Peer Groups Matter: Comparative Industry Effects in ESG Rating Disagreement

Alex G. Kim[1] Kyungmin Kim[2] Jongsub Lee[3]

January, 2026

Abstract

ESG ratings are widely used benchmarks but often diverge across providers. We identify the comparative industry effect—differences in how raters classify industries and construct peer groups for relative evaluation—as a key, overlooked driver of this divergence. Using variance decomposition and simulation analyses, we show that this effect explains 30% of industry-level differences in ESG scores, comparable in magnitude to measurement-based discrepancies. Divergence is especially pronounced for private and complex-ownership firms, where limited disclosure prompts greater reliance on industry benchmarks. Improving ESG comparability thus requires not only measurement standardization but also harmonization in industry classification and peer group construction.

Keywords: ESG Ratings, ESG Rating Disagreement, Industry Effect, Relative Performance Evaluation, Peer Effect

JEL Codes: G24, M14

[1]The University of Chicago, Booth School of Business. E-mail: alex.kim@chicagobooth.edu

[2]Seoul National University, Graduate School of Business. E-mail: kyungmin3157@snu.ac.kr

[3]Seoul National University, Graduate School of Business and University of Florida, Warrington College of Business. E-mail: jongsub.lee@snu.ac.kr

We appreciate insightful comments from Hans Christensen, Mark Kohlbeck, Florian Habermann, Bok Baik, Mee-ok Choi (discussant), Dong Beom Choi, Su-Keun Kwak, Scott Alan Robinson (discussant) and seminar participants at the 2026 Financial Accounting and Reporting Section (FARS) Conference, 2024 Korean Financial Management Association, 2023 American Accounting Association Conference, 2023 Korean Securities Association, 2023 Joint Conference with the Allied Korea Finance Associations, and 2022 Korean Accounting Association Winter Conference.

1 Introduction

As ESG ratings are widely used among both practitioners and academics, their reliability and informativeness have become increasingly important. While the number of raters has increased to more than 140 raters worldwide as of 2022, there exists significant disagreement in ESG ratings across raters (Berg et al., 2022b; Chen et al., 2025). A recent Wall Street Journal article by Sindreu and Kent (2018) highlights the contrasting results; R^2 between different ESG ratings is only 13%, while the same statistic for various credit ratings is nearly 82% (Figure 1). What explains this huge disagreement across ESG ratings with credit rating scores being generally consistent across similar set of raters? Despite this discrepancy being prevalent, a growing number of academic studies utilize ESG scores from different raters[1]. More importantly, the discrepancy in ESG ratings spurs confusion in the stock market (Serafeim and Yoon, 2022a; Berg et al., 2022a; Gibson Brandon et al., 2021) and discourages firms that aim to enhance their ESG profiles by making green technological innovations and/or improving social and governance related issues. It is, therefore, imperative to understand what causes rating discrepancy and how to mitigate such confusion in ESG ratings.

The literature on investigating potential causes of rating discrepancy is surprisingly scarce. Chatterji et al. (2016) provide initial evidence that significant heterogeneity exists across ESG raters. They warn that divergent ratings might adversely affect informed investment decisions in the stock market. Building on this finding, Berg et al. (2022b) show that the primary sources of ESG rating divergence are scope, weight, and measurement of each rater, all of which are firm-level evaluation metrics. As such, extant literature has so far documented the existence of rating divergence and identify some firm-level factors that contribute to this widely observed phenomenon.

If the rating discrepancy stems largely from idiosyncratic firm-level measurement noises, aggregating them to the industry level would likely reduce the error, thus leading to more harmonized industry-level evaluations across raters. Our observation suggests, however, the opposite. Our

[1] See for example Serafeim and Yoon (2022b), Lins et al. (2017), among many. One common solution to mitigate the rating discrepancy is to take the averaged score of different raters. However, such approach is problematic in that the averaged score may cancel out some additional information contained in divergent ratings themselves.

simple analysis reveals that raters tend to disagree even at the industry level and the magnitude of such disagreement is comparable to that of firm-level disagreement. What causes this initially puzzling industry-level disagreement? How large is the industry-level disagreement compared to other commonly known firm-level sources of rating disagreement?

To answer these questions, it should be noted that most of ESG rating agencies perform industry-based relative performance evaluation (see Table 1), whereas credit rating assessment is mostly on an absolute scale (e.g., a high leverage ratio takes its meaning in an absolute sense rather than on a relative basis in comparison to peers). What aggravates the problem is that ESG raters tend to define their own industry taxonomies or peer groups to assess the ESG scores within such subgroups[2]. Furthermore, materiality standards, which are known to have significant effects on stock returns, are also applied at the industry level (Khan et al., 2016).

Table 1 summarizes each rater's history, coverage, industry taxonomy, relative evaluation metrics, and industry-specific assessment criteria. As shown, each rater covers at least 8,500 firms and has undergone acquisitions. Note that ESG rating providers use proprietary industry classifications, such as the GICS (Global Industry Classification Standard) co-developed by MSCI and S&P. Refinitiv and Sustainalytics rely on their own proprietary industry classifications. Since ESG raters employ heterogeneous peer group classifications and apply a relative performance evaluation approach to these heterogeneous peer groups, the *comparative industry effect* emerges.

We define comparative industry effect as the systematic industry-level differences in ESG ratings that arise because of rating agencies' relative ESG performance evaluation practice. This effect encompasses both the assignment of a single firm to different industries by different raters (industry classification effect) and the use of different peer group compositions by different raters to evaluate a single firm (relative performance evaluation effect)[3], each of which may drive substantial divergence in ESG assessments.

[2] See Table 1 for detailed explanation about how raters treat industry-level evaluation differently.

[3] Although the comparative industry effect conceptually incorporates both the industry classification effect and the relative performance evaluation effect, our empirical analysis focuses on the aggregate comparative industry effect without disentangling the two channels, reflecting the practical difficulty of isolating these components given data constraints.

Consider British American Tobacco p.l.c as an illustrative example of industry classification effect. The firm is classified under the *Food & Beverages* industry by Refinitiv, but under the *Tobacco* industry by MSCI. The relative performance evaluation effect also arises as the firm has distinct peers even though they are assigned under the same industry classification. Specifically, S&P and MSCI both classify British American Tobacco p.l.c under the Tobacco industry. However, the firms constituting that industry differ substantially.[4]

We analyze widely covered U.S. and global firms that are commonly rated by the four major ESG rating agencies: MSCI, Refinitiv (also known as LSEG or Asset4), S&P Global, and Sustainalytics. Consistent with prior studies[5], we standardize the common sample to have zero mean and unit variance in the cross-section, allowing for direct and fair comparisons across raters.

Our simple analysis reveals that industry-level disagreement is persistent and has a non-trivial magnitude compared to firm-level disagreement. As an illustrative exercise, we calculate the standard deviation of ESG scores averaged by rater within each year–industry group (industry-level ESG scores), and document substantial heterogeneity in industry-level ESG scores across raters. Such heterogeneity is particularly salient for environmental scores and remains remarkably stable over time (see Figure 3).

To rigorously quantify the contribution of industry-level disagreement, we employ an R^2 decomposition. Applying variance decomposition techniques to standardized composite ESG scores at the SIC 2-digit industry level, we find that industry fixed effects alone explain over 50% of the variance in ESG scores, with explanatory power rising above 81% when industry-by-rater interactions are included.

We then extend our analysis to the firm level. Even in this setting, industry-by-rater effects remain substantial, explaining 4.68% of the variance, which is about 22% of the magnitude of the

[4]S&P's *Tobacco* industry in 2014 included Altria Group, Inc., British American Tobacco p.l.c., Imperial Brands p.l.c., ITC Ltd, Japan Tobacco Inc., Philip Morris International Inc., PT Gudang Garam Tbk, Swedish Match AB, and KT&G Corporation, while MSCI covered 36 peers. MSCI's *Tobacco* industry in 2014 included Alliance One International Inc., Altadis Emisiones Financieras S.A.U, ITC Ltd, JT International Berhad, Lorillard Tobacco Company, Multi Solutions II, Inc., Reynolds American Inc., Souza Cruz S.A., Swedish Match AB, The West Indian Tobaccom Company Limited, Universal Corporation, UST Inc., Vector Group Ltd., and more including S&P's 9 peers.

[5]See for example Berg et al. (2022b).)

firm-by-rater effects. Notably, this effect is even more pronounced in the environmental dimension, which is arguably least affected by subjective measurement and scope-related errors across raters (Berg et al., 2022b; Kim et al., 2025). The incremental R^2 from industry-by-rater interactions more than doubles to 10.04%, and its relative size rises to 52% of the firm-by-rater effect, nearly twice that observed for composite ESG scores. Collectively, our results demonstrate that industry-level divergence is not only persistent but also particularly consequential for environmental ratings.

We then separately investigate potential sources of industry-level rating divergence. To do so, we conduct a numerical simulation using firm-level data for which both Refinitiv and S&P Global provide raw values for identical ESG metrics. By focusing exclusively on the common samples and on indicators measured before any weights or scope adjustments are applied, we ensure that the analysis is based on directly comparable, unadjusted data inputs. This approach eliminates divergence that arises from differences in scope and weighting methodologies between the two rating agencies.

We decompose the simulated industry-level divergence into three components: (1) methodological differences across raters, (2) comparative industry effect, and (3) raw measurement discrepancies. Here, a methodological difference arises when one rater uses only one metric (i.e., performance) in the percentile rank scoring methodology, while the other rater averages multiple data points (i.e., disclosure, public documentation, or performance) (see Appendices A1 and A2 for more details). A raw measurement discrepancy results from using either absolute amount of emissions or revenue-normalized emission intensity in evaluating direct carbon emissions.

Our findings indicate that the comparative industry effect, arising from differences in industry classification and peer group construction across raters, constitutes approximately 30% of the total simulated divergence when discrepancies stemming from heterogeneous measurements are minimized. In this context, its magnitude is almost eight times greater than that of measurement divergence.

Even in settings with substantial measurement discrepancies, the comparative industry effect remains a significant component, accounting for roughly half the magnitude of measurement divergence. While these results are consistent with prior studies emphasizing measurement divergence, our analysis makes a novel contribution by identifying and quantifying the comparative industry effect as a distinct and economically meaningful driver, not explicitly addressed in earlier research.

Notably, this simulation result corroborates our variance decomposition, which shows that industry-level disagreement is more pronounced in the environmental dimension where measurement issues are generally less severe due to greater objectivity and data availability. The heightened disagreement observed in this context cannot be attributed solely to measurement divergence. These results collectively suggest the necessity of considering the comparative industry effect alongside measurement divergence to fully understand the sources of industry-level ESG rating disagreement.

To enhance the external validity of our simulation analysis, we also examine plausibly exogenous variation introduced by structural revisions to the Global Industry Classification Standard (GICS). In 2014, 2016, and 2018, GICS significantly reshaped peer group compositions by adding or eliminating industry groups and sub-industries. Although our theoretical simulation relies on a local sample and a fixed time period, we address this limitation by examining whether these exogenous shifts in industry classification amplify ESG rating disagreement across raters. Our regression results show that firms in industries affected by GICS revisions experience significantly higher industry-level rating disagreement, with the effect being more pronounced in the environmental dimension. In contrast, the impact of GICS revisions on firm-level disagreement is negligible.

Lastly, we study economic characteristics that are associated with ESG rating discrepancies. We show that industry-level disagreement tends to be higher among firms with lower levels of public disclosure and that operate in an environment that fosters complex organizational structures. In particular, private firms and firms that operate in countries where concentrated ownership is a common practice exhibit significantly greater industry-level disagreement compared to public firms and those with more dispersed ownership. This result suggests that ESG raters rely more heavily on industry-level benchmarks when firm-level disclosure is limited. Taken together, our findings imply that disclosure practices are associated with rater behavior and the consistency of

ESG assessments.

Our study contributes to the literature in two folds. First, we extend the literature by documenting that industry-level divergence constitutes the primary and persistent source of ESG rating disagreement across providers. While it is well recognized that ESG raters employ industry-level peer groups for relative performance evaluation, the magnitude and structural importance of industry-level disagreement in explaining cross-rater ESG rating divergence have been largely overlooked. Building on Schmidl and Muck (2025), who examine the role of peer group granularity in ESG score construction within a single rating framework, we move beyond within-rater peer group effects and highlight the importance of cross-rater differences in industry classification and peer group composition. By performing a transparent and robust disaggregation, we quantify the relative contributions of industry- and firm-level factors and demonstrate that industry-level disagreement constitutes a substantial and persistent share of ESG rating divergence.

Second, our findings provide novel and practical insights into the underlying sources of industry-level ESG rating disagreement. By introducing and empirically validating the concept of comparative industry effect, we identify a distinct and economically meaningful driver of rating divergence that has not been explicitly addressed in prior literature. Our simulation-based decomposition demonstrates that, when measurement divergence is minimized, the comparative industry effect becomes more pronounced than the measurement divergence itself, and remains significant even when measurement discrepancies are substantial. This insight is reinforced by our quasi-experimental analysis leveraging exogenous GICS structural revisions.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 describes the data and methodology. Section 4 reports the main empirical findings. Section 5 concludes.

Related Studies Several recent works document the existence of the disagreement in ESG ratings. Chatterji et al. (2016) provide initial evidence on the existence of rating disagreement. They find a lack of agreement across social ratings from six dominant raters (KLD, Asset4, Calvert, FTSE4Good, DJSI, and Innovest). In addition, they identify two potential sources of disagreement: theorization and commensurability. In a more recent study, Berg et al. (2022b) decompose the sources of rating divergence and analyze the relative importance of each source. They identify three primary sources of rating disagreement: measurement, scope, and weight. Among the three potential drivers, measurement divergence is identified as the main cause of rating divergence. However, these drivers are estimated based on pooled regression results at the firm level. We, on the other hand, take industry effects in to account motivated by the fact that ESG raters perform industry-level relative performance evaluation[6]. Relatedly, Schmidl and Muck (2025) show that peer group definitions within a single rating framework materially affect ESG score construction and their financial relevance. Building on this insight, we explicitly account for industry effects and show that heterogeneity in industry benchmarking across ESG providers constitutes an important and previously unexplored source of cross-provider ESG rating disagreement.

In this study, however, we focus on whether industry effects play a role in ESG rating disagreement. Despite its practical importance, prior works that investigate the effect of industry-level confusion on rating discrepancy is surprisingly scarce. One of the few prior works that acknowledges the role of industry-level approach is Khan et al. (2016). They study the materiality standards provided by the SASB and emphasize the differential importance of sustainability issues across industries[7]. While the authors do not provide direct evidence that industry-level materiality standards increase rating disagreement, their study highlights the importance of an industry-level approach in understanding highly divergent ESG scores. In particular, investors appear to recognize that

[6]Several recent papers study the economic consequences of ESG rating disagreement. Specifically, Serafeim and Yoon (2022a) and Berg et al. (2022a) investigate the effect of ESG rating disagreement on the capital market. Serafeim and Yoon (2022a) find that rating disagreement slows the incorporation of ESG-related news into stock prices. On the other hand, Berg et al. (2022a) show that reducing noise in ESG scores can lead to a stronger association between ESG scores and stock prices. Jointly, these studies suggest that rating discrepancy may aggravate the confusion in the stock market and thus distort the efficient allocation of resources.

[7]The Sustainability Accounting Standards Board (SASB) was established in 2011 with the objective of creating sustainability accounting standards to facilitate public companies in disclosing material and decision-relevant information to investors. By November 2018, SASB had formalized and published standards encompassing 77 industries across 11 sectors. Collectively, these codified standards delineate and offer guidance on 444 industry-specific disclosure topics. The development of SASB standards aims to identify sustainability issues that are reasonably anticipated to exert a material influence on the operational outcomes and financial condition of companies within their respective industries (Bochkay et al., 2025).

varying levels of materiality for sustainability issues should be assessed at the industry level (Khan et al., 2016; Grewal et al., 2021).

Chatterji et al. (2016) investigate whether the normalization of rating across industries affects discrepancy and find that four out of six raters (except KLD and Asset4) normalize their ratings by industries to relatively evaluate among industry peers. Although the paper does not explicitly address industry-level differences, its finding implies that industry-level evaluation is associated with rating divergence across raters.

2 Institutional Details: Industry Evaluation Models by ESG Raters

We use ESG ratings from four largest ESG raters, MSCI, Refinitiv (also referred to as LSEG or Asset4), Sustainalytics, and S&P Global. Each of these raters has a distinct historical background, which may influence the level of discrepancy observed in their ratings. In this section, we provide an overview of their evaluation metrics, with particular emphasis on their industry-level approach.

MSCI entered the ESG rating market through the acquisition of RiskMetrics in 2010 and GMI Ratings in 2014. MSCI currently offers the broadest coverage of firms among ESG raters. Refinitiv, which has provided ESG ratings under the Asset4 brand since 2002, acquired Asset4 from Thomson Reuters in 2009. On January 29, 2021, LSEG (London Stock Exchange Group) completed its all-share acquisition of Refinitiv[8]. Sustainalytics, founded in 2008, received a significant investment when Morningstar acquired approximately 40% of its shares in 2017. S&P Global initially offered ESG ratings based on Trucost carbon data and, in 2017, integrated ESG elements across its indices through S&P Dow Jones Indices (S&P DJI). S&P Global expanded its ESG rating offerings in 2019 with the acquisition of RobecoSAM [9] (SustainAbility, 2020).

[8]See London Stock Exchange Group (2021) for details on the LSEG acquisition of Refinitiv.

[9]Several recent papers study the quality of ESG ratings and the conflicts of interest that can arise from acquisitions of ESG providers.Specifically, Li et al. (2024) show that ESG ratings of the rated firms are likely to be upwardly biased when the acquirers are credit rating agencies and the rated firm is an existing-pay client in their credit rating business. Chen et al. (2025) find that the entry of the new ESG rating agency (i.e., Sustainalytics in 2010) decreases incumbents' ESG rating disagreements, indicating the improvement of the rating quality. Collectively, these findings suggest that competition affects the quality of ESG ratings.

Table 1 indicates that ESG rating providers employ distinct methodologies and industry classifications. These four ESG raters not only adopt different industry classifications but also employ relative performance evaluation at the industry level, as each emphasizes assessing companies' performance relative to their industry peers (see Table 1). This approach, widely mentioned in their methodologies or manuals, allows for a more contextualized evaluation of ESG factors within specific sectors.

In addition to their use of relative performance evaluation, the providers apply distinct industry-level measurement metrics. Refinitiv selects specific categories (e.g., emissions) and associated questions (e.g., Total CO2 equivalents, NOx emissions, or VOC emissions) at the industry level. Each category consists of multiple questions, with distinct question compositions and categories across industries, such as finance and manufacturing. Firms receive scores based on their industry classification, with category scores represented as percentile rankings of question scores within industry peers (see Appendix A for more details). Category weights for the Environmental and Social dimensions are provided at the industry level, while Governance score weights are determined at the country level. Additionally, MSCI applies category (or issue) scores and weights at the industry level for both the Environmental and Social dimensions. Sustainalytics adopts a unique approach to ESG performance evaluation[10]. Unlike other raters, Sustainalytics assesses a firm's overall unmanaged risk, which includes both unmanageable risks and manageable risks that remain unmanaged. Despite its distinctive measurement approach, Sustainalytics computes each firm's ESG rating based on risk exposure within subindustries. Finally, S&P Global calculates ESG scores using industry-specific questions, with varying weights assigned at the question, category, and dimension (E/S/G) levels across industries. The final ESG score is derived from the sum of the weighted scores at these levels.

As industry-specific components are reflected differently across rating agencies, discrepancies in the level of industry impact are inevitable. This paper seeks to uncover the sources of disagreement

[10] Sustainalytics has provided the ESG Risk Rating dataset since 2018. To address concerns regarding the purpose of the ESG Risk Rating, which is designed to assess the magnitude of a firm's unmanaged ESG risks, we use the ESG Legacy Rating, which evaluates a firm's overall ESG performance, for our analysis.

in ESG ratings at the industry level.

3 Data and Sample Selection

To assess the extent of rater disagreement and its cross-sectional determinants, we concentrate primarily on publicly traded firms operating in the United States and other global markets that receive comprehensive coverage from multiple ESG rating providers.[11]

The regional distribution shown in Panel A of Table 2 highlights the global scope of our sample composition. We determine the regional distribution using the country code extracted from the ISIN provided by MSCI. The Americas form the largest regional segment, comprising 50.73% of observations (4,424 firm-year observations), followed by Europe with 24.09% (2,101 observations) and Asia with 19.66% (1,714 observations). In Panel B, we present the distributions at the country level. The predominance of developed capital markets is even more apparent — the United States constitutes 39.90% of all observations (3,479 firm-years), with Japan as the second-largest country at 11.94% (1,041 observations), followed by the United Kingdom at 5.40% (471 observations).

In Panel C, we show the industry distribution in our sample based on two-digit SIC codes. Industries with high ESG relevance, like Metal Mining (2.41%, 210 observations) and Oil and Gas Extraction (3.03%, 264 observations), have noticeable representation. The largest industry groups include Chemicals and Allied Products (7.63%, 665 observations), Depository Institutions (7.45%, 650 observations), and Business Services (7.09%, 618 observations). Unlike the country-level distributions, sectoral diversity is generally more dispersed, suggesting that the rating disagreement is not likely driven by the several prominent industries.

We focus on four different raters (MSCI, Refinitiv – also known as LSEG or Asset4 –, Sustainalytics, and S&P) because they are one of the most highly cited ESG rating providers and use different industry classification regimes. All four raters provide a composite ESG score out of 100,

[11] Among the 1,742 firms covered by all four rating agencies, approximately 40% operate in the United States (see Table 2).

ranging from 0 (most negative) to 100 (most positive)[12]. Our final samples are commonly covered by the four raters at the end of each fiscal year. As a result, our final sample is restricted to 1,742 firms with 8,720 firm-year observations from 2013 to 2019. Since Sustainalytics provides legacy data only until 2019, we limit our sample period to 2019. We further require several financial fundamentals as control variables in regression analyses.

In Table 3, we report the descriptive details of ESG and environmental scores separately. The minimum values of ESG aggregate and environmental scores from Sustainalytics (which are 31[13] and 20, respectively) are relatively higher than those from the other three raters. Notably, the average scores in the merged common sample are systematically higher than those observed in each rater's original full coverage universe prior to merging (see Appendix Table A3). This pattern suggests that the merged sample across the four raters may be biased toward firms with relatively stronger ESG profiles. This result also indicates that some raters may be more lenient than others in grading, confirming the rater effects documented by Berg et al. (2022b). For this reason, following prior literature we use standardized Z-score throughout our analyses. The scores are standardized within cross-sections to have a zero mean and unit variance.

For regression analyses, we require firm-level financial fundamentals commonly associated with ESG scores, such as size, book-to-market ratio, leverage, and return on assets (ROA). We obtain firm fundamentals from the Compustat Annual database.

[12]For Refinitiv, ESG scores are provided on a scale from 0 (most negative) to 1 (most positive). To align this with a 100-point scale, we multiply the score by 100. Similarly, for MSCI, ESG scores are provided on a scale from 0 (most negative) to 10 (most positive), and we scale-up the MSCI ratings since they provide 10-point scale composite scores by multiplying the score by 10.

[13]The fact that Sustainalytics reports a minimum score of 31—substantially higher than other raters—may appear to be a merging artifact. However, Appendix Table A3 shows that even within each rater's original universe prior to merging, Sustainalytics's minimum score remains 31. Moreover, prior studies also report a similar pattern, with Sustainalytics's minimum score typically around 37 even after merging, confirming that this is not a merging error but a consistent feature of its scoring distribution.

4 The Existence of Industry-Level Disagreement

4.1 An Illustrative Example

We begin by illustrating how ESG raters diverge in their evaluation of industries. Figure 2 presents the 2014 ESG Z-scores assigned by MSCI and S&P for two sectors: Metal Mining (SIC 10; 28 firms) and Securities & Commodity Brokers, Dealers, Exchanges & Services (SIC 62; 28 firms). Given the highly tangible nature of mining operations and the relatively standardized environmental disclosures, one might expect ESG raters to exhibit substantial agreement in their assessments of Metal Mining firms. In contrast, ESG evaluations for securities firms whose business models are more opaque and governance-intensive would intuitively seem more subjective and thus more prone to inter-rater disagreement.

Empirically, however, we observe the opposite. For securities firms, the average ESG Z-scores assigned by MSCI and S&P are relatively aligned (–0.16 vs. –0.29). In contrast, the Metal Mining industry exhibits substantial divergence: MSCI assigns a markedly negative industry mean (–1.14), while S&P reports a positive average (+0.12), yielding a much larger inter-rater distance. This counterintuitive pattern motivates a central idea of our analysis: while transparent disclosures and data availability can promote measurement consistency across ESG raters, a significant portion of rating discrepancy remains unexplained.

To better understand the source of this puzzling inter-rater discrepancy in the Metal Mining industry, consider Barrick Mining Corporation, a leading precious metals producer. S&P classifies Barrick broadly under Metals & Mining and ranks it first among 26 peer firms (3.8th percentile). In contrast, MSCI places Barrick in the more narrowly defined Metals & Mining – Precious Metals category and ranks it third among 12 peers (25th percentile). Furthermore, S&P evaluates Metals & Mining industry positively overall whereas MSCI assesses Precious Metals category negatively. These differences result in a highly favorable ESG Z-score of +1.83 from MSCI and a sharply negative score of –1.17 from S&P. The difference in the two scores cannot be entirely attributed to data availability – Barrick voluntarily discloses ESG metrics in accordance with well-established SASB standards for the Metals & Mining sector, which are known for their high degree of standardization. The divergence, therefore, likely stems from differences in industry classification, which alter the composition and competitiveness of the peer group, and from subsequent relative performance evaluation (RPE) within those groups. In S&P's broader universe, Barrick's ESG profile stands out more favorably, whereas MSCI's narrower comparator set raises the benchmark and yields a less favorable relative position. This result shows how comparative industry effect, which is jointly driven by classification choices and peer-relative evaluation, can generate substantial rating discrepancies even for a highly transparent firm.

4.2 Market Consequences of ESG Disagreement

The above example well motivates our central research question: even in industries with public data and high-quality disclosures (i.e., less measurement noise), rater disagreement can be pronounced. Using the example of Barrick, we shed light on the potential importance of peer group selection in ESG score assessment. This raises an important question: does such disagreement have real consequences in financial markets?

Methodology To answer this question, we examine whether ESG rating disagreement increases abnormal stock return volatility, a widely used proxy for firm-specific information uncertainty. We measure disagreement using the absolute deviation of a given rater's ESG score from the consensus of the other raters for the same firm. Abnormal return volatility is defined as the sum of the absolute abnormal returns over the [1,+1] window around the last rater's announcement date. Abnormal returns are computed as residuals from a one-factor market model estimated over the pre-announcement window (200,11), where the CRSP value-weighted market return is used as the proxy for the market factor.

We further decompose this disagreement into industry-level and firm-level components to assess whether market uncertainty is driven primarily by idiosyncratic firm-level noise or by systematic disagreement shared across firms within the same industry. Industry-level component is the SIC-

2 digit-year average absolute disagreement, while firm-level component captures within-industry deviations from that average.

Empirical Results The results are reported in Table 10. Panel A of Table 10 shows that ESG rating disagreement has a positive and statistically significant effect on abnormal return volatility. When the last rater's ESG score deviates more from the consensus of the other raters, abnormal return volatility increases, indicating that ESG disagreement generates firm-specific uncertainty and heterogeneous investor interpretations.

Panel B decomposes this disagreement into industry-level and firm-level components. Strikingly, the increase in abnormal return volatility is driven by the industry-level component of disagreement. The coefficient on industry-level disagreement is positive and statistically significant across specifications, whereas the firm-level component is economically small and statistically indistinguishable from zero. This result implies that market uncertainty arises not from idiosyncratic noise in individual firms' ESG assessments, but from systematic disagreement shared across firms within the same industry. In other words, it is precisely the industry-level component of ESG disagreement that has economically meaningful consequences in financial markets.

Although abnormal return volatility is a market price–based measure, ESG ratings are primarily used by institutional investors and may convey limited new information to them about firm fundamentals. This raises the question of why ESG rating disagreement would nonetheless generate observable market reactions. We argue that ESG disagreement can matter not through an informational channel, but through a contracting channel that amplifies uncertainty about how ESG-related rules and constraints are applied. As documented in recent studies, the market for sustainability-linked loans (SLLs) has grown rapidly since 2017, making ESG ratings increasingly relevant as contractual inputs rather than informational signals (Kim et al. (2025)). In SLL contracts, borrowing costs are explicitly tied to ESG performance assessed by a designated rating agency, such that changes in ESG ratings directly affect loan pricing and contractual outcomes. For example, Kim et al. (2025) describe a general-purpose loan issued to Crown Holdings Inc., in which interest rates adjust by 2.5 basis points for each notch change in the ESG rating agency assigned by Sustainalytics. In such settings, ESG rating disagreement across agencies—where one rater signals improvement while others do not—creates uncertainty not about firm fundamentals, but about which ESG assessment will bind in contractual arrangements and related eligibility constraints. This contractual ambiguity can generate heterogeneous interpretations among market participants and increase short-window return volatility, even in the absence of systematic mispricing or new fundamental information.

4.3 A Quantitative Exercise

In this section, we perform a simple quantitative analysis to demonstrate the existence of rater disagreement at the industry level.

Methodology We follow prior literature (e.g., Serafeim and Yoon, 2022a) to measure ESG rating disagreement across raters in two different ways: the standard deviation (σ) and the Euclidean distance (d) of four ratings.

To ensure comparability, we assume that there exists a *true* industry classification regime. We use a widely available taxonomy, the two-digit SIC classifications. Although ESG rating providers often rely on proprietary industry taxonomies, this common framework allows us to abstract away from rater-specific classifications and focus on cross-rater variation in ESG scores within a unified industry structure.

Let R_{rijt} denote the ESG score of firm i in industry j in year t assigned by rater r. We define R_{rjt} as the industry-level ESG score of industry j in year t by rater r, computed as the arithmetic average of ESG scores for all firms in the same industry-year group, i.e., $R_{rjt} = \frac{1}{N_{jt}} \sum_{i \in (j,t)} R_{rijt}$, where N_{jt} is the number of firms in industry j in year t. We decompose the total score as $R_{rijt} = R_{rjt} + \epsilon_{rijt}$, where ϵ_{rijt} captures the idiosyncratic component, which is the deviation of firm i's score from the rater-specific industry mean.

This decomposition allows us to measure disagreement at two levels: (i) industry-level disagreement and (ii) firm-level disagreement. Industry-level disagreement is defined as the cross-rater standard deviation of R_{rjt} for each (j,t) group, while firm-level disagreement is defined as the cross-rater standard deviation of ϵ_{rijt} for each firm-year observation. Total disagreement is the standard deviation of R_{rijt}. When raters perfectly agree on industry-level assessments, the variance of R_{rjt} will be close to zero. Likewise, if they agree on firm-level assessments conditional on industry, the variance of ϵ_{rijt} will be close to zero. On the other hand, if raters do not agree on industry-level assessment, we expect the variance of R_{rjt} to be positive.

Similarly, we also use averaged pairwise Euclidean distance[14]. Throughout the paper, we use standard deviations to measure rating disagreement. To ensure the robustness of our results, we also include findings based on the Euclidean distance measure in Appendix E. The results are consistent regardless of which method we use.

Empirical Results The results are reported in Table 4. Our disaggregation model separates total disagreement into systematic industry-level differences in rating approaches and idiosyncratic firm-level assessment variations. Specifically, industry-level ESG disagreement (mean of 0.23 for standard deviation and 0.28 for Euclidean distance) represents approximately 42% of the corresponding firm-level disagreement measures (mean of 0.55 for standard deviation and 0.68 for Euclidean distance), respectively[15]. This finding indicates that raters exhibit differences in their industry-level assessment frameworks. The relative magnitude of industry-level disagreement becomes even more pronounced for environmental scores, where industry-level disagreement accounts for approximately 66% of firm-level disagreement (0.34 versus 0.52). The escalation in the

[14] Total disagreement (TD^d), industry-level disagreement (ID^d), and firm-level disagreement (FD^d) are as follows:

$$TD^d = \frac{|R_{1ijt} - R_{2ijt}| + |R_{1ijt} - R_{3ijt}| + |R_{1ijt} - R_{4ijt}| + |R_{2ijt} - R_{3ijt}| + |R_{2ijt} - R_{4ijt}| + |R_{3ijt} - R_{4ijt}|}{6} \tag{1}$$

$$ID^d = \frac{|R_{1jt} - R_{2jt}| + |R_{1jt} - R_{3jt}| + |R_{1jt} - R_{4jt}| + |R_{2jt} - R_{3jt}| + |R_{2jt} - R_{4jt}| + |R_{3jt} - R_{4jt}|}{6} \tag{2}$$

$$FD^d = \frac{|\epsilon_{1ijt} - \epsilon_{2ijt}| + |\epsilon_{1ijt} - \epsilon_{3ijt}| + |\epsilon_{1ijt} - \epsilon_{4ijt}| + |\epsilon_{2ijt} - \epsilon_{3ijt}| + |\epsilon_{2ijt} - \epsilon_{4ijt}| + |\epsilon_{3ijt} - \epsilon_{4ijt}|}{6}. \tag{3}$$

[15] One should note that, however, the sum of industry-level disagreement and firm-level disagreement do not equal the total disagreement. This is because industry-level assessments and the residual terms are not orthogonal in practice.

relative importance of industry-level disagreement is particularly noteworthy given that environmental metrics are widely regarded as having superior measurement objectivity and data availability compared to social and governance dimensions.

Figure 3 presents the time-series trends of industry-level, firm-level, and total disagreement across raters. Panel A plots the trends for composite ESG scores. Both firm-level and industry-level disagreement remain relatively stable over time, although industry-level disagreement shows a slight downward trend. Panel B repeats the analysis for environmental scores. While the average firm-level disagreement is comparable to that observed for composite ESG scores, the average industry-level disagreement is substantially higher than in the composite ESG case.

If firm-level disagreement were driven primarily by idiosyncratic measurement noise, one might expect aggregation to the industry level to dampen such noise, as firm-specific errors should partially offset each other. Our results, however, point in the opposite direction, suggesting that factors beyond random measurement error play an important role in shaping raters' judgments. One plausible explanation lies in the fact that many ESG rating methodologies benchmark firms against industry baselines. When these baselines differ across raters, they reflect deeper divergences in how raters conceptualize and operationalize the evaluation of relative ESG performance at the industry level, which are differences that can materially contribute to overall rating divergence.

5 Understanding Industry-Level Disagreement

So far, we have confirmed that raters tend to disagree at the industry level. In this section, we aim to understand the magnitude of such divergence in relation to other sources of rating disagreement.

5.1 Variance Decomposition

In this section, we employ a variance decomposition approach to understand the magnitude of rating dispersion at the industry level. This approach allows us to quantify the explanatory power of rater, year, industry, and their interactions in determining ESG score disagreement among raters.

Methodology We examine the degree to which these fixed effects improve explanatory power in the following series of regressions:

$$R_{rjt} = \alpha_r \mathbb{1}_r \qquad +\epsilon_{rjt,1} \tag{4}$$

$$R_{rjt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_t \qquad +\epsilon_{rjt,2} \tag{5}$$

$$R_{rjt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_t + \gamma_j \mathbb{1}_j \qquad +\epsilon_{rjt,3} \tag{6}$$

$$R_{rjt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_t + \gamma_j \mathbb{1}_j + \theta_{jr} \mathbb{1}_{j\times r} \qquad +\epsilon_{rjt,4} \tag{7}$$

R_{rjt} represents the ESG Z-score of industry j in year t as assigned by rater r. It is calculated as the arithmetic average of R_{rijt}, which is the ESG Z-score (or environmental Z-score) of firm i within industry j in year t by rater r. Here, $\mathbb{1}_r$, $\mathbb{1}_t$, and $\mathbb{1}_j$ denote dummy variables for each rater, year, and SIC-2 digit industries, respectively. $\mathbb{1}_{j\times r}$ is an interaction term between SIC-2 digit industries and rater fixed effects.

We further extend our analysis to the firm level by employing a nested variance decomposition. This approach enables us to separately identify and quantify the contributions of both industry- and firm-level effects within a unified framework. R_{rijt} represents the ESG Z-score assigned by rater r to firm i in industry j during year t. This score represents the rater's evaluation of firm i's ESG performance in a given industry-year context. We then sequentially include rater, year, industry, and their interactions, as well as firm-specific ($\mathbb{1}_i$) and firm-by-rater ($\mathbb{1}_{i\times r}$) fixed effects.

Specifically, we estimate the following six regressions:

$$R_{rijt} = \alpha_r \mathbb{1}_r \qquad +\epsilon_{rjt,1} \tag{8}$$

$$R_{rijt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_t \qquad +\epsilon_{rjt,2} \tag{9}$$

$$R_{rijt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_j + \gamma_{rt} \mathbb{1}_{r\times t} \qquad +\epsilon_{rijt,3} \tag{10}$$

$$R_{rijt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_j + \gamma_{rt} \mathbb{1}_{r\times t} + \delta_j \mathbb{1}_j \qquad +\epsilon_{rijt,4} \tag{11}$$

$$R_{rijt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_j + \gamma_{rt} \mathbb{1}_{r\times t} + \delta_j \mathbb{1}_j + \theta_{jr} \mathbb{1}_{j\times r} \qquad +\epsilon_{rijt,5} \tag{12}$$

$$R_{rijt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_j + \gamma_{rt} \mathbb{1}_{r\times t} + \delta_j \mathbb{1}_j + \theta_{jr} \mathbb{1}_{j\times r} + \eta_i \mathbb{1}_i \qquad +\epsilon_{rijt,6} \tag{13}$$

$$R_{rijt} = \alpha_r \mathbb{1}_r + \beta_t \mathbb{1}_j + \gamma_{rt} \mathbb{1}_{r\times t} + \delta_j \mathbb{1}_j + \theta_{jr} \mathbb{1}_{j\times r} + \eta_i \mathbb{1}_i + \lambda_{ir} \mathbb{1}_{i\times r} \qquad +\epsilon_{rijt,7} \tag{14}$$

This specification enables us to quantify the relative contribution of industry-level disagreement compared to firm-level disagreement across raters.

Results Panel A of Table 5 presents the results of the R^2 decomposition at the industry (SIC-2) average level. We regress the SIC2-level average ESG Z-score on a sequence of fixed effects for rater, year, industry, and the rater-by-industry interaction. The baseline model (Equation (4)) including only rater fixed effects yields a negligible R^2 (0.35%), indicating that differences across raters alone explain little of the variance in industry-level ESG Z-scores. However, the inclusion of industry fixed effects sharply increases the R^2 to 53.37%, with an incremental R^2 of 52.95%. This result highlights that industry characteristics are the dominant driver of systematic variation in average ESG scores at the industry level. Further, adding rater-by-industry interaction effects increases the R^2 to 81.38%, with an incremental R^2 of 28.01%. This substantial gain demonstrates that, beyond average industry effects, raters systematically diverge in their assessments of specific industries. In other words, how ESG raters assess the identical industry differently attributes the portion of ESG rating divergence.

The industry-level analysis (Panel A) captures only the variation in the industry average and does not disentangle the sources of disagreement at the industry and firm levels. As a result, this approach does not precisely quantify the relative contributions of each source. To address this limitation, we extend our analysis to the firm level. Panel B of Table 5 presents the results. Firm fixed effects explain 67.29% of the variance in ESG scores, with an incremental R^2 of 56.05%. Adding the firm-by-rater interaction term further increases the R^2 to 88.57%, with an incremental R^2 of 21.28%. This indicates that a significant portion of the total variance is driven by systematic differences in how raters assess specific firms, consistent with prior literature that documents the importance of firm-specific factors in ESG rating divergence (Berg et al., 2022b).

Importantly, the incremental R^2 attributable to industry-by-rater effects at the firm level is 4.68%, while the relative magnitude of industry-level effects to firm-level effects is approximately 22%. This result indicates that industry-level disagreement plays an economically significant role in the overall ESG rating divergence. Industry-level divergence remains a non-trivial and systematic component of rating divergence, even after accounting for the firm-level disagreement.

In Table 6, we turn to a more granular analysis using only environmental scores. We find that the magnitude of industry-level rater effects in both panels (the industry average level in Panel A and the firm level in Panel B) is amplified in the environmental dimension. Specifically, industry fixed effects increase the R^2 substantially to 48.03% (incremental R^2 of 47.10%), while adding industry-by-rater interaction effects further increases the R^2 to 83.07% (incremental R^2 of 35.04%).

Panel B also shows that the industry-level rater effect becomes notably more pronounced even at the firm level. The incremental R^2 attributable to the industry-by-rater interaction more than doubles from 4.68% in the composite ESG analysis to 10.04%. Similarly, the relative magnitude nearly doubles from 22% in the ESG score to 51.54% in the environmental score.

These findings demonstrate that industry-level rating divergence is present and becomes particularly pronounced in the environmental dimension. This pattern is notable, as it challenges the conventional assumption that environmental metrics are more objective and consistently measured than social and governance dimension. The results suggest that such divergence may be driven by heterogeneity in industry-level materiality frameworks or by differences in industry-specific benchmarking practices. We investigate these possibilities more formally in the next section.

5.2 The Source of Industry-Level Divergence

Our previous analyses highlight that industry-level disagreement exists in ESG ratings. The incremental R^2 from industry-rater interactions demonstrates that raters apply different evaluation approaches across industries. To identify the underlying sources of this disagreement, in this section, we implement a simulation-based decomposition using carbon emission metrics as a representative case. Specifically, we use absolute Scope 1 emissions and Scope 1 and 2 emission intensity normalized by revenue (see Table 7) as our underlying raw data. This approach enables us to isolate the relative contributions of three key channels: (i) methodological differences, (ii) comparative industry (peer group) effects, and (iii) raw measurement discrepancies.

Methodology We conduct an experiment using firm-level data for which both Refinitiv and S&P Global provide raw values and scores for identical ESG metrics.[16] By focusing on common indicators prior to the application of weights or scope adjustments, we control for selection bias and scope (or weight) divergence. For each rater, we replicate ESG scores using their raw values and proprietary methodologies. We then sequentially align methodologies and peer groups to isolate each effect.[17] The difference in explanatory power (incremental R^2) across these steps quantifies the contribution of each source to total industry-level confusion.

Results Panel A of Table 7 presents how the magnitude of the industry-by-rater interaction effect changes under different experimental settings. This identification allows us to decompose the sources of industry-level ESG rating divergence for direct (Scope 1) greenhouse gas emissions. In Columns (1) and (2), we report the results using each rater's proprietary methodology, peer group classification, and raw values within their own sample universes. These results serve as our

[16] Since S&P reports raw values from 2020 onward, this analysis uses data from 2020 to 2022.

[17] Detailed methodology for replicating Refinitiv and S&P scores is provided in the Appendix A and B. The correlations between the replicated scores and ESG metric (or question) scores from Refinitiv and S&P are 0.94 and 0.74, respectively (See Appendix Table A2).

baseline, capturing the full extent of divergence arising from differences in universe, methodology, peer groups, and measurement.

Columns (3) and (4) present results after harmonizing the methodology by applying Refinitiv's scoring approach to each rater's universe and peer groups, while still using each rater's proprietary raw data. The difference in the incremental R^2 for the industry-by-rater term between Columns (2) and (4) isolates the portion of industry-level confusion attributable to methodological differences. Since the replication similarity for S&P's methodology is relatively low (correlation = 0.74), we rely on Refinitiv's methodology (correlation = 0.94) for more robust identification of methodological effects.

Columns (5) and (6) further restrict the analysis to the common sample and aligned peer groups by using Refinitiv's industry classification, while continuing to use Refinitiv's methodology and each rater's proprietary raw data. Here, Refinitiv reports the Scope 1 & 2 emission intensity, which is normalized by revenue, while S&P reports the absolute emission of Scope 1. This difference may cause the divergence of raw measurement. The difference in the incremental R^2 between Columns (4) and (6) captures the comparative industry effect, i.e., the explanatory power attributable to differences in industry classification and peer composition across raters. This effect arises either when the same firm is assigned to different industries by different raters (industry classification effect) or when a firm is assigned to the same industry but with different peer group compositions (relative performance evaluation effect).

Panel A shows that total industry-level confusion, as measured by the incremental R^2 for the industry-by-rater term in the baseline, is 3.09%. When methodologies are harmonized, this falls to 1.00%, implying that methodological differences account for 2.09 percentage points (i.e., nearly 33% of the of the incremental R^2). When peer groups are further aligned, the industry-by-rater effect drops to 0.11%, indicating that the comparative industry effect explains 0.89 percentage points, which is approximately 30% of the total rater effect. The remaining 0.11% is attributable to raw measurement discrepancies. Notably, the magnitude of the comparative industry effect is roughly eight times larger than that of raw measurement, highlighting the substantial role of peer benchmarking and industry classification in ESG rating disagreement.

These results indicate that, alongside measurement divergence, which is identified in the prior literature as the dominant source of ESG rating divergence, the comparative industry effect stemming from differences in peer group construction and industry classification constitutes a nontrivial driver of ESG rating disagreement. We also validate this pattern using additional social indicator, employee turnover rate, which exhibits low measurement divergence across raters. As reported in Appendix Table A6, we find consistent evidence that the comparative industry effect is a primary driver of industry-level disagreement.

Panels B and C of Table 7 extend our analysis to settings with pronounced measurement divergence. In Panel B, we compare absolute Scope 1 carbon emissions from S&P Global with revenue-normalized Scope 1 and 2 emission intensity. And in Panel C, we compare absolute Scope 1 carbon emissions from S&P and absolute Scope 2 emissions from Refinitiv.

In Panel B, although raw measurement discrepancies account for the largest share of rater disagreement (1.68% out of 4.16%), the comparative industry effect remains economically significant at 1.24%. This magnitude represents approximately 30% of the total industry-level rater disagreement. In Panel C, the comparative industry effect is 0.5%, while measurement discrepancy accounts for 1.06%.

Overall, despite greater data misalignment, the comparative industry effect continues to explain a significant portion of the industry-level rater disagreement. Importantly, when measurement divergence is minimized (as shown in Panel A), the relative magnitude of the comparative industry effect becomes more pronounced. This pattern demonstrates that, regardless of the degree of measurement alignment, differences in peer group construction and industry classification across raters consistently account for a substantial portion of industry-level disagreement.

6 Economic Determinants of Industry-Level Rating Disagreements

While our earlier analyses provide evidence that comparative industry effect is a prominent source of ESG rating disagreement at the industry level, we aim to understand economic determinants of rating divergence in this section.

6.1 Exogenous Changes in Industry Classifications

Methodology and Institutional Details We employ a quasi-experimental approach that leverages exogenous structural revisions to the Global Industry Classification Standard (GICS) and study its relation with industry-level rating discrepancy.

MSCI and S&P announced the revisions to the GICS structure in 2014, 2016, and 2018 (S&P and MSCI, 2013, S&P and MSCI, 2015, and S&P and MSCI, 2017).[18] This framework consists of (i) sector, (ii) industry Group, (iii) industry, and (iv) sub-industry levels; however, our study focuses on the revisions of sub-industry level, as these changes are central to the composition of the peers (see Appendix Table A5 for more details).

At the sub-industry level, revisions include updates to industry definitions, names, and peer firms. The composition of firms within an industry may change when a new sub-industry is introduced or an existing one is discontinued. Since ESG ratings are constructed based on relative performance evaluations within industry peer groups, changes in the composition of firms may affect industry-level disagreement. If such factors contribute to industry-level discrepancies, the structural changes within GICS may exacerbate these disagreements.

Specifically, we partition the sample based on whether a firm's industry was affected by GICS peer group changes, which serve as a natural experiment for shifts in industry classification. Such revisions are expected to increase the likelihood of disagreement among raters regarding business classification and, consequently, peer group assignment, which is a critical factor in relative ESG

[18] The changes to the GICS structure were implemented after the close of business (ET) on Friday, February 28, 2014, August 31, 2016, and September 28, 2018, respectively.

evaluation and a key channel for the comparative industry effect. Our analysis corroborates the simulation exercise in Section 5.2, where we focus on a local sample with little time-series variation.

Results Table 8 presents regression results examining the association between ESG rating disagreement and an indicator for industries affected by the structural changes in GICS (*GICS*). As basic firm fundamentals, we control for firm size, book-to-market ratio, leverage, and return on assets. Book-to-market ratio is defined as the book value of equity divided by its market value. Leverage is calculated as the sum of long-term debt and debt in current liabilities, scaled by the book value of asset. We exclude firms with a zero leverage ratio. ROA is defined as earnings before interest, taxes, depreciation, and amortization (EBITDA), divided by total assets. Detailed variable definitions are provided in Appendix H. The dependent variables are the total, industry-level, and firm-level disagreements measured using the standard deviation of ESG Z-scores (Panel A) and environmental Z-scores (Panel B). All continuous variables are winsorized at the 1% and 99% levels, and standard errors are clustered at the firm level. Year and industry fixed effects are omitted, as the dependent variables are already demeaned by year and industry averages. Since the shock occurred in 2014, 2016, and 2018, including year fixed effects may lead to an underestimation of the treatment effect, as the year dummies can absorb some of the variation attributable to the GICS revisions. Nevertheless, for robustness, we also report the results with year fixed effects in columns (7), (8), and (9), and find that the main results remain qualitatively unchanged.

In Column (2) of Panel A, the coefficient on *GICS* is positive and highly significant (0.05, t = 3.39). This result indicates that firms in industries affected by GICS revisions experience substantially higher industry-level disagreement. Given that the standard deviation of industry-level disagreement is 0.15 (see Table 3), the estimated economic magnitude is 33% (=0.05/0.15). In other words, a GICS-affected industry is associated with a one-third standard deviation increase in the dispersion of ESG ratings at the industry level. In contrast, the association between *GICS* and firm-level disagreement (which is measured after demeaning the common industry component) is statistically insignificant (Columns (3), (6), and (9)), suggesting that structural changes to industry classification is not associated with firm-specific rater disagreement. The effect on total disagreement is positive but not statistically significant.

We repeat the analysis using the environmental Z-score as the dependent variable (Panel B of Table 8). In this specification, the effects of GICS structural revisions become more pronounced. The coefficient on *GICS* in Column (2) increases substantially to 0.23 (t = 19.03), indicating a much larger impact of GICS revisions on industry-level disagreement in environmental ratings compared to the composite ESG scores. The economic magnitude increases to 127% (=0.23/0.18), indicating that being classified as a GICS-affected industry is associated with a 1.27 standard deviation increase in the dispersion of environmental ratings at the industry level, relative to non-GICS-affected industries. This magnitude is more than double that observed for the total ESG score. This pronounced effect persists when control variables or fixed effects (Columns (5) and (8)) are added. While the effect on total disagreement is also positive and highly significant (0.16, t = 8.43 in Column (1)), the association between *GICS* and firm-level disagreement remains negligible in terms of their magnitudes (0.03, t = 2.31 in Column (3); 0.04, t = 2.66 in Column (6); 0.04, t = 2.65 in Column (9)).

These findings confirm our conjecture that industry-level disagreement among raters worsens when peer group compositions change. Notably, the negligible magnitude of firm-level disagreement implies that the impact of GICS revisions is concentrated at the industry aggregation level. The impact of the GICS changes on industry-level disagreement is also more pronounced in the environmental dimension than in the overall ESG score, consistent with previous analyses. This supports the view that peer group composition and industry classification are fundamental determinants of rating divergence.

6.2 Cross-Sectional Analysis by Disclosure Practice and Ownership Structure

Next, we examine several firm-level economic conditions that are likely associated with industry-level rater disagreement. Specifically, we examine whether rater disagreement is stronger for private firms whose disclosures are less standardized and more opaque than public firms and for firms with concentrated ownership structure.

Methodology First, we partition the sample based on the firm's disclosure status (i.e. public vs. private firms). We use company type data provided by S&P Global to classify firms as public or private. Our sample is constructed by merging coverage from four major ESG rating providers (MSCI, Refinitiv, S&P Global, and Sustainalytics) using Compustat's GVKEY identifier. Given this construction, the sample is naturally tilted toward firms that are, or have been, publicly listed. Among the 65 firms classified as private in our sample, 28 lack market value information in Compustat during the sample period, indicating that many of these firms may have once been publicly traded. Nonetheless, these firms have operated at some point under reduced disclosure requirements, and thus lower transparency. Then, for each group, we repeat the exercise in Tables 5 and 6 to gauge the magnitude of industry-level disagreement in overall rating discrepancy.

To further examine whether limited transparency affects the industry-level disagreement, we compare firms operating in countries with dispersed ownership structures (i.e., widely held firms) to those operating in countries with concentrated ownership structures, following the classification in La Porta et al. (1999). We classify the United Kingdom, Japan, and the United States as dispersed ownership structures, while Mexico, Belgium, and Sweden are classified as concentrated ownership firms, often via pyramidal structures (see Table IV of La Porta et al., 1999). Again, we separately perform the exercise in Tables 5 and 6 for each subgroup.

Results Panel A of Table 9 presents the results based on disclosure status. The incremental R^2 for industry fixed effects is 33.17% for private firms, compared to 6.46% for public firms, suggesting that when firm-specific information is limited, ESG raters rely more heavily on industry benchmarks to evaluate firm performance. Moreover, private firms exhibit a substantially higher incremental R^2 for industry-by-rater fixed effects (14.77%) than public firms (4.70%). The relative magnitude of industry-level disagreement compared to firm-level disagreement is also markedly greater for private firms (167% = 14.77%/8.86%) than for public firms (22% = 4.70%/21.42%). These findings indicate that, when structured corporate disclosures are more difficult to obtain, disagreement among ESG raters at the industry level becomes more pronounced.

The pattern becomes more pronounced in the environmental dimension. In Panel B, the incremental R^2 for industry-by-rater fixed effects remains substantially higher for private firms (16.86%) than for public firms (9.87%), and the relative magnitude of industry-level disagreement compared to firm-level disagreement is also greater for private firms (240% = 16.86%/7.04%) than for public firms (50% = 9.87%/19.76%).

In Panel C, we report the results using ownership structure as the partitioning variable. The incremental R^2 for industry-by-rater fixed effects is generally lower for firms operating in countries with dispersed ownership structure (e.g., 7.21% in the US, 9.56% in Japan, and 27.07% in the UK). This result suggests that raters depend less on industry benchmarks when ownership structure is dispersed and investor protection is stronger (La Porta et al., 1999). In contrast, countries with concentrated ownership exhibit substantially higher industry-by-rater fixed effects (e.g., 17.32% in Mexico, 18.16% in Belgium, and 26.74% in Sweden). This result indicates that when information environment is more transparent and stakeholder-oriented, the divergence in industry-level evaluation across raters becomes narrower. The relative magnitude of industry-level disagreement is also much higher for these firms (355% in Mexico, 318% in Belgium, and 296% in Sweden) compared to widely held firms (34% in the US, 50% in Japan, and 193% in the UK).

Panel D further confirms that this pattern is even more pronounced in the environmental dimension, consistent with our previous analyses. The relative magnitude of industry-level disagreement in environmental scores is notably higher for firms with complex ownership structures (285% in Mexico, 1692% in Belgium, and 747% in Sweden) than for those with dispersed ownership structures (57% in the US, 77% in Japan, and 160% in the UK).

Taken together, these cross-sectional analyses demonstrate that the industry-level rating disagreement is associated with the disclosure environment and ownership complexity. When information transparency at the firm level is limited, raters tend to rely more heavily on industry-level benchmarks in their assessments, which further aggravates the disagreement at the industry level.

7 Conclusion

Our study provides new evidence on the structure and sources of ESG rating disagreement, with a particular focus on industry-level effects and the role of comparative industry evaluation.

We demonstrate that industry-level disagreement is a stable and systematic feature of total ESG rating variation. Using a simple numerical exercise, we show that raters tend to disagree at the industry level and that this pattern is persistent for an extended period of time. If the measurement noise were solely attributable for the rating divergence, firm-level noise is expected to cancel out when aggregated at the industry level. We argue that our puzzling observation is likely due to *comparative industry effect*.

The comparative industry effect can arise from differences in industry classification (i.e., a single firm being classified in different industries by different raters) and relative performance evaluation (RPE) frameworks across raters (i.e., a firm's evaluation peers being different across different raters). Using an R^2 decomposition analysis, we show that comparative industry effect is indeed a significant driver of rating disagreement. We further study the potential sources of industry-level disagreement through a simulation-based decomposition and find that the comparative industry effect accounts for a magnitude nearly equivalent to measurement divergence.

We also show that industry-level disagreement is particularly pronounced among firms with limited firm-specific disclosure, such as privately held or concentrated ownership firms. In these settings, raters tend to rely more heavily on industry-level benchmarks, leading to greater divergence relative to publicly traded or widely held firms, where richer disclosures facilitate more granular, firm-specific evaluations.

Our study is the first to systematically document the persistent and systematic influence of industry-level factors on ESG ratings and rating discrepancies. While ESG raters extensively utilize industry-based relative performance evaluation, the impact of heterogeneous industry classifications and relative performance evaluation methodologies on rating disagreement has been largely underexplored in the literature. Our empirical findings confirm the significant and enduring role of industry-level effects in driving ESG rating divergence.

These findings carry important implications for both policymakers and researchers. First, they are highly relevant to standard setters and regulators given that ESG rating discrepancies can distort resource allocation. Recognizing the practical challenges in enforcing universal measurement standards across diverse raters, standard setters should prioritize harmonizing industry classifications and relative performance evaluation frameworks to reduce rating inconsistencies. Second, our research extends the framework of Berg et al. (2022b) by demonstrating that rating disagreement can be decomposed into distinct industry-level and firm-level components. Our analysis highlights the critical role of relative performance evaluation and the comparative peer effect in shaping industry-level rating divergence.

References

Berg, F., Koelbel, J. F., Pavlova, A., and Rigobon, R. (2022a). Esg confusion and stock returns: Tackling the problem of noise. Technical report, National Bureau of Economic Research.

Berg, F., Kölbel, J. F., and Rigobon, R. (2022b). Aggregate confusion: The divergence of esg ratings. *Review of Finance*, 26(6):1315–1344.

Bochkay, K., Hales, J., and Serafeim, G. (2025). Disclosure standards and communication norms: Evidence of voluntary sustainability standards as a coordinating device for capital markets. *Review of accounting studies*, pages 1–44.

Chatterji, A. K., Durand, R., Levine, D. I., and Touboul, S. (2016). Do ratings of firms converge? implications for managers, investors and strategy researchers. *Strategic Management Journal*, 37(8):1597–1614.

Chen, C., Dube, S., and Froymovich, S. (2025). Esg rating competition and rating quality. *Journal of Accounting Research*.

Gibson Brandon, R., Krueger, P., and Schmidt, P. S. (2021). Esg rating disagreement and stock returns. *Financial Analysts Journal*, 77(4):104–127.

Grewal, J., Hauptmann, C., and Serafeim, G. (2021). Material sustainability information and stock price informativeness. *Journal of Business Ethics*, 171(3):513–544.

Khan, M., Serafeim, G., and Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The accounting review*, 91(6):1697–1724.

Kim, S., Kumar, N., Lee, J., and Oh, J. (2025). Esg lending. *Journal of Financial Economics*.

La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (1999). Corporate ownership around the world. *The journal of finance*, 54(2):471–517.

Li, X., Lou, Y., and Zhang, L. (2024). Do commercial ties influence esg ratings? evidence from moody's and s&p. *Journal of Accounting Research*, 62(5):1901–1940.

Lins, K. V., Servaes, H., and Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *the Journal of Finance*, 72(4):1785–1824.

London Stock Exchange Group (2021). Lseg all-share acquisition of refinitiv expected to complete on 29 january 2021. https://www.lseg.com/content/dam/lseg/en_us/documents/investor-relations/mergers-and-acquisitions/refinitiv/lseg-all-share-acquisition-refinitiv-expected-completion-date-19jan2021.pdf. Accessed: 2025-06-23.

MSCI ESG RESEARCH LLC (2024). Esg ratings methodology. MSCI ESG RESEARCH LLC, April 2024, Retrieved from ESG RATINGS METHODOLOGY.

Refinitiv (2022). Environmental, social and governance scores from refinitiv. London Stock Exchange Group (LSEG), May 2022, Retrieved from Refinitiv ESG Scores Methodology document.

Schmidl, T. and Muck, M. (2025). Relative to whom? the impact of peer groups on esg ratings and financial performance. *The Impact of Peer Groups on ESG Ratings and Financial Performance (May 17, 2025)*.

Serafeim, G. and Yoon, A. (2022a). Stock price reactions to esg news: The role of esg ratings and disagreement. *Review of accounting studies*, pages 1–31.

Serafeim, G. and Yoon, A. (2022b). Which corporate esg news does the market react to? *Financial Analysts Journal*, 78(1):59–78.

Sindreu, J. and Kent, S. (2018). Why it's so hard to be an 'ethical'investor. *Wall Street Journal*, (September 1).

S&P and MSCI (2013). Msci and sp dow jones indices announce revisions to the global industry classification standard (gics®) structure in 2014. Accessed: 2025-04-03.

S&P and MSCI (2015). S&p dow jones indices and msci announce further revisions to the global industry classification standard (gics®) structure in 2016. Accessed: 2025-04-03.

S&P and MSCI (2017). Revisions to the global industry classification standard (gics®) structure in 2018. Accessed: 2025-04-03.

S&P Global Market Intelligence (2022). S&p global esg scores methodology guide. S&P Global, 2022, Retrieved from S&P Global ESG Scores Methodology Guide document.

Strebulaev, I. A. and Yang, B. (2013). The mystery of zero-leverage firms. *Journal of Financial Economics*, 109(1):1–23.

SustainAbility (March 2020). Rate the raters 2020: Investor survey and interview results. Accessed: 2025-08-11.

Sustainalytics (2017). Esg ratings methodology quick overview. Accessed: 2023-06-23.

Figure 1: **Credit Rating vs. ESG Rating**

This figure portrays how ESG raters disagree relative to credit rating agencies. A recent Wall Street Journal article by Sindreu and Kent (2018) highlights this contrasting results; R^2 between different ESG ratings is only 13%, while the same statistic for various credit ratings is nearly 82%.

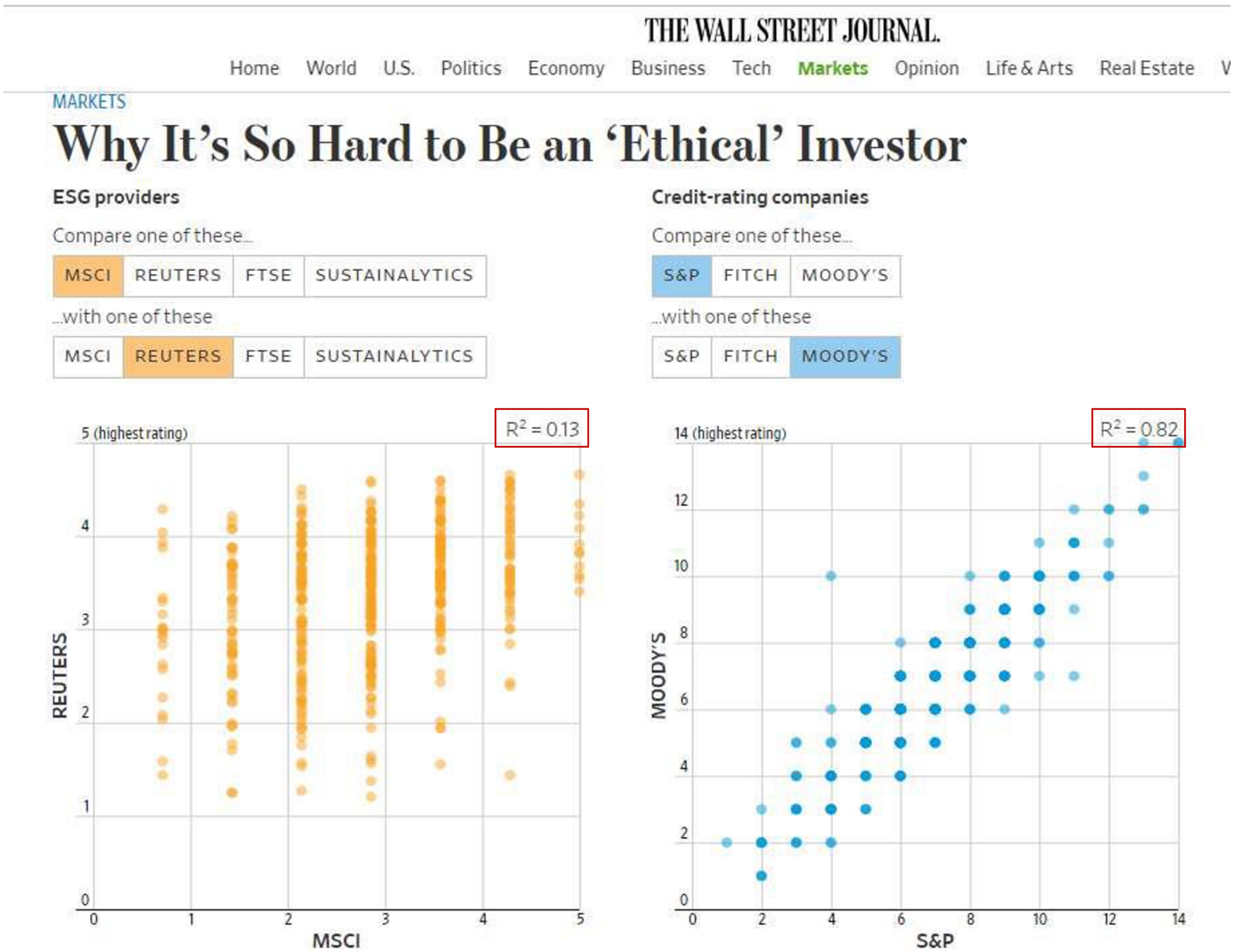


Figure 2: **Illustrative Evidence of Industry-Level Disagreement**

This figure presents the 2014 ESG Z-scores assigned by MSCI and S&P for two industries: Metal Mining (SIC 10) and Securities & Commodity Brokers, Dealers, Exchanges & Services (SIC 62).

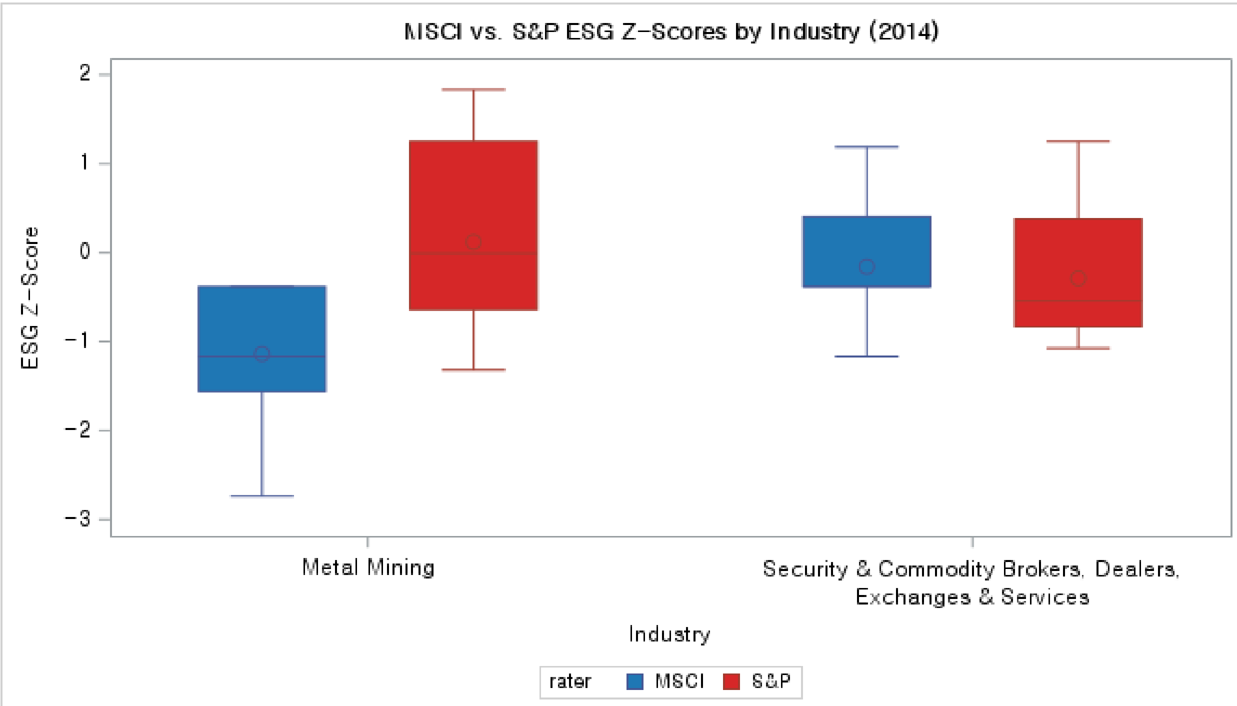
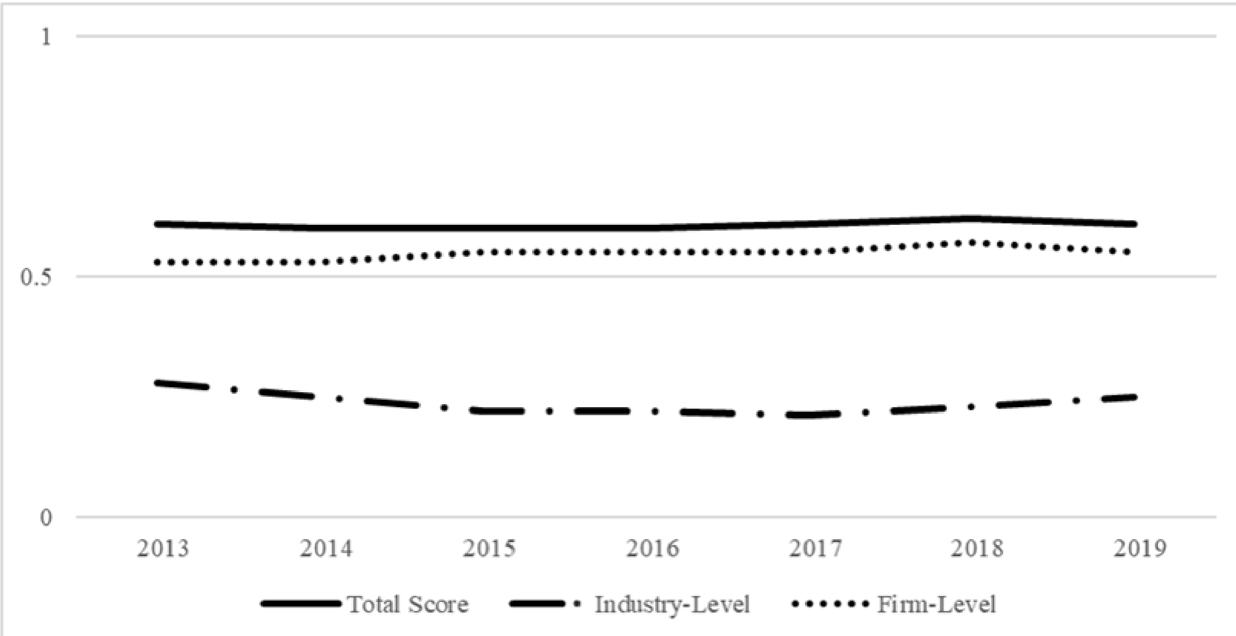


Figure 3: **Time Trend of Total, Industry-Level, and Firm-Level Disagreement**

This figure portrays the time trend of total, industry-level, and firm-level disagreement using standard deviation as a measure of disagreement. Refer to Table 4 for detailed explanation. Figure 3A shows the trend when we standardize ESG scores within the cross-section. Figure 3B shows the trend when we standardize environmental (ENV) scores.

Panel A. Time Trend of σ(ESG)



Panel B. Time Trend of σ(ENV)

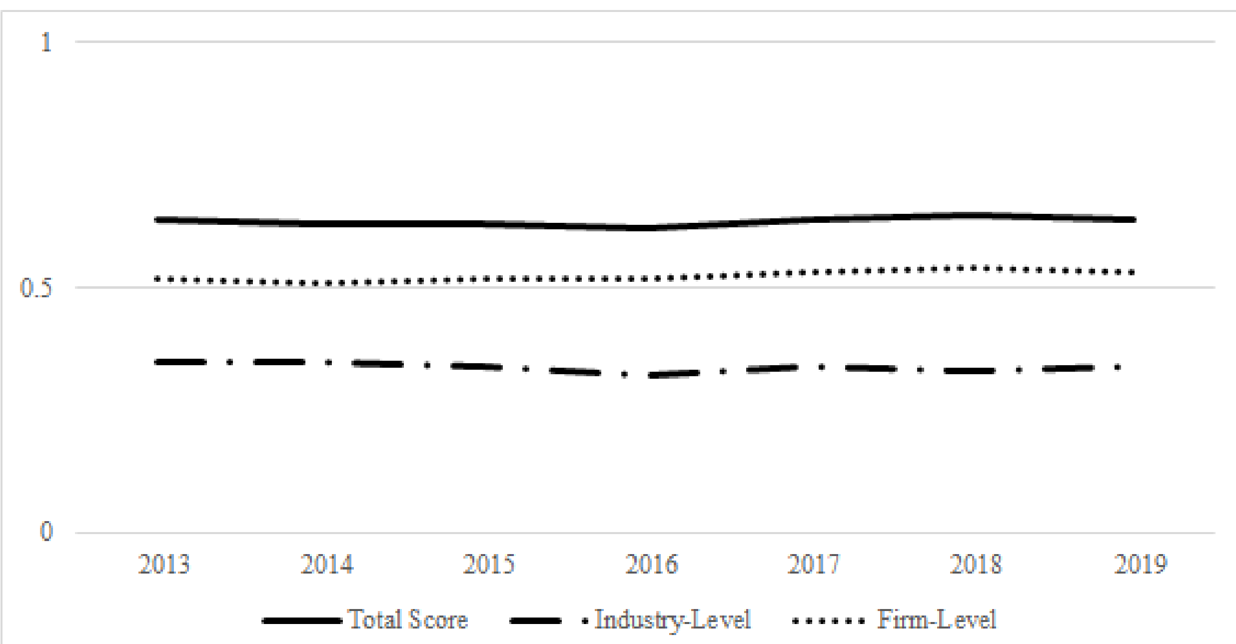


Table 1: **ESG Raters and Evaluation Metrics**

This table provides a summary of each rater's history, coverage, industry classification methods, relative evaluation metrics, and industry-specific assessment criteria.

Raters	**MSCI ESG RESEARCH LLC (2024)**	**Refinitiv (2022) (LSEG or Asset4)**	**Sustainalytics (2017)**	**S&P Global Market Intelligence (2022)**
History	Acquired RiskMetrics in 2010 and GMI Ratings in 2014	Acquired by Thomson Reuters in 2009; Spun-off financial risk department in 2018	Initiated in 2008; Morningstar acquired 40% of shares in 2017	Acquired RobecoSAM ESG department in 2019
Coverage	About 8,500 firms (14,000 issuers including subsidiaries)	About 9,000 firms	About 13,000 firms	About 11,000 firms
Industry Taxonomy	GICS (Global Industry Classification Standard)	TRBC (The Refinitiv Business Classifications)	Sustainalytics Classification	GICS (Global Industry Classification Standard)
Relative Performance Evaluation	MSCI ESG Ratings are *industry-relative measures* and are determined at the company level. ... The top-level assessment is the overall Company ESG Rating, an industry-relative seven-point letter rating scale from AAA to CCC. These assessments are not absolute but are explicitly intended to be interpreted *relative to a company's industry peers*.	The scores are based on *relative performance of ESG factors with the company's sector* (for environmental and social) and country of incorporation (for governance). Refinitiv does not presume to define what 'good' looks like; we let the data determine *industry-based relative performance* within the construct of our criteria and data model.	We analyze and rate the performance of companies worldwide, *across 42 different peer groups (comparable sub-industries)*. ... Performance against ESG issues is analyzed by looking at a comprehensive set of core and sector-specific metrics, which are scored and weighted to determine a company's overall ESG performance. Underlying each industry template is a customized weight matrix that defines *the relative importance of each indicator and reflects the emphasis on key ESG issues per industry*.	As relevant sustainability risks, opportunities and available technologies tend to be industry-specific, the scores measure a company's overall *sustainability performance relative to its peers* within its sub-industry. ... As data collection and reporting on such topics matures, performance-based scoring is often introduced to capture a trend or measure *a company's performance relative to its peers*.
Industry-specific				
Question score	✗	✓	✗	✓
Question weight	✗	✗	✗	✓
Category/Issue score	✓(E/S dimensions)	✓	✗	✓
Category/Issue weight	✓(E/S dimensions)	✓(E/S dimensions)	✗	✓
Dimension weight	✓(E/S dimensions)	✗	✗	✓

Table 2: **Sample Distribution**

Table 2 presents regional distributions (Panel A), country distributions (Panel B), and industry distributions (Panel C) of the sample we used in our analysis. The sample comprises 8,720 firm-year observations from 1,742 unique firms commonly covered by all four ESG rating providers (MSCI, Refinitiv, Sustainalytics, and S&P) over the period 2013-2019.

Panel A. Regional Distributions

Region	Count	Percent
Africa	138	1.58%
Americas	4424	50.73%
Asia	1714	19.66%
Europe	2101	24.09%
Oceania	343	3.93%
Total	8720	100%

Panel B. Country Distributions

Country	Region	Country Code	Count	Percent
Argentina	Americas	AR	1	0.01%
Australia	Oceania	AU	303	3.47%
Austria	Europe	AT	42	0.48%
Belgium	Europe	BE	58	0.67%
Bermuda	Americas	BM	75	0.86%
Brazil	Americas	BR	176	2.02%
Canada	Americas	CA	396	4.54%
Cayman Islands	Americas	KY	116	1.33%
Chile	Americas	CL	58	0.67%
China	Asia	CN	75	0.86%
Colombia	Americas	CO	25	0.29%
Czechia	Europe	CZ	6	0.07%
Denmark	Europe	DK	68	0.78%
Egypt	Africa	EG	6	0.07%
Finland	Europe	FI	59	0.68%
France	Europe	FR	320	3.67%
Germany	Europe	DE	297	3.41%
Greece	Europe	GR	12	0.14%
Hong Kong	Asia	HK	162	1.86%
Hungary	Europe	HU	1	0.01%
India	Asia	IN	47	0.54%
Indonesia	Asia	ID	51	0.58%
Ireland	Europe	IE	36	0.41%
Isle of Man	Europe	IM	3	0.03%
Israel	Asia	IL	39	0.45%
Italy	Europe	IT	77	0.88%
Japan	Asia	JP	1041	11.94%
Jersey	Europe	JE	18	0.21%
Korea, Republic of	Asia	KR	71	0.81%
Luxembourg	Europe	LU	9	0.10%
Malaysia	Asia	MY	20	0.23%
Mauritius	Africa	MU	6	0.07%
Mexico	Americas	MX	85	0.97%
Netherlands Antilles	Europe	AN	7	0.08%
Netherlands, Kingdom	Europe	NL	125	1.43%
New Zealand	Oceania	NZ	33	0.38%
Norway	Europe	NO	48	0.55%
Panama	Americas	PA	7	0.08%
Papua New Guinea	Oceania	PG	7	0.08%
Peru	Americas	PE	6	0.07%
Philippines	Asia	PH	14	0.16%
Poland	Europe	PL	2	0.02%
Portugal	Europe	PT	26	0.30%
Singapore	Asia	SG	83	0.95%
South Africa	Africa	ZA	126	1.44%
Spain	Europe	ES	123	1.41%
Sweden	Europe	SE	138	1.58%
Switzerland	Europe	CH	155	1.78%
Taiwan, Province of	Asia	TW	33	0.38%
Thailand	Asia	TH	30	0.34%
Turkiye	Asia	TR	48	0.55%
United Kingdom	Europe	GB	471	5.40%
United States of America	Americas	US	3479	39.90%
Total			8720	100%

Panel C. Industry Distributions by SIC 2-digit Code

SIC 2-digit	Industry Name	Count	Percent
10	Metal Mining	210	2.41%
12	Coal Mining	20	0.23%
13	Oil and Gas Extraction	264	3.03%
14	Mining and Quarrying of Nonmetallic Minerals, Except Fuels	20	0.23%
15	Construction - General Contractors & Operative Builders	86	0.99%
16	Heavy Construction, Except Building Construction	39	0.45%
17	Construction - Special Trade Contractors	9	0.10%
20	Food and Kindred Products	402	4.61%
21	Tobacco Products	35	0.40%
22	Textile Mill Products	7	0.08%
23	Apparel, Finished Products from Fabrics & Similar Materials	91	1.04%
24	Lumber and Wood Products, Except Furniture	15	0.17%
25	Furniture and Fixtures	21	0.24%
26	Paper and Allied Products	80	0.92%
27	Printing, Publishing and Allied Industries	30	0.34%
28	Chemicals and Allied Products	665	7.63%
29	Petroleum Refining and Related Industries	121	1.39%
30	Rubber and Miscellaneous Plastic Products	71	0.81%
31	Leather and Leather Products	27	0.31%
32	Stone, Clay, Glass, and Concrete Products	103	1.18%
33	Primary Metal Industries	131	1.50%
34	Fabricated Metal Products	77	0.88%
35	Industrial and Commercial Machinery and Computer Equipment	379	4.35%
36	Electronic & Other Electrical Equipment & Components	401	4.60%
37	Transportation Equipment	353	4.05%
38	Measuring, Photographic, Medical, & Optical Goods, & Clocks	381	4.37%
39	Miscellaneous Manufacturing Industries	45	0.52%
40	Railroad Transportation	55	0.63%
41	Local & Suburban Transit & Interurban Highway Transportation	21	0.24%
42	Motor Freight Transportation	63	0.72%
44	Water Transportation	22	0.25%
45	Transportation by Air	124	1.42%
46	Pipelines, Except Natural Gas	6	0.07%
47	Transportation Services	42	0.48%
48	Communications	349	4.00%
49	Electric, Gas and Sanitary Services	581	6.66%
50	Wholesale Trade - Durable Goods	124	1.42%
51	Wholesale Trade - Nondurable Goods	85	0.97%
52	Building Materials, Hardware, Garden Supplies & Mobile Homes	34	0.39%
53	General Merchandise Stores	97	1.11%
54	Food Stores	103	1.18%
55	Automotive Dealers and Gasoline Service Stations	37	0.42%
56	Apparel and Accessory Stores	79	0.91%
57	Home Furniture, Furnishings and Equipment Stores	16	0.18%
58	Eating and Drinking Places	68	0.78%
59	Miscellaneous Retail	89	1.02%
60	Depository Institutions	650	7.45%
61	Nondepository Credit Institutions	78	0.89%
62	Security & Commodity Brokers, Dealers, Exchanges & Services	219	2.51%
63	Insurance Carriers	359	4.12%
64	Insurance Agents, Brokers and Service	15	0.17%
65	Real Estate	123	1.41%
67	Holding and Other Investment Offices	179	2.05%
70	Hotels, Rooming Houses, Camps, and Other Lodging Places	18	0.21%
72	Personal Services	17	0.19%
73	Business Services	618	7.09%
75	Automotive Repair, Services and Parking	21	0.24%
78	Motion Pictures	9	0.10%
79	Amusement and Recreation Services	75	0.86%
80	Health Services	65	0.75%
81	Legal Services	7	0.08%
82	Educational Services	21	0.24%
83	Social Services	1	0.01%
87	Engineering, Accounting, Research, and Management Services	65	0.75%
99	Nonclassifiable Establishments	102	1.17%
Total		8720	100%

Table 3: **Descriptive Statistics**

Table 3 presents descriptive statistics for the main variables used in our analysis. The sample comprises 8,720 firm-year observations from 1,742 unique firms commonly covered by all four ESG rating providers (MSCI, Refinitiv, Sustainalytics, and S&P) over the period 2013-2019. ESG and ENV(Environmental) scores are standardized to a 0-100 scale across all providers, with Z-scores representing standardized versions having mean zero and standard deviation one. We follow prior literature (e.g., Serafeim and Yoon (2022a)) to measure ESG rating disagreement across raters in two different ways: the standard deviation (σ) and the Euclidean distance (d) of four ratings. See Appendix for the results of the Euclidean distance (d). Carbon emission data are sourced from two providers. Refinitiv reports both absolute and intensity metrics for Scope 1 and 2 emissions, where intensity refers to emissions normalized by revenue. In contrast, S&P Global provides only absolute emission levels for direct (Scope 1) emissions. Control variables are size, book-to-market ratio, leverage ratio, and ROA. Size, defined as the lagged logarithm of total assets, controls for the tendency that big firms have more resources and inputs to receive higher ESG scores. Book-to-market ratio is the book value of equity divided by the market value of equity. Leverage ratio is long-term debt plus debt in current liabilities divided by the book value of common equity. ROA is the ratio of earnings before income tax, depreciation, and amortization over lagged total assets. All control variables are winsorized in 1% and 99%.

Variables		N	Mean	Std	Min	Med	Max
Scores							
	MSCI ESG	8,720	45.99	11.60	0.00	49.00	100.00
	MSCI ENV	8,720	55.26	23.02	0.00	54.00	100.00
	Ref ESG	8,720	59.37	18.14	1.50	61.51	95.72
	Ref ENV	8,720	56.13	26.39	0.00	60.97	99.14
	Sus ESG	8,720	61.74	10.42	31.00	61.55	95.00
	Sus ENV	8,720	60.99	14.40	20.00	61.00	100.00
	S&P ESG	8,720	43.61	22.57	0.00	37.00	95.00
	S&P ENV	8,720	41.78	28.60	0.00	37.00	100.00
Z-Scores							
	MSCI ESG	8,720	0.00	1.00	−3.75	−0.11	4.22
	MSCI ENV	8,720	0.00	1.00	−2.44	−0.03	2.04
	Ref ESG	8,720	0.00	1.00	−3.37	0.12	2.01
	Ref ENV	8,720	0.00	1.00	−2.29	0.18	1.65
	Sus ESG	8,720	0.00	1.00	−3.12	−0.02	3.08
	Sus ENV	8,720	0.00	1.00	−2.87	0.01	2.89
	S&P ESG	8,720	0.00	1.00	−1.74	−0.35	2.32
	S&P ENV	8,720	0.00	1.00	−1.73	−0.16	2.05
Disagreement							
	σ(ESG Z-Score)	8,720	0.61	0.28	0.02	0.58	1.98
	σ(ESG Z-Score)_Industry	8,720	0.23	0.15	0.03	0.20	1.59
	σ(ESG Z-Score)_Firm	8,720	0.55	0.25	0.00	0.53	2.01
	σ(ENV Z-Score)	8,720	0.64	0.29	0.01	0.60	1.72
	σ(ENV Z-Score)_Industry	8,720	0.34	0.18	0.04	0.30	1.47
	σ(ENV Z-Score)_Firm	8,720	0.52	0.25	0.00	0.49	1.83
	d(ESG Z-Score)	8,720	0.75	0.34	0.03	0.71	2.38
	d(ESG Z-Score)_Industry	8,720	0.28	0.18	0.03	0.25	1.82
	d(ESG Z-Score)_Firm	8,720	0.68	0.31	0.00	0.64	2.45
	d(ENV Z-Score)	8,720	0.78	0.34	0.01	0.73	2.15
	d(ENV Z-Score)_Industry	8,720	0.41	0.21	0.05	0.37	1.68
	d(ENV Z-Score)_Firm	8,720	0.64	0.30	0.00	0.60	2.05
ES Indicators							
	Ref_Scope 1&2 Intensity	11,142	13,071	1,293,068	0	32	136,455,422
	Ref_Scope 1 Emission	10,619	3,089,484	63,844,039	0	25,357	4,079,000,000
	Ref_Scope 2 Emission	10,510	654,729	17,641,383	0	50,540	1,333,000,000
	S&P_Scope 1 Emission	11,142	2,298,564	25,234,967	0	25,714	2,354,736,000
	Ref_Employee Turnover Rate	4,362	16.37	18.77	0.03	13.5	599.86
	S&P_Employee Turnover Rate	4,362	15.79	11.58	0.03	13.5	106
Control Variables							
	Size	6,972	10.316	2.010	6.995	9.910	16.470
	Book-to-Market ratio	6,972	0.739	0.380	0.112	0.714	2.384
	Leverage	6,972	0.266	0.151	0.003	0.258	0.649
	ROA	6,972	0.124	0.070	−0.036	0.115	0.357

Unique firms: 1,742

Table 4: **Disaggregation of ESG Rating Disagreement**

Table 4 reports the disaggregation of ESG and ENV (Environmental) rating disagreement into industry-level disagreement and firm-level disagreement. For each rater, we compute the industry-level ESG score as the average of firm-level scores within each year–industry group, using SIC 2-digit classifications. Industry-level disagreement is defined as the standard deviation across raters of these industry-level scores. Firm-level disagreement is defined as the standard deviation across raters of the firm-level deviations from their respective industry-year averages. Panel A and B report results based on the standard deviation of the four ESG scores and ENV scores, respectively.

Panel A. ESG Score

σ(ESG Z-Score) of	(1) N	(2) Total Score	(3) Industry-Level	(4) Firm-Level
Full Sample	8720	0.61	0.23	0.55
By year				
2013	847	0.62	0.27	0.55
2014	1115	0.61	0.25	0.54
2015	1148	0.61	0.22	0.55
2016	1283	0.60	0.22	0.55
2017	1331	0.60	0.21	0.56
2018	1370	0.62	0.23	0.57
2019	1626	0.61	0.25	0.55

Panel B. ENV Score

σ(ENV Z-Score) of	(1) N	(2) Total Score	(3) Industry-Level	(4) Firm-Level
Full Sample	8720	0.64	0.34	0.52
By year				
2013	847	0.64	0.36	0.52
2014	1115	0.63	0.35	0.52
2015	1148	0.63	0.35	0.51
2016	1283	0.62	0.33	0.52
2017	1331	0.64	0.34	0.53
2018	1370	0.65	0.34	0.54
2019	1626	0.64	0.34	0.53

Table 5: R^2 Decomposition - Total ESG Score

Table 5 presents the R^2 decompositions for ESG scores in both industry-level average and firm-level. The analysis employs the 2-digits SIC industry classification as the industry fixed effects. In each panel, fixed effects are introduced cumulatively: each subsequent specification adds the listed fixed effect(s) to all previously included fixed effects. Panel A presents results from regressions of the SIC2-level average standardized ESG score, sequentially adding rater, year, industry, and industry-by-rater fixed effects. Panel B presents analogous results at the firm level, where the standardized ESG score is regressed on rater, year, year-by-rater, industry, industry-by-rater, firm, and firm-by-rater fixed effects, cumulatively. The incremental R^2 column reports the increase in explanatory power attributable to each additional set of fixed effects.

Panel A. Industry Average-level ESG Score

	(1) R^2	(2) Incremental R^2
Rater FE	0.35%	
Year FE	0.42%	0.07%
Industry FE	53.37%	52.95%
Industry x Rater FE	81.38%	28.01%
Number of SIC industries	65	
N	1788	

Panel B. Firm-level ESG Score

	(1) R^2	(2) Incremental R^2
Rater FE	0.00%	
Year FE	0.00%	0.00%
Year x Rater FE	0.00%	0.00%
Industry FE	6.56%	6.56%
Industry x Rater FE	11.24%	4.68%
Firm FE	67.29%	56.05%
Firm x Rater FE	88.57%	21.28%
Industry-level Confusion	4.68%	
Relative magnitude	22%	
Number of firms	1742	
N	34880	

Table 6: R^2 **Decomposition - Environmental Dimension**

Table 6 presents the R^2 decompositions for ENV (Environmental) scores in both industry-level average and firm-level. The analysis employs the 2-digits SIC industry classification as the industry fixed effects. In each panel, fixed effects are introduced cumulatively: each subsequent specification adds the listed fixed effect(s) to all previously included fixed effects. Panel A presents results from regressions of the SIC2-level average standardized ENV score, sequentially adding rater, year, industry, and industry-by-rater fixed effects. Panel B presents analogous results at the firm level, where the standardized ENV score is regressed on rater, year, year-by-rater, industry, industry-by-rater, firm, and firm-by-rater fixed effects, cumulatively. The incremental R^2 column reports the increase in explanatory power attributable to each additional set of fixed effects.

Panel A. Industry Average-level ENV Score

	(1) R^2	(2) Incremental R^2
Rater FE	0.90%	
Year FE	0.93%	0.03%
Industry FE	48.03%	47.10%
Industry x Rater FE	83.07%	35.04%
Number of SIC2 industries	65	
N	1788	

Panel B. Firm-level ENV Score

	(1) R^2	(2) Incremental R^2
Rater FE	0.00%	
Year FE	0.00%	0.00%
Year x Rater FE	0.00%	0.00%
Industry FE	6.86%	6.86%
Industry x Rater FE	16.90%	10.04%
Firm FE	69.83%	52.93%
Firm x Rater FE	89.31%	19.48%
Industry-level Confusion	10.04%	
Relative magnitude	52%	
Number of firms	1742	
N	34880	

Table 7: **The Source of the Industry-Rater Effect**

Panel A of Table 7 presents how the magnitude of the industry-by-rater interaction effect changes under different experimental settings, allowing us to decompose the sources of industry-level ESG rating divergence for direct (Scope 1) greenhouse gas emissions. In Columns (1) and (2), we report the results using each rater's proprietary methodology, peer group classification, and raw values within their own sample universes. Columns (3) and (4) present results after harmonizing the methodology by applying Refinitiv's scoring approach to each rater's universe and peer groups, while still using each rater's proprietary raw data. The difference in the incremental R^2 for the industry-by-rater term between Columns (2) and (4) isolates the portion of industry-level confusion attributable to methodological differences.Panels B and C of Table 7 extend our analysis by incorporating more pronounced measurement divergence. Specifically, Panel B compares Scope 1 absolute carbon emissions from S&P Global with Refinitiv's revenue-normalized Scope 1 and 2 carbon intensity metrics, while Panel C contrasts S&P Global's Scope 1 absolute emissions with Refinitiv's Scope 2 absolute emissions.

Panel A. Direct Greenhouse Gas Emissions (Scope 1)

	(1) R^2	(2) Incremental R^2	(3) R^2	(4) Incremental R^2	(5) R^2	(6) Incremental R^2
Rater FE	5.90%		0.27%		0.82%	
Year FE	6.27%	0.37%	0.31%	0.04%	0.84%	0.02%
Rater × Year FE	6.37%	0.10%	0.32%	0.01%	0.87%	0.03%
Industry FE	10.00%	3.63%	7.94%	7.62%	9.69%	8.82%
Industry × Rater FE	13.09%	3.09%	8.94%	1.00%	9.80%	0.11%
Firm FE	55.91%	42.82%	91.08%	82.14%	95.63%	85.83%
Firm × Rater FE	91.29%	35.38%	96.98%	5.90%	97.08%	1.45%
Industry-level Confusion	3.09%		1.00%		0.11%	
Decomposition						
Methodology			2.09%			
Comparative Industry Effect					0.89%	
Raw Measurement					0.11%	
N	21238		21238		21238	
Score constructed from	Universe / Universe		Universe / Universe		Universe / Common Sample	
Methodology	Refinitiv / S&P		Refinitiv / Refinitiv		Refinitiv / Refinitiv	
Peers	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / Refinitiv	
Raw value	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / S&P	

Panel B. Scope 1 and 2 Greenhouse Gas Emission Intensity

	(1) R^2	(2) Incremental R^2	(3) R^2	(4) Incremental R^2	(5) R^2	(6) Incremental R^2
Rater FE	2.19%		1.78%		0.02%	
Year FE	2.42%	0.23%	1.79%	0.01%	0.02%	0.00%
Rater × Year FE	2.68%	0.26%	1.79%	0.00%	0.02%	0.00%
Industry FE	7.72%	5.04%	9.09%	7.30%	9.02%	9.00%
Industry × Rater FE	11.88%	4.16%	12.01%	2.92%	10.70%	1.68%
Firm FE	58.19%	46.31%	72.58%	60.57%	76.04%	65.34%
Firm × Rater FE	90.20%	32.01%	96.35%	23.77%	96.40%	20.36%
Industry-level Confusion	4.16%		2.92%		1.68%	
Decomposition						
Methodology			1.24%			
Comparative Industry Effect					1.24%	
Raw Measurement					1.68%	
N	22284		22284		22284	
Score constructed from	Universe / Universe		Universe / Universe		Universe / Common Sample	
Methodology	Refinitiv / S&P		Refinitiv / Refinitiv		Refinitiv / Refinitiv	
Peers	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / Refinitiv	
Raw value	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / S&P	

Panel C. Indirect Greenhouse Gas Emission (Scope 2)

	(1) R^2	(2) Incremental R^2	(3) R^2	(4) Incremental R^2	(5) R^2	(6) Incremental R^2
Rater FE	7.01%		0.11%		1.21%	
Year FE	7.40%	0.39%	0.17%	0.06%	1.24%	0.03%
Rater × Year FE	7.48%	0.08%	0.18%	0.01%	1.27%	0.03%
Industry FE	10.92%	3.44%	7.23%	7.05%	9.03%	7.76%
Industry × Rater FE	14.02%	3.10%	8.79%	1.56%	10.09%	1.06%
Firm FE	53.91%	39.89%	80.67%	71.88%	82.70%	72.61%
Firm × Rater FE	91.39%	37.48%	96.97%	16.30%	97.07%	14.37%
Industry-level Confusion	3.10%		1.56%		1.06%	
Decomposition						
Methodology			1.54%			
Comparative Industry Effect					0.50%	
Raw Measurement					1.06%	
N	21020		21020		21020	
Score constructed from	Universe / Universe		Universe / Universe		Universe / Common Sample	
Methodology	Refinitiv / S&P		Refinitiv / Refinitiv		Refinitiv / Refinitiv	
Peers	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / Refinitiv	
Raw value	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / S&P	

Table 8: **Determinants of ESG Rating Disagreements**

Table 8 provides the results of regression analyses investigating the determinants of rating disagreement. Panel A presents estimates for ESG rating disagreement, whereas Panel B reports corresponding results for environmental (ENV) rating disagreement. We use total disagreement in ESG scores (measured in Standard Deviation), industry-level disagreement, and firm-level disagreement in Columns (1) to (9). Control variables include the natural log of total assets (*Size*), book-to-market ratio (*BtoM*), the ratio of total liability to total equity (*Leverage*), return-on-assets (*ROA*), and an indicator that equals one when a firm is in industries affected by revisions to the GICS structure (*GICS*). All continuous variables are winsorized at extreme 1% and 99% level. Standard errors are clustered at firm level. ***, **, and * denote statistical significance at 1%, 5%, and 10% level respectively.

Panel A. ESG Score

Dep Var=	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
σ(ESG Z-Score)	Total	Industry	Firm	Total	Industry	Firm	Total	Industry	Firm
GICS	0.02	0.05***	0.00	0.02	0.04***	0.00	0.02	0.04***	0.00
	(1.20)	(3.39)	(0.05)	(1.29)	(3.36)	(0.20)	(1.27)	(3.40)	(0.14)
Size				0.01**	-0.00	0.01**	0.01**	-0.00	0.01**
				(2.21)	(-1.10)	(2.21)	(2.21)	(-1.05)	(2.22)
BtoM				0.04**	0.03**	0.04**	0.04**	0.03**	0.04**
				(2.01)	(2.32)	(2.20)	(1.97)	(2.16)	(2.19)
Leverage				0.13	0.04	0.06	0.12	0.03	0.06
				(1.40)	(0.68)	(0.66)	(1.36)	(0.59)	(0.67)
ROA				0.09**	-0.00	0.04	0.09**	-0.00	0.04
				(2.39)	(-0.19)	(1.10)	(2.36)	(-0.20)	(1.01)
Intercept	0.61***	0.23***	0.56***	0.46***	0.23***	0.43***	0.46***	0.27***	0.42***
	(88.29)	(55.19)	(85.02)	(10.85)	(8.70)	(10.55)	(10.74)	(9.98)	(10.07)
Year FE	No	No	No	No	No	No	Yes	Yes	Yes
Ind FE	No	No	No	No	No	No	No	No	No
Cluster	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
N	6972	6972	6972	6972	6972	6972	6972	6972	6972
Adj. R^2	0.001	0.012	-0.000	0.010	0.015	0.008	0.009	0.027	0.008

Panel B. ENV Score

Dep Var= *σ(ESG Z-Score)*	(1) Total	(2) Industry	(3) Firm	(4) Total	(5) Industry	(6) Firm	(7) Total	(8) Industry	(9) Firm
GICS	0.16***	0.23***	0.03**	0.16***	0.23***	0.04***	0.16***	0.23***	0.04***
	(8.43)	(19.03)	(2.31)	(8.50)	(19.19)	(2.66)	(8.48)	(19.26)	(2.65)
Size				0.01*	0.00	-0.00	0.01*	0.00	-0.00
				(1.67)	(0.45)	(-0.23)	(1.67)	(0.41)	(-0.22)
BtoM				-0.04*	0.01	-0.05***	-0.04*	0.01	-0.05***
				(-1.77)	(0.74)	(-2.90)	(-1.82)	(0.63)	(-2.91)
Leverage				-0.08	0.14**	-0.33***	-0.08	0.13**	-0.33***
				(-0.79)	(2.17)	(-3.86)	(-0.83)	(2.06)	(-3.87)
ROA				-0.00	0.02	-0.10***	-0.08	0.02	-0.11***
				(-0.08)	(0.76)	(-2.91)	(-0.08)	(0.88)	(-2.92)
Intercept	0.61***	0.30***	0.52***	0.58***	0.26***	0.63***	0.60***	0.29***	0.63***
	(85.18)	(65.39)	(83.27)	(14.25)	(9.80)	(16.46)	(14.26)	(10.60)	(16.09)
Year FE	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Ind FE	No	No	No	No	No	No	No	No	No
Cluster	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm	Firm
N	6972	6972	6972	6972	6972	6972	6972	6972	6972
Adj. R^2	0.044	0.218	0.003	0.046	0.219	0.013	0.046	0.222	0.013

Table 9: **Cross-Sectional Analysis by Disclosure and Ownership Structure**

Table 9 reports R^2 decompositions for ESG and ENV (Environmental) scores, segmented by disclosure environment (public vs. private firms) and ownership structure (dispersed ownership vs. concentrated ownership). The analysis applies 2-digit SIC industry classification as industry fixed effects. Panels A and B present results for public and private firms, with Panel A showing the decomposition for the composite ESG Z-score and Panel B for the ENV Z-score. Panels C and D extend the analysis to widely held and family firms, classified according to La Porta et al. (1999), across six countries. Panel C displays the decomposition for the composite ESG Z-score, and Panel D for the ENV Z-score. In all panels, the reported values are based on sequential regressions of the standardized ESG or ENV score at the firm level on rater, year, industry (SIC-2 digits), firm, and their respective interaction terms.

Panel A. ESG Z-Score

	Public Firms		Private Firms	
	(1) R^2	(2) Incremental R^2	(3) R^2	(4) Incremental R^2
Rater FE	0.01%		3.36%	
Year FE	0.02%	0.01%	3.75%	0.39%
Year × Rater FE	0.02%	0.00%	3.99%	0.24%
Industry FE	6.48%	6.46%	37.16%	33.17%
Industry × Rater FE	11.18%	4.70%	51.93%	14.77%
Firm FE	67.13%	55.95%	80.56%	28.63%
Firm × Rater FE	88.55%	21.42%	89.42%	8.86%
Industry-level Confusion	4.70%		14.77%	
Relative magnitude	21.94%		166.70%	
Number of firms	1599		65	
N	32776		1116	

Panel B. ENV Z-Score

	Public Firms		Private Firms	
	(1) R^2	(2) Incremental R^2	(3) R^2	(4) Incremental R^2
Rater FE	0.01%		1.44%	
Year FE	0.02%	0.01%	1.79%	0.35%
Year × Rater FE	0.02%	0.00%	2.45%	0.66%
Industry FE	6.96%	6.94%	29.20%	26.75%
Industry × Rater FE	16.83%	9.87%	46.06%	16.86%
Firm FE	69.54%	52.71%	85.22%	39.16%
Firm × Rater FE	89.30%	19.76%	92.26%	7.04%
Industry-level Confusion	9.87%		16.86%	
Relative magnitude	49.95%		239.49%	
Number of firms	1599		65	
N	32776		1116	

Panel C. ESG Z-Score

	Dispersed Ownership						Concentrated Ownership					
	United States		Japan		United Kingdom		Mexico		Belgium		Sweden	
	(1) R^2	(2) Incr.	(3) R^2	(4) Incr.	(5) R^2	(6) Incr.	(7) R^2	(8) Incr.	(9) R^2	(10) Incr.	(11) R^2	(12) Incr.
Rater FE	0.58%		1.05%		0.54%		8.99%		1.13%		10.23%	
Year FE	0.62%	0.04%	1.39%	0.34%	1.10%	0.56%	9.46%	0.47%	4.38%	3.25%	12.57%	2.34%
Year × Rater FE	0.82%	0.20%	2.09%	0.70%	1.39%	0.29%	10.71%	1.25%	5.76%	1.38%	13.34%	0.77%
Industry FE	11.35%	10.53%	15.99%	13.90%	22.07%	20.68%	50.58%	39.87%	55.52%	49.76%	34.87%	21.53%
Industry × Rater FE	18.56%	7.21%	25.55%	9.56%	49.14%	27.07%	67.90%	17.32%	73.68%	18.16%	61.61%	26.74%
Firm FE	65.84%	47.28%	65.18%	39.63%	72.71%	23.57%	84.87%	16.97%	81.72%	8.04%	73.30%	11.69%
Firm × Rater FE	87.07%	21.23%	84.32%	19.14%	86.72%	14.01%	89.75%	4.88%	87.43%	5.71%	82.34%	9.04%
Relative magnitude	34%		50%		193%		355%		318%		296%	
Number of firms	784		167		90		14		10		23	
Ownership Concentration	0		0		-		0.25		0.79		0.53	
N	13916		4164		1884		340		232		552	

Panel D. ENV Z-Score

	Dispersed Ownership						Concentrated Ownership					
	United States		Japan		United Kingdom		Mexico		Belgium		Sweden	
	(1) R^2	(2) Incr.	(3) R^2	(4) Incr.	(5) R^2	(6) Incr.	(7) R^2	(8) Incr.	(9) R^2	(10) Incr.	(11) R^2	(12) Incr.
Rater FE	1.67%		1.76%		0.74%		0.78%		0.56%		8.96%	
Year FE	1.71%	0.04%	2.11%	0.35%	1.29%	0.55%	1.68%	0.90%	3.96%	3.40%	10.01%	1.05%
Year × Rater FE	1.73%	0.02%	2.30%	0.19%	1.72%	0.43%	3.12%	1.44%	5.27%	1.31%	11.47%	1.46%
Industry FE	10.77%	9.04%	19.93%	17.63%	27.13%	25.41%	29.12%	26.00%	59.30%	54.03%	27.65%	16.18%
Industry × Rater FE	21.99%	11.22%	34.75%	14.82%	48.58%	21.45%	54.70%	25.58%	83.83%	24.53%	70.78%	43.13%
Firm FE	69.21%	47.22%	67.40%	32.65%	73.76%	25.18%	79.46%	24.76%	89.65%	5.82%	81.07%	10.29%
Firm × Rater FE	88.73%	19.52%	86.59%	19.19%	87.18%	13.42%	88.44%	8.98%	91.10%	1.45%	86.84%	5.77%
Relative magnitude	57%		77%		160%		285%		1692%		747%	
Number of firms	784		167		90		14		10		23	
Ownership Concentration	0		0		-		0.25		0.79		0.53	
N	13916		4164		1884		340		232		552	

Table 10: **Market Consequences of ESG Disagreement**

Table 10 reports the results of regression analyses investigating the market consequences of rating disagreement. We measure disagreement using the absolute deviation of a given rater's ESG score from the consensus of the other raters for the same firm. Abnormal return volatility is defined as the sum of the absolute abnormal returns over the [1,+1] window around the last rater's announcement date. Abnormal returns are computed as residuals from a one-factor market model estimated over the pre-announcement window (200,11), where the CRSP value-weighted market return is used as the proxy for the market factor. Control variables include the natural log of total assets (*Size*), book-to-market ratio (*BtoM*), the ratio of total liability to total equity (*Leverage*), and return-on-assets (*ROA*). All continuous variables are winsorized at extreme 1% and 99% level. Standard errors are clustered at firm level. ***, **, and * denote statistical significance at 1%, 5%, and 10% level respectively.

Panel A. Disagreement on Raw ESG Score

Dep. var:	Abnormal Return Volatility
	(1)
$\lvert \Delta Rating_{\text{last}} \rvert$	0.0011**
	(2.03)
Size	-0.0021***
B/M	0.0095***
ROA	-0.0161***
Leverage	0.0014
N	3005
Year FE	Yes
Industry FE	Yes
Adj. R^2	0.122

Panel B. Disentanglement of Industry-level and Firm-level Components

Dep. var: Abnormal Return Volatility			
	(1)	(2)	(3)
Industry $\lvert\Delta Rating\rvert$	0.0002**		0.0002*
	(2.08)		(1.89)
Firm $\lvert\Delta Rating\rvert$		0.0001	0.0001
		(1.54)	(1.15)
Consensus (avg_others)	-0.0007*	-0.0007*	-0.0007*
	(-1.92)	(-1.77)	(-1.85)
Size	-0.0021***	-0.0022***	-0.0021***
	(-7.45)	(-7.46)	(-7.45)
BtoM	0.0115***	0.0117***	0.0115***
	(5.99)	(5.98)	(6.00)
ROA	-0.0046	-0.0038	-0.0045
	(-0.66)	(-0.54)	(-0.64)
Leverage	-0.0008	-0.0008	-0.0008
	(-0.38)	(-0.40)	(-0.40)
Intercept	0.0246***	0.0264***	0.0245***
	(7.11)	(7.83)	(7.11)
N	3605	3605	3605
Year FE	Yes	Yes	Yes
Adj. R^2	0.073	0.072	0.073

Table 11:

Table 11 reports the results of regression analyses examining whether underreporting in carbon emissions disclosure gives rise to systematic industry-level disagreement across ESG raters. The dependent variable is industry-level disagreement (*Ind_std*), measured as the standard deviation of ESG scores across raters within a given two-digit SIC industry and year. *Average˙coverage* is constructed to capture the extent of carbon emissions disclosure availability across ESG raters at the industry level. This is the simple average of these rater-specific SIC2–year coverage rates across all ESG raters. Standard errors are clustered at firm level. ***, **, and * denote statistical significance at 1%, 5%, and 10% level respectively.

Panel A.

Dependent var. *Ind_std*	
Average_coverage	-0.0884
	(0.0669)
	[-1.32]
Intercept	0.3953***
	(0.0480)
	[8.23]
Observations	447
R-squared	0.0039
Adj. R-squared	0.0017
Root MSE	0.2301

Appendix

Appendix A. Replication Methodology of Refinitiv Score

The ESG score evaluates a company's environmental, social, and governance (ESG) performance using verifiable, publicly available data. Refinitiv's ESG score is based on 10 category scores[19], which are aggregated into three pillar scores representing environmental, social, and corporate governance performance. These category scores are derived from 182 specific ESG metrics, which can be either boolean or numeric. Boolean metrics are typically binary responses ('Yes,' 'No,' or 'Null'). This paper focuses on numeric data, as the primary aim is to analyze divergence across raters caused by relative performance evaluation. When a company reports a numeric data point, the score for that indicator is calculated based on a relative percentile ranking among its industry peers.

The percentile rank scoring methodology is used to calculate the 10 category scores. This method ranks data points[20] within a group, making it less sensitive to outliers. The percentile rank score is calculated using the following formula:

$$\text{score} = \frac{\text{no. of companies with a worse value} + \frac{\substack{\text{no. of companies with the same value} \\ \text{included in the current one}}}{2}}{\text{no. of companies with a value}}$$

This method ranks a company relative to its peers, accounting for both companies with worse values and those with the same value, ensuring a more balanced assessment across industries. More detailed methodology is referred to Refinitiv (2022).

[19]The 10 categories are organized as follows: In the *environmental dimension*, they include (1) Resource Use, (2) Emissions, and (3) Innovation. In the *social dimension*, they cover (4) Workforce, (5) Human Rights, (6) Community, and (7) Product Responsibility. Finally, in the *governance dimension*, they encompass (8) Management, (9) Shareholders, and (10) CSR Strategy.

[20]Each metric has a polarity indicating whether a higher value of the data point is positive or negative. For example, more water recycled is considered positive, while higher emissions are negative. If the polarity is negative (positive), the raw value of the data point should be multiplied by -1 (+1) when constructing the ranking.

Appendix B. Replication Methodology of S&P Score

S&P's ESG score is based on 68 criteria scores[21], using both public and private company data. The criterion score is the weighted average of the individual question scores based on their respective weights. For example, within the Talent Attraction & Retention criterion score, multiple ESG metrics (or questions) are assessed, such as Employee Turnover Rate and Long-term Incentives for Employees. While Refinitiv requires just one data point to evaluate the score of a particular ESG metric, S&P requires multiple data points (i.e., assessment focus, as shown in Appendix Table A1).In the case of the Employee Turnover Rate question, the data points related to disclosure/transparency, public documentation, and performance are used. Additional credits are awarded if relevant publicly available evidence is provided, along with specific breakdowns (e.g., by age, gender, management level, or racial, ethnic, national, or cultural background) as part of the disclosure/transparency assessment. In this paper, we calculate the question score by averaging the data points, which range from 0 to 100. Our focus is primarily on the data points for disclosure/transparency, public documentation, performance, and external verification. The full 100 points are awarded if the information is publicly available, third-party verified, and includes at least two disclosed breakdowns. The performance measure follows a similar score calculation methodology to Refinitiv's numeric data points. After applying a percentile rank score to evaluate performance, we use min-max normalization to scale the performance score between 0 and 100 as a linear peer-group comparison. The final question score is the average of these four assessment focuses. Further details on the methodology can be found in S&P Global Market Intelligence

[21]The 68 criteria scores are: (1) Anti-crime Policy & Measures Score, (2) Brand Management Score, (3) Codes of Business Conduct Score, (4) Compliance with Applicable Export Control Regimes Score, (5) Customer Relationship Management Score, (6) Efficiency Score, (7) Strategy for Emerging Markets Score, (8) Financial Stability and Systemic Risk Score, (9) Fleet Management Score, (10) Energy Mix Score, (11) Health & Nutrition Score, (12) Information Security, Cybersecurity & System Availability Score, (13) Innovation Management Score, (14) Privacy Protection Score, (15) Market Opportunities Score, (16) Materiality Score, (17) Marketing Practices Score, (18) Network Reliability Score, (19) Policy Influence Score, (20) Principles for Sustainable Insurance Score, (21) Product Quality and Recall Management Score, (22) Reliability Score, (23) Risk & Crisis Management Score, (24) Supply Chain Management Score, (25) Sustainable Construction Score, (26) Tax Strategy Score, (27) Biodiversity Score, (28) Building Materials Score, (29) Sustainable Finance Score, (30) Climate Strategy Score, (31) Co-Processing Score, (32) Electricity Generation Score, (33) Environmental Policy & Management Systems Score, (34) Environmental Reporting Score, (35) Fuel Efficiency Score, (36) Genetically Modified Organisms Score, (37) Hazardous Substances Score, (38) Low Carbon Strategy Score, (39) Mineral Waste Management Score, (40) Operational Eco-Efficiency Score, (41) Packaging Score, (42) Product Stewardship Score, (43) Raw Material Sourcing Score, (44) Recycling Strategy Score, (45) Resource Conservation and Resource Efficiency Score, (46) Financial Risks of Climate Change Score, (47) Sustainable Forestry Practices Score, (48) Transmission & Distribution Score, (49) Water Operations Score, (50) Water Related Risks Score, (51) Strategy to Improve Access to Drugs or Products Score, (52) Addressing Cost Burden Score, (53) Asset Closure Management Score, (54) Responsibility of Content Score, (55) Corporate Citizenship and Philanthropy Score, (56) Financial Inclusion Score, (57) Human Capital Development Score, (58) Human Rights Score, (59) Labor Practice Indicators Score, (60) Local Impact of Business Operations Score, (61) Occupational Health and Safety Score, (62) Passenger Safety Score, (63) Social Impacts on Communities Score, (64) Social Integration & Regeneration Score, (65) Social Reporting Score, (66) Stakeholder Engagement Score, (67) Partnerships Towards Sustainable Healthcare Score, and (68) Talent Attraction & Retention Score.

(2022).

Assessment Focus	Description of Information Sought
Disclosure/Transparency	Disclosure of qualitative/quantitative information
Documents	Document supporting company's response
Public Documents	Publicly available document supporting company's response
Exposure/Coverage	Coverage of measures implemented, or data reported
Trend	Trend of key indicators in the last three / four years
Performance	Performance of key indicators in comparison to SAM's expected threshold
Awareness	Awareness about internal and external issues and measures taken
External Verification	Third-party verification of data or of processes

Appendix Table 1: Assessment Focus of S&P

Appendix C. Comparison of Replicated Scores and Reported ESG Ratings Correlation

	Refinitiv Category Score			S&P Category Score		
Universe	N	Mean	Std	N	Mean	Std
Reported Score	17901	50.40	28.67	29644	32.17	37.25
Replicated Score	17901	50.00	28.86	29644	37.16	31.72
	Correlation		P-value	Correlation		P-value
	0.94		<.0001	0.74		<.0001

Appendix Table 2: Correlation Between Replicated Scores and Reported ESG Scores

Following the replication methodologies outlined in Appendices A and B, we replicated the ESG score using the raw values of carbon emission data from each rater. The correlation between the reported scores from Refinitiv and S&P and the replicated scores are 0.94 and 0.74 (see Appendix Table A2), respectively. The relatively low correlation for the S&P replication can be attributed to the multiple assessment focuses included in the S&P rating (see Appendix Table A1), with each focus being addressed in greater detail through their own specialized expertise. In this paper, we adopt the replication methodology of Refinitiv, as the replication scores exhibit a high correlation of 95% with the reported scores.

Appendix D. Sample Comparison with Berg et al. (2022b)

We compare our 2014 sample with that used in Berg et al. (2022b). While our study examines the intersection of four raters, Berg et al. (2022b) investigate the intersection of six raters (Refinitiv, Sustainalytics, S&P Global, KLD, Moody's ESG, and MSCI). Despite differences in the timing of data acquisition, our descriptive statistics are generally consistent with those reported in Berg et al. (2022b). One notable difference is that we report a lower average score for Refinitiv (Asset4), which may be attributed to variations in ESG scores and coverage across different data download platforms[22]. Since our analysis is based on the intersection of four raters, we include data from S&P, which has the smallest coverage among the raters examined in Berg et al. (2022b). Consequently, the distribution of our common sample is unlikely to deviate significantly from prior literature. To mitigate the variation introduced by the merging process, we conduct an experiment that matches the full sample of Refinitiv and S&P data (Table 7). Moreover, the pairwise correlations of MSCI, Refinitiv (Asset4), Sustainalytics, and S&P ESG scores in our sample are 99.99%, 99.94%, 99.99%, and 99.98%, respectively. Taken together, these findings indicate that our sample is broadly representative of those used in previous studies, mitigating concerns that sample selection biases may be driving our main results.

[22]Refinitiv provides ESG data through two platforms: LSEG Eikon Datasets and Wharton Research Data Services (WRDS).

Panel A. Full Sample Comparison

As of 2014	Full Sample in This Paper				Berg et al.			
	MSCI	Refinitiv	Sustainalytics	S&P	MSCI	Refinitiv	Sustainalytics	S&P
N	11,263	4,352	4,852	3,003	9,662	4,013	4,531	1,665
Mean	4.25	41.88	57.44	43.49	4.70	50.90	56.40	47.19
Std	1.22	20.57	9.81	20.64	1.19	30.94	9.46	21.06
Min	0.00	0.65	31.00	13.00	0.00	2.78	29.00	13.00
Median	4.00	40.24	56.00	35.00	4.70	53.15	55.00	40.00
Max	9.00	92.89	100.00	94.00	9.80	97.11	89.00	94.00

Panel B. Common Sample Comparison

As of 2014	Common Sample used in this paper				Berg et al.			
	MSCI	Refinitiv	Sustainalytics	S&P	MSCI	Refinitiv (Asset4)	Sustainalytics	S&P
N	1,115	1,115	1,115	1,115	924	924	924	924
Mean	4.49	55.36	62.07	47.18	5.18	73.47	61.86	50.49
Std	1.27	19.05	9.97	20.63	1.22	23.09	9.41	20.78
Min	0.00	1.60	31.00	16.00	0.60	3.46	37.00	14.00
Median	4.00	57.56	62.00	40.00	5.10	81.48	62.00	47.00
Max	9.00	92.89	87.00	93.00	9.80	97.11	89.00	94.00

Appendix Table 3: Sample Comparison

Appendix E. Disaggregation Model - Euclidean Distance Measure

This table presents the disaggregation of ESG and ENV (Environmental) rating disagreement into industry-level and firm-level components. We regress ESG (and ENV) scores on year-industry fixed effects, using the residuals as firm-level scores. Industry fixed effects are defined using SIC 2-digit classifications. Panels A and B report the results when using the Euclidean distance of the ESG and ENV scores, respectively, as the measure of rating disagreement.

The results remain consistent with those based on the standard deviation approach. As shown in Panel A of Table A4, the year-by-year decomposition reveals a stable disaggregation pattern over the sample period from 2013 to 2019. Industry-level disagreement ranges from 0.26 (in 2015 and 2017) to 0.33 (in 2013), while firm-level disagreement remains similarly consistent, ranging from 0.67 to 0.70. This temporal stability confirms our interpretation that industry-level disagreement is a persistent and systematic feature of ESG rating methodologies, rather than a result of transitory variation or noise.

Panel A. ESG Score

σ(ESG Z-Score) of	(1) N	(2) Total Score	(3) Industry-Level	(4) Firm-Level
Full Sample	8720	0.75	0.28	0.68
By year				
2013	847	0.76	0.33	0.67
2014	1115	0.74	0.29	0.66
2015	1148	0.74	0.26	0.68
2016	1283	0.74	0.27	0.67
2017	1331	0.74	0.26	0.68
2018	1370	0.76	0.28	0.70
2019	1626	0.75	0.30	0.68

Panel B. ENV Score

σ(ENV Z-Score) of	N (1)	Total Score (2)	Industry-Level (3)	Firm-Level (4)
Full Sample	8720	0.78	0.41	0.64
By year				
2013	847	0.79	0.43	0.64
2014	1115	0.77	0.42	0.63
2015	1148	0.77	0.42	0.63
2016	1283	0.76	0.40	0.63
2017	1331	0.78	0.42	0.64
2018	1370	0.79	0.41	0.66
2019	1626	0.78	0.41	0.64

Appendix Table 4: Disaggregation Model using Euclidean Distance Measure

Appendix F. The Structural Change in GICS (Global Industry Classification Structure)

Every November, MSCI Inc. and S&P Dow Jones Indices, prominent providers of ESG ratings, announce the final results of revisions to the Global Industry Classification Standard (GICS) for the years 2014, 2016, and 2018 (S&P and MSCI (2013), S&P and MSCI (2015), and S&P and MSCI (2017)). These revisions may involve the creation of new sub-industries, the merging of existing sub-industries, or the discontinuation of certain sub-industries. For example, a new sub-industry, Renewable Electricity, was introduced. Such changes can significantly impact the composition of firms within the broader industry group. Not only does the firm composition change, but the naming conventions and definitions of sub-industries are also updated to reflect market trends. For instance, the Independent Power Producers & Energy Traders industry was renamed to Independent Power & Renewable Electricity Producers to incorporate the newly introduced Renewable Electricity sub-industry.

The updated definitions for the relevant sub-industries are as follows:

Independent Power Producers & Energy Traders: Companies that operate as Independent Power Producers (IPPs), Gas & Power Marketing & Trading Specialists and/or Integrated Energy Merchants. Excludes producers of electricity using renewable sources, such as solar power, hydropower, and wind power. Also excludes electric transmission companies and utility distribution companies classified in the Electric Utilities Sub-Industry.

Renewable Electricity: Companies that engage in generation and distribution of electricity using renewable sources, including, but not limited to, companies that produce electricity using biomass, geothermal energy, solar energy, hydropower, and wind power. Excludes companies manufacturing capital equipment used to generate electricity using renewable sources, such as manufacturers of solar power systems and installers of photovoltaic cells and companies involved in the provision of technology, components, and services mainly to this market.

Through these revisions in sub-industry composition, naming, and definitions, the components of firms within a given sub-industry may vary, which can influence the overall industry composition and the way firms are classified within the GICS structure.

Year	Announcement Date	Affected Industry	Composition Changes in Sub-Industry
2014	2013-11-05	The Independent Power Producers & Energy Traders	The industry is being renamed to Independent Power & Renewable Electricity Producers and a new Sub-Industry is being created for Renewable Electricity.
		Metals & Mining	A new Sub-Industry is being created for Silver in the Metals & Mining Industry.
		Machinery	The Construction & Farm Machinery & Heavy Trucks Sub-Industry is being split into two Sub-Industries: Construction Machinery & Heavy Trucks and Agricultural & Farm Machinery.
		REITs	Two new REITs Sub-Industries are being created. Hotel & Resort REITs and Health Care REITs are being carved out from Specialized REITs.
2016	2015-11-02	Diversified Financial Services & Capital Markets	A new sub-industry for Financial Exchanges & Data is being created. The new sub-industry is being moved from the Diversified Financial Services industry and into the Capital Markets industry in the Financials sector.
		Internet & Direct Marketing Retail	Due to the increasing overlap between online retailers and catalog retailers, the Catalog Retail sub-industry is being discontinued. The companies are being combined into the Internet Retail sub-industry, which is being renamed to Internet & Direct Marketing Retail.
		Equity Real Estate Investment Trusts	Real Estate is being moved out from under the Financials sector and is now being promoted to its own sector. The Real Estate Investment Trusts industry is being renamed to Equity Real Estate Investment Trusts (REITs) and excludes Mortgage REITs.
		Metals & Mining	In order to distinguish between copper and other metals in the Diversified Metals & Minerals sub-industry, a Copper sub-industry is being created in the Metals & Mining industry.
2018	2017-11-15	Media	The Media Industry Group will move out of Consumer Discretionary and into the Communication Services Sector, and will be renamed Media & Entertainment. The Media Industry will contain four existing Sub-Industries: Advertising, Broadcasting, Cable & Satellite, and Publishing.
		Entertainment	The Entertainment Industry will contain two Sub-Industries: Movies & Entertainment and Interactive Home Entertainment. The Movies & Entertainment Sub-Industry will now include online entertainment streaming companies in addition to companies currently classified in the Movies & Entertainment Sub-Industry.
		Interactive Media & Services	The Interactive Home Entertainment Sub-Industry will be comprised of companies from the current Home Entertainment Software Sub-Industry in the Information Technology Sector, as well as producers of mobile gaming applications.
		IT Services	The existing Internet Software & Services Industry and Sub-Industry will be discontinued. A new Sub-Industry will be created under the IT Services Industry called Internet Services & Infrastructure.

Appendix Table 5: Structural Change in GICS

Appendix G. The Source of the Industry-Rater Effect - Social Indicator: Employee Turnover Rate

	(1) Adj. R^2	(2) Incremental R^2	(3) Adj. R^2	(4) Incremental R^2	(5) Adj. R^2	(6) Incremental R^2
Rater FE	2.93%		0.38%		0.03%	
Year FE	3.71%	0.78%	0.37%	−0.00%	0.02%	−0.01%
Rater × Year FE	4.27%	0.56%	0.36%	−0.01%	0.02%	−0.01%
Industry FE	6.14%	1.87%	2.50%	2.13%	3.50%	3.49%
Firm FE	31.22%	25.08%	79.77%	77.27%	83.70%	80.19%
Industry × Rater FE	32.61%	1.39%	80.87%	1.10%	83.99%	0.29%
Firm × Rater FE	61.91%	29.30%	83.86%	2.99%	83.97%	−0.02%
Industry-level Confusion	1.39%		1.10%		0.29%	
Decomposition						
Methodology			0.29%			
Comparative Industry Effect					0.81%	
Raw Measurement					0.29%	
N	8724		8724		8724	
Score constructed from	Universe / Universe		Universe / Universe		Universe / Common sample	
Methodology	Refinitiv / S&P		Refinitiv / Refinitiv		Refinitiv / Refinitiv	
Peers	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / Refinitiv	
Raw value	Refinitiv / S&P		Refinitiv / S&P		Refinitiv / S&P	

Appendix Table 6: The Source of the Industry-Rater Effect - Social Indicator

Appendix H. Definition of variables

This table provides detailed definitions of control variables employed in the empirical analysis. The variable formulas follow standard Compustat mnemonics. Consistent with Strebulaev and Yang (2013), we construct these variables following their methodology, and measure stock prices (PRCC_F) as of fiscal year-end (DATADATE). For firms in the Compustat Global database, PSTKL is replaced with PSTK, and CSHO with CSHOi, to ensure comparability.

Variable	Description	Definition
Size	Natural logarithm of total book assets	ln(AT)
Book-to-Market Ratio	Ratio of book assets (Tobin's q) to market assets	$\frac{\text{AT}}{\text{LT+PSTKL−TXDITC+(CSHO×PRCC\_F)}}$
Return on Assets (ROA)	Ratio of earnings before interest, taxes, and depreciation to book assets	$\frac{\text{OIBDP}}{\text{AT}}$
Leverage Ratio	Ratio of total book debt (long-term and current) to total assets	$\frac{\text{DLTT+DLC}}{\text{AT}}$

Appendix Table 7: Definition of variables