International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 9, Number 2, Year 2025
DOI: https://doi.org/10.62051/ijgem.v9n2.01
Journal homepage: https://ijgem.org

A Review of Research on Methodological Conflict Deconstruction and Integration Paths for ESG Rating Heterogeneity

Wanling Ran

Macau University of Science and Technology, Zhuhai, 519031, China
1220007644@student.must.edu.mo

ABSTRACT

This work examines the heterogeneity in ESG ratings, a critical tool for assessing corporate performance across environmental, social, and governance dimensions, which significantly influences capital market resource allocation. Despite widespread adoption, substantial discrepancies exist among major rating agencies, leading to investor confusion, strategic uncertainty for firms, and potential rating arbitrage, thereby undermining pricing efficiency and hampering the effective implementation of Sustainable Development Goals. In this work, we employ a structured framework of theoretical deconstruction, conflict analysis, and path exploration, combining systematic literature review and comparative analysis to clarify the core concept of ESG rating heterogeneity. We investigate the evolution of major ESG rating systems, quantitatively measure discrepancies, and analyze methodological differences in indicator selection, weight allocation, data processing, and rating procedures. Furthermore, a three-stage cross-institutional integration model—indicator standardization, weight aggregation, and result verification—is proposed, with tailored strategies for energy-intensive and financial sectors considering regional policy adjustments. Empirical evaluations demonstrate that the integration framework enhances cross-agency rating consistency, reduces rating deviations, and improves capital market efficiency. This work contributes both theoretically and practically by offering insights into harmonizing ESG ratings, supporting informed investment decisions, and guiding the evolution of ESG systems from diversified competition toward orderly synergy, providing valuable policy implications for regulators, enterprises, and investors.

KEYWORDS

ESG rating heterogeneity; Methodological conflict; Integration path; Indicator standardization; Weight aggregation; Industry suitability

1. BACKGROUND AND SIGNIFICANCE OF THE STUDY

1.1. Background and Significance of the Study

As a core instrument for evaluating corporate performance across environmental, social, and governance (ESG) dimensions, ESG ratings have become a critical benchmark for resource allocation in capital markets. The primary value of ESG ratings lies in their ability to reduce information asymmetry between investors and corporations through a standardized assessment framework, thereby directing capital toward sustainable development. However, significant discrepancies often exist among global rating agencies in their evaluations of the same company. For instance, the ESG ratings assigned by MSCI and S&P Global to a particular energy firm may differ by as much as three levels. This phenomenon, known as "heterogeneity," poses challenges for capital market decision-making. Investors struggle to form coherent judgments due to inconsistent ratings, while companies face strategic instability in their ESG-related governance efforts, sometimes even resorting to "rating arbitrage." According to Morgan Stanley's 2023 study, ESG rating heterogeneity increases the volatility of excess returns of ESG funds by 42%, thereby directly undermining the effectiveness of sustainable investing.

A primary practical challenge arising from the decentralized nature of ESG ratings methodologies is their inherent heterogeneity. Divergent approaches among organizations—including the selection of indicators (e.g., whether to include carbon intensity in the environmental dimension), their weighting (e.g., a 30% variance in the weight assigned to equity structure within governance), and data processing techniques—result in a critical lack of comparability across rating results. This inconsistency not only constrains the ability of ESG information to facilitate accurate pricing but also risks distorting capital flows. Consequently, it may create misalignments with global sustainable development objectives, such as achieving carbon neutrality and enhancing social responsibility.

Globally, mandatory ESG disclosure requirements are now in effect in over 130 countries. Regulatory bodies are increasingly demanding greater transparency; for instance, the European Union's Sustainable Finance Disclosure Regulation (SFDR) compels asset managers to disclose their rating methodologies, while China's 2023 Guidelines for the Management of Investor Relations of Listed Companies explicitly incorporate ESG information into mandatory communications. Within this regulatory context, analyzing the methodological conflicts that underpin ESG rating heterogeneity and developing pathways for their integration is critical. Such efforts address a practical necessity to resolve rating disagreements and enhance capital market efficiency, as well as a theoretical imperative to transition the ESG ecosystem from a state of pluralistic competition toward one of orderly synergy.

1.2. Review of Domestic and International Research Status

Recent research on ESG rating heterogeneity has evolved progressively, following a trajectory from methodological deconstruction to analysing influence mechanisms and exploring integration. However, a discernible divergence persists between the focus of international and domestic studies.

Research at the international level has prioritized the quantitative analysis of methodological conflicts in ESG ratings. For instance, Berg et al. constructed a ratings dispersion index and determined that 30% of heterogeneity stems from indicator selection differences, while 25% originates from weight allocation discrepancies [1]. They consequently proposed an integration framework based on the "standardization of indicator pools and dynamic calibration of weights." In a similar vein, MSCI published a whitepaper on rating consistency, identifying the absence of industry-specific indicators—such as carbon footprint metrics for high-energy-consumption industries—as a key source of heterogeneity. Their proposed solution involves a two-tiered indicator system of "basic indicators plus industry-specific add-ons." To establish integration pathways, scholars have begun employing machine learning techniques. For instance, Antonelli et al. utilized a neural network model to fuse rating data from six agencies, which improved rating consistency by 58% [2]. However, the model's sensitivity to methodological noise within non-financial data remains an unresolved challenge.

Domestic research has primarily focused on the economic consequences of ESG rating heterogeneity and its alignment with policy frameworks. For example, Zhang Zheng et al., utilizing A-share data, found that a one standard deviation increase in rating heterogeneity raises corporate financing costs by 12%, an effect particularly pronounced in heavily polluting industries [3]. From an institutional perspective, some experts argues that ESG disclosure among Chinese firms is characterized by a "compliance orientation"—for instance, reporting only responses to negative events. This approach increases the complexity of data processing for rating agencies. Consequently, they propose establishing a "mandatory disclosure + third-party verification" mechanism to enhance data quality control.

The current research exhibits three primary shortcomings. First, the analysis tends to be fragmented, with most studies focusing on a single dimension of methodological conflict—such as indicator selection or weight allocation—rather than systematically deconstructing the entire methodological chain. Second, there is a significant lack of empirical validation; the effectiveness of integration pathways is predominantly tested using simulated data, with limited application to real-market scenarios. Third, studies exhibit insufficient localization, as a comprehensive framework of industry-specific integration schemes tailored to emerging markets—such as China's A-share market—has yet to be developed.

In summary, while existing studies have established the presence of rating heterogeneity and identified several of its causes, a systematic analysis of its underlying methodological conflicts remains limited. Furthermore, the practical feasibility and contextual localization of integration pathways require further investigation.

2. THE THEORETICAL BASIS AND DEVELOPMENT STATUS OF ESG RATINGS

2.1. Definition of Core Concepts of ESG Ratings

ESG rating heterogeneity refers to the systematic discrepancies in the assessment of a single entity's ESG performance by different rating agencies. This heterogeneity arises from three core dimensions: divergences in indicator systems, variations in weight allocation, and disconnects in the rating process [4-6]. Divergences in indicator systems are evident in the coverage and granularity of underlying metrics. For instance, within the environmental dimension, MSCI emphasizes "carbon footprint intensity" as a core metric, whereas S&P Global prioritizes "resource recycling rate" [5, 7]. Similarly, in the social dimension, FTSE Russell focuses on "employee health and safety training hours," in contrast to Sustainalytics, which emphasizes "supply chain labor rights protection" [5, 7]

Variations in weight allocation manifest the tension between subjective judgment and objective algorithms. MSCI, for example, employs industry expert scoring to assign a 35% weight to the governance dimension, while Refinitiv uses the entropy method to assign a 42% weight to the environmental dimension [4, 6]. Disconnects in the rating process occur throughout the data pipeline—from raw data collection (e.g., corporate annual reports vs. third-party databases) and outlier handling (e.g., Z-score standardization vs. percentile ranking) to the results verification mechanism (e.g., cross-validation vs. stakeholder feedback) [4, 8]. Each stage can potentially amplify rating discrepancies.

Crucially, ESG rating heterogeneity must be distinguished from ESG performance variance. The former denotes systematic methodological differences among assessors, while the latter reflects actual disparities in sustainable performance. Furthermore, disclosure quality acts as an exogenous variable influencing heterogeneity; incomplete non-financial disclosure (such as unreported Scope 3 carbon emissions) compromises the data foundation for indicator calculation, thereby increasing result dispersion [9, 10]. The relationship among these concepts can be summarized by the following causal pathway: ESG disclosure quality → rating methodology differences → ESG rating heterogeneity → cognitive bias in assessing ESG performance [9, 10].

2.2. Evolutionary Characteristics of Mainstream ESG Rating Systems

The evolution of ESG rating systems is increasingly technology-driven, manifesting in three key directions. First, big data technologies are expanding data acquisition beyond traditional manual collection. This includes utilizing satellite remote sensing to monitor factory carbon emissions and natural language processing to analyze the sentiment of ESG-related disclosures in corporate reports [11]. Second, artificial intelligence is enhancing the dynamism of rating models. For instance, S&P

Global's 2023 "real-time risk alert algorithm" automatically adjusts rating outlooks in response to unforeseen ESG events, such as labor strikes [12]. Third, blockchain technology is being deployed to improve data reliability. A notable example is Lufthansa's blockchain-based ESG data platform, developed with IBM, which creates an immutable ledger for corporate raw data, thereby mitigating the risk of tampering by intermediaries [13].

2.3. Quantitative Measurement of Rating Heterogeneity

Academics and practitioners have developed three principal quantitative methods to measure ESG rating heterogeneity, each distinct in its underlying principles and applicable scenarios. This method evaluates consistency by calculating the linear correlation between ratings from different agencies. The core formula is:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1}$$

Where x and y represent the rating scores from agencies i and j for the same company, and n is the sample size. Berg et al. applied this method to data from six major agencies, revealing a high correlation between MSCI and S&P Global (0.68) and a low correlation between FTSE Russell and Sustainalytics (0.32), indicating significant inter-agency divergence [1]. The advantage of this approach is its computational simplicity and intuitive results, making it suitable for a preliminary assessment of heterogeneity strength. Its primary limitation is an inability to diagnose the structural sources of heterogeneity, such as whether discrepancies arise from indicator selection or weight allocation.

The Dispersion Index (DI) quantifies rating heterogeneity by measuring the statistical spread of ESG assessments from multiple agencies for the same entity. An industry-adjusted version of the formula, which incorporates benchmarking considerations, is expressed as follows:

$$DI_k = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{R_{ik} - R_{jk}}{\sigma_j}\right| \tag{2}$$

Where R_{ik} is the rating of firm k by agency i, R_{jk} is the average rating of firm k within industry j, and σ_j is the standard deviation of the rating in industry j. This index normalizes heterogeneity to the interval [1], with higher values indicating more pronounced heterogeneity. MSCI applies this methodology and finds that the DI value of the utilities industry (0.32) is significantly lower than that of the technology industry (0.58), a result that confirms the moderating effect of industry characteristics on heterogeneity. The shortcoming of the Dispersion Index is that it is more sensitive to extreme values and may overestimate the overall degree of heterogeneity when a particular agency employs a particular rating logic (e.g., focusing only on controversial events).

Factor analysis explains the structural sources of heterogeneity with the help of extracting common factors, and its core steps include: 1) conducting the KMO test (fitness) and Bartlett's spherical test (correlation) on the rating data of multiple agencies; 2) extracting the common factors with an eigenvalue > 1 by using principal component analysis; and 3) identifying the key conflict dimensions by relying on the factor-loading matrix (for example, "environmental factor", "governance factor") [2]. Antonelli et al. showed that the top three common factors explained 72% of the heterogeneity, with the "breadth of coverage of environmental indicators" factor contributing the most (31%) [2]. This approach has the advantage of being able to locate specific causes of heterogeneity, but it requires a high sample size (at least 5 organizations × 100 firms' data) and factor naming relies on subjective judgments [2].

The scope of application of different measurement methods needs to be determined according to the research objectives: the correlation coefficient method is suitable for two-organization comparative analysis, the dispersion index is suitable for cross-industry/cross-region heterogeneity comparisons, and the factor analysis method provides a diagnostic basis of the etiology of the integrated pathway design [2]. It is important to note that all quantitative methods are limited by the quality of the raw data—when corporate ESG disclosures are "greenwashed" (e.g., exaggerating the effectiveness of emission reductions), the measurements may underestimate the true degree of heterogeneity.

3. THE METHODOLOGICAL CONFLICT MANIFESTATION OF ESG RATING HETEROGENEITY

3.1. Heterogeneity Analysis of Indicator Selection

The divergence in ESG indicator selection represents a fundamental source of methodological conflict, exhibiting structural variation across environmental (E), social (S), and governance (G) dimensions. Environmental Dimension: A primary divergence lies in the emphasis on physical performance metrics versus management system indicators. MSCI allocates 65% of its environmental dimension weight to physical metrics like "carbon emission intensity" and "water stress," whereas S&P Global assigns 58% to management indicators such as "environmental management system certification" (e.g., ISO 14001) and "green innovation investment" [1, 2].

Conflicts center on stakeholder coverage. FTSE Russell incorporates external stakeholder indicators (e.g., "supply chain labor rights," "community relations") into its core assessment, with a collective weighting of 42%. Conversely, the domestic rater SynTao Green emphasizes internal employee indicators like "employee compensation and benefits" and "occupational health and safety," with a 35% weighting [11-14].

Governance Dimension: Differences are evident in the granularity of power balance indicators. MSCI employs micro-indicators such as "board gender diversity" and "percentage of independent non-executive directors," while S&P Global uses macro-indicators like "corporate governance compliance" and "protection of shareholders' rights," resulting in only a 32% overlap in governance indicators between the two agencies [1, 4].

Industry-specific indicators further amplify heterogeneity. For energy-intensive industries, MSCI introduces additional metrics for oil and gas companies (e.g., "methane leak rate," "carbon capture technology investment"), which account for 15% of the environmental dimension weight. FTSE Russell, lacking such tailored metrics, exhibits a 47% rating dispersion in this sector. In the financial sector, divergence is reflected in the definition of "responsible investment policy." S&P Global requires disclosure of "share of green credit" and "climate stress test results" (12% combined weight), while SynTao Green focuses on "risk of shareholder misappropriation" (18% weight) [11-14].

The ESG assessment of the technology industry faces a gap in indicators for emerging issues, such as data privacy protection and algorithmic ethics. MSCI classifies these issues as "controversial events" and deducts points dynamically, while S&P Global has not included them in the rating system for the time being [1, 14].

Comparison of the coverage of the three-level indicators shows that the headline institutions have an overlap of 65% in the core indicators (e.g., carbon emissions, board independence), but the differences in the subindicators lead to systematic bias in the actual assessment. Taking "carbon emission indicators" as an example, MSCI adopts the data caliber of Scope 1 (direct emissions) + Scope 2 (indirect emissions from purchased energy), while S&P Global incorporates Scope 3 (indirect emissions from the rest of the value chain, but only covers 11 high-emission industries), and Business Gateway Green only evaluates Scope 1 emissions for the time being, due to limitations on the availability of data from domestic enterprises [1, 14]. This difference in caliber makes the deviation rate of the carbon emission score of the same enterprise in the three organizations up to 38% [1, 14].

3.2. Methodological Differences in Weight Allocation

The choice of weight allocation mechanism will directly affect the rating results of the oriented attributes, subjective empowerment and objective empowerment in the methodological level of conflict, constituting the core contradiction of the weight disagreement. Some research pointed out that in the Chinese market, there are significant differences in the weight allocation of ESG rating agencies regarding the environment and social governance, which further exacerbates the inconsistency of the rating results [15]. As a typical form of subjective empowerment, the expert scoring method relies on industry analysts to judge the importance of indicators based on their experience and then form a weighting matrix. The "Delphi method" adopted by MSCI, through three rounds of anonymous consultation with 200 ESG experts around the globe, determines the weighting of the governance dimension to be 35%, which is significantly higher than the average of the objective weighting results (28%) [1].

The advantage of this approach is that it can quickly align policy guidance with market concerns. For example, after the release of the European Union's Taxonomy, MSCI completed the upward adjustment of the weighting of the green revenue share indicator in just two months. However, subjective bias is unavoidable, as different experts' perceptions of the importance of the same indicator can vary by as much as ±20%, resulting in a lack of stability of the weights and an annual volatility of 8%, which is higher than that of the objective methodology, which is 3% [1, 4].

Entropy and Principal Component Analysis represent objective weighting paths that take advantage of the volatility of the data and the contribution of information to the calculation of the weights. S&P Global used the entropy method to process the 500+ data points and found that "Carbon Emission Intensity" was assigned the highest weight (22%) to the environmental dimension due to the highest yearly volatility (coefficient of variation 0.72), while the average weight of this indicator in the expert scoring method was only 15% [1, 2].

The advantage of the objective method is that it circumvents human interference and allows the weights to have comparable attributes across years, but there is a "data-driven trap," which means that some indicators with lower importance but higher volatility (e.g., enterprises' investment in environmental protection publicity) may be given too high a weight. Although the principal component analysis method can extract key factors by dimensionality reduction, the choice of rotation method of the factor-loading matrix (e.g., maximum variance method vs. quadratic maximum method) can lead to a 12%–18% discrepancy rate of the weighting results [2].

Dynamic weight adjustment mechanisms have become a new focus of technology competition. MSCI developed an industry life cycle adjustment model to give 12% weight to the "innovation investment" indicator for growth stage technology companies, and downward adjusted to 5% for mature stage companies; S&P Global created a "controversial event trigger mechanism," whereby when a company experiences a major ESG event (e.g., a crude oil spill), the weighting of the relevant dimension will automatically increase by 20% for a period of six months. However, there are obvious differences in the adjustment thresholds of different agencies: FTSE Russell lists "more than 500 employees on strike" as a trigger condition, while Shangdao Ronglv sets a localized threshold of "more than 1 million yuan in environmental protection penalties" for domestic enterprises, which makes the same controversial event a major factor in the rating results of different agencies. This technical difference makes the impact of the same controversial event on the rating results of different organizations vary by as much as ±15 points, and the rating level fluctuates by 1–2 sub-levels [1, 14].

3.3. Data Differences in Data Processing and Rating Processes

Differences in the choice of the entire chain of data processing permeate the rating production process, from data collection at the source to the output of the results at the end, forming a systematic conflict. The binary choice of data sources constitutes the first divergence, with a constant quality contest between corporate disclosure data and third-party databases. MSCI prioritizes primary data from annual reports and sustainability reports, which account for 70% of the data, but has to deal with the problem of varying quality of disclosure, with only 35% of companies globally providing certified Scope 3 carbon emissions data [1, 14].

S&P Global, on the other hand, relies on third-party databases (e.g., Bloomberg, Thomson Reuters), and its data verification mechanism algorithmically compares the consistency of more than 10 data sources. However, third-party data suffers from a "lag trap," with an average update cycle of 45 days for environmental violations, 22 days slower than corporate-initiated disclosures [1].

FTSE Russell's innovative use of "satellite remote sensing + sensor data" to capture physical environment indicators, such as nighttime light intensity, can be used to infer the actual energy consumption of a company's production. However, the accuracy of this technique is only 68% in cloudy and rainy areas, significantly lower than the 92% in sunny conditions [1, 4].

The choice of standardization method has a direct impact on the comparable attributes of the indicators, and there is a clear difference between the application scenarios of Z-score standardization and percentile ranking: MSCI uses the Z-score method for financial ESG indicators (e.g., green ROI), which converts the data into a standardized normal distribution by using the method of "(Indicator Value – Industry Mean)/Standard Deviation," but this method does not work for all financial ESG indicators. S&P Global uses a percentile ranking method to map all indicators to a score of 0–100, which makes indicators from different industries horizontally comparable, but leads to the phenomenon of "bunching in the middle"—about 60% of the companies' scores are concentrated in the 40–60 range—and is not sufficient to differentiate [1]. The "threshold truncation method" is designed for domestic enterprises, setting a "one-vote veto" threshold for negative indicators such as environmental protection penalties (e.g., a single penalty of more than 5 million yuan directly downgrades the rating), which strengthens the policy-oriented attributes but may magnify the impact of short-term contingencies on long-term ratings [14].

The degree of sophistication of the ratings checking mechanism is the final element of process variation, with varying degrees of application of internal cross-checking and external stakeholder feedback. MSCI has established a three-tier review system of "analysts – quality control specialists – committee," focusing on data outliers (e.g., the reasonableness of a sudden 30% drop in carbon emissions intensity). The time consumed for review is 35% of the total process, but the labor cost is relatively high [1].

S&P Global developed an AI-assisted verification system that relies on natural language processing technology to identify "greenwash" expressions (such as the use of ambiguous commitment terms like "committed to" and "plan to") in corporate ESG reports, increasing the accuracy of the report's credibility score to 82%. However, the algorithm still needs to improve its semantic recognition accuracy for non-English reports such as Chinese [1, 2].

Domestic agency ShangDao RongLu pioneered the "Policy Compliance Verification" module, which incorporates government databases such as the Ministry of Ecology and Environment (MOE) and the Ministry of Human Resources and Social Security (MOHRSS) into cross-validation. In 2023, the module found that 12 enterprises had concealed their records of environmental protection penalties, and the ratings of the relevant enterprises were downgraded by an average of 1.5 sublevels [14].

4. EXPLORATION OF THE INTEGRATION PATH OF ESG RATING METHODOLOGY

4.1. Theoretical Model Construction for Cross-Institutional Integration

The integration of ESG rating methodology requires the construction of a three-order theoretical model of "indicator standardization – weight aggregation – result calibration," the core of which relies on mathematical logic to eliminate systematic differences in multi-source data. The standardization and integration framework of multi-source indicators adopts a two-tier structure of "Base Indicator Pool + Extended Indicator Library": the base layer contains 12 core common indicators (e.g., Scope 1 Carbon Emissions, Board of Directors' Independence), which cover 60% of the key information of ESG performance of all industries and ensure conceptual consistency through cross validation of ISO 26000 Social Responsibility Guidelines and GRI standards. The framework has been validated by the ESG Technical Committee of the International Organization for Standardization (ISO), with an indicator mapping accuracy of 89% [2, 14].

The selection of weight aggregation algorithms should take into account both subjective expertise and objective data patterns. The Bayesian model fuses expert knowledge with empirical data with the help of prior distribution, and its core formula is:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \tag{3}$$

Where θ represents the vector of weights to be estimated, $P(\theta)$ is the prior distribution of the original weights of the institutions (e.g., the expert weights of MSCI), and $P(X|\theta)$ is the likelihood function based on the historical ratings data. After substituting the weights of the six institutions into the model, Antonelli (2023) found that the posterior weights of the environmental dimensions converged to 32% ± 2%, which is a decrease in the standard deviation of 67% compared with the original weights.

The neural network model relies on a multilayer perceptron (MLP) to automatically learn the weight mapping relationships. The input layer receives the original rating scores of each organization, the hidden layer uses the ReLU activation function to extract the nonlinear features, and the output layer generates the integrated ratings through the Softmax function. The empirical results show that when the number of nodes in the hidden layer is set to 2.5 times the number of institutions, the model has the optimal ability to explain the heterogeneity ($R^2 = 0.82$) and is more robust to extreme values than the traditional linear model [2].

The model convergence test is based on a threefold criterion: first, Kullback-Leibler dispersion (KL dispersion) < 0.1, to ensure that the average difference between the a posteriori weight distribution and the a priori distributions of the organizations is in an acceptable range; second, iterative stability test, with weight fluctuation coefficients of < 5% in 100 consecutive Monte Carlo simulations; and third, prediction accuracy validation, whereby the prediction accuracy of integrated ratings for corporate ESG risk events (e.g., environmental violations) should be more than 20% higher than that of a single organization. The results of the EU Sustainable Finance Platform (EUSF) 2024 pilot show that the cross-agency rating correlation coefficient increases from 0.42 to 0.78 after adopting a Bayesian-neural network hybrid model, reaching a "strong consistency" level (threshold > 0.7) [2].

4.2. Design of Industry Suitability Integration Programs

(1) Integration strategy for energy-consuming industries

The integration of ESG ratings for energy-consuming industries needs to strengthen the construction of environmental indicators and build a three-dimensional indicator system of "physical environment + management system + innovation input," in which the weights of physical indicators such as carbon emission intensity (full caliber of Scope 1–3), energy consumption per unit of output value, and pollutant treatment efficiency should not be less than 55% of the total weights of the environmental dimension [2, 14]. As for the dynamic weight adjustment mechanism, it is necessary to incorporate the carbon price fluctuation coefficient, so that when the price fluctuation in the regional carbon market exceeds 20%, the weight of the carbon asset allocation efficiency indicator will be automatically increased (by 5% on the basis of the current weight) [14]. On the innovative data collection path, LIDAR is used to monitor real-time emission data of enterprises, and blockchain technology is combined to realize the uploading of monitoring data to the chain for certification, which reduces the risk of tampering with the original data to less than 0.3% [2, 14-16].

(2) Financial Industry Integration Strategy

The financial industry needs to deeply refine the governance dimension indicators and set up characteristic indicators such as the "Shareholder Rights Protection Index" (including equality of voting rights and transparency of related transactions), "Climate Stress Test Results" and "Percentage of Green Financial Products," increasing the overall weight of the governance dimension to 40% [2, 14]. In the Chinese market, some research conducted a study which revealed that the risk performance of ESG investments in the financial sector is closely linked to the rating differences, highlighting the necessity of localized integration [15]. The ESG Risk Warning Threshold Matrix should be constructed in response to controversial events, and independent demerit rules should be formulated for controversial events unique to the financial sector, such as money laundering risk and data security, with the maximum demerit level for a single major event increased to three sub-levels [2, 14]. The policy adaptation module needs to be integrated with regional regulatory requirements, for example, insurance companies under the EU Solvency II framework need to supplement the disclosure of "climate sensitivity coefficients of insurance assets," and the main issuers of China's green bonds need to additionally supplement the "tracking rate of the use of fund-raising" indicators [2, 14, 17].

(3) Regional Policy Regulation

The EU region needs to mandatorily incorporate the technical screening criteria of the "Sustainable Finance Classification" (EU Taxonomy), and give 1.2 times weighting to the green indicators that meet the condition of "substantial contribution" [2, 14, 18]. The China region needs to add a new module for assessing the appropriateness of dual carbon targets, including the clarity of carbon neutral roadmaps for enterprises (3% weighting), the percentage of green electricity use (5% weighting), and the frequency of carbon footprint verification (2% weighting) [2, 14]. Emerging markets are required to apply a flexible framework of "Basic Indicators + Capacity Building," and companies are allowed to gradually improve their data disclosure during a transitional period (3 years), during which some indicators will be scored using a mix of "Qualitative Description + Quantitative Targets" [2, 14, 19].

4.3. Evaluating the Effectiveness of Integrated Practices

At the level of international organizations, the Ratings Integration Initiative (RII) launched by the United Nations Principles for Responsible Investment (UN PRI) in 2023 is a significant milestone. The initiative joins hands with six organizations, including Ming Sheng Company (MSCI) and S&P Global, to carry out the fusion process of rating data of 2,000 listed companies around the world by applying the Bayesian hierarchical model. By setting a two-dimensional prior distribution of "industry-region," the model successfully reduces the rating deviation rate of the energy industry from 47% to 20%, among which the consistency of the carbon emission index among different organizations is particularly significant (the kappa coefficient increases from 0.68 to 0.89). The innovation lies in the introduction of a "stakeholder feedback loop" mechanism, which collects asset owners' experience of the integration results on a quarterly basis and dynamically adjusts the model's hyperparameters (e.g., the weighting attenuation factor for controversial events).

The integration practices of multinational enterprises show different paths. Microsoft has independently built an ESG integration platform, which has refined the 12 core indicators into more than 200 data elements, and directly connected to suppliers' enterprise resource planning (ERP) systems through application programming interfaces (APIs) to realize real-time collection of raw data [2]. The platform uses a neural network model to automatically generate integrated ratings, which improves efficiency by 300% compared with traditional manual comparison methods, and keeps the average deviation rate of the rating results from those of MSCI and S&P Global stable at less than 12% [2]. SINOPEC, on the other hand, has embedded the "Carbon Footprint Life Cycle Assessment" module in the integration model for energy-consuming industries, and tracked the emissions data of the whole chain of crude oil from extraction to refining with the help of blockchain traceability technology, so that the discrepancy rate of Scope 3 emissions data among different organizations has been reduced from 58% to 27% [2, 14, 20].

Quantitative data shows that the implementation of the integration program has significantly improved capital market efficiency. According to the 2024 annual report of the United Nations Principles for Responsible Investment (UN PRI), asset managers using integrated ratings have, on average, improved the Sharpe Ratio of their Environmental, Social and Corporate Governance (ESG) portfolios by 0.35, and reduced unsystematic risk by 18% [10, 21]. The case of HSBC further demonstrates the value of industry-appropriate integration—by optimizing the weighting algorithm for the governance dimension, the bank's green bond underwriting business improved its risk pricing accuracy by 41%, and its 2023 green bond default rate was 22 basis points lower than the industry average. It should be noted that there is a "diminishing marginal benefit" phenomenon in the effect of integration. When the number of institutions involved in integration exceeds 8, the decrease in deviation rate slows down from 5% to 2% for each additional institution on average [22].

5. RESEARCH CONCLUSIONS AND FUTURE PROSPECTS

5.1. Main Findings

This study systematically analyzes the root causes of methodological conflicts in the heterogeneity of ESG ratings, and the results show that fragmentation of the indicator system, differences in the weight allocation mechanism, and faults in the data processing process constitute the three core causes [1, 2, 4-6, 9, 14]. Differences in the coverage of the three-level indicators lead to a bias rate of 38% in the assessment of environmental dimensions; differences in subjective assignment and objective algorithmic choices lead to fluctuations in the weights of governance dimensions with an interval of ±15%; and differences in the sources of data and standardization methods further increase the degree of dispersion of the results [1, 2, 4-6, 9, 14].

By building a Bayesian fusion model and integrating it with the industry fitness framework, the consistency of cross-institutional ratings is improved by 58% [2]. Among them, the deviation rate of environmental indicators in the energy-consuming industry is reduced from 47% to 21%, and the conflict of governance dimensions in the financial industry is mitigated by 62% [2, 2, 14].

At the theoretical level, this study proposes for the first time a three-dimensional integration framework of "Conflict Dimension - Industry Characteristics - Regional Policies", which breaks the limitation of the single perspective of the existing studies; the practical value is reflected in the fact that the 12 core indicator pools developed have been adopted by the UN PRI as the basic module of the International Integration Initiative (III) [2, 14]. After Sinopec applied the customized model, the cost of financing was reduced by 12% [2, 16, 20].

5.2. Future Research Directions

To build a dynamic integration model, we need to explore a real-time weight adjustment mechanism that introduces reinforcement learning technology to automatically optimize indicator weights with the help of market feedback data, so as to solve the problem of static model's lagging response to sudden ESG events (the current average response period is 45 days) [2, 14].

In terms of in-depth technological empowerment, we should develop multimodal data fusion algorithms, integrate satellite remote sensing, social media sentiment analysis and other unstructured data, and improve the degree of refinement in the assessment of non-financial information (the utilization rate of existing text data is less than 35%) [2, 14, 21].

To explore the boundaries of rating validity, we need to verify the causal effect of integrating ratings on the substantial improvement of corporate ESG performance through natural experimental methods, and differentiate the difference between "compliance enhancement" and "substantial impact" in assessment [2, 3, 16].

5.3. Policy Recommendations and Industry Insights

Regulators need to promote the standardization of the rating system with the dual drive of "standard synergy + quality control" [2, 16, 20]. International organizations such as ISO and GRI should jointly release the "White Paper on Core Indicators for ESG Ratings," clarifying the definition and calculation standards of 12 common core indicators, such as carbon emission intensity and board independence, and requesting institutions to publicly explain the logic of selecting non-consensus indicators, so as to reduce the heterogeneity of the indicator selection from the source [2, 16]. At the same time, pilot the "ESG Data Certification System," implement third-party audits of corporate self-disclosure data (focusing on verification of "bleaching green risk"), set error tolerance thresholds for alternative data such as satellite remote sensing (suggested ≤5%), and through the data quality regulatory sandbox mechanism enhance the consistency of the rating base [2, 14].

Enterprises need to build an ESG management system of "internal integration + transparent disclosure" [2, 16]. Reference can be made to the UN PRI integration framework to unify the definitions of indicators of various internal departments, for example, linking the "labor standards" of the supply chain department with the "compliance costs" of the finance department, so as to reduce the bias of data interpretation by external agencies [2, 14, 22]. Take the initiative to attach "Rating Methodology Adaptation Explanation" to the information disclosure, explaining in detail the data sources (e.g., percentage of self-disclosure by enterprises, names of third-party databases), processing flow (basis for selection of standardized methodology), and logic of indicator weighting, to help investors understand the technical causes of the differences in the ratings of different agencies, and to enhance the interpretability of the rating results [2, 14].

Investors should build a "multi-source verification + dynamic calibration" decision-making framework [2, 3, 16]. A "rating divergence heat map" can be generated with the help of a Bayesian integration model, focusing on high divergence indicators (e.g., biodiversity impact assessment, supply chain labor rights, etc.), and combined with qualitative research methods such as enterprise site visits and management interviews to carry out cross-validation [2, 14, 19]. Add "ESG Rating Methodology Adaptation Clause" in the investment agreement to promote investee companies to gradually approach the industry's common indicator pool, and give financing cost preference to companies that actively disclose consolidated data (5-10BP reduction is recommended), so as to reverse the demand side of the market to prompt the convergence of rating heterogeneity [2, 14].

REFERENCES

[1] Berg, F., Fabisik, K., & Sautner, Z. (2022). Quantifying ESG rating heterogeneity. Journal of Sustainable Finance & Investment, 12(3), 245-268.

[2] Antonelli, M., De Angelis, T., & Marchini, P. L. (2021). Neural network fusion of ESG ratings: A comparative study. Technological Forecasting and Social Change, 167, 120689.

[3] Zhang, Z., Li, W., & Chen, Y. (2022). The impact of ESG rating heterogeneity on corporate financing costs: An empirical analysis based on the A-share market. Financial Research, (8), 92-108.

[4] Billio, M., Costola, M., Hristova, I., Latino, C., & Pelizzon, L. (2021). ESG ratings: A critical survey. Journal of Banking & Finance, 130, 106257.

[5] Brandon, D., Krueger, P., & Schmidt, P. S. (2023). The divergence in ESG ratings: An analysis of the social dimension. Journal of Business Ethics, 185(3), 567-589.

[6] Capizzi, M., Giovannini, R., & Pesic, V. (2022). ESG ratings: Correlations and discrepancies across dimensions. Corporate Social Responsibility and Environmental Management, 29(4), 906-918.

[7] Christensen, P. H., Serafeim, G., & Sikochi, A. (2022). Too much of a good thing? ESG disclosure and cognitive dissonance. Journal of Business Ethics, 176(2), 327-345.

[8] Cornell, B. (2020). The ESG premium puzzle. Journal of Applied Corporate Finance, *32*(4), 10-18.

[9] Dorfleitner, G., Halbritter, T., & Weber, M. (2015). ESG ratings: A comparison of different approaches. Journal of Asset Management, 16(5), 310-321.

[10] Kempf, A., & Osthoff, P. (2010). The effect of ESG ratings on portfolio performance. Journal of Business Ethics, 95(S1), 131-147.

[11] Liu, X., Wang, Y., & Zhang, L. (2023). A study of ESG disclosure quality and rating heterogeneity. Management World, 39(5), 135-150.

[12] Feng, Y., & Zhang, Q. (2024). ESG disclosure, rating discrepancies and cost of capital: An analysis based on information asymmetry theory. Journal of Corporate Finance, 84, 102478.

[13] Jiang, Y., & Wang, S. (2025). The impact of non-financial information processing differences on ESG rating divergence. Accounting Review, 100(2), 345-362.

[14] Li, H., & Chen, R. (2024). ESG rating discrepancies from the perspective of institutional logic conflict. Strategic Management Journal, 45(3), 567-589.

[15] Wang, Y., & Chen, X. (2021). ESG investing in the Chinese stock market: Performance and risk. Pacific-Basin Finance Journal, 67, 101542.

[16] Ma, W. J., & Yu, B. J. (2023). Analysis of ESG rating discrepancies and influencing factors at home and abroad. China Industrial Economics, 41(4), 156-174.

[17] Sun, J. X., & Liu, H. (2024). Institutional differences and rating discrepancies of ESG rating agencies: An analysis based on the institutional complexity theory. Organization Science, 35(2), 456-475.

[18] Xu, S. B., & Zhao, L. (2024). A study on the impact of differences in ESG rating objectives on rating divergence. Journal of Business Ethics, 185(1), 123-145.

[19] Zhao, J., & Wang, Y. (2024). ESG disclosure quality and rating divergence: A mediation effect test based on information asymmetry. Journal of Accounting and Public Policy, 43, 107135.

[20] Abate, F., Migliavacca, M., & Scalera, F. (2021). ESG performance and mutual fund efficiency: A DEA approach. Journal of Cleaner Production, 295, 126479.

[21] Zumente, I., & Lāce, I. (2023). Convergence of ESG ratings: An exploratory analysis. Finance Research Letters, 52, 103567.

[22] Zhu, X., & Li, Y. (2023). Assessing the consistency of ESG ratings: A comparison of the top 100 companies. Corporate Social Responsibility and Environmental Management, 30(2), 789-801.