESG incidents

	Dependent variable:			
	Incidents	× Severity	× Novelty	× Reach
	(1)	(2)	(3)	(4)
Predicted Raw score	-0.00	-0.01	-0.01	-0.01
	(0.02)	(0.01)	(0.02)	(0.02)
Residual Raw score	0.01	0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.85	0.87	0.85	0.85
N	407,327	407,327	407,327	407,327

Panel C: Governance ESG incidents

	Dependent variable:			
	Incidents	× Severity	× Novelty	× Reach
	(1)	(2)	(3)	(4)
Predicted Raw score	0.01	0.00	0.01	0.00
	(0.03)	(0.03)	(0.02)	(0.02)
Residual Raw score	0.01	0.01	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Adjusted R^2	0.80	0.79	0.79	0.80
N	418,489	418,489	418,489	418,489

Panel D: Cross-cutting ESG incidents

	Dependent variable:			
	Incidents	× Severity	× Novelty	× Reach
	(1)	(2)	(3)	(4)
Predicted Raw score	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Residual Raw score	0.02	0.01	0.02	0.01
	(0.01)**	(0.01)**	(0.01)**	(0.01)*
Adjusted R^2	0.85	0.86	0.84	0.85
N	1,205,733	1,205,733	1,205,733	1,205,733

Table V
Do Weight-Driven Ratings Predict Real ESG Outcomes?

Each panel of this table displays results from 21 second-stage regressions. In the first stage, we generate predicted and residual values from the following regression.

$$\text{Raw score}_{i,c,t(y)} = \alpha + \beta \text{ Weight}_{i,c,t(y)} + \varepsilon_{i,c,t(y)}$$

where *i* denotes firm, *c* denotes environmental criteria, and *t* denotes month (and indexes year, *y*, for control variables in the second-stage regression). Panel A uses all firm-month-criteria observations associated with the environmental pillar. Panel B restricts this sample to performance-related (as opposed to preparedness- or disclosure-related) firm-month-criteria observations associated with the environmental pillar. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. In the second stage, we regress measures of toxic releases on the predicted and residual values of *Raw score* as follows:

$$\text{Toxic releases}_{i,c,t(y)} = \alpha + \beta_1 \text{ Predicted Raw score}_{i,c,t(y)} + \beta_2 \text{ Residual Raw score}_{i,c,t(y)}$$
$$+ \text{ Firm-year controls}_{i,y} + \text{Firm}_i \times \text{Criteria}_c \text{ FE} + \text{Criteria}_c \times \text{Month FE}_{t(y)} + \varepsilon_{i,c,t(y)}$$

We standardize toxic release measures, *Predicted Raw score*, and *Residual Raw score* to follow mean-zero, unit-variance distributions. We suppress firm-year controls to conserve space. Firm-year controls include *Firm size* (the natural log of the firm's total assets), *Dividends* (the firm's total dividends divided by total assets), *Cash* (the firm's cash divided by total assets), *Leverage* (the firm's total debt including current divided by total assets), *ROA* (the firm's net income divided by total assets), *BTM* (the firm's book value per share divided by its price close at the end of the calendar year), *CAPX* (the firm's capital expenditures divided by total revenue), *SG&A* (the firm's selling, general, and administrative expense divided by total revenue), *R&D* (the firm's research and development expense divided by total revenue, which takes a value of zero if it is missing), and *Analysts* (the natural logarithm of one plus the number of equity analysts covering the firm). We impute firm-year balance sheet, income statement, and analyst data to firm-month-criteria observations in the same calendar year. Data for raw scores and criteria weights are from Sustainalytics. The sample includes firm-month-criteria observations from August 2009 through September 2019. Toxic release data are from the U.S. Environmental Protection Agency's Toxic Releases Inventory database. Balance sheet and income statement data are from Compustat. Analyst data are from IBES. Standard errors are clustered at the peer group, month, and criteria levels. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: All Firm-month-criteria Observations Associated with the Environmental Pillar

	Dependent variable (contemporaneous year):						
	Total releases	Onsite releases	Offsite releases	Offsite recycled	Offsite recovery	Offsite treated	Total transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Raw score	-0.0063	-0.0061	-0.0029	0.0015	0.0009	0.0007	-0.0029
	(0.0045)	(0.0045)	(0.0022)	(0.0022)	(0.0008)	(0.0015)	(0.0021)
Residual Raw score	-0.0033	-0.0039	0.0016	0.0005	-0.0002	0.0008	-0.0009
	(0.0020)	(0.0024)	(0.0026)	(0.0006)	(0.0008)	(0.0017)	(0.0015)
Adjusted R^2	0.85	0.85	0.93	0.97	0.93	0.91	0.91
N	639,296	639,296	639,296	639,296	639,296	639,296	639,296
	Dependent variable (+1 year):						
	Total releases	Onsite releases	Offsite releases	Offsite recycled	Offsite recovery	Offsite treated	Total transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Raw score	-0.0067	-0.0068	-0.0019	0.0021	0.0002	0.0006	-0.0014
	(0.0055)	(0.0053)	(0.0029)	(0.0029)	(0.0012)	(0.0013)	(0.0011)
Residual Raw score	-0.0015	-0.0024	0.0038	0.0019	-0.0006	-0.0008	-0.0011
	(0.0030)	(0.0035)	(0.0021)*	(0.0017)	(0.0014)	(0.0013)	(0.0019)
Adjusted R^2	0.90	0.91	0.84	0.97	0.93	0.91	0.91
N	490,081	490,081	490,081	490,081	490,081	490,081	490,081
	Dependent variable (+2 years):						
	Total releases	Onsite releases	Offsite releases	Offsite recycled	Offsite recovery	Offsite treated	Total transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Raw score	-0.0084	-0.0093	0.0021	0.0030	-0.0003	0.0000	-0.0000
	(0.0072)	(0.0071)	(0.0062)	(0.0045)	(0.0005)	(0.0014)	(0.0008)
Residual Raw score	-0.0043	-0.0046	0.0002	0.0033	-0.0022	-0.0021	-0.0006
	(0.0045)	(0.0045)	(0.0027)	(0.0031)	(0.0021)	(0.0018)	(0.0018)
Adjusted R^2	0.86	0.86	0.83	0.97	0.93	0.91	0.91
N	441,436	441,436	441,436	441,436	441,436	441,436	441,436

Panel B: Performance-related Firm-month-criteria Observations Associated with the Environmental Pillar

	Dependent variable (contemporaneous year):						
	Total releases	Onsite releases	Offsite releases	Offsite recycled	Offsite recovery	Offsite treated	Total transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Raw score	-0.0038	-0.0036	-0.0018	0.0032	0.0011	0.0002	-0.0007
	(0.0029)	(0.0030)	(0.0016)	(0.0037)	(0.0009)	(0.0013)	(0.0021)
Residual Raw score	-0.0046	-0.0056	0.0034	0.0006	0.0005	-0.0033	-0.0037
	(0.0046)	(0.0047)	(0.0037)	(0.0005)	(0.0013)	(0.0013)	(0.0019)
Adjusted R^2	0.79	0.79	0.83	0.97	0.93	0.90	0.92
N	354,876	354,876	354,876	354,876	354,876	354,876	354,876
	Dependent variable (+1 year):						
	Total releases	Onsite releases	Offsite releases	Offsite recycled	Offsite recovery	Offsite treated	Total transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Raw score	-0.0033	-0.0031	-0.0019	0.0039	0.0020	-0.0018	-0.0012
	(0.0024)	(0.0021)	(0.0019)	(0.0039)	(0.0010)*	(0.0012)	(0.0015)
Residual Raw score	0.0006	-0.0000	0.0030	0.0024	0.0031	-0.0047	-0.0045
	(0.0042)	(0.0043)	(0.0033)	(0.0014)*	(0.0027)	(0.0022)**	(0.0026)*
Adjusted R^2	0.90	0.91	0.83	0.97	0.93	0.91	0.91
N	262,472	262,472	262,472	262,472	262,472	262,472	262,472
	Dependent variable (+2 years):						
	Total releases	Onsite releases	Offsite releases	Offsite recycled	Offsite recovery	Offsite treated	Total transfers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Predicted Raw score	-0.0040	-0.0039	-0.0012	0.0036	0.0002	-0.0030	-0.0051
	(0.0037)	(0.0035)	(0.0045)	(0.0044)	(0.0011)	(0.0016)*	(0.0028)*
Residual Raw score	-0.0064	-0.0062	-0.0029	0.0043	-0.0019	-0.0037	-0.0064
	(0.0049)	(0.0048)	(0.0040)	(0.0026)	(0.0018)	(0.0014)**	(0.0028)**
Adjusted R^2	0.85	0.85	0.83	0.96	0.93	0.90	0.91
N	238,994	238,994	238,994	238,994	238,994	238,994	238,994

Internet Appendix

for

ESG Ratings Management

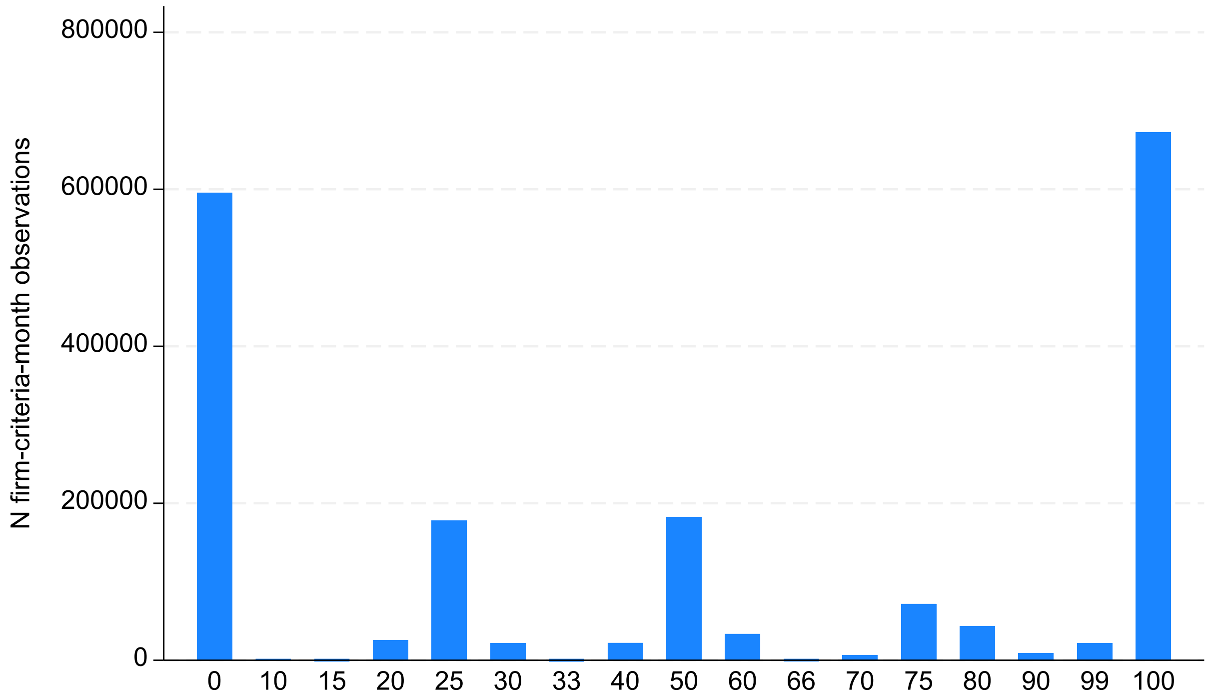


Figure A.1. Distribution of firm-criteria-month raw scores. This figure displays the frequency of raw scores for the firm-month-criteria observations in the main sample. The sample runs from August 2009 through September 2019. Data are from Sustainalytics.

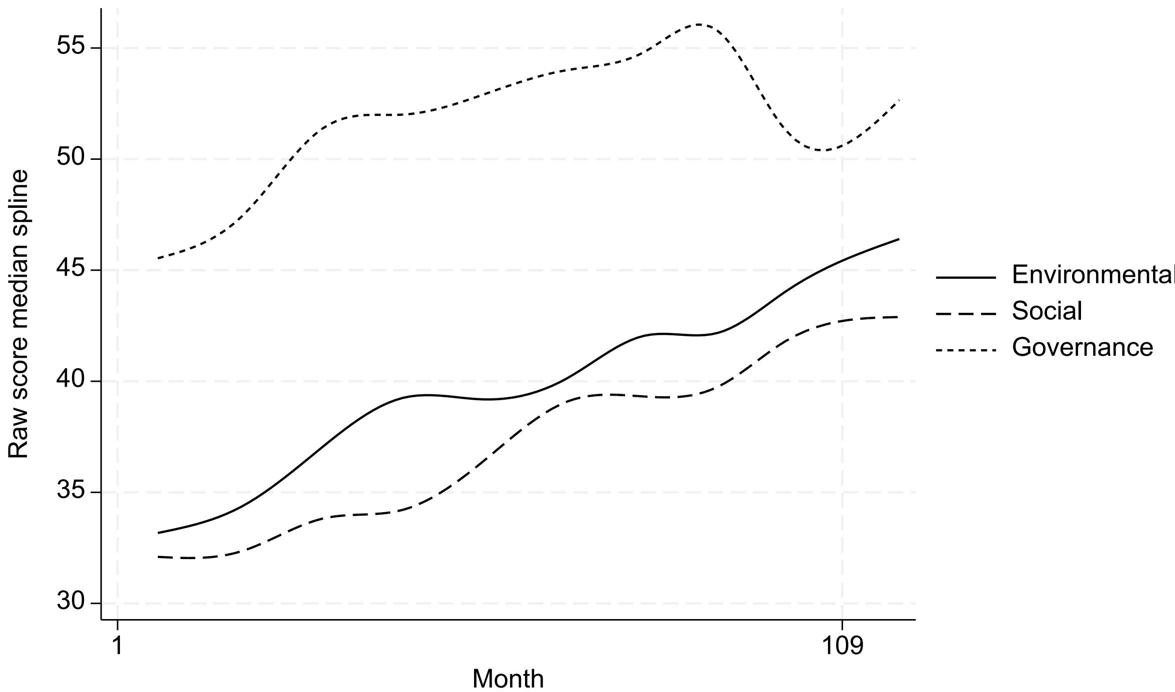


Figure A.2. Raw score drift. We compute mean raw scores among environmental (social, governance) criteria each month from August 2009 through September 2019. This figure displays median splines of these values. Data are from Sustainalytics.

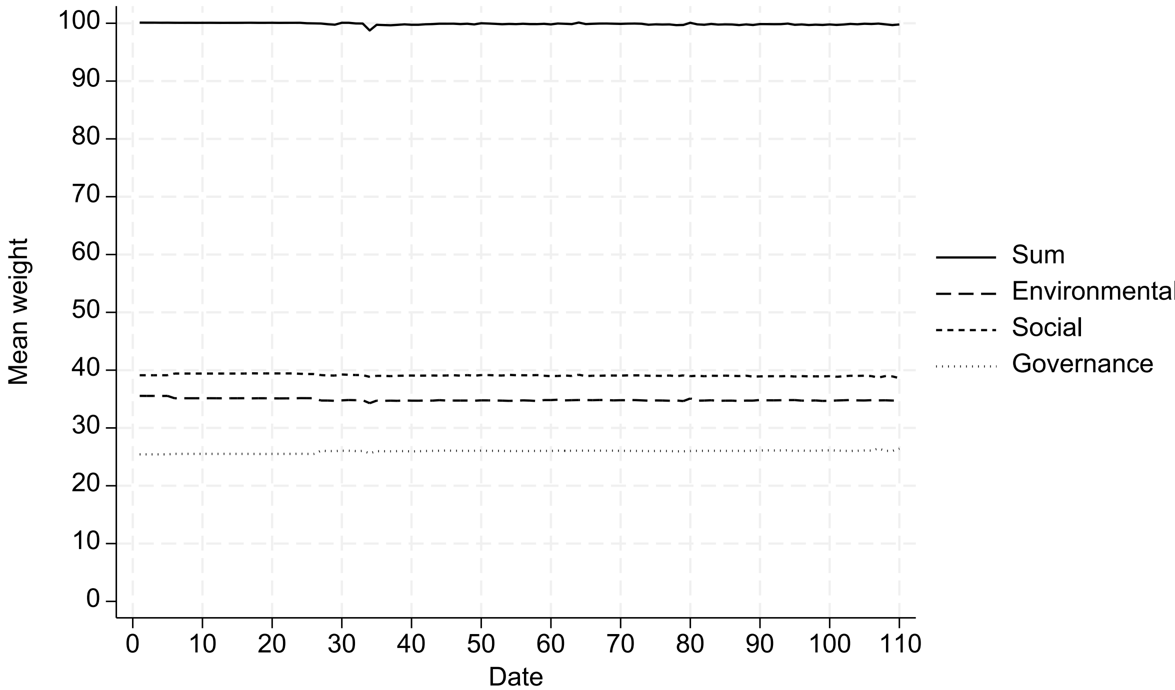
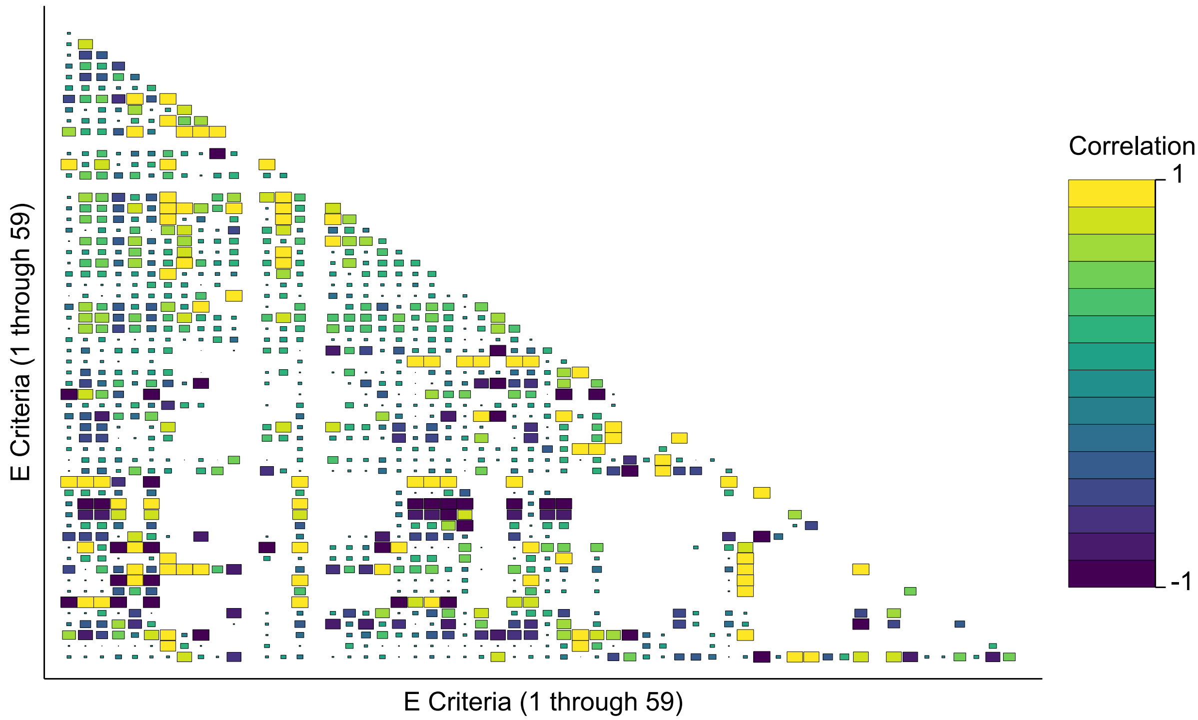
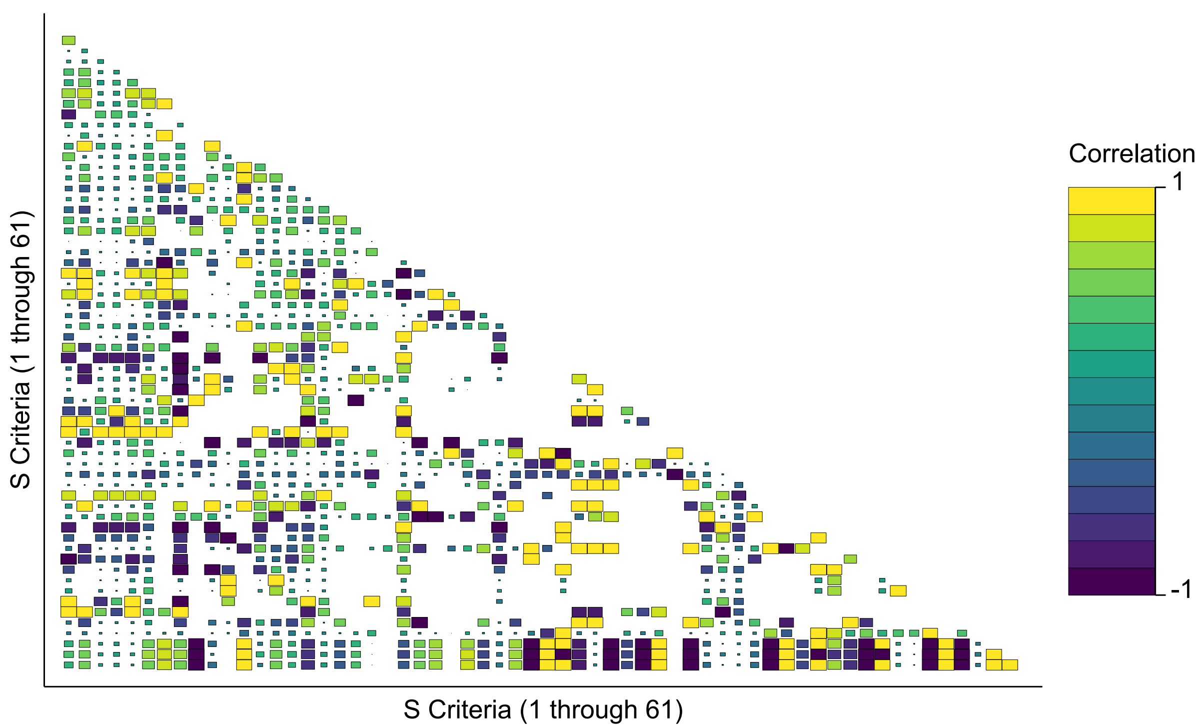


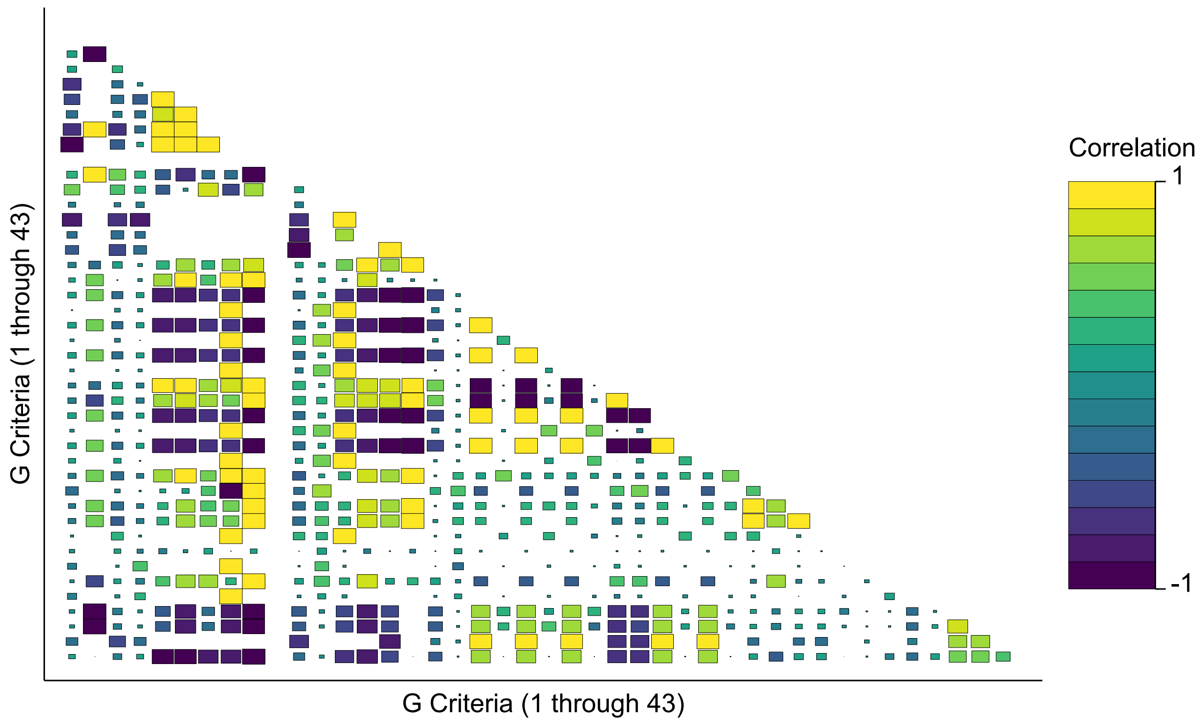
Figure A.3. Describing criteria weights through time. This figure displays means of criteria weights after summing them by firm-pillar-month. The sample runs from August 2009 through September 2019. Data are from Sustainalytics.



Panel A – Environmental criteria change correlation matrix

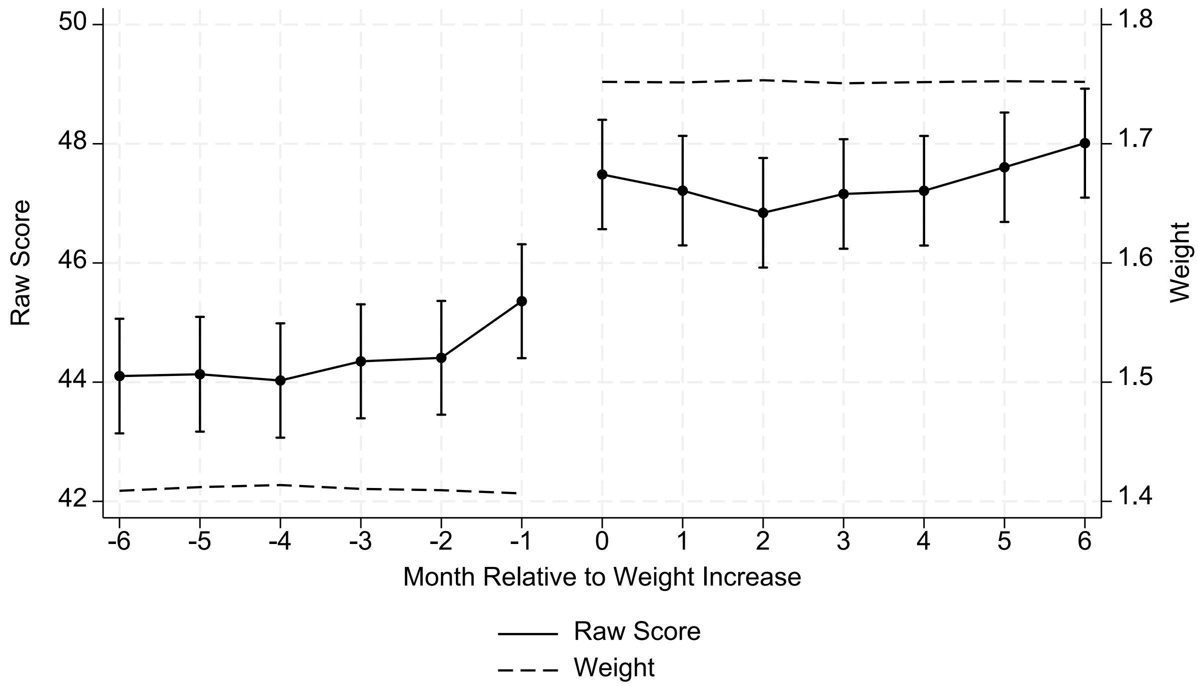


Panel B – Social criteria change correlation matrix

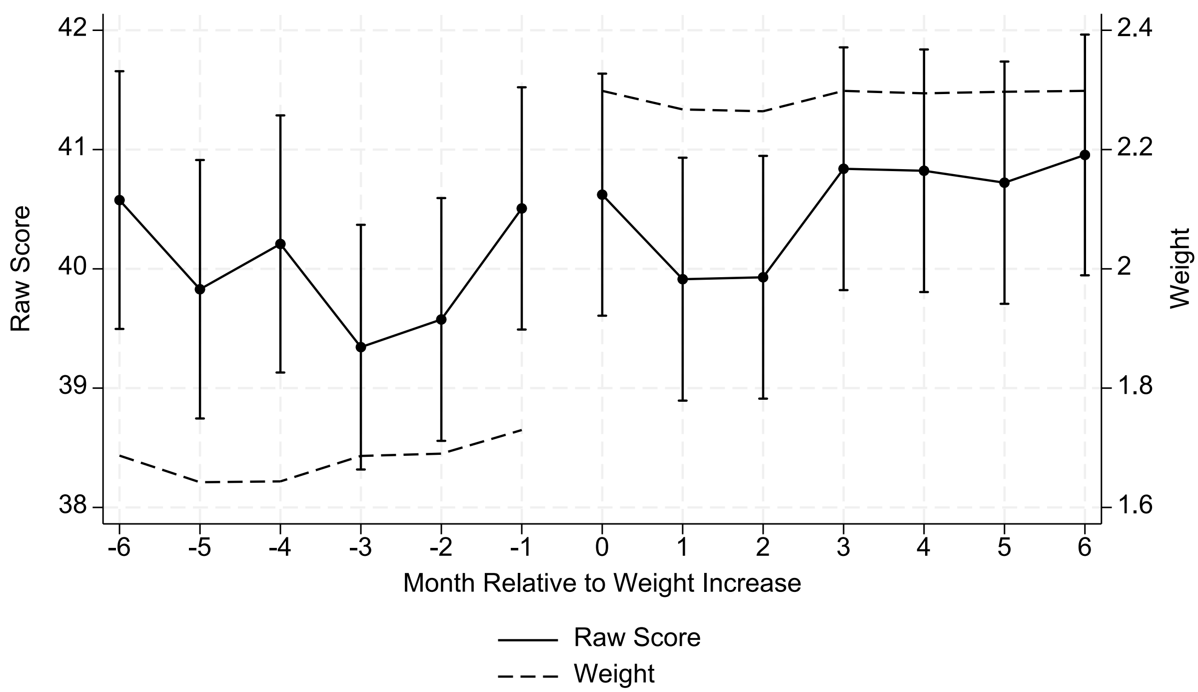


Panel C – Governance criteria change correlation matrix

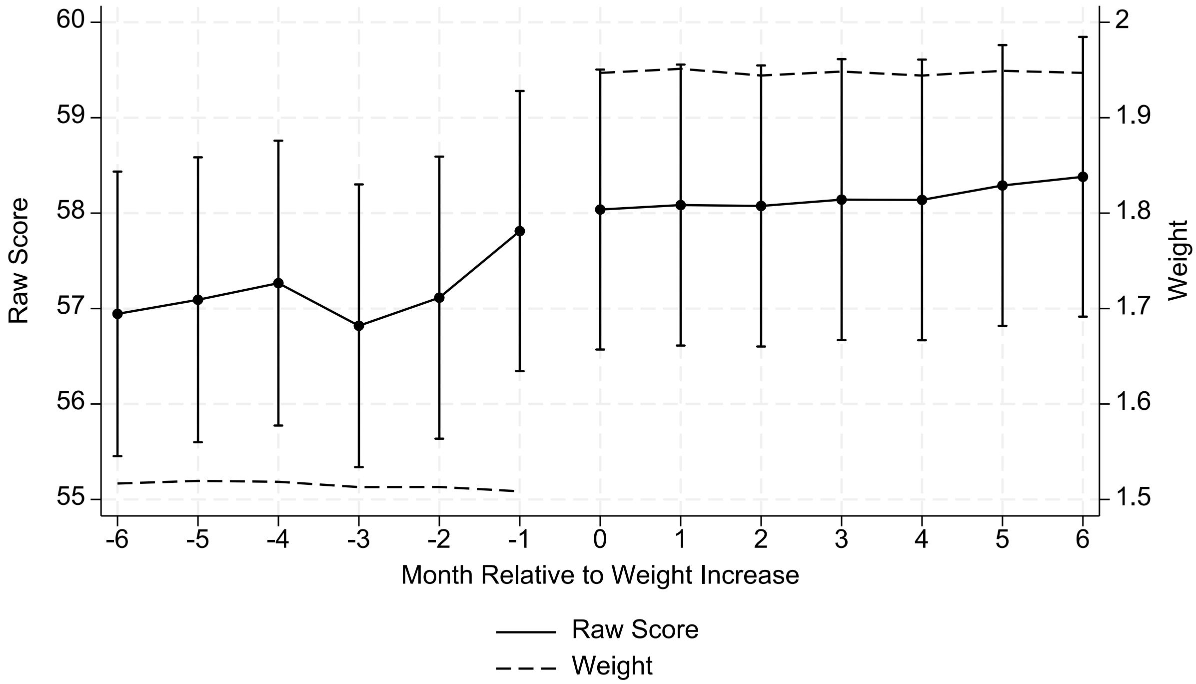
Figure A.4. Correlation matrices of changes in criteria weights. This figure displays correlations of changes in criteria weights from one firm-month observation to the next. The sample runs from August 2009 through September 2019. Panel A (B, C) displays correlations among criteria for the environmental (social, governance) pillar. Data are from Sustainalytics.



Panel A – Environmental pillar criteria



Panel B – Social pillar criteria



Panel C – Governance pillar criteria

Figure A.5. Raw scores around criteria weight increases by pillar. Panel A (B, C) of this figure displays mean raw scores around firm-criteria-months when weights associated with the Environmental (Social, Governance) pillar increase. Range caps represent 95% confidence intervals. Data are from Sustainalytics.

Table A.I
Data Example

This table provides an example of firm-criteria-month observations associated with an anonymized firm for the months of September and October, 2011. Shading indicates a contemporaneous increase in weight and raw score. Data are from Sustainalytics.

			September		October	
Pillar	Criteria	Code	Weight	Raw score	Weight	Raw score
Environmental	Formal Environmental Policy	e_1_1	.5%	25	1%	25
Environmental	Carbon Intensity Trend	e_1_10	2%	0	.5%	0
Environmental	% Primary Energy Use from Renewables	e_1_11	1%	0	.5%	0
Environmental	Operations Related Controversies or Incidents	e_1_12	3%	100	6.5%	100
Environmental	Environmental Management System	e_1_2	.5%	60	2%	60
Environmental	External Certification of EMS	e_1_3	2%	0	2%	0
Environmental	Programmes & Targets to Reduce Water Use	e_1_3_4	2%	25	2%	100
Environmental	Other Programmes to Reduce Key Environmental Impacts	e_1_3_5	1%	25	1.5%	50
Environmental	Environmental Fines and Non-monetary Sanctions	e_1_4	1%	100	1%	100
Environmental	Participation in Carbon Disclosure Project	e_1_5	.5%	25	.5%	25
Environmental	Scope of Corporate Reporting on GHG Emissions	e_1_6	1%	0	.5%	0
Environmental	Programmes and Targets to Reduce GHG Emissions from own operations	e_1_7	1.5%	50	1%	50
Environmental	Programmes and Targets to Increase Renewable Energy Use	e_1_8	1%	25	.5%	25
Environmental	Carbon Intensity	e_1_9	1%	0	.5%	0
Environmental	Formal Policy or Programme on Green Procurement	e_2_1	1%	30	1%	30
Environmental	Programmes to Improve the Environmental Performance of Suppliers	e_2_1_1	1%	0	2%	25
Environmental	External Environmental Certification Suppliers	e_2_1_2	1%	0	2%	0
Environmental	Programmes and Targets to Stimulate Sustainable Agriculture	e_2_1_3	2%	0	2%	25
Environmental	Programmes and Targets to Stimulate Sustainable Aquaculture/Fisheries	e_2_1_4	2%	25	1%	50
Environmental	Environmental Supply Chain Incidents	e_2_2	3%	100	4%	100
Environmental	Sustainability Related Products & Services	e_3_1_1	2%	0	3%	0
Environmental	Organic Products	e_3_1_8	2%	0	3%	0
Environmental	Products & Services Related Controversies or Incidents	e_3_2	3%	100	2%	100
	Sum		*35%*		*40%*	
Social	Policy on Freedom of Association	s_1_1	2%	0	1%	0
Social	Formal Policy on Working Conditions	s_1_1_1	2%	25	1%	25
Social	Formal Policy on the Elimination of Discrimination	s_1_2	1.5%	50	1%	50
Social	Programmes to Increase Workforce Diversity	s_1_3	1.5%	25	1%	25
Social	Percentage of Employees Covered by Collective Bargaining Agreements	s_1_4	2%	0	1%	0
Social	Employee Turnover Rate	s_1_5	2%	0	.5%	0

(*Table continued next page*)

(*Table continued from previous page*)

Social	Percentage of Temporary Workers	s_1_5_1	2%	0	1%	0
Social	Top Employer Recognition	s_1_6	2%	75	.5%	75
Social	Employee Related Controversies or Incidents	s_1_7	3%	50	5%	50
Social	Scope of Social Supply Chain Standards	s_2_1	1%	50	1%	50
Social	Quality of Social Supply Chain Standards	s_2_1_1	1%	25	2%	25
Social	Supply Chain Monitoring System	s_2_2	2.5%	100	2%	100
Social	Fair Trade Products	s_2_2_4	2.5%	0	3%	0
Social	Social Supply Chain Incidents	s_2_3	3%	80	5%	80
Social	Customer Related Controversies or Incidents	s_3_3	3%	80	3%	80
Social	Activities in Sensitive Countries	s_4_1	1%	100	1%	100
Social	Society & Community Related Controversies or Incidents	s_4_3	3%	100	3%	100
Social	Guidelines for Philanthropic Activities and Primary Areas of Support	s_5_1	1%	75	.75%	75
Social	Corporate Foundation	s_5_2	2%	100	.75%	100
Social	Percent Cash Donations of NEBT	s_5_3	2%	25	1.5%	25
	Sum		*40%*		*35%*	
Governance	Policy on Bribery and Corruption	g_1_1	1%	100	1%	100
Governance	Whistleblower Programmes	g_1_2	1.5%	50	2%	50
Governance	Signatory to UN Global Compact	g_1_3	1%	0	1%	0
Governance	Tax Transparency	g_1_4	1.5%	0	2%	0
Governance	Business Ethics Related Controversies or Incidents	g_1_5	3%	100	4%	100
Governance	CSR Reporting Quality	g_2_1	1.25%	25	1%	25
Governance	Audit Committee Independence	g_2_10	.5%	100	.25%	100
Governance	Non-Audit Fees Relative to Audit Fees	g_2_11	.5%	100	.25%	100
Governance	Compensation Committee Independence	g_2_12	.5%	100	.25%	100
Governance	Governance Related Controversies or Incidents	g_2_13	3%	100	4.25%	100
Governance	External Verification of CSR Reporting	g_2_2	.5%	0	1%	0
Governance	Disclosure of Directors' Remuneration	g_2_3	.5%	100	.25%	100
Governance	Disclosure of Directors' Biographies	g_2_4	.5%	100	.25%	100
Governance	Oversight of ESG Issues	g_2_5	1.25%	50	1%	50
Governance	Executive Compensation Tied to ESG Performance	g_2_6	1.25%	0	1%	0
Governance	Board Diversity	g_2_7	1.25%	60	1%	60
Governance	Separation of Board Chair and CEO Roles	g_2_8	.5%	0	.5%	0
Governance	Board Independence	g_2_9	.5%	100	1%	100
Governance	Policy on Political Involvement and Contributions	g_3_1	1%	25	.75%	25
Governance	Total Value of Political Contributions or Political Spending	g_3_2	1%	50	.75%	50
Governance	Public Policy Related Controversies or Incidents	g_3_4	3%	100	1.5%	100
	Sum		*25%*		*25%*	

Table A.II
Summary Statistics Split by ESG Pillar

This table displays summary statistics for firm-month-criteria observations from August 2009 through September 2019. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Δ *Raw score* (Δ *Weight*) is the change in *Raw score* (*Weight*) from the preceding firm-month-criteria. Data for raw scores and weights are from Sustainalytics. *Preparedness* (*Disclosure*, *Performance*) is an indicator variable taking a value of one if the criteria relates to ESG preparedness (disclosure, performance). *Firm size* is the natural log of the firm's total assets. *Dividends* is the firm's total dividends divided by total assets. *Cash* is the firm's cash divided by total assets. *Leverage* is the firm's total debt including current divided by total assets. *ROA* is the firm's net income divided by total assets. *BTM* is the firm's book value per share divided by its price close at the end of the calendar year. *CAPX* is the firm's capital expenditures divided by total revenue. *SG&A* is the firm's selling, general, and administrative expense divided by total revenue. *R&D* is the firm's research and development expense divided by total revenue. *R&D* takes a value of zero if it is missing. Balance sheet and income statement data are from Compustat. *Analysts* is the natural logarithm of one plus the number of equity analysts covering the firm. Analyst data are from IBES. We impute firm-year balance sheet, income statement, and analyst data to firm-month-criteria observations in the same calendar year. Panel A (B, C) reports summary statistics for the observations from the full sample associated with the Environmental (Social, Governance) pillar.

Panel A: Environmental Pillar Firm-month-criteria Observations

	N	Mean	Median	SD	10^{th} pctl	90^{th} pctl
Raw score	568,174	46.45	40	42.60	0	100
Δ Raw score	568,174	0.11	0.00	6.83	0.00	0.00
Weight	568,174	1.75	1.11	1.55	0.50	3.00
Δ Weight	568,174	0.00	0.00	0.17	0.00	0.00
Preparedness	568,174	0.33	0.00	0.47	0.00	1.00
Disclosure	568,174	0.12	0.00	0.32	0.00	1.00
Performance	568,174	0.55	1.00	0.50	0.00	1.00
Firm size	568,174	9.52	9.43	1.22	8.22	11.13
Dividends	568,174	0.03	0.02	0.04	0.00	0.06
Cash	568,174	0.10	0.08	0.08	0.01	0.21
Leverage	568,174	0.27	0.25	0.18	0.06	0.49
ROA	568,174	0.06	0.06	0.08	0.00	0.14
BTM	568,174	0.39	0.32	0.43	0.09	0.75
CAPX	568,174	0.11	0.04	0.58	0.01	0.19
SG&A	568,174	0.22	0.19	0.23	0.04	0.39
R&D	568,174	0.04	0.00	0.10	0.00	0.13
Analysts	568,174	1.37	1.79	1.15	0.00	2.71

Panel B: Social Pillar Firm-month-criteria Observations

	N	Mean	Median	SD	10^{th} pctl	90^{th} pctl
Raw score	602,371	47.11	50	42.02	0	100
Δ Raw score	602,371	0.10	0.00	5.95	0.00	0.00
Weight	602,371	1.85	1.50	1.28	0.60	3.19
Δ Weight	602,371	0.00	0.00	0.14	0.00	0.00
Preparedness	602,371	0.30	0.00	0.46	0.00	1.00
Disclosure	602,371	0.02	0.00	0.13	0.00	0.00
Performance	602,371	0.68	1.00	0.47	0.00	1.00
Firm size	602,371	9.51	9.40	1.23	8.20	11.13
Dividends	602,371	0.03	0.02	0.03	0.00	0.06
Cash	602,371	0.10	0.08	0.09	0.01	0.21
Leverage	602,371	0.26	0.24	0.17	0.06	0.48
ROA	602,371	0.06	0.06	0.08	0.00	0.14
BTM	602,371	0.39	0.32	0.43	0.10	0.75
CAPX	602,371	0.11	0.04	0.61	0.01	0.18
SG&A	602,371	0.23	0.20	0.24	0.05	0.41
R&D	602,371	0.04	0.00	0.12	0.00	0.14
Analysts	602,371	1.38	1.79	1.16	0.00	2.71

Panel C: Governance Pillar Firm-month-criteria Observations

	N	Mean	Median	SD	10^{th} pctl	90^{th} pctl
Raw score	616,683	60.36	80	42.81	0	100
Δ Raw score	616,683	0.10	0.00	6.91	0.00	0.00
Weight	616,683	1.20	1.00	1.06	0.25	3.00
Δ Weight	616,683	0.00	0.00	0.08	0.00	0.00
Preparedness	616,683	0.34	0.00	0.47	0.00	1.00
Disclosure	616,683	0.10	0.00	0.30	0.00	0.00
Performance	616,683	0.57	1.00	0.50	0.00	1.00
Firm size	616,683	9.51	9.42	1.23	8.21	11.13
Dividends	616,683	0.03	0.02	0.03	0.00	0.06
Cash	616,683	0.10	0.08	0.09	0.01	0.21
Leverage	616,683	0.27	0.25	0.18	0.06	0.49
ROA	616,683	0.06	0.06	0.08	0.00	0.14
BTM	616,683	0.38	0.32	0.44	0.10	0.74
CAPX	616,683	0.10	0.04	0.58	0.01	0.18
SG&A	616,683	0.23	0.19	0.24	0.04	0.41
R&D	616,683	0.04	0.00	0.11	0.00	0.14
Analysts	616,683	1.36	1.79	1.15	0.00	2.71

Table A.III
The Relation between Scores and Weights Controlling for Additional Leads/Lags

This table displays OLS regression results with *Raw Score* as the dependent variable and *Weight* as the independent variable of interest. *Raw score* (*Weight*) is the raw ESG score (weight expressed as a percentage) for a firm-month-criteria. Data for raw scores and criteria weights are from Sustainalytics. The sample includes firm-month-criteria observations from August 2009 through September 2019. Standard errors are clustered at the peer group, month, and criteria levels. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
$Weight_{-6}$		0.07
		(0.17)
$Weight_{-5}$		0.09
		(0.08)
$Weight_{-4}$		0.01
		(0.15)
$Weight_{-3}$	0.01	-0.03
	(0.16)	(0.10)
$Weight_{-2}$	-0.10	-0.11
	(0.11)	(0.12)
$Weight_{-1}$	-0.19	-0.22
	(0.12)	(0.14)
Weight	**0.76**	**0.65**
	(0.34)**	**(0.26)****
$Weight_{+1}$	-0.23	-0.28
	(0.10)**	(0.12)**
$Weight_{+2}$	-0.28	-0.22
	(0.16)*	(0.14)
$Weight_{+3}$	0.10	0.08
	(0.08)	(0.12)
$Weight_{+4}$		-0.10
		(0.13)
$Weight_{+5}$		0.17
		(0.13)
$Weight_{+6}$		-0.06
		(0.13)
Firm × Criteria FE	Yes	Yes
Firm × Month FE	Yes	Yes
Criteria × Month FE	Yes	Yes
Adjusted R^2	0.85	0.86
N	1,501,033	1,205,466

Table A.IV
Criteria Classifications by Preparedness, Disclosure, or Performance

This table displays classifications of Sustainalytics 163 criteria into categories of preparedness (1), disclosure (2), or performance (3). Criteria related to the environmental (social, governance) pillar begin with the prefix "e" ("s", "g"). Criteria codes and definitions are available from Sustainalytics.

Code	Category	Code	Category	Code	Category	Code	Category	Code	Category
e_1_1	1	e_2_1_5	3	s_1_1_1	1	s_3_1_12	3	g_1_4_3	1
e_1_2	3	e_2_1_6	1	s_1_5_1	3	s_3_2_1	3	g_1_4_4	1
e_1_3	3	e_2_1_7	3	s_1_6_1	3	s_4_1	3	g_1_4_5	1
e_1_4	3	e_2_1_8	3	s_1_6_2	1	s_4_3	3	g_1_4_6	1
e_1_5	2	e_2_1_9	1	s_1_6_3	1	s_4_2_1	1	g_2_1	2
e_1_6	2	e_2_1_10	3	s_1_6_4	3	s_4_2_2	1	g_2_2	1
e_1_7	1	e_3_2	3	s_1_6_5	3	s_4_2_3	1	g_2_3	1
e_1_8	1	e_3_1_1	3	s_1_6_6	3	s_4_2_4	1	g_2_4	1
e_1_9	3	e_3_1_2	3	s_1_6_2_1	3	s_4_2_5	1	g_2_5	1
e_1_10	3	e_3_1_3	3	s_2_1	3	s_4_2_6	1	g_2_6	3
e_1_11	3	e_3_1_4	3	s_2_2	3	s_4_2_7	1	g_2_7	3
e_1_12	3	e_3_1_5	3	s_2_3	3	s_4_2_8	1	g_2_8	3
e_1_1_1	2	e_3_1_6	3	s_2_1_1	3	s_4_2_9	1	g_2_9	3
e_1_2_1	1	e_3_1_7	1	s_2_1_2	3	s_4_2_10	1	g_2_10	3
e_1_2_2	2	e_3_1_8	3	s_2_1_3	1	s_4_2_11	1	g_2_11	3
e_1_2_3	2	e_3_1_9	1	s_2_1_3_1	1	s_4_2_12	1	g_2_12	3
e_1_2_4	2	e_3_1_10	3	s_2_2_1	2	s_4_2_13	1	g_2_13	3
e_1_2_6	3	e_3_1_11	3	s_2_2_2	2	s_4_2_14	3	g_2_5_1	3
e_1_2_7	3	e_3_1_12	3	s_2_2_3	3	s_5_1	1	g_3_1	1
e_1_2_8	3	e_3_1_13	1	s_2_2_4	3	s_5_2	3	g_3_2	3
e_1_3_2	1	e_3_1_14	3	s_2_2_2_1	3	s_5_3	3	g_3_4	3
e_1_3_3	1	e_3_1_15	3	s_3_3	3	g_1_1	1	g_3_3_1	2
e_1_3_4	1	e_3_1_16	3	s_3_1_1	1	g_1_2	1	g_1_3_6	3
e_1_3_5	1	e_3_1_17	3	s_3_1_2	1	g_1_3	3	g_2_3_1	3
e_1_7_0	1	e_1_2_6_2	3	s_3_1_3	1	g_1_4	2	g_2_4_1	3
e_1_7_1	1	e_1_3_1	3	s_3_1_4	1	g_1_5	3	g_2_7_1	3
e_1_7_2	1	s_1_1	1	s_3_1_5	3	g_1_1_1	1	g_2_8_1	3
e_2_1	1	s_1_2	1	s_3_1_6	1	g_1_3_1	3	g_2_9_1	3
e_2_2	3	s_1_3	1	s_3_1_7	1	g_1_3_2	1	g_2_10_1	3
e_2_1_1	1	s_1_4	3	s_3_1_8	3	g_1_3_3	3	g_2_11_1	3
e_2_1_2	3	s_1_5	3	s_3_1_9	1	g_1_3_4	3	g_2_12_1	3
e_2_1_3	1	s_1_6	3	s_3_1_10	3	g_1_3_5	2		
e_2_1_4	1	s_1_7	3	s_3_1_11	1	g_1_4_1	1		