

Systematic Review

Understanding ESG Ratings: A Systematic Literature Review of Methodologies, Divergences, Impact, Standardization, Disclosure Quality, Technology, and Global Financial Implications (2020–2025)

Hannan Vilchis Zubizarreta * and Delfor Tito Aquino

School of Architecture, Universidad de Navarra, 31009 Pamplona, Spain; dtitoaquino@alumni.unav.es

* Correspondence: hvilchis-zu@alumni.unav.es

Abstract

Purpose: This paper aims to systematically synthesize academic research published between 2020 and 2025 that investigates environmental, social, and governance (ESG) ratings and scores, with a focus on their methodologies, comparative performance, and impact on firm outcomes. **Design/methodology/approach:** A systematic literature review (SLR) was conducted using the Lens.org scholarly database. A structured title search retrieved 334 open access journal articles published between 2020 and May 2025 containing the terms “ESG Score”, “ESG Rating”, or “ESG Rater”. The PRISMA 2020 protocol guided the selection and screening process. **Findings:** The literature exhibits growing concern about the divergence among ESG ratings, the methodological opacity of rating providers, and the variable financial implications of ESG scores. Common themes include score disagreements, rating agency biases, and emerging models for standardizing ESG assessments. **Originality:** This review provides the most up-to-date synthesis of the ESG rating literature, focusing exclusively on articles explicitly addressing ESG ratings or scores in their titles. It contributes clarity to the fragmented ESG measurement space by organizing findings around key methodological and evaluative debates.

Keywords: ESG rating; ESG score; ESG rater; systematic review; sustainability assessment; ESG divergence

1. Introduction

Environmental, Social, and Governance (ESG) ratings have become pivotal instruments in the global push toward responsible investment and corporate accountability. Their increasing adoption by institutional investors, asset managers, and regulators is transforming both capital allocation and corporate behavior, reflecting a broader societal demand for sustainable business practices [1–3]. However, the proliferation of ESG ratings has also exposed critical tensions at the heart of sustainability assessment. Chief among these is the growing evidence of substantial divergence between rating providers—where different agencies assign markedly different scores to the same firms, resulting in confusion, inefficiency, and skepticism regarding the reliability and comparability of ESG signals [1,4,5].

From an interpretative standpoint, this review adopts a critical realist perspective, acknowledging that ESG ratings are socially constructed instruments shaped by methodological choices, institutional contexts, and stakeholder interests rather than objective



Academic Editor: Pierfrancesco De Paola

Received: 26 December 2025

Revised: 10 April 2026

Accepted: 16 April 2026

Published: 16 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

measures of corporate sustainability. The authors position ESG ratings as boundary objects that mediate between corporate practices and market expectations, recognizing that their meaning and utility are contingent upon the frameworks and assumptions embedded within rating methodologies.

This divergence is not merely a technical curiosity but has tangible consequences for market participants and policymakers. For investors, inconsistent ESG ratings complicate portfolio construction, risk management, and engagement strategies, while for firms, they introduce reputational uncertainty and potentially distort access to capital [6–8]. From a regulatory perspective, the lack of alignment among rating systems poses a significant obstacle to efforts at harmonizing disclosure standards and ensuring the integrity of sustainable finance initiatives [2,9].

A critical distinction that warrants explicit attention is the dual function of ESG ratings: they serve both as measures of financial risk (assessing how environmental, social, and governance factors may affect firm value and investment returns) and as indicators of real sustainability impact (evaluating a firm's actual contributions to environmental and social outcomes). These two conceptualizations are not synonymous—an ESG rating may effectively signal financial risk while providing limited insight into genuine sustainability performance, and vice versa. This review explicitly examines both dimensions, recognizing that conflating them contributes to methodological confusion and undermines the practical utility of ESG assessments.

Compounding these challenges is the fact that the theoretical and empirical relationship between ESG performance and firm financial outcomes remains highly contested. While some studies report a positive link between high ESG scores and favorable financial metrics such as lower cost of capital or higher firm value, others point to null or even negative associations, often influenced by context, sector, and the construction of ESG measures themselves [10–12]. Methodological opacity, inconsistent data sources, and varying approaches to weighting ESG pillars further muddy the landscape, raising concerns about greenwashing, rating bias, and the efficacy of current frameworks [13–15].

As ESG ratings become increasingly central to regulatory frameworks—particularly in jurisdictions such as the European Union, where initiatives like the CSRD and EU Taxonomy set ambitious new standards for sustainability disclosure—the need for methodological clarity, transparency, and contextual sensitivity has never been more urgent [3,16,17]. These dynamics are especially pronounced in emerging markets and among small and medium-sized enterprises (SMEs), where disclosure practices and rating methodologies must adapt to local realities in order to avoid misrepresentation and promote meaningful impact [18–20].

2. Methodology

The decision to restrict the search to articles explicitly containing “ESG Score,” “ESG Rating,” or “ESG Rater” in their titles was motivated by the need to ensure thematic precision and relevance. Given the expansive nature of ESG-related scholarship, a title-specific search strategy minimizes the inclusion of peripheral studies that address ESG only tangentially, thereby maintaining focus on research that directly engages with rating methodologies and scoring mechanisms. This approach prioritizes depth over breadth, enabling a more rigorous synthesis of the specific literature addressing ESG measurement and evaluation.

This review adopts the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) framework to ensure transparency and rigor throughout the selection and synthesis process. The literature search was conducted using the Lens.org scholarly database, employing a structured query targeting articles with the terms “ESG

Score,” “ESG Rating,” or “ESG Rater” in their titles. The search parameters were further refined to include only peer-reviewed journal articles, open access publications, and works published in English between 2020 and 15 May 2025. This initial search yielded a total of 334 articles. After removing duplicates, titles and abstracts were screened to assess their relevance to the topic. Full-text review was subsequently performed, with strict inclusion criteria requiring that studies directly address ESG ratings, scoring mechanisms, or the role of rating providers. Articles focusing solely on general ESG investing or sustainability themes without substantive discussion of rating methodologies were excluded from the final analysis. This process ensured a focused and thematically coherent corpus, enabling robust synthesis and critical evaluation of recent developments in ESG rating scholarship.

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [21]. The PRISMA 2020 checklist was completed to ensure transparent reporting of all review stages, including identification, screening, eligibility assessment, and inclusion. Figure 1 presents the PRISMA 2020 flow diagram, illustrating the study selection process from initial database search through to final inclusion.

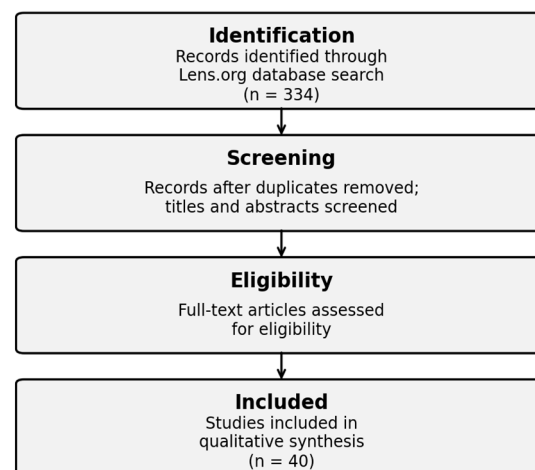


Figure 1. PRISMA 2020 flow diagram of the study identification, screening, eligibility assessment, and inclusion process.

This review was not registered in a prospective register (e.g., PROSPERO), as it was conducted specifically for journal submission purposes without prior access to systematic review registration infrastructure. The review protocol was internally documented prior to commencement of the literature search.

Figure 1 presents the PRISMA 2020 flow diagram, illustrating the study selection process from initial database search through to final inclusion. The diagram shows the number of records identified ($n = 334$), screened, assessed for eligibility, and included in the qualitative synthesis ($n = 40$).

3. Thematic Findings from the Literature

3.1. Divergence and Disagreement Among ESG Ratings

A substantial body of recent scholarship has converged on the problem of ESG rating divergence—the phenomenon where different ESG rating agencies produce conflicting scores for the same firm. This inconsistency, as highlighted across multiple empirical and qualitative studies, poses significant challenges for investors, analysts, and regulators seeking to integrate sustainability metrics into financial decision-making.

Quantitative analyses such as Berg et al. [1] offer foundational insight, employing data from six major rating agencies to demonstrate that average inter-rater correlations range from only 0.38 to 0.71, with disagreement primarily driven by differences in measurement (56%), followed by scope (38%) and weighting (6%). Notably, the presence of a “rater effect” or halo bias—where an agency’s overall perception influences all category assessments—compounds these inconsistencies. The authors recommend the adoption of harmonized taxonomies and decomposition methods to increase transparency, echoing calls for standardization throughout the literature [1,4,22].

The practical implications of this divergence are manifold. Institutional investors, as Tukiainen [8] reports from a Nordic context, are acutely aware of rating inconsistencies and thus rarely accept ESG ratings at face value. Instead, they supplement ratings with their own interpretive frameworks or triangulate across multiple providers, treating divergence as a signal of broader methodological weaknesses—particularly the lack of standardized data sources and measurement techniques.

Several studies interrogate the market-level consequences of rating disagreement. Yu and Wu [23], using a panel of over 13,000 Chinese A-share company observations, find that greater ESG rating disparities directly reduce investor confidence, exacerbate information asymmetry, and negatively affect corporate valuations. Media amplification and analyst attention further intensify these effects, while adherence to international standards (e.g., GRI) and third-party verification help to mitigate them. Similarly, Billio et al. [4] show that despite widespread adoption of ESG integration in investment, portfolios based on consensus ESG ratings offer no significant performance advantage after adjusting for risk, highlighting how fragmented ratings dilute the efficacy of sustainable investing.

Empirical findings are reinforced by Mathisen and Sem [24], who analyze portfolios sorted by ESG disagreement and uncover an “uncertainty premium”: high disagreement, especially in governance scores, is associated with positive alpha, as investors demand compensation for bearing additional informational risk. However, the effect varies by ESG component, being absent for environmental factors and negative for social factors—suggesting nuanced impacts across ESG dimensions and further undermining any simplistic reliance on aggregate ESG scores.

Other studies have taken a methodological approach to resolving divergence. Bissoondoyal-Bheenick et al. [5] examine multiple aggregation techniques (e.g., equal-weighting, PCA, voting, optimized scores) and find that such methods can increase reliability, particularly in markets with lower baseline disagreement. Notably, close alignment between certain providers (e.g., Asset4 and Sustainalytics) contrasts sharply with negative correlations for others (e.g., MSCI), especially for environmental measures. This supports arguments that combining or “triangulating” ratings may reduce bias and increase comparability, but also highlights persistent challenges, particularly in heterogeneous markets like the U.S.

The lack of a standardized framework underpins much of the current debate. Jacobs and Levy [25] point out that the same company can occupy top and bottom positions in sector rankings depending on the provider—Tesla is emblematic—while empirical work by Brandon et al. [26] links rating disagreement to higher stock returns, possibly reflecting an “information asymmetry risk premium.” Efforts such as the SASB’s industry-specific standards and MIT’s Aggregate Confusion Project represent attempts to improve comparability, yet scholars continue to call for greater transparency and harmonization in both rating construction and underlying data collection [1,25].

Disclosure quality is a contested lever in resolving divergence. While Benedetti [27] finds that higher-quality ESG reporting correlates with lower rating disagreement, she also documents outliers—firms with high-quality disclosure but persistent rater divergence—

suggesting that even robust disclosure standards cannot fully resolve the issue given continued methodological fragmentation among agencies. Wang, Chen, and Jiang [28] deepen this analysis for the Chinese context, constructing a multidimensional ESG report informativeness index and showing that richer ESG disclosures reliably reduce rater divergence, especially in high-transparency environments and for larger firms.

The implications extend to capital market intermediaries as well. Liu et al. [7] reveal that greater ESG rating disagreement leads to higher analyst forecast errors and dispersion, particularly in settings with low accounting disclosure quality. This “noise effect” is attenuated in firms with higher transparency and among more diligent analysts, suggesting that information asymmetry from ESG disagreement undermines, rather than enhances, the usefulness of non-financial metrics for analysts and investors alike.

Finally, studies like Jørgensen and Ellingsen [29] highlight that rating divergence is especially pronounced in the social and governance pillars due to subjective interpretation, while the environmental pillar benefits from greater quantifiability and thus higher alignment. Disagreement is shown to have tangible market effects—social score dispersion is penalized in stock pricing, while environmental dispersion may even be rewarded, reflecting optimism bias and the complexity of investor preferences.

3.2. ESG Scores and Financial Performance

The relationship between ESG scores and firm financial performance remains a central, yet highly contested, theme in the academic literature. While numerous studies point to a positive association, recent research has highlighted persistent ambiguities, contextual nuances, and methodological shortcomings that complicate straightforward conclusions.

A key advance is provided by Lee, Raschke, and Krishen [11], who examine whether the divergence between ESG scores reported by rating agencies and a theoretically balanced ESG score influences firm outcomes. Drawing on organizational ambidexterity theory, they find that only balanced ESG scores—where environmental, social, and governance pillars receive equal weighting—are reliably linked to superior financial performance among large U.S. firms. In contrast, agency-weighted ESG scores, and positive discrepancies between agency and balanced scores (potential greenwashing), are penalized by the market. This underscores the need for methodological transparency and multidimensional balance in ESG evaluation.

Halid et al. [10] reinforce the empirical ambiguity in the ESG–performance relationship, reviewing global evidence that ranges from positive to negative or neutral effects. Their stakeholder theory-driven conceptual model advocates for integrating composite ESG measures with traditional accounting metrics to better capture the strategic value of ESG investments. The authors highlight that the fragmented research landscape—often analyzing E, S, and G components in isolation—contributes to inconsistent findings and advocate for more holistic and standardized approaches.

Adding further nuance, Bissoondoyal-Bheenick et al. [30] provide disaggregated evidence from G20 countries, showing that social scores (S) most consistently boost firm value (especially in the retail sector), while governance scores (G) can negatively impact excess returns, perhaps due to market concerns about overinvestment in governance mechanisms. These results reinforce the value of examining ESG’s sub-dimensions separately, echoing broader calls to avoid simplistic composite indices that may obscure important sectoral or informational effects.

The cost of capital is a focal point for several studies. Priem and Gabellone [31] find that high ESG scores lower the weighted average cost of capital (WACC) for European firms, but primarily in countries with weaker legal environments, supporting the notion that ESG can substitute for institutional deficiencies. The environmental and social pillars

are particularly influential, while governance scores unexpectedly correlate with higher WACC. Löffler [32] similarly reports that higher ESG performance is linked to lower equity and debt costs (but not to lower market risk/beta) among S&P 500 firms, supporting the economic case for ESG integration via reduced financing costs.

Evidence from specific regions and industries highlights contextual complexity. Jama and Horstad [33] show that in the Nordic region, ESG scores do not predict profitability (ROE) but do enhance firm value and reduce risk, particularly through the social and combined ESGC scores. Meanwhile, in the European tourism sector, Demiraj et al. [34] find a negative association between ESG scores and profitability (ROA), raising questions about short-term costs, inefficient integration, and the possibility of greenwashing.

Finally, studies from emerging markets such as India paint a more skeptical picture. Shobhwani and Lodha [12], analyzing NSE-100 companies, find no significant association between ESG risk scores (either overall or by component) and firm performance across operational, financial, or market-based measures. Their findings suggest that under voluntary or weakly regulated ESG disclosure regimes, ESG risk may not serve as a meaningful differentiator for firm outcomes, and align with prior critiques [35] about the limited materiality of ESG risk absent robust stakeholder engagement and regulatory support.

3.3. Methodological Transparency and Bias

The recent literature has increasingly scrutinized the methodological foundations and transparency of ESG (Environmental, Social, Governance) ratings, exposing both advances in measurement and persistent sources of bias and inconsistency. This focus is critical, as methodological opacity and subjectivity undermine the credibility and utility of ESG scores for investors, researchers, and regulators.

Gamlath et al. [36] propose a machine learning-driven framework for ESG rating generation that integrates both financial and textual data to reduce subjective human bias and inconsistency in traditional ESG scoring. Their multi-model architecture enables automated ESG estimation, forecasting, and directional classification, demonstrating the potential of scalable, data-driven systems to provide timely, robust sustainability benchmarks aligned with modern investment analytics. Similar innovations are seen in Tseng et al. [37], who introduce the DynamicESG dataset—enabling real-time, news-driven ESG scoring via annotated media narratives and machine learning. This dynamic approach addresses a key limitation of static, disclosure-based ratings: the inability to capture rapid changes in corporate behavior or public perception. Both studies reflect a growing trend towards leveraging AI and big data for more responsive, granular ESG analytics.

In contrast to these technological advances, Sahin et al. [14] critically assess the transparency of Refinitiv's ESG scoring, revealing that scores can be retroactively altered without disclosure, compromising data stability and research replicability. Their findings show that minor changes in sub-score weights or data points can significantly alter firm rankings, distorting both academic studies and financial decision-making. This lack of transparency, echoed in Rönnerberg [13], who identifies scope, measurement, and weight divergences as root causes of cross-agency disagreement, highlights a persistent challenge: the absence of standardization and the influence of agency origin and social context on rating outcomes. Calls for transparent, replicable methodologies and clear reporting of data extraction and scoring practices are echoed across the literature as essential for research credibility and comparability.

Acknowledging the limitations of current ESG scoring frameworks for smaller firms, Murè et al. [38] present a tailored self-assessment tool for European SMEs, aligned with evolving EU regulations and sustainability goals. Their model emphasizes transparency, sector specificity, and actionability, addressing the regulatory and practical gaps faced

by SMEs in accessing sustainable finance and integrating into responsible supply chains. The approach stands in contrast to the often opaque, one-size-fits-all nature of legacy ESG ratings.

Recent methodological work also focuses on addressing subjective bias in ESG weighting. Yu et al. [23] introduce a group decision-making adaptation of the TOPSIS model that mathematically integrates heterogeneous preference structures from multiple decision-makers, reducing reliance on arbitrary weight assignments and thus enhancing both objectivity and flexibility in ESG scoring. This approach directly addresses one of the major contributors to cross-provider rating disagreement and reflects a broader movement toward more participatory and adaptive ESG evaluation frameworks.

Stocco [15] uses machine learning to demystify proprietary ESG scoring, uncovering that a significant share of ESG ratings (up to 60%) are based on aspirational commitments rather than historical performance, raising concerns about “greenwashing” and the reliability of ratings as indicators of actual sustainability outcomes. His research further demonstrates that ESG mandates in investment strategies can have mixed financial impacts, and that social media reputation moderates market responses to ESG controversies. Giese, Nagy, and Lee [39] add that governance indicators tend to have more immediate impacts on risk, while environmental and social dimensions affect performance over longer horizons—implying that time sensitivity and sectoral relevance must be factored into ESG methodologies for them to be meaningful.

Finally, Ang et al. [40] highlight the importance of dynamic, network-based variables—such as shared directors or common investors—in enhancing the predictive accuracy of ESG ratings. Their work suggests that static, firm-level approaches are insufficiently sensitive to the real-time evolution of stakeholder relationships and information flows, and that the inclusion of dynamic inter-firm and event-based data can significantly improve the interpretability and timeliness of ESG assessments.

3.4. Rating Providers and Standardization Efforts

The proliferation and influence of ESG rating providers have both propelled and complicated the evolution of global sustainability reporting. A persistent theme in the recent literature is the tension between regulatory ambition and the practical realities of achieving cross-border standardization in ESG disclosure and assessment.

Frecautan and Nita [2] identify the European Union’s regulatory framework—anchored in the CSRD, EU Taxonomy, and NFRD—as the most comprehensive but also the most complex globally. By institutionalizing the principle of double materiality (addressing both financial and broader societal/environmental impacts), the EU approach goes beyond investor-centric standards like SASB or TCFD. This breadth promises global influence but also brings challenges: high compliance costs, the risk of “taxonomy shopping” by multinational firms, and the ongoing danger of regulatory fragmentation. The EU framework may become a global benchmark due to its legal enforceability and ambition, but its complexity raises real concerns for consistent, practical adoption across markets.

The literature reveals substantial heterogeneity among ESG frameworks, both regionally and by sector. Liang et al. [3] highlight the GRI’s universal and sector-specific standards, SASB’s investor-oriented focus, the climate-centric TCFD, and the CDP’s grading on environmental disclosure, each serving distinct user groups. Sjöberg and Östling [41] show how even within Europe, alignment between frameworks can be asymmetric: the BREEAM-SE certification in construction meets and often exceeds the EU Taxonomy’s requirements, yet the reverse is not true. This illustrates how frameworks interact in practice, with sector-specific certifications providing greater operational detail but also adding complexity to the regulatory mosaic.

Despite strong regulatory momentum, the absence of a unified global ESG disclosure standard continues to hinder the reliability, comparability, and decision-usefulness of ESG data [9,42]. Key sources of friction include divergent definitions of materiality, inconsistent metrics, and fragmented institutional initiatives. Notably, Zaid [9] calls for an inclusive, multi-stakeholder process to bridge the gap between financial and societal materiality, transparency, and technical harmonization. Chen [42] further distinguishes between “risk” and “performance” ratings, emphasizing that a lack of clarity on what ESG scores are intended to measure undermines their value. Both studies highlight the need for principle-based, adaptable standards supported by regulated, independent agencies.

Schütze and Sandbaek [16] underscore the EU Taxonomy’s significance as a “common language” for sustainability, especially within the financial sector. Its objectives—enhancing transparency, fighting greenwashing, and channeling capital toward sustainable activities—are ambitious. However, implementation remains challenging: initial compliance costs are high, regulatory uncertainty persists, and data gaps, especially for SMEs and non-EU firms, limit its full effectiveness. Despite these obstacles, the Taxonomy provides strategic opportunities for the EU to set global norms in sustainable finance—if ongoing refinements, clearer guidance, and more granular sectoral approaches can address its limitations.

Most traditional ESG frameworks are designed with large corporations in mind, often marginalizing SMEs. Cakir et al. [18] respond to this gap with *esg2go*, a new rating system explicitly calibrated for SMEs. By minimizing size bias, applying sector-specific benchmarks, and blending “handprint” (positive impact) and “footprint” (risk) measures, *esg2go* demonstrates that scalable, pragmatic alternatives can deliver more equitable and robust assessments for smaller firms. This approach, while still aligned with larger frameworks, addresses practical realities often ignored by mainstream ESG raters and regulators.

Ye [17] synthesizes regulatory developments across major economies, noting the EU’s move to mandatory, high-quality ESG disclosure, the US’s gradual shift toward more prescriptive standards, and China’s evolving approach. The main obstacles remain the lack of unified standards, the variability of enforcement, and the difficulty regulators face in managing compliance and data quality. Ye advocates for a combination of unified standards, improved data quality, and a mix of regulatory incentives and penalties to foster more authentic, comparable, and decision-useful ESG reporting.

3.5. ESG Ratings in Emerging Markets

The literature on ESG ratings in emerging markets reveals a landscape marked by methodological challenges, pillar-specific effects, and the pressing need for context-sensitive approaches. Across diverse geographies, recent studies converge on a central finding: while ESG integration can enhance firm outcomes, the reliability and comparability of ratings remain highly contingent on local disclosure standards, regulatory maturity, and the granularity of the assessment tools employed.

A consistent theme in empirical research is the asymmetry of ESG pillar impacts. Using proprietary CRISIL ratings for Indian firms during the COVID-19 crisis, Narula et al. [20] demonstrate that governance scores (G) are robustly associated with improved operational and market-based performance (ROA, ROCE, Tobin’s Q), while environmental (E) and social (S) pillars exert weaker or insignificant effects. This pillar-level heterogeneity is attributed to the relative maturity of governance frameworks versus the nascent development of environmental and social practices within the Indian context, a finding mirrored in the divergence between financial and non-financial sectors. The study employs hierarchical OLS regressions and t-tests, reinforcing the value of nuanced, pillar-specific analysis and underlining the limitations of aggregated ESG scores in emerging markets.

Complementing this, Yadav et al. [43] provide further evidence from the Indian market, using panel data and GMM regression to show that ESG scores—particularly when viewed holistically—can positively and significantly influence stock returns, even during systemic shocks such as the COVID-19 pandemic. Importantly, their results suggest that ESG factors may outperform traditional financial indicators as predictors of market resilience, especially under conditions of heightened uncertainty. The use of robust econometric methods, including controls for macro-financial variables, strengthens the credibility of their findings. This strand of research highlights not only the potential for ESG to act as a signal of risk management and firm quality, but also the context-dependent nature of its financial relevance.

Beyond India, scholars have begun to adapt ESG evaluation frameworks to account for regional and sectoral variation. Anikin et al. [44] address the absence of a unified Russian ESG standard by developing a multi-layered, locally adapted rating system. Their model combines qualitative and quantitative indicators, employs industry-adjusted normalization, and incorporates Data Envelopment Analysis (DEA) for benchmarking efficiency. This methodological innovation reflects a growing consensus: the direct transplantation of global ESG models to emerging markets risks misrepresentation, unless recalibrated for local disclosure norms and stakeholder priorities.

A particularly critical perspective is offered by Mobius and Ali [19], who argue that reliance on generic, “off-the-shelf” ESG ratings is especially problematic in emerging economies. Through high-profile case studies—Sunny Optical (China) and Yes Bank (India)—they demonstrate how standard ratings can fail to capture both latent risks and strengths, leading to mispriced investment opportunities. Their work stresses the importance of deep, bottom-up research, active engagement with management, and culturally attuned due diligence. Notably, their critique aligns with calls for an “owner’s mindset” among investors, shifting ESG assessment from a box-ticking exercise to a process of ongoing, trust-based engagement.

Taken together, these studies highlight several areas of methodological agreement. First, they affirm the necessity of pillar-level analysis and the dangers of over-reliance on composite scores. Second, they identify regulatory fragmentation and disclosure quality as persistent obstacles to meaningful ESG integration. Third, they underscore the critical importance of context-specific models and active ownership, challenging the sufficiency of externally imposed standards.

At the same time, notable gaps remain. Few studies systematically compare the predictive validity of alternative rating frameworks across emerging markets, and there is limited longitudinal research capturing the evolution of ESG relevance as regulatory reforms and disclosure practices mature. There is also a growing recognition that qualitative engagement and forensic analysis—while resource-intensive—may be especially indispensable in environments where formal ratings lag behind real corporate practice.

4. Discussion and Implications

The present literature review highlights the evolution of the ESG ratings field, revealing a complex and often fragmented landscape characterized by both emerging consensus and persistent disagreement. A major point of convergence among recent scholarship is the widespread recognition that ESG rating divergence—driven by methodological, contextual, and sectoral factors—remains one of the most significant barriers to the integration of sustainability metrics into mainstream financial decision-making [1,4,22]. Yet, even as scholars broadly agree on the urgency of standardization and transparency, the literature documents notable differences in perspectives regarding the sources of divergence, its consequences, and possible solutions.

Across the literature, there is robust agreement that ESG rating divergence is fundamentally rooted in differences in measurement practices, data sources, and indicator weightings [1,13,14]. Studies consistently demonstrate that measurement differences account for the largest share of disagreement, with scope and weighting playing smaller but non-trivial roles. This has direct implications for the comparability and reliability of ESG assessments across providers, and undermines their utility for investors and regulators [4,25].

Another area of consensus is the moderating role of disclosure quality and transparency in reducing rating disagreement. Both Benedetti [27] and Wang, Chen, and Jiang [28] empirically confirm that more granular and transparent ESG reporting tends to narrow divergence among raters, although they caution that disclosure alone cannot fully resolve methodological fragmentation.

Particularly in emerging markets, studies underscore the importance of adapting ESG frameworks to local institutional contexts and sectoral characteristics [19,20,44]. Composite or “off-the-shelf” ratings are widely criticized for their inability to capture local nuances, reinforcing calls for more granular, bottom-up evaluation methods and active ownership strategies.

The call for harmonization and global standards is echoed across almost all the recent literature [2,9,42]. There is broad support for the development of unified frameworks that can bridge the gap between financial materiality and broader societal impacts, as exemplified by the EU’s adoption of double materiality and the evolving architecture of global reporting standards [3,17].

Despite consensus on methodological issues, the literature remains divided on the link between ESG scores and financial performance. Some studies report positive associations, particularly when ESG scores are balanced across pillars [11], or in specific contexts such as the cost of capital in weak institutional environments [31,32]. Others find negative or neutral effects, especially in sectors or regions where ESG integration remains nascent or is not well understood [12,34]. This divergence is amplified by sectoral, geographic, and regulatory heterogeneity, as well as differences in the operationalization of ESG metrics.

The literature reveals a fundamental tension between two conceptualizations of ESG ratings: as instruments for financial risk assessment versus as indicators of genuine sustainability impact. Studies examining financial outcomes predominantly treat ESG ratings as signals of risk exposure or management quality, finding variable associations with cost of capital, stock returns, and firm value depending on institutional context and rating methodology. Conversely, research focused on real-world impact questions whether current rating frameworks adequately capture firms’ environmental and social contributions, with evidence suggesting that ratings may reflect aspirational commitments rather than measurable outcomes. This distinction is critical for both theory and practice, as it shapes how stakeholders interpret and act upon ESG information.

The literature is also split on whether combining or aggregating ESG scores from multiple providers meaningfully enhances their reliability [5]. While some aggregation methods increase alignment, especially in markets with lower baseline disagreement, persistent divergence among providers—particularly regarding the environmental and social pillars—suggests that methodological convergence is still out of reach for many regions and sectors.

Recent advances in machine learning and real-time analytics have been heralded as potential game-changers for ESG evaluation [36,37,40]. However, their widespread adoption remains limited by issues of data availability, transparency, and the continued reliance of many rating providers on subjective or static evaluation models [14,15]. This technological optimism is thus tempered by caution regarding the persistence of bias and opacity in rating methodologies.

The reviewed literature demonstrates an increasing methodological sophistication in addressing ESG rating challenges. The deployment of advanced econometric techniques (e.g., system GMM, propensity score matching), the rise of dynamic, news-based, and network-oriented data models, and the movement toward participatory, group decision-making approaches all reflect a maturing research agenda [23,39,40]. However, concerns about data stability, replicability, and the dominance of large-firm bias—particularly in traditional frameworks—remain salient, especially for SMEs [18,38].

From the authors' interpretative perspective, the fragmentation observed in ESG rating methodologies reflects deeper epistemological tensions within the field. The attempt to quantify complex, context-dependent phenomena such as corporate sustainability inevitably involves value judgments about what matters, how it should be measured, and whose interests are served. This review recognizes that ESG ratings are not neutral technical instruments but rather embody particular worldviews about corporate responsibility and stakeholder accountability. Acknowledging this interpretative dimension is essential for advancing both scholarly understanding and practical applications of ESG assessment.

A notable methodological gap persists regarding the longitudinal and comparative validity of competing ESG frameworks, especially in emerging markets. Few studies provide direct, side-by-side assessments of rating systems over time or systematically investigate how regulatory reforms and increased disclosure alter the predictive value of ESG scores.

5. Conclusions

This review demonstrates that the landscape of ESG ratings remains highly fragmented and contested, despite rapid advances in both methodological sophistication and regulatory ambition. The persistent divergence among ESG ratings—stemming primarily from inconsistencies in measurement, data sources, and weighting—continues to undermine the comparability, credibility, and practical utility of ESG scores for investors, regulators, and researchers. While disclosure quality and transparency have been shown to partially mitigate rater disagreement, they are not panaceas, as structural and methodological fragmentation persists even among firms with high-quality reporting.

The relationship between ESG scores and financial performance is particularly complex, with evidence varying widely by context, sector, and the construction of ESG metrics.

While balanced and transparent ESG measurement is associated with stronger firm outcomes in some settings, the lack of consistent positive effects across studies suggests that a “one-size-fits-all” approach remains elusive. The literature also makes clear that standardization efforts, such as those led by the EU, offer promise but face significant barriers related to complexity, compliance costs, and the challenge of cross-market adoption—especially for SMEs and in emerging economies.

Overall, the field is moving toward greater recognition of the need for context-specific models, active ownership approaches, and methodological innovation, including dynamic and participatory frameworks. However, the enduring gaps in global comparability, the limited longitudinal assessment of rating validity, and the frequent disconnect between rating frameworks and real-world sustainability outcomes signal that significant work remains.

Future research on ESG ratings should move beyond the identification of divergence and focus on systematically comparing the predictive validity, consistency, and real-world impact of different ESG frameworks across geographies and over time. There is a particular need for longitudinal and comparative studies that capture how evolving regulatory environments and disclosure standards influence the usefulness of ESG ratings in both developed and emerging markets. Scholars should also prioritize disaggregated, pillar-specific analyses—especially in under-researched contexts such as SMEs and sector-specific

settings—to ensure that rating models are truly responsive to local realities and avoid the pitfalls of “one-size-fits-all” approaches. Methodological innovation remains crucial, with machine learning, big data, and dynamic network analysis offering promising tools to enhance the granularity, timeliness, and replicability of ESG scores, provided that transparency and interpretability are maintained. In addition, research should address the unique challenges faced by SMEs in ESG assessment, developing inclusive frameworks that account for firm size, supply chain participation, and stakeholder engagement. Finally, as global efforts toward regulatory harmonization accelerate, future work should investigate not only the mechanisms for achieving international standardization, but also the value of qualitative, engagement-driven, and “owner’s mindset” approaches to ESG evaluation, especially where data limitations and contextual complexity persist. By pursuing these avenues, the field can move toward a more robust, credible, and decision-useful ESG ratings ecosystem.

Author Contributions: Conceptualization, methodology, formal analysis, investigation, data curation, and writing—original draft preparation, H.V.Z.; validation, supervision, and writing—review and editing, D.T.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created in this study. The data supporting the findings are the published articles identified through the Lens.org database search described in Section 2, all of which are publicly accessible.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Berg, F.; Kölbel, J.F.; Rigobon, R. Aggregate confusion: The divergence of ESG ratings. *Rev. Financ.* **2022**, *26*, 1315–1344. [\[CrossRef\]](#)
2. Frecautan, I.; NITA, A. Who is going to win: The EU ESG regulation or the rest of the world?—A critical review. *Ann. Univ. Oradea Econ. Sci. Ser.* **2022**, *31*, 109–120. [\[CrossRef\]](#)
3. Liang, L.; Ding, T.; Tan, R.; Song, M. International ESG Standards and Frameworks. In *The ESG Systems in Chinese Enterprises: Theory and Practice*; Springer Nature: Singapore, 2025; pp. 25–61.
4. Billio, M.; Costola, M.; Hristova, I.; Latino, C.; Pelizzon, L. Inside the ESG ratings:(Dis) agreement and performance. *Corp. Soc. Responsib. Environ. Manag.* **2021**, *28*, 1426–1445. [\[CrossRef\]](#)
5. Bissoondoyal-Bheenick, E.; Bennett, S.; Lehnher, R.; Zhong, A. ESG rating disagreement: Implications and aggregation approaches. *Int. Rev. Econ. Financ.* **2024**, *96*, 103532. [\[CrossRef\]](#)
6. Liao, Y.; Wu, Y. An analysis of the influence of ESG rating divergence on investor sentiment. *Int. J. Glob. Econ. Manag.* **2024**, *3*, 162–177. [\[CrossRef\]](#)
7. Liu, X.; Dai, J.; Dong, X.; Liu, J. ESG rating disagreement and analyst forecast quality. *Int. Rev. Financ. Anal.* **2024**, *95*, 103446. [\[CrossRef\]](#)
8. Tukiainen, S. ESG Ratings and Their Divergence from an Investor Perspective. Master’s Thesis, Aalto University School of Business, Helsinki, Finland, 2021.
9. Zaid, M.A.; Issa, A. A roadmap for triggering the convergence of global ESG disclosure standards: Lessons from the IFRS foundation and stakeholder engagement. *Corp. Gov. Int. J. Bus. Soc.* **2023**, *23*, 1648–1669. [\[CrossRef\]](#)
10. Halid, S.; Rahman, R.A.; Mahmud, R.; Mansor, N.; Wahab, R.A. A literature review on ESG score and its impact on firm performance. *Int. J. Acad. Res. Account. Financ. Manag. Sci.* **2023**, *13*, 272–282. [\[CrossRef\]](#)
11. Lee, M.T.; Raschke, R.L.; Krishen, A.S. Understanding ESG scores and firm performance: Are high-performing firms E, S, and G-balanced? *Technol. Forecast. Soc. Change* **2023**, *195*, 122779. [\[CrossRef\]](#)
12. Shobhwani, K.; Lodha, S. Impact of ESG risk scores on firm performance: An empirical analysis of NSE-100 companies. *Asia-Pac. J. Manag. Res. Innov.* **2023**, *19*, 7–18. [\[CrossRef\]](#)
13. Rönnerberg, V. An Analysis of the Precision and Validity of Environmental, Social, and Governance (ESG) Rating Systems. Master’s Thesis, Lahti University of Technology, Lappeenranta, Finland, 2024.
14. Sahin, Ö.; Bax, K.; Paterlini, S.; Czado, C. The pitfalls of (non-definitive) Environmental, Social, and Governance scoring methodology. *Glob. Financ. J.* **2023**, *56*, 100780. [\[CrossRef\]](#)

15. Stocco, D. Three Essays on ESG Ratings and Quantitative Finance. Doctoral Dissertation, Politecnico Di Milano, Milano, Italy, 2024.
16. Schütze, F.; Sandbaek, B. Taxonomy Disclosure in the EU—A Useful Framework, Despite Current Challenges. *Econ. Voice* **2025**, *22*, 161–172. [\[CrossRef\]](#)
17. Ye, Z. Comparative Analysis of Global ESG Information Disclosure Regulatory Frameworks and Challenges. In *2024 10th International Conference on Humanities and Social Science Research (ICHSSR 2024)*, Xiamen, China, 26–28 April 2024; Atlantis Press: Dordrecht, The Netherlands, 2024; pp. 942–950.
18. Cakir, I.; Aerni, P.; Bergman, M.M.; Cakir, B. esg2go: A method to reduce bias, improve coherence, and increase practicality of ESG rating and reporting. *Sustainability* **2023**, *15*, 16872. [\[CrossRef\]](#)
19. Mobius, M.; Ali, U. ESG in emerging markets: The value of fundamental research and constructive engagement in looking beyond ESG ratings. *J. Appl. Corp. Financ.* **2021**, *33*, 112–120. [\[CrossRef\]](#)
20. Narula, R.; Rao, P.; Kumar, S.; Matta, R. ESG scores and firm performance-evidence from emerging market. *Int. Rev. Econ. Financ.* **2024**, *89*, 1170–1184. [\[CrossRef\]](#)
21. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Serafeim, G.; Yoon, A. Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Rev. Account. Stud.* **2023**, *28*, 1500–1530.
23. Yu, Y.; Zhang, H.; Cao, C.; Zhao, Q. Addressing ESG Rating Divergence: A Group Decision-Making Approach with Individual Preferences. *SSRN* **2024**. [\[CrossRef\]](#)
24. Mathisen, A.M.; Sem, S. ESG Rating Disagreement: An Investigation of Alpha in the Presence of ESG Rating Disagreement. Master's Thesis, Copenhagen Business School, Frederiksberg, Denmark, 2022.
25. Jacobs, B.I.; Levy, K.N. The challenge of disparities in ESG ratings. *J. Impact ESG Invest.* **2022**, *2*, 107–111. [\[CrossRef\]](#)
26. Gibson Brandon, R.; Krueger, P.; Schmidt, P.S. ESG rating disagreement and stock returns. *Financ. Anal. J.* **2021**, *77*, 104–127. [\[CrossRef\]](#)
27. Benedetti, G. ESG Disclosure Quality And Ratings' Divergence: An Empirical Analysis. Master's Thesis, Università Ca' Foscari, Venezia, Italy, 2024.
28. Wang, J.; Chen, B.; Jiang, C. Does ESG Report Information Content Reduce ESG Rating Divergence? *SSRN* **2025**. [\[CrossRef\]](#)
29. Jørgensen, E.N.J.; Ellingsen, T.H. ESG Disagreement: Determining Factors and Impact on Stock Performance. Master's Thesis, Handelshøyskolen BI, Oslo, Norway, 2021.
30. Bissoondoyal-Bheenick, E.; Brooks, R.; Do, H.X. ESG and firm performance: The role of size and media channels. *Econ. Model.* **2023**, *121*, 106203. [\[CrossRef\]](#)
31. Priem, R.; Gabellone, A. The impact of a firm's ESG score on its cost of capital: Can a high ESG score serve as a substitute for a weaker legal environment. *Sustain. Account. Manag. Policy J.* **2024**, *15*, 676–703. [\[CrossRef\]](#)
32. Löffler, M. The Impact of ESG Scores on a Firm's Cost of Capital. Doctoral Dissertation, ZHAW Zürcher Hochschule für Angewandte Wissenschaften, Zurich, Switzerland, 2023.
33. Jama, K.; Horstad, A. ESG Scores and Firm Performance in the Nordics: An Empirical Study of the Link Between ESG Scores and Profitability, Firm Value and Cost of Capital in the Nordics. Master's Thesis, BI Norwegian Business School, Oslo, Norway, 2022.
34. Demiraj, R.; Dsouza, S.; Demiraj, E. ESG scores relationship with firm performance: Panel data evidence from the European tourism industry. *Press. Procedia* **2023**, *16*, 116–120.
35. Lys, T.; Naughton, J.P.; Wang, C. Signaling through corporate accountability reporting. *J. Account. Econ.* **2015**, *60*, 56–72. [\[CrossRef\]](#)
36. Gamlath, M.; Gunathilaka, C.; Wijesinghe, A.; Ahangama, S.; Perera, I.; Sivaneasharajah, L. An Integrated Approach to ESG Index Construction with Machine Learning. In *2023 Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lankam, 9–11 November, 2023; IEEE: New York, NY, USA; pp. 252–257.
37. Tseng, Y.M.; Chen, C.C.; Huang, H.H.; Chen, H.H. Dynamicsesg: A dataset for dynamically unearthing esg ratings from news articles. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham, UK, 21–25 October 2023; pp. 5412–5416.
38. Murè, P.; Giorgio, S.; Antonelli, V.; Crisafulli, A. “ESG score” vs. “ESG rating”: A conceptual model for the sustainability assessment and self-assessment of European SMEs. *Front. Environ. Econ.* **2024**, *3*, 1452416. [\[CrossRef\]](#)
39. Giese, G.; Nagy, Z.; Lee, L.E. Deconstructing ESG ratings performance: Risk and return for E, S, and G by time horizon, sector, and weighting. *J. Portf. Manag.* **2021**, *47*, 94–111.
40. Ang, G.; Guo, Z.; Lim, E.P. On predicting ESG ratings using dynamic company networks. *ACM Trans. Manag. Inf. Syst.* **2023**, *14*, 1–34. [\[CrossRef\]](#)
41. Sjöberg, K.; Östling, M. Frameworks for Sustainable Construction: Comparing the EU Taxonomy, ESG Ratings and BREEAM-SE. Master's Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2024.

42. Chen, Z. Developing a Global ESG Rating System: Challenges, Strategies, and Future Prospects. *Financ. Econ. Res.* **2025**, *2*, 1–10. [[CrossRef](#)]
43. Yadav, M.; Dhingra, B.; Batra, S.; Saini, M.; Aggarwal, V. ESG scores and stock returns during COVID-19: An empirical analysis of an emerging market. *Int. J. Soc. Econ.* **2025**, *52*, 390–405.
44. Anikin, D.; Finogenova, Y.; Subbota, K.; Ermakova, A. Developing the ESG rating methodology for Russian companies. *J. Law Sustain. Dev.* **2023**, *11*, e0308. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.