Divergent Measures of Sustainability: A Comparative Analysis of ESG Rating Consistency and Methodological Roots in China

Abstract

The significant divergence in ESG ratings across providers poses a fundamental challenge to their reliability as metrics of corporate sustainability. While this phenomenon is well-documented globally, the consistency of major ESG databases in China remains unexamined. This study conducts the first comprehensive, firm-level comparison of ESG ratings from China's two leading data providers, CSMAR and CNRDS, using a matched panel of A-share listed firms from 2015 to 2020. Our multi-method analysis reveals a profound lack of agreement. We document a persistent level gap, with CSMAR assigning systematically higher scores, and find alarmingly low correlation and reliability coefficients across all ESG pillars—even exhibiting a negative relationship in the Social dimension. Crucially, common standardization techniques fail to reconcile these discrepancies, demonstrating that the divergence is structural, stemming from fundamental methodological differences in their data-driven versus framework-driven philosophies, rather than a scaling artifact. The divergence is most pronounced in industries with complex ESG profiles, such as Finance and Information Technology. These findings underscore that the ratings are non-interchangeable, carrying critical implications for the robustness of empirical research, the validity of investment strategies, and the strategic focus of corporations in China.

1. Introduction

The integration of Environmental, Social, and Governance (ESG) criteria into investment decisions and corporate strategy has become a global paradigm. At the heart of this trend lie ESG ratings, which aim to quantify corporate sustainability performance and guide capital allocation. However, a critical challenge undermines their reliability: substantial divergence in ratings across different providers. A well-documented body of international literature reveals that major rating agencies often produce weakly correlated scores, with pairwise correlations typically ranging from 0.3 to 0.6 (Berg, Koelbel & Rigobon, 2022; Christensen, Serafeim & Sikochi, 2022[1]). This divergence, stemming from fundamental differences in measurement scope, indicator quantification, and weighting aggregation, systematically challenges the validity of ESG scores as objective, comparable benchmarks.

While this "rating divergence" phenomenon is globally recognized, its manifestation and implications in emerging markets remain comparatively underexplored. In China, the world's largest emerging economy, the ESG ecosystem is rapidly evolving, characterized by a unique interplay of policy orientation, improving corporate disclosure, and the rise of domestic data providers. Among these, the China Stock Market & Accounting Research (CSMAR) and China National Research Data Services (CNRDS) have emerged as two leading and widely used ESG databases in academic research and investment practice. [2]Yet, existing studies on the Chinese market have primarily focused on the economic consequences of ESG performance, largely overlooking a foundational question: To what extent are these two dominant rating systems consistent with each other?

Preliminary evidence suggests that, akin to the global landscape, Chinese ESG ratings may suffer from significant inconsistencies. The methodological frameworks of CSMAR and CNRDS differ substantially in their underlying philosophy—CSMAR often described as a data-driven, bottom-up aggregator, while CNRDS aligns more closely with international frameworks

[1] These ranges summarize empirical correlation values reported across multiple large-sample studies, including Berg et al. (2022), Christensen et al. (2022), and Gibson et al. (2021).
[2] For example, recent empirical studies using Chinese ESG data include Li & Wang (2023) and Zhao & Xiao (2022) based on CSMAR, and Cao & Wang (2022) and Hu & Zheng (2023) using CNRDS.

in a top-down manner.[3] Such methodological heterogeneity, as evidenced in mature markets, can lead to contradictory risk assessments, distorted capital allocation, and biased empirical inferences (Gibson, Krueger & Schmidt, 2021; Dimson et al., 2020). Consequently, the growing reliance on these databases for research and investment in China necessitates a systematic evaluation of their comparability. Without such an assessment, findings based on one database may not be generalizable, raising concerns about the robustness and reproducibility of a significant stream of literature.

This study aims to fill this critical gap by conducting the first comprehensive, firm-level comparison of ESG ratings from CSMAR and CNRDS for Chinese A-share listed companies from 2015 to 2020. Our research is designed to address four core questions: whether the two ratings exhibit systematic differences in levels, trends, and distributions; what the degree of correlation and reliability is between the systems, both in absolute scores and relative firm rankings; if common standardization techniques can reconcile the observed discrepancies; and how the patterns of divergence vary across industries and individual ESG pillars.

To address these questions, we construct a matched panel dataset and employ a multi-method analytical framework. We begin by comparing rating levels and trends over time. We then apply a battery of statistical tests, including paired difference tests, Pearson and Spearman correlations, Intraclass Correlation Coefficients (ICC), and Lin's Concordance Correlation Coefficients (CCC). Crucially, we rigorously test the efficacy of normalization techniques—min-max scaling and percentile ranking—to determine if the divergence is a mere artifact of scaling or a deeper, structural issue.

Our findings reveal a profound and persistent lack of agreement between the two databases. We document a significant and stable level gap, with CSMAR consistently assigning higher overall ESG scores. More critically, we find alarmingly low correlation and reliability coefficients across all ESG dimensions, with the Social pillar even exhibiting a negative relationship. These discrepancies are most pronounced in industries with complex, qualitative ESG disclosures, such as Finance and Information Technology. Most importantly, all standardization attempts fail to produce meaningful agreement, unequivocally demonstrating that

[3] Characterizations of the two systems are based on publicly available methodological documents released by CSMAR and CNRDS, including their ESG Technical Manuals (various years).

the divergence is rooted in fundamental methodological differences rather than superficial scaling variations.

The contributions of this study are threefold. First, it provides the first systematic evidence on the scale and structure of ESG rating divergence within China's domestic data ecosystem, a crucial context given the market's global significance. Second, it moves beyond documenting differences to empirically demonstrating the limitations of common harmonization techniques, offering a methodological caution for researchers. Third, it traces the empirical discrepancies back to the core philosophical and architectural divides between the rating systems, providing a coherent explanation for the observed divergence.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature on ESG rating divergence. Section 3 describes the data and methodology. Section 4 presents the empirical comparison and results. Section 5 discusses the methodological roots and implications of our findings, and Section 6 concludes.

2. Literature Review

2.1 ESG Rating Divergence and Consistency Studies

A growing body of literature documents substantial divergence among ESG rating agencies, reflecting underlying differences in information production, data construction, and evaluation methodologies. Early evidence from the global market shows that ratings from major international providers—such as MSCI, Sustainalytics, Refinitiv, and Bloomberg—are often weakly correlated, with average pairwise correlations typically ranging between 0.3 and 0.6 (Berg, Koelbel & Rigobon, 2022; Christensen, Serafeim & Sikochi, 2022). This divergence undermines the reliability of ESG scores as objective measures of corporate sustainability and introduces significant challenges for investors, regulators, and researchers seeking consistent benchmarks.

Berg et al. (2022) categorize the sources of disagreement into three main channels: scope differences (which issues are rated), measurement differences (how indicators are quantified), and weighting differences (how sub-scores are aggregated). Similarly, Gibson, Krueger & Schmidt (2021) emphasize that methodological heterogeneity leads to inconsistent risk pricing in ESG-based portfolios. From an econometric perspective, Dimson et al. (2020) and Christensen et al. (2022) highlight that ESG data inconsistency weakens empirical inference, leading to model instability and biased coefficients in cross-country studies.

In summary, the existing international research agrees that ESG rating divergence is systematic rather than random, and that cross-database comparability cannot be achieved merely through normalization or rescaling. Such findings raise concerns about the validity of ESG scores as comparable measures of corporate sustainability across markets and datasets.

2.2 Methodological Sources of Divergence

The methodological foundations of ESG ratings vary substantially across rating agencies, reflecting differences in how non-financial information is conceptualized and quantified. Even when agencies draw from overlapping data sources, divergence arises from variation in indicator interpretation, qualitative assessment, and aggregation rules (Christensen et al., 2022; Liang & Renneboog, 2020). Berg et al. (2022) provide empirical evidence that measurement differences account for more than half of the overall cross-agency variation, while weighting explains only a

small fraction, suggesting that subjective judgment in indicator coding is the primary source of inconsistency.

From an economic perspective, these methodological discrepancies reflect differences in the production function of information—each rating agency effectively transforms raw environmental, social, and governance disclosures into proprietary "signals" under its own weighting technology. This heterogeneity in information transformation mirrors cross-model differences in how financial analysts or credit rating agencies process firm-level data. In practice, ESG models differ along three interrelated dimensions: (1) Indicator scope and coverage, including whether environmental scores capture resource use, carbon intensity, or climate policies; (2) Data transformation and scaling, such as the normalization of raw metrics, percentile adjustments, or treatment of missing observations; and (3) Aggregation and weighting schemes, ranging from fixed to adaptive weighting, and from equal-weighted to industry-adjusted frameworks.

In financial markets, such methodological divergence generates information asymmetry across investors who rely on different ESG databases, leading to inconsistent risk assessments and distorted capital allocation. When ESG signals differ systematically, asset pricing models incorporating them may yield biased estimates of expected returns, firm valuation, or cost of capital (Gibson et al., 2023). Consequently, identifying and quantifying methodological heterogeneity is not merely a measurement. Concentrating ESG factors into empirical finance and for ensuring comparability in cross-market investment research.

2.3 ESG Rating Divergence in Emerging Markets

Compared with mature markets, ESG data in emerging economies remain fragmented and less standardized. Empirical studies show that disclosure quality, regulatory oversight, and data transparency strongly affect ESG data reliability (Liu et al., 2023; Li & Gao, 2022). In emerging markets such as China, ESG ratings are often influenced by policy orientation, government-driven disclosure incentives, and asymmetric information between firms and data providers. While several domestic data vendors—e.g., CNRDS, CSMAR, and SynTao Green Finance—have developed ESG scoring frameworks, their underlying methodologies differ substantially in indicator selection, weighting, and industry adjustments.

Existing Chinese studies (e.g., Fang et al., 2021; Zhang & Liu, 2022) primarily focus on the impact of ESG performance on firm value or risk, but rarely examine the consistency of the rating systems themselves. Few works have quantitatively compared CSMAR and CNRDS at the firm level using robust agreement metrics such as ICC, CCC, or paired significance tests. Therefore, there remains a clear empirical gap in understanding the extent and structure of rating divergence within China's ESG data ecosystem.

In summary, prior literature has established that ESG rating divergence is a global phenomenon, primarily driven by methodological heterogeneity. However, the degree, direction, and structure of such divergence in the Chinese ESG market remain underexplored. Given the growing use of CSMAR and CNRDS data in academic research and investment analysis, a systematic evaluation of their consistency, bias, and methodological alignment is both timely and necessary. This study fills that gap by conducting a comprehensive firm-level comparison between the two leading Chinese ESG databases—CNRDS and CSMAR—from 2015 to 2020. By jointly analyzing score levels, inter-system correlations, and reliability metrics, it offers the first systematic evidence on the scale and structure of ESG rating divergence in the Chinese market.

3. Data and Methodology

3.1 Data and Sample Selection

I constructed a firm-level panel dataset of Chinese A-share listed companies covering the period from 2015 to 2020. The ESG data were obtained from two major domestic databases — the China Stock Market and Accounting Research (CSMAR) ESG module and the China National Research Data Services (CNRDS) ESG database. While both databases aim to quantify firms' environmental, social, and governance (ESG) performance, their coverage and methodologies differ substantially. The CNRDS database primarily focuses on the constituents of the CSI 300 Index and other large-cap firms with relatively high disclosure quality, whereas the CSMAR ESG module includes all A-share listed firms, offering broader but more heterogeneous coverage.

To ensure comparability, I restricted the sample to firms that were simultaneously covered by both databases in a given year. This intersection yields a balanced and representative subset of the A-share market where both rating systems provide ESG scores. Firm-level financial and industry information were retrieved from the CSMAR financial statements and CSRC industry classification modules. After data cleaning and matching, the final sample contains approximately 280–320 firms per year, resulting in about 1,800 firm-year observations. These firms span 11 primary industry categories defined by the China Securities Regulatory Commission (CSRC), including manufacturing, finance, energy, and information technology. It is important to note that the two databases adopt distinct ESG evaluation methodologies, including differences in indicator selection, scaling, and weighting. While this paper's primary focus is on quantifying and comparing such differences, detailed methodological contrasts are further discussed in Section 6.

3.2 Variable Construction and Standardization

In this section, I define the key variables used in the analysis and describe how I standardized the ESG scores to ensure comparability between the two databases. The primary variables are the raw ESG scores and two standardized versions — min–max normalized scores and percentile-standardized scores — derived from the CSMAR and CNRDS databases, respectively.

For each firm i in year t, I denote the raw ESG scores from the two databases as $ESG_{it}^{CSMAR}$and ESG_{it}^{CNRDS}. These raw scores reflect each provider's original rating methodology, which may differ in scale (e.g., 0–100 vs. 0–10), weighting structure, and temporal calibration across the Environmental (E), Social (S), and Governance (G) pillars.

To address these differences, I first applied min–max normalization within each industry–year cell, rescaling scores between 0 and 1 for each database:

$$ESG_{it}^{MM} = \frac{ESG_{it} - \min(ESG_{kt})}{\max(ESG_{kt}) - \min(ESG_{kt})} \quad (1)$$

where kindexes the firm's industry classification. This normalization preserves the relative score distribution but eliminates scale and unit disparities, ensuring that firms are evaluated relative to their peers in the same industry and year. By controlling for industry composition and annual recalibration effects, this transformation mitigates biases caused by sector-specific score inflation or time-varying rating standards.

Next, I constructed percentile-standardized scores, also within each industry–year group, to further examine whether divergence persists under rank-based comparisons. Specifically, I computed each firm's percentile rank as:

$$ESG_{it}^{pct} = \frac{\text{rank}(ESG_{it})}{N_{kt}} \quad (2)$$

where N_{kt}denotes the number of firms in industry kand year t.

This double-dimension rank standardization places both datasets on a common 0–1 scale, removing not only scale-related bias but also distortions driven by uneven sectoral representation or annual changes in reporting practices. It ensures that each firm's position reflects its relative ESG standing within its contemporaneous industry context.
CCC).

Finally, Table 3 and Table 11 reports the descriptive statistics for the raw, min–max, and percentile-standardized scores. As shown, CSMAR consistently assigns higher ESG and sub-pillar scores than CNRDS, confirming the presence of systematic level gaps even before formal consistency testing. By combining both scale-based (min–max) and rank-based (percentile) standardization within industry–year cells, I ensure that the ensuing comparisons capture methodological divergence rather than artifacts of scale, industry structure, or temporal variation.

3.3 Consistency Tests and Analytical Framework

3.3.1 Consistency Evaluation between Rating Systems

To provide an initial assessment of consistency between the CNRDS and CSMAR ESG rating systems, I begin by evaluating whether the two databases assign statistically different ESG levels to the same firms. For each firm–year observation, I compare the overall ESG score and the three sub-pillars—Environmental (E), Social (S), and Governance (G)—using paired t-tests and Wilcoxon signed-rank tests. These tests are conducted under two scoring conventions: (i) raw ESG scores, and (ii) min–max normalized scores, constructed using the year-specific industry–level transformation defined earlier in Equation (1).

Because ESG score distributions differ systematically across both industries and years, I also implement an extended comparison using standardized scores within **industry–year** groups. These standardized values follow the percentile definitions in Equation (2), which remove scale and distributional distortions by placing firms on a common 0–1 rank scale relative to peers in the same industry and year. This adjustment ensures that cross-database comparisons are not driven by industry composition or temporal changes in reporting environments.

To quantify the magnitude and direction of disagreement between the two databases, I compute firm-level score differences as

$$\Delta ESG_{it} = ESG_{it}^{CSMAR} - ESG_{it}^{CNRDS} \qquad (3)$$

which serve as the primary outcome variable for consistency evaluations. A positive value indicates that CSMAR rates a firm more favorably, while a negative value implies higher ratings in CNRDS. These Δ variables are subsequently used in reliability tests and robustness analyses presented in Sections 3.3.2 and 3.3.3.

In the next subsection, I introduce the full correlation and reliability framework—including Pearson and Spearman correlations, the Intraclass Correlation Coefficient (ICC), and Lin's Concordance Correlation Coefficient (CCC)—which builds directly on these difference measures and standardized scores.

3.3.2 Correlation and Reliability Tests

To complement the mean-difference analysis, I next evaluate the statistical consistency between the CNRDS and CSMAR rating systems using a comprehensive set of correlation- and reliability-based measures. These tests assess whether the two databases produce comparable

information in terms of (i) linear co-movement of score levels, (ii) similarity in firm rankings, and (iii) absolute agreement in assigning ESG values. All tests are conducted for the overall ESG score as well as for the Environmental (E), Social (S), and Governance (G) sub-dimensions, and are repeated under three scoring schemes: raw scores, min–max normalized scores, and industry–year percentile-standardized scores.

I begin by estimating Pearson and Spearman correlation coefficients[4]. Pearson correlations capture the linear co-movement of raw ESG levels, while Spearman correlations evaluate whether the two systems assign similar relative rankings to firms. Following standard practice in ESG consistency research, I compute these correlations both pooled across the full sample and year-by-year[5], and I additionally calculate industry-level correlations after standardizing each firm's ESG score within its specific industry–year cell. This multi-layer correlation structure is designed to distinguish scale-driven differences from systematic methodological divergence across industries and over time.

Next, I assess reliability using two widely adopted metrics in multi-source rating evaluation. First, I compute the Intraclass Correlation Coefficient (ICC) using the absolute-agreement specification[6], which assesses whether CNRDS and CSMAR assign similar ESG levels to the same firm in a given year. Second, I apply Lin's Concordance Correlation Coefficient (CCC) to jointly measure correlation and deviation from the 45-degree line of perfect agreement[7]. Similar to the correlation analysis, both ICC and CCC are computed for each ESG dimension, using raw scores as well as alternative standardized versions to test robustness.

Finally, to ensure that correlations and reliability measures are not confounded by scale differences or industry composition effects, I repeat all tests after applying (i) year-specific min–max normalization (Equation (1))[8], and (ii) industry–year percentile standardization (Equations

[4] Pearson correlation captures linear dependence, while Spearman correlation evaluates monotonic ranking consistency; see Kendall (1970) *Rank Correlation Methods* for classical foundations.
[5] Industry classification follows the CSI industry taxonomy used in CSMAR and CNRDS. Industry–year normalization is recommended in multi-market ESG studies to control industry composition bias (see Berg et al., 2022).
[6] I adopt the ICC(1,1) "absolute agreement" form following McGraw & Wong (1996), which evaluates whether two measurement systems provide interchangeable numerical values rather than merely correlated rankings.
[7] Lin (1989) proposed the CCC specifically for assessing measurement agreement by combining precision (correlation) and accuracy (closeness to the 45° line).
[8] Min–max scaling preserves the relative shape of the distribution and is widely used in ESG harmonization frameworks; see OECD (2020) "ESG Data Harmonization Report".

(2)–(3))[9]. These transformations eliminate differences in scoring range, remove industry-driven dispersion, and harmonize cross-year variability, allowing me to isolate whether the two systems disagree because of numerical scaling or because they embed fundamentally distinct rating methodologies.

Taken together, this combined framework—Pearson and Spearman correlations, ICC, CCC, and multi-level standardized tests—provides a rigorous basis for evaluating the degree of agreement between CNRDS and CSMAR ratings. It allows me to determine whether the two systems move together in levels, agree on firm rankings, or converge toward a common underlying ESG assessment once scale and industry-year effects are accounted for.

3.3.3 Robustness and Validity Checks

To ensure that the empirical comparisons between CNRDS and CSMAR are not driven by differences in score scale, distributional shape, or industry composition, I conduct a series of robustness checks that re-estimate all consistency and agreement analyses under alternative standardization schemes and firm classifications. These procedures allow me to assess whether the observed discrepancies between the two systems persist after harmonizing scale and adjusting for industry–year heterogeneity.

I begin by re-evaluating all firm-level comparisons using percentile-standardized ESG scores. For each database, firms are ranked within each year according to their ESG score, and their percentile positions are computed as in Equation (2). To further control for variation in disclosure intensity and rating dispersion across industries, I extend this transformation to industry–year groups, where each firm is ranked only among peers within its industry and year, as in Equation (3). This industry–year percentile transformation ensures that the two systems are compared in terms of their relative rankings within comparable informational and regulatory environments.

Using these standardized measures, I repeat all agreement tests—including paired t-tests, Wilcoxon signed-rank tests, Spearman correlations, intraclass correlation coefficients (ICC), and concordance correlation coefficients (CCC)—and compute the cross-database difference

$$\Delta ESG_{it} = ESG_{it,\text{CSMAR}}^{pct(ind)} - ESG_{it,\text{CNRDS}}^{pct(ind)} \qquad (4)$$

[9] Rank-based normalization eliminates scale and distributional differences and is especially suitable for ESG ratings given non-standardized scoring systems (Dimson et al., 2020).

Although the percentile transformation removes level differences by construction, correlation and reliability measures remain extremely weak, with ICC values below 0.10 and CCC values near zero across all ESG dimensions. These results indicate that the lack of agreement between the two systems is not driven by differences in scale or industry composition, but instead reflects deeper methodological divergence.

Next, I evaluate robustness to an alternative scaling transformation by applying min–max normalization within each year, as described in Equation (1). This approach rescales scores from both systems to a common $[0,1]$interval while preserving within-year relative rankings. Even under this scheme, CSMAR continues to assign higher normalized ESG scores than CNRDS, and cross-database correlations remain persistently low. This confirms that the observed discrepancies are not due to score-range differences or the influence of extreme observations.

Finally, I compare correlations and reliability statistics across the three scoring schemes—raw, min–max normalized, and industry–year percentile standardized. Across all transformations, Spearman correlations remain weak (typically $| \rho | < 0.15$), ICC values remain below 0.10, and CCC values remain effectively zero. The stability of these results across multiple standardization methods demonstrates that the divergence between CNRDS and CSMAR is structural rather than an artifact of scale or distributional properties.

Taken together, these robustness and validity checks show that the low level of agreement between the two ESG rating systems persists even after extensive harmonization and adjustment procedures. The consistency of these findings across specifications provides strong methodological support for the conclusion that CNRDS and CSMAR should not be treated as substitutes in firm-level empirical analysis.

4. Empirical Comparison on Cross-System Differences

4.1 System-Level Differences: Scale, Trend, and Stability

4.1.1 Rating Scale and Level Differences

To assess whether CNRDS and CSMAR ratings capture the same underlying ESG construct, we begin by examining their temporal evolution from 2015 to 2020. If both systems measure identical firm-level sustainability performance, we would expect to observe parallel trends and stable level relationships over time. However, as summarized in Table 1 and Figure 1, the two databases exhibit fundamental divergences in both scale and trajectory across all ESG dimensions.

My analysis reveals, first, a systematic and persistent scale difference between the two systems. As shown in Table 1, CSMAR consistently assigns ESG scores approximately 20 points higher than CNRDS throughout the sample period. This significant level gap remains remarkably stable, even as both systems capture the general upward trend in corporate ESG performance. This persistence indicates that the divergence stems from fundamental differences in scaling conventions rather than from transient measurement errors. Such a pervasive discrepancy implies that the two systems are built upon divergent baseline assumptions regarding the absolute ESG standing of Chinese A-share firms. The magnitude of this gap is substantial enough to materially influence the outcomes of any empirical analysis relying on raw ESG levels—such as cross-sectional regressions, portfolio formation, or industry benchmarking—leading to different conclusions depending on the database selected.

Second, although both datasets document an overall improvement in firms' ESG performance over time, the shape and timing of these improvements differ markedly. CNRDS exhibits a smooth, monotonic increase across all dimensions, a pattern consistent with gradual enhancements in ESG disclosure regulations and corporate sustainability practices in China. In contrast, CSMAR's trajectories are characterized by pronounced structural breaks. Specifically, its Environmental and Social scores remain relatively flat from 2015 to 2019, followed by an abrupt escalation in 2020. Such sudden, synchronized shifts are unlikely to reflect genuine, large-scale economic phenomena, as ESG fundamentals typically evolve incrementally rather than discontinuously. Instead, these jumps are highly consistent with a substantial methodological overhaul, potentially involving changes in indicator definitions, weighting schemes, disclosure coverage, or the integration of newly digitized data sources.

Third, the governance dimension reveals an even sharper contrast in evaluative philosophy. Prior to 2019, CSMAR's Governance scores were extremely low—often less than half the corresponding CNRDS values—suggesting that its pre-2019 governance assessment relied on a narrower or more stringent set of indicators. However, in 2020, CSMAR's G scores surged to levels comparable with CNRDS. This suggests a significant recalibration of governance metrics around that time, possibly linked to the adoption of more comprehensive board-structure indicators or an effort to harmonize governance criteria with international ESG frameworks.

In summary, while both datasets capture the macro-level improvement in ESG performance, their underlying scoring logics diverge sharply. CSMAR appears to incorporate episodic and substantial methodological updates, whereas CNRDS provides a smoother and potentially more consistent long-term assessment. These patterns strongly suggest that the two datasets are unsuitable for research designs requiring high temporal consistency—such as event studies or pre-post regulatory evaluations—unless they are carefully and explicitly harmonized.

4.1.2 Industry-Level Differences in Rating Levels

To further explore the structure of cross-system discrepancies, we conduct an industry-level comparison using multi-year averages. Table 2 reports the mean ESG scores for 16 industries, while Figure 2 visualizes the temporal evolution **of the top three industries**, illustrating representative patterns of cross-system divergence.

First, as detailed in Table 2, CSMAR consistently assigns higher ESG scores across nearly all sectors, with nearly uniform directionality of differences. The average disparity of 19–20 points echoes the system-wide level gap observed earlier, suggesting that industry composition alone cannot explain the discrepancies. Even industries with relatively strong regulatory oversight—such as utilities, transportation, and finance—exhibit wide gaps, indicating that the discrepancies stem from the databases' scoring frameworks rather than sector-specific reporting heterogeneity.

Second, Table 2 confirms that the magnitude of cross-system divergence is strongly industry-dependent, with high-technology and high-regulation industries... displaying the largest differences. These sectors have complex ESG reporting environments and feature indicators that are more sensitive to changes in methodology—such as cybersecurity, data governance,

emissions reporting, and environmental compliance metrics. Thus, the larger discrepancies in these industries likely reflect higher methodological elasticity within CSMAR's rating system, especially after its 2019–2020 revision.

Third, CNRDS exhibits greater cross-industry variance, suggesting it better discriminates between sectors with inherently different ESG risk profiles. CSMAR, on the other hand, presents a more compressed cross-industry distribution, potentially due to scaling choices that disproportionately elevate the scores of lower-rated industries. This pattern implies that CNRDS may better capture genuine sectoral heterogeneity by distinguishing among industries with different ESG risk profiles. CSMAR, on the other hand, presents a more compressed distribution, potentially due to scaling choices that raise lower-scoring industries more than higher-scoring ones.

Finally, the temporal evolution of industry-level scores (Figure 2) reinforces these differences. CNRDS maintains stable industry rankings, reflecting consistent evaluation criteria. CSMAR, however, displays widening gaps across industries post-2018, with some sectors experiencing rapid score inflation. These widening cross-industry disparities are indicative of a scoring system that is more sensitive to methodological shifts than to underlying firm behavior. To delve deeper into the nature of these temporal shifts, Figure 3 decomposes the trends by ESG sub-pillars for the top five industries. The visualization reveals that CSMAR's abrupt score inflation in 2020 was pervasive across Environmental, Social, and Governance dimensions, affecting industries like Health and Social Work most prominently. This granular, sub-pillar evidence underscores that the methodological recalibration was systemic, reshaping not just overall scores but the fundamental composition of ESG assessments across sectors.

Taken together, the industry-level analyses highlight that the two databases not only differ in absolute scale but also in cross-sectoral structure. As a consequence, empirical work relying on industry-level ESG variation—such as sector-adjusted ESG regressions, industry-matched portfolios, or sectoral ESG benchmarking—may produce contradictory results depending on the chosen data provider.

4.2 Industry-Level Divergence: Patterns, Sources, and Implications

4.2.1 Cross-Industry Divergence Patterns

To move beyond system-wide averages and uncover the structural sources of disagreement, we analyze rating differences across industries. This comparison reveals that the divergence is not random but follows predictable patterns tied to industry characteristics, underscoring the methodological schism between the two rating systems.

First, we find that the direction and magnitude of divergence are systematically related to industry type. As detailed in Table 4, CSMAR tends to assign higher ratings to many industrial, capital-intensive, and resource-extraction sectors—such as ferrous metal smelting, support activities for mining, and water transport—particularly on environmental and operational metrics. Conversely, service-oriented and traditional sectors like Postal Services and Textile & Apparel are consistently assigned lower scores by CSMAR relative to CNRDS. It is important to note that this pattern is not absolute, as some heavy industries also show mixed results; this indicates that the methodological bias interacts with specific industry characteristics in a complex way. Overall, the pattern suggests that CSMAR's framework is structurally more favorable to many heavy industries, while CNRDS applies comparatively lenient criteria to several service sectors.

Second, the divergence is not uniform across ESG pillars, revealing dimension-specific methodological priorities. The Environmental (E) pillar exhibits the most extreme dispersion, with massive positive gaps in pollution-intensive industries. The Social (S) pillar shows consistent but complex directional differences, with CSMAR awarding higher scores in industrial sectors but lower scores in service industries. In contrast, the Governance (G) pillar displays smaller and more mixed differences, suggesting less consensus on how to evaluate corporate governance across datasets. This pillar-level heterogeneity confirms that the systems disagree not only on overall performance but on the fundamental constituents of ESG quality.

Third, as shown in Panel B of Table 2, CNRDS exhibits greater cross-industry variance (Std. Dev. = 4.84) than CSMAR (Std. Dev. = 3.21). This suggests that CNRDS better discriminates between sectors with inherently different ESG risk profiles, while CSMAR's more compressed distribution implies a scaling choice that disproportionately elevates the scores of lower-rated industries.

In summary, the cross-industry analysis demonstrates that the choice of database does not merely shift all scores by a constant but actively reshapes the landscape of sectoral ESG performance. Consequently, any research reliant on industry-level variation—including sector-

adjusted regressions, industry-matched portfolios, or sectoral benchmarking—is highly susceptible to database-driven conclusions.

Finally, the temporal evolution of industry-level scores (Figure 2) reinforces these differences. CNRDS maintains stable industry rankings, reflecting consistent evaluation criteria. CSMAR, however, displays widening gaps across industries post-2018, with some sectors experiencing rapid score inflation. These widening cross-industry disparities are indicative of a scoring system that is more sensitive to methodological shifts than to underlying firm behavior.

Taken together, the industry-level analyses highlight that the two databases not only differ in absolute scale but also in cross-sectoral structure. As a consequence, empirical work relying on industry-level ESG variation—such as sector-adjusted ESG regressions, industry-matched portfolios, or sectoral ESG benchmarking—may produce contradictory results depending on the chosen data provider.

4.2.2 Within-Industry Consistency and Reliability

Having established patterns of divergence across industries, we now investigate whether the two systems achieve any consistency *within* industries at the firm level. We employ a trio of agreement metrics—linear correlation, intraclass correlation (ICC), and concordance correlation (CCC)—to test for reliability. The results, summarized in Table 5, decisively indicate a lack of any meaningful agreement.

First, linear regressions of CNRDS scores on CSMAR scores yield slopes that are negligible—even negative in the case of the Social pillar—along with R^2 values approaching zero. This fundamental lack of a linear relationship means that knowing a firm's CSMAR score provides virtually no information for predicting its CNRDS score, and vice versa. The two systems are capturing unrelated signals at the firm level.

Second, measures of absolute agreement are equally poor. Intraclass correlation coefficients (ICC) are consistently below 0.1 across all dimensions, falling squarely within the "poor reliability" range. This indicates that the two systems cannot be used interchangeably as measures of the same underlying ESG construct. Even the dimension with the highest agreement (Environmental) fails to demonstrate acceptable reliability.

Third, Lin's concordance correlation coefficients (CCC), which assess both precision and accuracy, are near zero. A CCC near zero signifies a complete lack of agreement in both the

ranking of firms and the absolute level of their scores. Notably, the Social pillar even shows negative concordance in some industries, implying that the systems not only disagree but sometimes rank firms in opposite directions.

Finally, we test whether standardizing scores—using percentile ranks or min-max scaling—can reconcile these differences. As shown in Table 6, the answer is a definitive no. Neither transformation meaningfully improves linear correlation, ICC, or CCC values. The persistent disagreement after normalization proves that the root cause is not a simple difference in score distribution, but profound discrepancies in indicator selection, weighting, and the very definition of what constitutes good ESG performance.

The collective evidence from these reliability tests confirms that the divergence is structural and fundamental. CNRDS and CSMAR ratings are not merely noisy measures of the same thing; they are effectively measuring different constructs. This renders them non-interchangeable for firm-level analysis within industries.

4.3 Limitations of Standardization Approaches

4.3.1 Comparison of Standardization Methods

A critical question is whether the documented divergences can be reconciled through common standardization techniques. We evaluate this by applying min-max scaling, percentile transformation, and industry-adjusted normalization. The results, summarized in Tables 7, 8 and 9, unequivocally demonstrate that these methods fail to produce meaningful agreement.

First, normalizing scores to a common [0, 1] scale does not eliminate systemic biases. The directional discrepancy persists unchanged: across all ESG dimensions and normalization schemes, CSMAR consistently produces higher standardized values than CNRDS. The Environmental and Social pillars continue to exhibit especially large gaps, often exceeding 0.20, which underscores a systematic inflation in CSMAR's assessment of these disclosures. This persistence indicates that the core issue is not one of distributional range but of fundamental scoring levels.

Second, and more critically, the two systems show virtually no co-movement even after normalization. As detailed in Table 8, cross-database correlations are negligible across all dimensions ($|\rho| < 0.10$), with the Social pillar again exhibiting a negative relationship. This means that knowing a firm's relative standing in one system provides no information for

predicting its standing in the other, confirming that the discordance extends beyond simple level differences to the very ordering of firms.

Third, industry-adjusted normalization offers only marginal improvement. While controlling for sectoral composition reduces the mean difference in the composite ESG score to statistically insignificant levels, significant discrepancies persist within the E, S, and G sub-pillars (Table 7). This key finding implies that industry effects account for only a minor portion of the overall divergence. The primary drivers lie deeper, in the indicator-level methodological choices—such as disclosure scoring, weight assignment, and dimension definitions—that are applied uniformly across sectors.

Fourth, transforming scores into ordinal ranks or quintiles reveals a profound lack of categorical agreement. The exact quintile agreement rate remains below 22% across all dimensions (Table 9), a level barely exceeding random chance. Chance-adjusted agreement (Cohen's kappa) is effectively zero, and rank-based correlations (Spearman) are weak. This has dire implications for research relying on portfolio sorts or categorical analyses, as the resulting firm groupings are entirely database dependent.

The persistent level differences across all normalization schemes are visually summarized in the Bland-Altman plots (Figure 6), which clearly show a positive mean bias and wide limits of agreement, confirming the failure of these techniques to harmonize the two systems. In conclusion, the collective failure of level normalization, correlation structure, and categorical alignment confirms that the divergence is not a superficial scaling artifact. It is rooted in the structural architecture of the rating systems themselves.

4.3.2 Post-Standardization Correlation and Agreement

After establishing that standardization cannot align the score levels, we investigate whether it at least produces coherent *patterns* of co-movement or enables a practical linear mapping between the systems. The evidence on both fronts is negative.

An examination of the correlation structure of standardized scores provides further evidence of conceptual misalignment While each system displays strong internal coherence—as vividly illustrated in the correlation heatmap (Figure 5)—with CSMAR in particular showing high correlations (>0.6) between its composite score and sub-pillars, indicating a tightly integrated model—the cross-system correlations are virtually nonexistent. This lack of alignment

is persistently observed year-on-year, as shown in Figure 4. Pairwise correlations between CNRDS and CSMAR standardized scores rarely exceed 0.10 and are often close to zero. The Social dimension even displays slightly negative correlation, implying that firms praised for their social performance by one system may be penalized by the other. This suggests a fundamental disagreement on what constitutes good social performance.

Finally, we directly test the feasibility of a simple linear conversion between the two systems. Regressing CNRDS standardized scores on their CSMAR counterparts yields slopes that are negligible in magnitude—including those that are statistically significant—and R^2 values consistently below 1%. The Social dimension again presents the most striking evidence, with a significantly negative slope, definitively ruling out any stable, monotonic relationship. These results, detailed in Table 10, make it clear that no simple formula can translate a CSMAR score into its CNRDS equivalent, or vice versa. The fundamental lack of a linear relationship is vividly captured in the scatterplots (Figure 8), where the data points form a diffuse cloud far from the line of perfect agreement, rather than a tight, diagonal band.

Collectively, the evidence in this section —from tabular analysis (Tables 8, 10) to visual trends (Figures 4, 5)— confirms that normalization cannot bridge the methodological chasm between CNRDS and CSMAR. The divergence is robust to scaling, persists in ordinal rankings, is absent in cross-system correlations, and cannot be resolved through linear mapping. This underscores that the two providers are not merely rating the same construct with different scales, but are capturing fundamentally different constructs altogether.

4.4 Distributional Properties and Structural Differences

4.4.1 Comparative Distributional Characteristics

The preceding analyses establish divergence in levels, trends, and correlations. We now probe deeper into the distributional properties of the standardized scores to uncover the underlying structural disparities. The evidence confirms that the differences are not merely parametric but reflect fundamentally distinct shapes and dispersions in the score distributions.

First, beyond the well-documented mean differences, the two systems produce distributions with markedly different variances and shapes. As detailed in Table 11, CNRDS exhibits greater cross-sectional variance in the Environmental and Social dimensions, as shown by their higher standard deviations (0.193 vs. 0.173 and 0.192 vs. 0.154, respectively). The

systematic upward shift and compression of CSMAR's distributions, as summarized numerically in Table 11, are clearly visualized in the Figures 9 – 12. The distributions for the Governance pillar, however, show identical dispersion (0.169 for both systems), while CSMAR maintains a higher mean. This wider dispersion suggests a scoring model capable of discriminating more finely between firms with varying ESG performance. In contrast, CSMAR's distribution is notably more compressed, with lower overall variance. This pattern is consistent with a scoring framework that applies upward adjustments, particularly to lower-performing firms, thereby reducing heterogeneity and compressing the score range. This fundamental difference in discriminative power has direct implications for research seeking to explain cross-sectional variation in ESG performance.

Second, the distributional differences are robust across normalization techniques, including percentile ranks and min-max scaling. The persistence of these patterns after removing scale effects indicates that the root cause lies in the core rating methodologies—such as how indicators are weighted, how disclosure quality is scored, and how points are aggregated across categories. The systems are not just calibrated differently; they are built to measure different concepts of ESG quality.

In summary, the distributional evidence moves beyond establishing *that* the systems differ, to illustrating *how* they differ in their fundamental architecture. CNRDS portrays a world of greater ESG heterogeneity, while CSMAR presents a more homogenized landscape. This reinforces the conclusion that they cannot be used interchangeably.

4.4.2 Industry Heterogeneity Patterns

The structural divergence between the rating systems is not uniform across the economy but is instead mediated by industry characteristics. A fine-grained analysis of industry-level differences, as shown in Tables 12, 13, and 14, reveals a clear hierarchy of discrepancy that is logically tied to the nature of ESG measurement in each sector.

We find that the magnitude of cross-system divergence is strongly predicted by industry attributes. Sectors with complex, qualitative, and rapidly evolving ESG disclosures—such as Finance, Information Technology, and Real Estate—consistently exhibit the largest discrepancies. In these "high-divergence" industries, the lack of standardized, quantitative metrics grants rating providers greater discretion in indicator selection and weighting, thereby

amplifying methodological differences. The Environmental and Social pillars are the primary sources of disagreement here, displaying a near-uniform upward bias in CSMAR's scores. In contrast, the Governance pillar shows a more mixed pattern.

Conversely, industries with more established, quantitative, and physical ESG metrics—such as Mining and Raw Materials Manufacturing—show the smallest discrepancies. In these "low-divergence" sectors, the reliance on more objective data points (e.g., emissions, energy use, safety records) constrains methodological elasticity, leading to greater alignment between the two providers.

Furthermore, the pattern of differences across ESG pillars within industries is highly revealing. The Governance pillar shows the most symmetric and smallest differences, and even exhibits reversals where CNRDS assigns higher scores (e.g., in Construction and Real Estate). This suggests that CNRDS may employ a stricter or more comprehensive governance evaluation in certain contexts. Ultimately, the within-industry rank-order agreement remains dismally low (rarely exceeding 0.30), proving that the systems disagree on firm rankings even within homogenous sectoral contexts.

The complex pattern of industry- and dimension-specific divergence is synthesized in the heatmap (Figure 13), which immediately reveals that Environmental and Social scores in sectors like Finance and IT are the primary sources of cross-system discrepancy. This systematic heterogeneity implies that the sensitivity of empirical research to database choice is itself industry-dependent. Conclusions about the ESG performance of financial or technology firms are far more fragile than those concerning mining or utilities. This final layer of analysis underscores that the non-interchangeability of CNRDS and CSMAR is a pervasive, structural feature of the Chinese ESG rating landscape, with implications that vary predictably across the economy.

5. Robustness Test

5.1 Temporal Stability of Correlation Patterns

We first investigate whether the weak association between the two rating systems varies across time periods, particularly given the apparent methodological shift in CSMAR's approach around 2020. If the divergence were primarily driven by this recalibration episode, we would expect substantially lower correlations in 2020 compared to earlier years.

The results contradict this temporal-specificity hypothesis. As shown in Figure 7 and detailed in Appendix B, both Pearson (level-based) and Spearman (rank-based) correlations remain exceptionally weak throughout the 2015-2020 period, rarely exceeding 0.20 in any year or dimension. The Social pillar consistently demonstrates the weakest alignment, frequently exhibiting negative correlations. Most importantly, we observe no systematic improvement in correlation coefficients during the pre-2020 period, indicating that the disagreement represents a persistent feature of the two rating systems rather than a temporary disruption. These findings withstand percentile standardization within each year, confirming that the low correlations reflect fundamental methodological differences rather than year-specific scaling variations.

5.2 Firm-Level Difference Tests

To complement the correlation analysis, we implement rigorous paired difference tests at the firm-year level. Appendix C provides comprehensive evidence from both parametric (t-test) and non-parametric (Wilcoxon signed-rank test) approaches, examining whether the mean differences observed in our primary analysis hold under alternative statistical methods.
The results robustly confirm systematic directional differences across all ESG dimensions ($p < 0.001$). The two methodological approaches yield virtually identical conclusions: CSMAR assigns significantly higher scores for environmental performance (mean difference ≈ 11.7 points) and overall ESG assessment (≈ 20 points), while CNRDS assigns higher scores for social (≈ 3.2 points) and governance (≈ 15.5 points) dimensions. The remarkable consistency of these directional patterns across different statistical methods strengthens the conclusion that the observed discrepancies reflect conceptual differences in how the two systems weight various ESG components, rather than statistical artifacts or outlier-driven effects.

Critically, when we apply percentile standardization to remove scale differences (Appendix C, Panel B), the mean differences vanish as expected, but this does not indicate true agreement. Rather, it confirms that the raw-score discrepancies originate from distinct scaling frameworks, while the underlying methodological divergence persists in rank-based measures.

5.3 Industry-Level Aggregation Tests

We further examine whether the documented discrepancies persist when analyzing data at the industry level, thereby addressing concerns about firm-level noise or compositional effects. If the differences were merely idiosyncratic to specific firms, industry-level aggregation should substantially reduce the observed discrepancies.

The industry-level analysis in Appendix E reveals precisely the opposite pattern. The directional differences observed at the firm level persist—and in some cases intensify—when examining industry averages. The systematic pattern of CSMAR assigning higher environmental ratings and CNRDS assigning higher governance ratings holds across nearly all sectors. Appendix E, Panel A shows that industries such as Finance, IT Services, and Water Production & Supply exhibit the largest absolute differences (exceeding 20 points in raw scores), confirming that discrepancies are most pronounced in sectors where rating discretion is greatest.

Moreover, the magnitude of disagreement follows a logical pattern: largest in sectors with complex, qualitative ESG indicators and smallest in sectors with standardized, quantitative metrics. This industry gradient, detailed in both Appendix E and our earlier cross-sectional analyses, provides compelling evidence that methodological differences are most pronounced where rating discretion is greatest, further supporting the interpretation that the divergence stems from fundamental measurement approaches.

5.4 Comprehensive Standardization Assessment

We extend our earlier analysis of standardization techniques by examining their efficacy within temporal and industry subgroups. Appendix D provides a rigorous year-by-year decomposition of both raw and percentile-standardized differences, offering insights into the temporal evolution of cross-system discrepancies.

The results uniformly demonstrate the limitations of standardization approaches. While percentile transformation within year-industry cells mechanically eliminates mean differences (as

confirmed by the near-zero differences in Appendix D, Panel B), rank-based agreement remains exceptionally weak (correlations consistently below 0.10). The 2020 reversal pattern observed in Appendix D, Panel A is particularly informative: CSMAR's Social and Governance scores surge above CNRDS levels in that year, suggesting a methodological realignment rather than random variation. This pattern holds across all years, industries, and ESG dimensions, indicating that the systems disagree not only on absolute performance levels but also on relative rankings within well-defined peer groups. The failure of standardization to produce meaningful agreement even under these controlled conditions provides particularly compelling evidence that the divergence originates from conceptual rather than statistical differences.

The battery of robustness tests conducted in this section—supported by Appendices B through E—systematically addresses potential alternative explanations for the documented discrepancies between CNRDS and CSMAR ESG ratings. The consistency of findings across temporal periods, methodological approaches, aggregation levels, and normalization techniques provides strong evidence that the observed divergence represents a structural feature of the two rating systems rather than a methodological artifact.

Three key conclusions emerge from this comprehensive robustness assessment. First, the temporal analysis reveals persistent weak correlations throughout 2015-2020, indicating that the disagreement is chronic rather than episodic. Second, the firm-level and industry-level tests confirm directional consistency, with each system emphasizing different ESG components in a conceptually coherent pattern. Third, standardization techniques consistently fail to achieve meaningful harmonization, demonstrating that the differences are deeply embedded in the rating methodologies themselves. These findings have important implications for empirical research. The non-interchangeability of CNRDS and CSMAR ratings appears to be a fundamental characteristic that researchers must acknowledge rather than a methodological inconvenience that can be easily remedied through statistical adjustments. Future research should explicitly account for these systematic differences when designing studies or interpreting results based on Chinese ESG ratings.

6. Discussion

6.1 Methodological Divergence: The Roots of Empirical Discrepancies

6.1.1 Core Philosophical Divide: Framework-Driven vs. Data-Driven Models

The systematic evidence from this study demonstrates that the ESG ratings from CNRDS and CSMAR exhibit fundamental differences in level, trend, structure, and concordance. These discrepancies are not random noise but a direct manifestation of their underlying methodological philosophies—framework-driven (CNRDS) versus data-driven (CSMAR)—which permeate their data sources, methodological pathways, indicator construction, and weighting schemes.

CSMAR adheres to a "data-driven" bottom-up logic. Its model originates from the collection of approximately 700 underlying data points, which are aggregated upwards into over 200 indicators and finally categorized into 14 key issues. This approach predisposes its scores to reflect the completeness of management processes and the existence of policies (CSMAR, 2022). CNRDS follows a "framework-driven" top-down design. Its system prioritizes alignment with international frameworks such as ISO 26000, GRI, and SASB, using the Analytic Hierarchy Process to determine indicator weights before sourcing supporting data (CNRDS, n.d.). This approach focuses its scores more on performance outcomes, risk exposure, and adherence to international standards.

This philosophical divide directly manifests in the core empirical results. The data-driven aggregation and industry-weighting of CSMAR likely contribute to score distributions that often approximate normality. Conversely, CNRDS's framework-driven, standard-referenced approach appears to create a stricter benchmark, resulting in more asymmetrical, left-skewed distributions across the E and S dimensions, as illustrated in Figure 9. Furthermore, the scatterplots in Figure 8 visually confirm the absence of a strong linear relationship, most notably the weak negative correlation in the Social (S) dimension. These patterns are quantified at the industry level in Table 14, which reveals that sectors like Finance and IT Services exhibit the highest divergence, often driven by the environmental and social pillars.

The definitive explanation for these statistical patterns lies in a side-by-side comparison of the two systems' indicator compositions, which reveals that CNRDS and CSMAR are not merely "measuring" ESG differently; they are, in effect, "defining" it through distinct conceptual lenses. This is most acute in the Social (S) dimension, where the core of CSMAR's construct is "Social Responsibility within Business Operations," focusing on internal and partner

relationships through themes like Human Capital Development and Supply Chain Management. CNRDS, however, constructs its S dimension around "External Social Impact and Contribution," prioritizing broader societal welfare through themes like Social Contribution and Health and Safety. This fundamental dichotomy explains the negative correlation observed in Figure 8, as a firm excelling in CSMAR's operational metrics may perform poorly on CNRDS's impact-oriented benchmarks. A parallel divergence exists in Governance (G), where CSMAR centers on "Foundational Compliance and Internal Governance," while CNRDS emphasizes "Market Communication and Risk Governance." This clarifies the significant negative differences for industries like Real Estate in Table 14, where a firm may have sound internal structures (scoring well in CSMAR) but lackluster disclosure practices (scoring poorly in CNRDS). Although not negatively correlated, the Environmental (E) dimension follows a similar pattern, with CSMAR emphasizing management systems (e.g., Environmental Policy) and CNRDS stressing performance and risk (e.g., Circular Economy).

These conceptual divergences are further amplified by the systems' different approaches to industry customization. CSMAR's development of 51 customized models with a "core + sector-specific indicators" structure allows its scores to reflect deep industry characteristics. CNRDS, using a broader 19-primary-industry classification and adjusting weights via expert pairwise comparison, applies a different granularity and adjustment logic. This discrepancy causes their scoring priorities to diverge even within the same industry, an effect maximized in sectors with complex profiles, such as Finance and IT, as highlighted in Table 14, Panel B.

In summary, the roots of all systemic discrepancies—from the distributional shapes in Figure 9 and the correlation patterns in Figure 8 to the industry-level divergences in Table 14—are now clear. This is not a matter of measurement error but is the inevitable result of the two agencies constructing distinct "social realities" of ESG. CNRDS and CSMAR provide two different ESG edifices, built from different theoretical blueprints and data bricks.

6.2 Theoretical Implications: The Social Construction of ESG Ratings

Our findings provide strong evidence for the profoundly socially constructed nature of ESG ratings. There is no single, objective "ESG truth" for rating systems to discover. Instead, CNRDS and CSMAR, through their respective methodological lenses, actively construct distinct versions of ESG reality. This fact challenges a fundamental assumption in much of the

mainstream literature: treating the score from any single rating system as a straightforward proxy for a firm's "true" ESG performance is theoretically untenable. Consequently, future research should shift its focus from the quest to identify the "most accurate" rating towards a deeper understanding of the specific realities constructed by different systems. Investigating how these divergent constructions influence capital allocation and corporate strategic behaviour represents a critical new direction for the field.

6.3 Practical Implications and Recommendations

This study serves as a clear warning: ESG ratings from CNRDS and CSMAR are not interchangeable. Investment decisions based on a mix of ratings from the two systems are inherently flawed. Investors must become informed users—understanding the methodological "why" behind the scores. If an investment strategy prioritizes the robustness of management systems and processes, CSMAR may offer more relevant signals. Conversely, for strategies focused on tangible performance outcomes and risk exposure, CNRDS is likely to be a more appropriate reference.

Corporations must confront the reality of "one set of disclosures, multiple interpretations." The strategic imperative should shift from the futile effort to cater to all rating systems towards a focused commitment to understand the key material ESG issues within their specific industry. Substantive improvement and transparent disclosure on these material issues are paramount. Achieving relative leadership within an industry peer group is a more credible and sustainable goal than chasing an absolute high score from any single provider.

Based on the findings of this study, we call for the development of a more consistent and transparent ESG assessment methodology framework. Regulators should focus on standardizing foundational data, rating agencies must enhance methodological transparency, and academic research needs to treat rating system selection as a core methodological issue. Only through these concerted efforts can we establish comparability and dialogue across differently constructed ESG realities and advance the field toward greater maturity.

Conclusion

This systematic comparison demonstrates that CNRDS and CSMAR ESG ratings, driven by fundamentally different methodological philosophies, construct divergent realities of corporate sustainability in China. Our empirical evidence, robust across firm-level, industry-level, and temporal analyses, solidifies this core argument with three key findings:

First, a profound methodological schism defines the systems. CSMAR's data-driven, bottom-up approach emphasizes management processes and policy completeness, whereas CNRDS's framework-driven, top-down methodology prioritizes performance outcomes and risk exposure. Second, this philosophical divide manifests in stark empirical divergences. Our empirical analysis reveals a consistent pattern of fundamental divergence: CSMAR systematically assigns significantly higher absolute scores, particularly in the environmental and social dimensions, while the two systems exhibit exceptionally weak correlation—even displaying a negative relationship in the social (S) pillar, which serves as definitive evidence that they are measuring distinct constructs. Furthermore, these discrepancies are most pronounced in sectors with complex and qualitative ESG profiles, such as Finance and IT, indicating that methodological discretion substantially amplifies cross-system divergence. Third, and most critically, statistical standardization fails to reconcile these differences. Neither percentile ranking nor industry adjustment produces meaningful agreement. This failure proves that the root cause is not a superficial scaling issue but a deep, conceptual disagreement on the very definition of ESG performance.

These findings necessitate a paradigm shift for all stakeholders. Investors must treat the ratings as non-interchangeable, selecting databases strategically—CSMAR for management quality, CNRDS for performance and risk. Corporations should abandon the futile goal of pleasing all systems and instead focus on substantive leadership on industry-material issues. For the broader ecosystem, we call for coordinated action: regulators must prioritize foundational data standardization, rating agencies must enhance methodological transparency, and researchers must treat rating system choice as a core methodological decision. Ultimately, this study establishes that the Chinese ESG landscape is not awaiting the discovery of a single "correct" rating. Instead, market maturity will come from acknowledging this multi-faceted reality, building transparency, and developing the sophistication to critically interpret these different constructions of corporate sustainability.

References

Berg, F., Koelbel, J. F., & Rigobon, R. (2022). *Aggregate confusion: The divergence of ESG ratings.* Review of Finance, 26(6), 1315–1344. https://doi.org/10.1093/rof/rfac033

Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). *Why is corporate virtue in the eye of the beholder? The case of ESG ratings.* The Accounting Review, 97(1), 147–175. https://doi.org/10.2308/TAR-2019-0506

Cao, Y., & Wang, J. (2022).
ESG disclosure and investment efficiency: Evidence from Chinese listed firms. *China Journal of Accounting Research*, 15(3), 247–263.

CSMAR. (2022). *商道融绿 ESG 评级研究数据库使用说明书* (pp. 18–19). Shenzhen CSMAR Data Technology Co., Ltd.

CNRDS. (n.d.). *上市公司 ESG 评级数据库说明书* (pp. 1, 5–6). CNRDS.

Dimson, E., Marsh, P., & Staunton, M. (2020). *Divergent ESG ratings and their implications for investors.* Credit Suisse Global Investment Returns Yearbook 2020, 1–20.

Fang, Z., Lu, W., & Zhang, Y. (2021). *ESG performance and corporate risk: Evidence from China.* China Journal of Accounting Research, 14(3), 333–352. https://doi.org/10.1016/j.cjar.2021.05.002

Gibson, R., Krueger, P., & Schmidt, P. S. (2021). *ESG rating disagreement and stock returns.* Financial Analysts Journal, 77(4), 104–127. https://doi.org/10.1080/0015198X.2021.1983403

Gibson, R., Glossner, S., Krueger, P., Matos, P., & Steffen, T. (2023). *Do investors care about ESG disagreement?* Journal of Financial Economics, 149, 110–132. https://doi.org/10.1016/j.jfineco.2023.01.011

Hu, Q., & Zheng, Y. (2023). ESG performance and financial constraints: Evidence from China. *Finance Research Letters*, 54, 103657.

Kendall, M. G. (1970). *Rank correlation methods*. Charles Griffin.

Li, Y., & Gao, J. (2022). *ESG information disclosure, investor attention, and market reaction: Evidence from China.* Emerging Markets Finance and Trade, 58(11), 3251–3267. https://doi.org/10.1080/1540496X.2022.2062871

Li, Z., & Wang, T. (2023). ESG performance and corporate innovation: Evidence from China. *Journal of Corporate Finance*, 82, 102518.

Liang, H., & Renneboog, L. (2020). *On the foundations of corporate social responsibility.* Journal of Finance, 75(4), 1715–1768. https://doi.org/10.1111/jofi.1280

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45*(1), 255–268.

Liu, Q., Wang, J., & Chen, X. (2023). *Assessing the consistency of ESG ratings in China: An empirical analysis.* Sustainability, 15(5), 3924. https://doi.org/10.3390/su15053924

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46.

OECD (2020). *ESG Investing: Practices, Progress and Challenges*. OECD Publishing.

OECD (2020). *ESG Data Harmonization Report*. OECD.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.

SynTao Green Finance. (2021). *China ESG Rating Landscape 2021.* Retrieved from https://www.syntaogf.com/

Zhang, L., & Liu, J. (2022). *Does ESG disclosure improve firm value? Evidence from Chinese listed companies.* Journal of Cleaner Production, 335, 130–178. https://doi.org/10.1016/j.jclepro.2022.130178

Zhao, X., & Xiao, L. (2022). Does ESG affect corporate financing costs? Evidence from the Chinese bond market. *Pacific-Basin Finance Journal*, 74, 101798.

Appendix B. Robustness Test: Year-by-Year Correlation of ESG Ratings

This panel reports year-by-year correlations between CSMAR and CNRDS ESG ratings under both raw and percentile-standardized forms. Across all years and dimensions, both Spearman (rank-based) and Pearson (level-based) correlations remain consistently weak, typically below 0.20. The Environmental and Governance pillars occasionally exhibit modest positive association (e.g., E ≈ 0.22 and G ≈ 0.19 in 2019), but these values remain far below conventional thresholds for moderate correlation ($\rho = 0.5$). Social scores, in contrast, frequently display negative or near-zero correlations, suggesting divergent assessments of firms' social performance across systems. Importantly, the percentile transformation does not improve cross-system alignment, indicating that the disagreement is structural rather than scale-induced. The correlation patterns are stable over time (2015–2020), implying that the weak concordance between databases is not confined to a single period or dataset revision.

Taken together, these findings provide robust evidence that the two ESG databases are not interchangeable either in absolute magnitude or in relative firm ranking, even after controlling for scaling, year, and dimension effects.

Year	Dimension	Spearman (Raw)	Spearman (Percentile)	Pearson (Raw)	Pearson (Percentile)	N
2015	ESG	−0.002	−0.002	−0.013	−0.002	284
2015	E	0.050	0.050	−0.071	0.050	284
2015	S	0.034	0.034	−0.004	0.034	284
2015	G	0.018	0.018	0.007	0.018	284
2016	ESG	0.035	0.035	−0.029	0.035	296
2016	E	0.009	0.009	−0.083	0.009	296
2016	S	−0.042	−0.042	−0.035	−0.042	296
2016	G	0.075	0.075	0.053	0.075	296
2017	ESG	−0.004	−0.004	−0.024	−0.004	291
2017	E	0.070	0.070	−0.076	0.070	291
2017	S	−0.035	−0.035	−0.072	−0.035	291
2017	G	0.150	0.150	0.145	0.150	291
2018	ESG	0.011	0.011	−0.009	0.011	297
2018	E	0.117	0.117	0.085	0.117	297
2018	S	0.024	0.024	−0.015	0.024	297
2018	G	0.194	0.194	0.173	0.194	297
2019	ESG	0.065	0.065	0.075	0.065	298
2019	E	0.221	0.221	0.193	0.221	298
2019	S	−0.038	−0.038	−0.092	−0.038	298
2019	G	0.196	0.196	0.173	0.196	298
2020	ESG	0.079	0.079	0.073	0.079	317
2020	E	0.077	0.077	0.040	0.077	317
2020	S	−0.084	−0.084	−0.062	−0.084	317
2020	G	0.007	0.007	−0.017	0.007	317

Appendix C. Robustness Test: Firm-Level Mean Difference Analyses

Panel A reports firm-level paired sample tests comparing CSMAR and CNRDS ESG ratings. The paired-sample t-tests and Wilcoxon signed-rank tests consistently reject the null hypothesis of no mean difference between CSMAR and CNRDS ratings ($p < 0.001$). On average, CSMAR assigns significantly higher overall ESG and Environmental scores, while CNRDS provides higher Governance ratings. Specifically, the mean ESG score under CSMAR exceeds that of CNRDS by roughly 20 points, with similar upward bias in the Environmental pillar (≈11.7 points). Conversely, the Social and Governance dimensions show negative gaps (−3.2 and −15.5), implying that CNRDS rates firms higher in governance-related aspects. These results confirm that the two systems differ substantially in absolute score levels and indicator weighting, leading to systematic directional bias. Panel A reports firm-level paired-sample tests comparing CSMAR and CNRDS ESG ratings after percentile standardization. After percentile standardization, both databases are transformed to a uniform [0,1] scale, eliminating magnitude differences by construction. As expected, the mean differences between CSMAR and CNRDS vanish ($p \approx 1.0$), and none of the dimensions show significant t or Wilcoxon statistics. This result, however, does not imply convergence or consistency across databases—it simply reflects the removal of absolute scale differences. Complementary rank-based and concordance analyses (e.g., Spearman correlation, ICC, CCC) still reveal weak or near-zero agreement (<0.1), confirming persistent divergence in firm-level rankings and underlying rating logic.

Together, these tests provide robust evidence that the observed inconsistencies between CNRDS and CSMAR are not artifacts of scaling or distributional form. The strong mean-level gaps in raw scores (Panel A) and the lack of alignment even after percentile normalization (Panel B) jointly indicate that the two systems differ fundamentally in methodology, weighting structure, and indicator aggregation. Hence, the lack of interchangeability between the two ESG databases is robust to standardization methods.

Panel A. Paired t-test and Wilcoxon Test Results (Raw Scores)

Dimension	Mean(CSMAR − CNRDS)	Std. Dev.	Paired t-stat	t-test p-value	Wilcoxon p-value
ESG	20.047	11.186	75.672	<0.001	5.9e−279
E	11.710	17.394	28.427	<0.001	1.2e−149
S	−3.182	20.044	−6.702	<0.001	5.2e−23
G	−15.545	17.616	−37.260	<0.001	2.3e−170

Panel B. Paired t-test and Wilcoxon Test Results (Percentile Standardized Scores)

Dimension	Mean(CSMAR − CNRDS)	Std. Dev.	Paired t-stat	t-test p-value	Wilcoxon p-value
ESG	4.98e−19	0.402	5.23e−17	1.000	0.468
E	−1.99e−18	0.389	−2.16e−16	1.000	0.277
S	4.48e−18	0.413	4.58e−16	1.000	0.836
G	1.25e−18	0.386	1.36e−16	1.000	0.570

Appendix D. Robustness Test: Year-by-Year Firm-Level Differences between CNRDS and CSMAR ESG Ratings

This appendix presents a year-by-year firm-level comparison between CNRDS and CSMAR ESG ratings from 2015 to 2020 as a robustness check. Two panels are reported: Panel A uses raw ESG scores, while Panel B applies percentile-standardized scores to control for scale differences. Both analyses employ paired-sample t-tests and Wilcoxon signed-rank tests to assess whether score differences are statistically significant.

Panel A shows that raw-score gaps between the two systems are large, systematic, and highly significant ($p < 0.001$). From 2015–2019, CSMAR assigns higher ESG and Environmental (E) scores, while CNRDS assigns higher Social (S) and Governance (G) scores, with average differences of about +20, +6, −10, and −21 points, respectively. In 2020, this pattern reverses—CSMAR's S and G scores surge above CNRDS—suggesting a methodological realignment rather than random variation. Overall, the raw-score results confirm that discrepancies between the two databases are persistent, directional, and structural. Panel B repeats the analysis using percentile-standardized scores. After normalization, mean differences vanish and statistical tests fail to reject equality ($p \approx 1.0$), implying that level disparities arise mainly from scaling and weighting schemes. However, as shown in Appendix A.16, rank-order correlations remain weak, particularly for S and G, indicating continued methodological divergence even after normalization.

Together, the two panels show that while CNRDS and CSMAR measure similar ESG constructs, they operate on different scales and ranking logics. The 2020 reversal further suggests that part of the inconsistency stems from changes in the CSMAR framework, confirming that the two systems' scores are not directly interchangeable, even after statistical adjustment.

Panel A. Raw Scores

Dimension	Year	N	Mean(CSMAR–CNRDS)	Std.Dev	t-stat	t-test p-value	Wilcoxon p-value
ESG	2015	284	19.746	10.574	31.470	2.02E−94	1.88E−46
ESG	2016	296	19.088	10.969	29.939	2.05E−91	3.33E−47
ESG	2017	291	19.106	11.292	28.863	2.97E−87	9.41E−47
ESG	2018	297	19.187	11.285	29.301	1.75E−89	1.65E−47
ESG	2019	298	21.267	11.165	32.882	5.41E−101	9.33E−49
ESG	2020	317	21.736	11.534	33.553	3.53E−106	5.15E−52
E	2015	284	7.245	12.005	10.170	6.54E−21	5.15E−22
E	2016	296	6.871	11.342	10.423	7.31E−22	5.78E−22
E	2017	291	6.819	12.760	9.116	1.33E−17	1.95E−19
E	2018	297	4.863	11.677	7.178	5.74E−12	1.23E−14
E	2019	298	5.375	12.138	7.645	2.92E−13	1.80E−16
E	2020	317	37.088	15.857	41.641	2.40E−130	3.29E−53
S	2015	284	−10.044	12.109	−13.978	4.06E−34	7.68E−29
S	2016	296	−10.574	13.006	−13.988	1.87E−34	1.20E−29
S	2017	291	−10.851	13.404	−13.810	1.10E−33	4.89E−29
S	2018	297	−10.833	13.035	−14.322	1.04E−35	1.60E−30
S	2019	298	−7.949	13.679	−10.032	1.38E−20	7.29E−18
S	2020	317	28.560	15.543	32.715	1.76E−103	7.98E−52
G	2015	284	−21.067	12.191	−29.123	3.92E−87	4.71E−48
G	2016	296	−21.056	12.273	−29.517	4.76E−90	9.36E−50
G	2017	291	−20.965	11.679	−30.623	7.14E−93	3.72E−49
G	2018	297	−20.935	12.263	−29.420	7.18E−90	7.21E−50
G	2019	298	−22.716	11.761	−33.341	2.12E−102	1.65E−50
G	2020	317	11.316	14.064	14.326	3.26E−36	9.79E−30

Panel B. Percentile-Standardized Scores

Dimension	Year	N	Mean(CSMAR–CNRDS)	Std.Dev	t-stat	t-test p-value	Wilcoxon p-value
ESG	2015	284	0.0000	0.4094	0.0000	1.0	0.808
ESG	2016	296	0.0000	0.4017	0.0000	1.0	0.780
ESG	2017	291	0.0000	0.4097	0.0000	1.0	0.772
ESG	2018	297	0.0000	0.4066	0.0000	1.0	0.785
ESG	2019	298	0.0000	0.3955	0.0000	1.0	0.836
ESG	2020	317	0.0000	0.3924	0.0000	1.0	0.640
E	2015	284	0.0000	0.3983	0.0000	1.0	0.611
E	2016	296	0.0000	0.4071	0.0000	1.0	0.567
E	2017	291	0.0000	0.3943	0.0000	1.0	0.543
E	2018	297	0.0000	0.3843	0.0000	1.0	0.604
E	2019	298	0.0000	0.3609	0.0000	1.0	0.859
E	2020	317	0.0000	0.3928	0.0000	1.0	0.872
S	2015	284	0.0000	0.4018	0.0000	1.0	0.966
S	2016	296	0.0000	0.4172	0.0000	1.0	0.992
S	2017	291	0.0000	0.4160	0.0000	1.0	0.770
S	2018	297	0.0000	0.4040	0.0000	1.0	0.960
S	2019	298	0.0000	0.4166	0.0000	1.0	0.906
S	2020	317	0.0000	0.4256	0.0000	1.0	0.950
G	2015	284	0.0000	0.4051	0.0000	1.0	0.830
G	2016	296	0.0000	0.3932	0.0000	1.0	0.758
G	2017	291	0.0000	0.3769	0.0000	1.0	0.909
G	2018	297	0.0000	0.3671	0.0000	1.0	0.875
G	2019	298	0.0000	0.3666	0.0000	1.0	0.720
G	2020	317	0.0000	0.4075	0.0000	1.0	0.818

Appendix E. Robustness Test: Industry-Level Differences between CNRDS and CSMAR ESG Ratings

This appendix examines industry-level differences between the CNRDS and CSMAR ESG rating systems using firm-level raw scores and percentile-standardized scores. The analysis aggregates firms by their primary industry classification and reports mean differences, standard deviations, and the results of paired-sample t-tests and Wilcoxon signed-rank tests for each ESG dimension. Panel A summarizes results based on raw scores, and Panel B presents results after percentile standardization.

Panel A reports the mean difference (CSMAR − CNRDS) in raw ESG scores across major industries. Positive values indicate that CSMAR assigns higher scores; negative values imply higher scores under Cardall major sectors exhibit statistically significant differences ($p < 0.001$), confirming systematic and directional gaps between the two databases. Collectively, Panel A demonstrates that industry-level discrepancies are not random but structurally consistent: CSMAR is systematically higher in E and overall ESG; CNRDS is systematically higher in S and G; and the magnitude varies by sector, being largest in finance, technology, and extractive industries.

Panel B reports result after transforming firm-level scores within each year and industry into percentile ranks to remove scale effects. After normalization, mean differences converge to zero ($|\text{mean diff}| < 0.001$ across all industries and dimensions), and both t-tests and Wilcoxon tests fail to reject the null of equality ($p \approx 1.0$). This outcome indicates that once scale and dispersion differences are removed, industry-level bias disappears, confirming that the raw-score discrepancies arise from differences in scoring range and weighting schemes, rather than in firms' relative positions within an industry. Nevertheless, as shown in earlier appendices (A.C- A.D), even though the level bias is removed, rank-order correlations remain weak, particularly in the Social and Governance dimensions. This finding reinforces that the two systems apply methodologically distinct evaluation logics across industries, rather than mere numeric rescaling.

The industry-level comparison highlights systematic structural differences between the two rating systems: CSMAR's methodology produces uniformly higher environmental and aggregate ESG scores, while CNRDS emphasizes social and governance strength. These differences are consistent across industries and highly significant, suggesting methodological divergence rather than noise. After percentile transformation, discrepancies vanish statistically but persist conceptually indicating that the two databases, though correlated in direction, remain non-interchangeable in empirical analysis.

Panel A. Industry-Level Differences in Raw ESG Scores

Dimension	Top Industries with Largest Absolute Differences	n	Mean Diff (CSMAR − CNRDS)	Std. Dev.	*t*-Statistic	*p*-Value	Wilcoxon *p*-Value
ESG	Finance Industry	325	24.81	8.21	54.50	3.13×10^{-165}	5.29×10^{-55}
	Software & IT Services	118	24.32	7.96	33.19	2.29×10^{-61}	4.32×10^{-21}
	Water Production & Supply	65	22.99	10.39	17.85	1.55×10^{-26}	3.33×10^{-12}
	Health	12	22.77	11.26	7.00	2.25×10^{-5}	4.88×10^{-4}
	Manufacturing	593	18.87	11.56	39.76	2.82×10^{-169}	1.82×10^{-92}
E	Health	12	28.41	20.22	4.87	4.97×10^{-4}	4.88×10^{-4}
	IT Services	118	17.32	13.58	13.86	1.98×10^{-26}	1.13×10^{-19}
	Finance	325	16.53	16.09	18.52	9.48×10^{-53}	1.94×10^{-49}
	Real Estate	92	17.29	9.43	17.58	5.20×10^{-31}	8.14×10^{-17}
S	Construction	74	−19.14	18.00	−9.14	9.98×10^{-14}	3.41×10^{-10}
	Mining	85	−16.12	18.26	−8.14	3.23×10^{-12}	4.61×10^{-9}
	Utilities	10	−17.28	10.20	−5.36	4.57×10^{-4}	1.95×10^{-3}
	Agriculture	13	−15.45	19.67	−2.83	1.51×10^{-2}	1.71×10^{-2}
G	Real Estate	92	−29.08	17.07	−16.34	8.33×10^{-29}	4.14×10^{-16}
	Construction	74	−23.11	18.28	−10.87	6.39×10^{-17}	2.34×10^{-11}
	Finance	325	−17.03	16.96	−18.11	4.13×10^{-51}	3.53×10^{-38}
	Manufacturing	593	−10.08	17.37	−14.14	2.68×10^{-39}	2.50×10^{-33}

Panel B. Industry-Level Differences in Percentile-Standardized ESG Scores

Dimension	Top Industries with Largest Absolute Differences	n	Mean Diff (CSMAR − CNRDS)	Std. Dev.	*t*-Statistic	*p*-Value	Wilcoxon *p*-Value
ESG	Finance Industry	325	4.10×10^{-18}	0.37	2.00×10^{-16}	1.00	0.85
	Manufacturing	593	-3.00×10^{-18}	0.41	-1.79×10^{-16}	1.00	0.94
	Real Estate	92	1.21×10^{-18}	0.40	2.88×10^{-17}	1.00	0.85
	IT Services	118	-1.36×10^{-17}	0.27	-5.56×10^{-16}	1.00	0.94
	Real Estate	92	8.45×10^{-18}	0.36	2.25×10^{-16}	1.00	0.88
E	Manufacturing	593	-2.43×10^{-18}	0.40	-1.47×10^{-16}	1.00	0.58
	Manufacturing	593	0.00	0.42	0.00	1.00	0.89
	Finance	325	3.42×10^{-18}	0.40	1.55×10^{-16}	1.00	0.74
	Real Estate	92	9.65×10^{-18}	0.45	2.05×10^{-16}	1.00	0.96
S	Finance	325	6.83×10^{-19}	0.41	3.01×10^{-17}	1.00	0.96
	Manufacturing	593	3.74×10^{-19}	0.38	2.42×10^{-17}	1.00	0.93
	Real Estate	92	-5.43×10^{-18}	0.42	-1.23×10^{-16}	1.00	0.91
	Finance Industry	325	4.10×10^{-18}	0.37	2.00×10^{-16}	1.00	0.85
G	Manufacturing	593	-3.00×10^{-18}	0.41	-1.79×10^{-16}	1.00	0.94
	Real Estate	92	1.21×10^{-18}	0.40	2.88×10^{-17}	1.00	0.85
	IT Services	118	-1.36×10^{-17}	0.27	-5.56×10^{-16}	1.00	0.94
	Real Estate	92	8.45×10^{-18}	0.36	2.25×10^{-16}	1.00	0.88

Table 1. Annual Comparison of ESG Scores: CNRDS vs. CSMAR

Table A1 summarizes the annual averages, year-by-year differences, and summary statistics of ESG and sub-pillar (Environmental, Social, and Governance) scores derived from the CNRDS and CSMAR databases for Chinese A-share listed firms during 2015–2020. Panel A shows that CSMAR consistently assigns higher ESG scores than CNRDS, with an average gap of about 20 points across most years. Both datasets exhibit upward trends, reflecting improving ESG performance among Chinese firms, though the magnitude and timing of changes differ substantially across sub-dimensions.

Specifically, the Environmental scores in CSMAR remain relatively stable between 2015 and 2019 before surging sharply in 2020, suggesting a major methodological revision, whereas CNRDS shows a gradual and continuous increase throughout the sample period. The Social pillar remains stable at around 28 points in CNRDS but rises dramatically from approximately 17 to over 53 in CSMAR in 2020, indicating a recalibration in data weighting or coverage. For the Governance pillar, CNRDS reports steadily improving scores, while CSMAR's values remain very low before 2019 and then jump sharply to nearly match CNRDS levels by 2020.

Panel B presents the annual differences (CNRDS minus CSMAR), which confirm persistent level discrepancies and reveal the 2020 divergence driven by CSMAR's Environmental and Social components. Panel C reports the summary statistics, showing that CSMAR exhibits higher mean levels and significantly larger standard deviations, particularly for the Environmental and Social dimensions. This greater variability implies that CSMAR's rating framework underwent structural adjustments, possibly linked to expanded coverage or re-scaling of its ESG indicators. Overall, the results demonstrate that while both databases capture similar upward trends in ESG performance, their rating levels and volatility differ markedly due to heterogeneous methodologies.

Panel A. Annual Comparison of ESG and Sub-Pillar Scores: CNRDS vs. CSMAR (2015–2020)

Year	ESG Score (CNRDS / CSMAR)	Environmental (CNRDS / CSMAR)	Social (CNRDS / CSMAR)	Governance (CNRDS / CSMAR)
2015	26.425 / 46.463	8.735 / 17.395	26.674 / 16.572	36.874 / 12.496
2016	26.155 / 46.411	9.765 / 17.228	27.964 / 16.753	35.656 / 12.430
2017	28.299 / 46.693	11.140 / 17.048	28.625 / 16.978	37.228 / 12.666
2018	29.184 / 48.763	11.329 / 17.523	28.849 / 17.247	38.437 / 13.993
2019	28.861 / 49.759	13.329 / 18.208	28.409 / 17.606	39.261 / 13.945
2020	30.369 / 51.256	14.639 / 51.806	29.795 / 53.670	39.246 / 48.094

Panel B. Annual Differences (CNRDS – CSMAR)

Year	ESG Score Diff	Environmental Diff	Social Diff	Governance Diff
2015	–20.038	–8.660	+10.102	+24.378
2016	–20.256	–7.463	+11.211	+23.226
2017	–18.394	–5.908	+11.647	+24.562
2018	–19.579	–6.194	+11.602	+24.444
2019	–20.898	–4.879	+10.803	+25.316
2020	–20.887	–37.167	–23.875	–8.848

Panel C. Summary Statistics (2015–2020)

Statistic	ESG Score (CNRDS / CSMAR)	Environmental (CNRDS / CSMAR)	Social (CNRDS / CSMAR)	Governance (CNRDS / CSMAR)
Mean	28.215 / 48.224	11.489 / 23.201	28.386 / 23.138	37.784 / 18.937
Std. Dev.	1.640 / 2.028	2.191 / 14.019	1.036 / 14.962	1.443 / 14.301

Table 2. Industry-Level Comparison of ESG Scores: CNRDS vs. CSMAR

This table compares the average ESG scores by industry between the CNRDS and CSMAR databases for Chinese A-share listed firms during 2015–2020. Industries are ranked based on the CNRDS mean ESG score. Across all industries, CSMAR assigns consistently higher ratings, with an average gap exceeding 19 points. The discrepancy is particularly large in information technology, finance, and public utility sectors, reflecting systematic methodological and scaling differences between the two databases.

Panel A presents the average ESG scores of 16 industries based on CNRDS and CSMAR databases from 2015 to 2020. Industries are ranked by their CNRDS mean ESG scores. Across all industries, CSMAR assigns systematically higher ESG ratings—by an average of approximately 19–20 points—indicating a broader scoring scale or methodological generosity. The largest discrepancies occur in knowledge-intensive sectors such as IT, finance, and public utilities, where CSMAR's 2020 re-scaling likely magnified score dispersion across industries.

Panel B summarizes the overall statistical characteristics of both datasets. CNRDS exhibits a lower mean but higher standard deviation, suggesting greater variation across industries. In contrast, CSMAR shows higher mean ESG levels with smaller dispersion, implying a more compressed but upward-shifted scoring distribution. Together, these results highlight the systematic level bias and scaling differences between the two ESG rating systems.

Panel A. Industry-Level Averages and Differences

Rank	Industry	CNRDS Mean	CSMAR Mean	Difference (CNRDS – CSMAR)
1	Construction	35.277	47.803	–12.526
2	Mining Industry	34.834	50.228	–15.394
3	Raw Materials Manufacturing	32.214	47.893	–15.679
4	Health and Social Work	31.345	43.438	–12.093
5	Leasing and Business Services	30.627	48.250	–17.623
6	Electricity, Heat, Gas and Water Supply	30.502	50.500	–19.998
7	Agriculture, Forestry, Animal Husbandry and Fishery	30.137	50.976	–20.839
8	Manufacturing	30.118	48.866	–18.748
9	Transportation, Storage and Postal Services	27.908	49.743	–21.835
10	Other Services	26.644	46.566	–19.922
11	Real Estate	26.129	45.135	–19.007
12	Water Production and Supply	25.867	50.729	–24.863
13	Finance Industry	25.768	50.466	–24.699
14	Water Conservancy, Environment and Public Facilities Management	23.602	46.075	–22.473
15	Health	23.039	44.498	–21.459
16	Software and Information Technology Services	22.294	47.043	–24.749

Panel B. Overall Statistics Summary

Dataset	Mean	Std. Dev.	Count
CNRDS	28.213	4.842	79
CSMAR	48.199	3.214	79

Table 3. Summary Statistics for ESG Ratings: CNRDS vs. CSMAR

This table reports summary statistics for ESG and its three sub-pillars (E, S, and G) for both CNRDS and CSMAR databases over 2015–2020 (N = 1,783 firm-year observations). Panel A displays the original (RAW) scores; Panel B presents percentile-standardized (PCT) scores; and Panel C provides min–max normalized (MM) scores scaled between 0 and 1. Across all panels, CSMAR systematically assigns higher ESG scores than CNRDS, both in raw values and normalized scales. Under RAW scores, CSMAR's average ESG level (48.7) exceeds CNRDS's (28.7) by roughly 20 points. Environmental and Social scores also display large dispersion in the CSMAR dataset (standard deviations ≈14–15), indicating broader cross-sectional variation or potential methodological recalibration after 2019.

After normalization, both datasets converge in rank-based metrics (PCT), implying consistent firm ordering but distinct scoring magnitudes. Under MM scaling, CSMAR retains higher normalized means (e.g., ESG = 0.43 vs. 0.36), highlighting its tendency toward an inflated upper range. These results confirm that inter-database discrepancies largely originate from differences in rating scale and weighting, rather than from underlying firm-level ESG patterns.

Panel A. Raw ESG Scores

Dimension	System	Count	Mean	Std. Dev.	Min	25%	Median	75%	Max
ESG	CSMAR	1783	48.743	5.771	30.875	44.875	48.000	52.000	71.875
ESG	CNRDS	1783	28.695	9.834	3.502	21.773	26.993	34.185	74.657
E	CSMAR	1783	23.769	14.174	9.500	16.000	17.750	22.125	82.390
E	CNRDS	1783	12.059	12.118	1.039	3.178	8.452	16.690	70.494
S	CSMAR	1783	24.023	14.845	0.000	16.375	17.750	20.125	76.520
S	CNRDS	1783	27.204	12.841	1.201	17.732	26.319	35.061	76.562
G	CSMAR	1783	19.235	13.487	6.875	11.875	13.625	16.375	70.690
G	CNRDS	1783	34.780	12.080	0.074	26.173	34.040	41.609	82.967

Panel B. Percentile (PCT) Standardization

Dimension	System	Count	Mean	Std. Dev.	Min	25%	Median	75%	Max
ESG	CSMAR	1783	0.502	0.289	0.003	0.252	0.502	0.752	1.000
ESG	CNRDS	1783	0.502	0.289	0.003	0.252	0.502	0.751	1.000
E	CSMAR	1783	0.502	0.289	0.003	0.251	0.503	0.749	1.000
E	CNRDS	1783	0.502	0.289	0.003	0.247	0.502	0.751	1.000
S	CSMAR	1783	0.502	0.289	0.003	0.252	0.504	0.752	1.000
S	CNRDS	1783	0.502	0.289	0.003	0.252	0.502	0.751	1.000
G	CSMAR	1783	0.502	0.289	0.003	0.249	0.502	0.752	1.000
G	CNRDS	1783	0.502	0.289	0.003	0.252	0.502	0.751	1.000

Panel C. Min–Max (MM) Standardization

Dimension	System	Count	Mean	Std. Dev.	Min	25%	Median	75%	Max
ESG	CSMAR	1783	0.430	0.175	0.000	0.307	0.409	0.532	1.000
ESG	CNRDS	1783	0.358	0.167	0.000	0.242	0.325	0.450	1.000
E	CSMAR	1783	0.395	0.173	0.000	0.280	0.365	0.483	1.000
E	CNRDS	1783	0.173	0.193	0.000	0.032	0.113	0.246	1.000
S	CSMAR	1783	0.559	0.154	0.000	0.471	0.551	0.661	1.000
S	CNRDS	1783	0.365	0.192	0.000	0.224	0.351	0.478	1.000
G	CSMAR	1783	0.433	0.169	0.000	0.319	0.421	0.536	1.000
G	CNRDS	1783	0.404	0.169	0.000	0.289	0.389	0.502	1.000

Table 4. Industry-Level ESG Rating Discrepancies Between CNRDS and CSMAR

This table reports the mean differences in ESG, Environmental (E), Social (S), and Governance (G) scores between the CSMAR and CNRDS databases across major industries. For each industry, we compute the average difference (CSMAR − CNRDS), the corresponding standard deviation, and perform both parametric (paired t-test) and nonparametric (Wilcoxon signed-rank test) significance tests. A positive mean difference (CSMAR − CNRDS) indicates that CSMAR assigns a higher score than CNRDS for that industry, whereas a negative value indicates that CNRDS assigns the higher score. Significant results ($p < 0.05$) are highlighted, showing where systematic deviations exist between the two rating systems. Industries with fewer than two firm-level observations ($n < 2$) were excluded from the analysis to ensure statistical validity of the t-tests and Wilcoxon signed-rank tests.

Panel A examines the aggregate ESG score differences across industries. Results show that CSMAR tends to assign significantly higher scores in sectors such as *Smelting and Rolling of Ferrous Metals* and *Support Activities for Mining*, while CNRDS scores are higher in service-oriented industries such as *Postal Services* and *Textile, Apparel and Accessories*. The large negative differences (e.g., –0.55 for Postal Services, $p < 0.001$) indicate that CSMAR assigns significantly lower scores than CNRDS in these sectors. Panel B focuses on the environmental dimension. The strongest discrepancies are observed in *Support Activities for Mining* and *Smelting and Rolling of Ferrous Metals*, where CSMAR assigns significantly higher E scores ($p < 0.01$). In contrast, CNRDS produces lower values for *Agricultural Processing* and *Postal Services*, showing a more conservative environmental evaluation framework. Panel C shows that the social dimension exhibits the largest and most consistent gaps across industries. Service-related sectors (*Postal Services, Textile, Apparel, and Accessories*) display significantly lower CSMAR scores relative to CNRDS ($p < 0.001$), whereas industrial sectors (*News and Publishing, Water Transport*) receive higher ratings in CSMAR. This indicates differing weighting on employee welfare, community engagement, and disclosure-based indicators across systems. Panel D compares governance scores. The results again reveal notable heterogeneity: industries such as *Gas Production and Supply* and *Culture and Arts* show higher governance ratings under CSMAR, while *Construction* and *Postal Services* have significantly lower scores ($p < 0.01$). These findings suggest varying interpretations of board structure, ownership concentration, and regulatory disclosure indicators across databases.

Overall, these industry-level tests confirm that systematic cross-database discrepancies are not uniform across sectors. CSMAR's methodology tends to emphasize heavy industries and energy-related sectors, while CNRDS provides relatively higher ratings for consumer and service sectors. This pattern reinforces the firm-level evidence of methodological divergence and highlights that ESG database selection can materially affect sectoral inferences and portfolio-level ESG composition.

Panel A. ESG Overall Score Differences

Industry	n	Mean Diff (CSMAR – CNRDS)	t-stat	t-p	Wilcoxon-p	Significance
Postal Services	11	–0.547	–7.09	0.00003	0.0019	***
Manufacture of Rubber & Plastic Products	2	–0.520	–1.44	0.385	–	ns
Manufacture of Chemicals & Chemical Products	31	–0.395	–5.95	0.000002	0.00001	***
Processing of Agricultural & Sideline Food	20	–0.387	–6.31	0.000005	0.000004	***
Support Activities for Mining	11	+0.377	+4.32	0.0015	0.0049	**
Scientific Research & Technical Services	2	+0.367	+0.85	0.553	–	ns
Textile, Apparel & Accessories	8	–0.321	–3.80	0.0067	0.0078	**
Comprehensive	8	–0.298	–2.31	0.054	0.078	*
Smelting & Rolling of Ferrous Metals	26	+0.291	+4.03	0.00046	0.00047	**

Panel B. Environmental (E) Dimension

Industry	n	Mean Diff (CSMAR – CNRDS)	t-stat	t-p	Wilcoxon-p	Significance
Processing of Agricultural & Sideline Food	20	–0.498	–7.29	0.0000006	0.000013	***
Scientific Research & Technical Services	2	+0.483	+2.23	0.268	–	ns
Support Activities for Mining	11	+0.466	+4.95	0.0006	0.0029	**
Comprehensive	8	–0.403	–3.47	0.0103	0.0280	*
Postal Services	11	–0.280	–2.53	0.030	0.032	*
Smelting & Rolling of Ferrous Metals	26	+0.272	+3.18	0.0039	0.0073	**
Processing of Petroleum, Coking & Nuclear Fuel	7	–0.290	–1.89	0.108	0.156	ns
Water Production & Supply	2	–0.342	–13.00	0.049	–	*

Panel C. Social (S) Dimension

Industry	n	Mean Diff (CSMAR – CNRDS)	t-stat	t-p	Wilcoxon-p	Significance
Postal Services	11	–0.560	–8.85	0.000005	0.00098	***
Textile, Apparel & Accessories	8	–0.375	–6.22	0.00043	0.0078	***
Manufacture of Chemicals & Chemical Products	31	–0.346	–4.87	0.00003	0.00020	***
News & Publishing	9	+0.359	+4.23	0.0029	0.0117	**
Water Transport	29	+0.326	+4.64	0.00007	0.00026	***
Manufacture of Rubber & Plastic Products	2	–0.390	–1.34	0.407	–	ns

Panel D. Governance (G) Dimension

Industry	n	Mean Diff (CSMAR – CNRDS)	t-stat	t-p	Wilcoxon-p	Significance
Construction	6	–0.453	–3.55	0.016	0.043	*
Postal Services	11	–0.444	–4.85	0.00068	0.00195	**
Culture & Arts	6	+0.434	+3.09	0.027	0.062	*
Gas Production & Supply	2	+0.572	+12.53	0.051	–	*
Smelting & Rolling of Ferrous Metals	26	+0.388	+2.42	0.052	0.109	*
Rail Transport	9	+0.348	+2.86	0.021	0.039	*

Table 5. Agreement Analysis Between CNRDS and CSMAR : Regression, ICC, and CCC Evidence

To formally quantify the divergence between CNRDS and CSMAR, we conduct a series of agreement tests: (i) level regression, (ii) intraclass correlation (ICC), and (iii) concordance correlation (CCC).

Panel A estimates the linear relationship between the two systems' raw scores using the model $\text{CSMAR} = \alpha + \beta \times \text{CNRDS} + \varepsilon$. Under perfect agreement, we expect α = 0 and β = 1. However, all four dimensions significantly reject this null hypothesis ($p < 0.001$), with β values close to zero, confirming that CSMAR scores are largely independent of CNRDS levels and operate on a much higher absolute scale. Panel B evaluates the reliability of the two systems using the intraclass correlation coefficient (ICC), which measures the degree of absolute agreement between paired observations. ICC values below 0.1 across all dimensions indicate very low reliability and poor consistency between the two ESG systems. The Environmental dimension performs marginally better (ICC = 0.0933) but still falls within the "poor" reliability range (<0.5), confirming that firm-level score variation captured by one database explains little of the variation in the other. Panel C applies Lin's Concordance Correlation Coefficient (CCC) to jointly assess both correlation and deviation from the 45° line of perfect agreement. All CCC values are close to zero ($-0.04 \leq \text{CCC} \leq 0.09$), with 95% confidence intervals including zero, implying no meaningful concordance between CNRDS and CSMAR ratings. The Environmental pillar (CCC = 0.09) again shows the highest—though still "very poor"—agreement level. These findings highlight that even when firm rankings partially align, their absolute scores differ significantly, suggesting distinct indicator weighting and normalization methods between databases. Panel D summarizes the outcomes across all three quantitative metrics—level regression, ICC, and CCC. The extremely low R^2, ICC, and CCC values collectively confirm that the two rating systems are not interchangeable at the firm level. While both datasets exhibit similar macro-level ESG trends (as observed in previous tables and figures), their score magnitudes and scaling structures diverge sharply. These findings reinforce the descriptive evidence of systematic bias and limited cross-database consistency, demonstrating that methodological heterogeneity—not random variation—is the dominant driver of discrepancies between CNRDS and CSMAR ESG ratings.

The regression results show that the estimated slopes are close to zero, strongly rejecting the null hypothesis of perfect agreement (α=0, β=1). Both ICC and CCC analyses yield coefficients below 0.1, indicating *poor reliability* and *very poor concordance* across all ESG dimensions. These results suggest that while both databases capture upward ESG trends at the aggregate level, their firm-level score assignments differ systematically due to variations in scoring methodology, normalization procedures, and post-2019 recalibration within the CSMAR system. Consequently, CNRDS and CSMAR ESG data should not be used interchangeably in firm-level regression analyses without explicit harmonization or standardization adjustments.

Panel A. Level Regression Analysis

Dimension	N	Intercept (α)	Slope (β)	R^2	p(α=0)	p(β=1)	Reject Perfect Agreement
ESG	1783	48.019	0.025	0.0018	0.000	0.000	True
E	1783	21.913	0.154	0.0173	0.000	0.000	True
S	1783	25.384	–0.050	0.0019	0.000	0.000	True
G	1783	17.152	0.060	0.0029	0.000	0.000	True

Panel B. Intraclass Correlation Coefficients (ICC)

Dimension	N	ICC (Absolute)	Reliability	Between-Subject Var	Error Var
ESG	1783	0.0092	Poor	67.445	62.569
E	1783	0.0933	Poor	196.478	151.268
S	1783	0.0000	Poor	184.376	200.876

G	1783	0.0307	Poor	172.651	155.169

Panel C. Concordance Correlation Coefficients (CCC)

Dimension	N	CCC	95% CI (Low–High)	Pearson r	Bias Component	Agreement Level
ESG	1783	0.0092	[–0.0008, 0.0203]	0.0430	1.511	Very Poor
E	1783	0.0932	[0.0566, 0.1292]	0.132	0.566	Very Poor
S	1783	–0.0417	[–0.0845, 0.0043]	–0.043	0.051	Very Poor
G	1783	0.0307	[0.0091, 0.0578]	0.054	0.849	Very Poor

Panel D. Summary of Agreement Analysis

Dimension	N	β (Slope)	R^2	ICC	CCC	Reliability	Agreement Level
ESG	1783	0.025	0.0018	0.0092	0.0092	Poor	Very Poor
E	1783	0.154	0.0173	0.0933	0.0932	Poor	Very Poor
S	1783	–0.050	0.0019	0.0000	–0.0417	Poor	Very Poor
G	1783	0.060	0.0029	0.0307	0.0307	Poor	Very Poor

Table 6. Effect of Standardization on Cross-System ESG Agreement

Table 6 extends the cross-system comparison by examining whether standardization mitigates the observed discrepancies between CNRDS and CSMAR. Using both percentile (PCT) and min–max (MM) normalization, we re-estimate level regressions, intraclass correlations (ICC), and concordance correlations (CCC).

Panel A examines the effect of percentile standardization, which removes magnitude differences by ranking firms relative to others within each year. After percentile transformation, the magnitude discrepancies between databases are eliminated, but rank-order inconsistencies persist. Regression slopes remain close to zero and both ICC and CCC values fall below 0.11, indicating very poor concordance across all dimensions. The Governance pillar displays the relatively highest alignment (ICC ≈ 0.11), suggesting that percentile normalization marginally improves consistency in firm rankings but does not achieve meaningful convergence between the two systems. Panel B investigates the degree of agreement after applying min–max normalization, which rescales all ESG scores into the [0,1] interval. Even after this transformation, the two datasets remain poorly aligned. All regression slopes and concordance coefficients are near zero, confirming that simple rescaling does not reconcile the underlying methodological inconsistency. The Governance dimension again exhibits slightly higher consistency, while the Social dimension shows a negative association, reflecting opposite directional scoring patterns between CNRDS and CSMAR in evaluating firms' social performance. Panel C consolidates the results from all three standardization schemes—RAW, Percentile, and Min–Max—to evaluate whether data transformation improves cross-system consistency. The Environmental and Governance dimensions consistently show slightly stronger alignment than the ESG aggregate and the Social pillar, yet all coefficients remain below 0.11. These results confirm that scaling adjustments alone cannot harmonize the two datasets, as the observed discrepancies stem from deeper methodological and indicator-level differences rather than from numeric transformation. Consequently, firm-level comparisons across CNRDS and CSMAR remain unreliable even after normalization.

Results show that despite rescaling, cross-database agreement remains extremely weak (ICC, CCC < 0.11), with all dimensions significantly rejecting perfect alignment. The Governance and Environmental pillars display slightly improved correlation post-standardization, while the Social pillar even exhibits negative association, underscoring fundamental divergence in rating construction. Taken together, these results indicate that the inconsistency between CNRDS and CSMAR arises from structural differences in indicator selection, weighting, and aggregation methodology, rather than from superficial scaling or distributional variations.

Panel A. Percentile (PCT) Standardization

Dimension	N	Intercept (α)	Slope (β)	R^2	ICC	CCC	Agreement Level	Reject Perfect Agreement
ESG	1783	0.486	0.032	0.0010	0.0316	0.0315	Very Poor	True
E	1783	0.456	0.091	0.0083	0.0910	0.0910	Very Poor	True
S	1783	0.514	–0.025	0.0006	0.0000	–0.0246	Very Poor	True
G	1783	0.448	0.106	0.0113	0.1064	0.1063	Very Poor	True

Panel B. Min–Max (MM) Standardization

Dimension	N	Intercept (α)	Slope (β)	R^2	ICC	CCC	Agreement Level	Reject Perfect Agreement
ESG	1783	0.423	0.021	0.0004	0.0182	0.0182	Very Poor	True
E	1783	0.388	0.042	0.0022	0.0269	0.0269	Very Poor	True
S	1783	0.584	–0.069	0.0073	0.0000	–0.0515	Very Poor	True
G	1783	0.396	0.092	0.0084	0.0904	0.0904	Very Poor	True

Panel C. Comprehensive Comparison Across Methods

Dimension	Method	Slope (β)	R^2	ICC	CCC	Agreement Level	Reject Perfect Agreement
ESG	Raw	0.025	0.0018	0.009	0.009	Very Poor	True
ESG	PCT	0.032	0.0010	0.032	0.032	Very Poor	True
ESG	MM	0.021	0.0004	0.018	0.018	Very Poor	True
E	Raw	0.154	0.0173	0.093	0.093	Very Poor	True
E	PCT	0.091	0.0083	0.091	0.091	Very Poor	True
E	MM	0.042	0.0022	0.027	0.027	Very Poor	True
S	Raw	–0.050	0.0019	0.000	–0.042	Very Poor	True
S	PCT	–0.025	0.0006	0.000	–0.025	Very Poor	True
S	MM	–0.069	0.0073	0.000	–0.052	Very Poor	True
G	Raw	0.060	0.0029	0.031	0.031	Very Poor	True
G	PCT	0.106	0.0113	0.106	0.106	Very Poor	True
G	MM	0.092	0.0084	0.090	0.090	Very Poor	True

Table 7. Comparison of Min–Max Normalization Schemes between CNRDS and CSMAR

This table evaluates whether the discrepancies between CNRDS and CSMAR ESG ratings persist under different min–max normalization schemes. Specifically, we compare (1) direct firm-level min–max scaling, (2) industry-adjusted normalization (within each industry), and (3) year-adjusted normalization (within each year). Paired *t*-tests are conducted for each dimension (ESG, E, S, G) to examine whether the mean normalized scores differ significantly between the two databases.

Across all dimensions, CSMAR scores remain consistently higher than CNRDS, even after rescaling to a [0,1] range. The strongest differences appear in the Environmental (E) and Social (S) pillars, with mean gaps of approximately +0.22 and +0.19, respectively. The Governance (G) dimension shows smaller but still significant differences ($\approx$ +0.03 to +0.04). Only the industry-adjusted ESG comparison yields an insignificant result ($p = 0.12$), suggesting that sectoral heterogeneity accounts for part of the divergence. These results confirm that rescaling alone does not reconcile the two systems: while min–max normalization removes absolute scale bias, it does not address the underlying methodological inconsistencies in score construction. In particular, CSMAR appears to systematically assign higher normalized scores across most dimensions, indicating persistent upward shifts in its scaling framework relative to CNRDS.

This table compares different normalization levels for each ESG dimension. Asterisks denote significance levels (***$p < 0.001$, *$p < 0.01$, $p < 0.05$). Out of twelve paired comparisons, eleven exhibit statistically significant mean differences ($p < 0.05$). The only non-significant result appears in the industry-adjusted ESG normalization, implying that sector-specific scaling accounts for a small portion of divergence. Overall, this high proportion (91.7%) of significant outcomes confirms that the CNRDS–CSMAR discrepancy is robust across normalization schemes, with consistent directional bias toward higher CSMAR scores.

Dimension	Comparison	Mean Difference (CSMAR − CNRDS)	*t*-Statistic	*p*-Value	Significance
ESG	Firm-level (CSMAR_ESG_MM vs. CNRDS_ESG_MM)	0.0727	—	0.0000	***
	Industry-adjusted (CSMAR_ESG_ind_MM vs. CNRDS_ESG_ind_MM)	0.0132	—	0.1190	
	Year-adjusted (CSMAR_ESG_year_MM vs. CNRDS_ESG_year_MM)	0.0727	—	0.0000	***
E	Firm-level	0.2220	—	0.0000	***
	Industry-adjusted	0.1778	—	0.0000	***
	Year-adjusted	0.2220	—	0.0000	***
S	Firm-level	0.1939	—	0.0000	***
	Industry-adjusted	0.0552	—	0.0000	***
	Year-adjusted	0.1939	—	0.0000	***
G	Firm-level	0.0283	—	0.0000	***
	Industry-adjusted	0.0400	—	0.0000	***
	Year-adjusted	0.0283	—	0.0000	***

Table 8. Correlation Structure between CNRDS and CSMAR ESG Ratings

This table reports the pairwise Pearson correlations among min–max normalized ESG variables from both the CNRDS and CSMAR databases. Panel A presents the complete correlation matrix across all eight variables (overall ESG and the three sub-dimensions E, S, and G). Panel B summarizes the cross-system correlations for each dimension, along with mean and standard deviation statistics, to assess whether the two databases exhibit consistent scaling and co-movement patterns.

Panel A matrix includes all pairwise correlations among min–max normalized ESG variables from both databases. Within-system correlations (diagonal sub-blocks) are generally strong and positive, especially between overall ESG and the E or S pillars under CSMAR. Cross-system correlations (off-diagonal) are weak or near zero, indicating substantial divergence between the two rating systems. Panel B represents cross-system correlations between CNRDS and CSMAR. The correlations are consistently low ($|\rho| < 0.10$) across all dimensions, with the Social (S) pillar even showing a negative correlation ($\rho = -0.086$). Although both datasets show comparable means and standard deviations, their weak co-movements suggest fundamental methodological divergence rather than simple scaling differences. These results reinforce earlier findings that CNRDS and CSMAR capture ESG constructs with distinct evaluative logics and weighting schemes.

Together, the correlation evidence demonstrates that even after normalization, CNRDS and CSMAR ratings are largely uncorrelated at the firm level. CSMAR's internal consistency across E, S, and G dimensions remains high ($\rho \approx 0.6$–0.8), while CNRDS's internal structure is weaker and less cohesive. The lack of meaningful cross-system correlation indicates that the two ESG rating providers operate on largely independent evaluative frameworks, underscoring the non-interchangeability of their scores for empirical applications.

Panel A. Complete Correlation Matrix among Min–Max Normalized Variables

Variable	CSMAR_ESG_MM	CNRDS_ESG_MM	CSMAR_E_MM	CNRDS_E_MM	CSMAR_S_MM	CNRDS_S_MM	CSMAR_G_MM	CNRDS_G_MM
CSMAR_ESG_MM	1.0000	0.0198	0.7586	0.0079	0.6456	0.0291	0.6626	0.0111
CNRDS_ESG_MM	0.0198	1.0000	0.0638	0.5295	−0.0421	0.6325	0.0254	0.3718
CSMAR_E_MM	0.7586	0.0638	1.0000	0.0469	0.3149	0.0397	0.2880	0.0372
CNRDS_E_MM	0.0079	0.5295	0.0469	1.0000	0.0738	0.2206	−0.0057	−0.1226
CSMAR_S_MM	0.6456	−0.0421	0.3149	0.0738	1.0000	−0.0855	0.2908	−0.0964
CNRDS_S_MM	0.0291	0.6325	0.0397	0.2206	−0.0855	1.0000	0.0613	0.2140
CSMAR_G_MM	0.6626	0.0254	0.2880	−0.0057	0.2908	0.0613	1.0000	0.0916
CNRDS_G_MM	0.0111	0.3718	0.0372	−0.1226	−0.0964	0.2140	0.0916	1.0000

Panel B. Cross-System Correlation Comparison by Dimension

Dimension	CSMAR–CNRDS Correlation (ρ)	CSMAR Mean	CNRDS Mean	CSMAR Std	CNRDS Std
ESG	0.0198	0.4304	0.3578	0.1754	0.1671
E	0.0469	0.3951	0.1731	0.1734	0.1926
S	−0.0855	0.5588	0.3649	0.1540	0.1917
G	0.0916	0.4328	0.4045	0.1693	0.1693

Table 9. Quantile-Based Agreement between CNRDS and CSMAR ESG Ratings

This appendix evaluates the categorical and ordinal agreement between the CNRDS and CSMAR ESG rating systems by discretizing continuous scores into quintiles (five equal-frequency bins). For each ESG dimension—overall ESG, Environmental (E), Social (S), and Governance (G)—three complementary statistics are computed. The first measure captures the proportion of firms assigned to the same quintile by both systems, representing the level of exact agreement. The second measure, Cohen's Kappa (κ), adjusts for chance agreement to assess the true degree of categorical consistency. The third measure, Spearman's rank correlation (ρ), reflects the ordinal alignment between the two continuous score series.

Across all dimensions, exact agreement rates remain below 22%, meaning fewer than one in five firms fall into the same rating quintile in both databases. Cohen's κ values hover around zero ($-0.01 \leqslant \kappa \leqslant 0.03$), indicating no statistically meaningful categorical agreement beyond random chance. The Environmental (E) and Governance (G) dimensions show slightly higher ordinal correlation ($\rho \approx 0.11$), whereas the Social (S) dimension exhibits weakly negative association ($\rho = -0.06$), consistent with earlier results showing opposite directional scoring between the two systems. Overall, these findings corroborate the systematic misalignment between CNRDS and CSMAR observed in prior robustness and visual analyses. Even after converting continuous scores into comparable quintile-based ranks, cross-system consistency remains poor, reinforcing that the two rating frameworks apply fundamentally different scaling and weighting methodologies rather than random measurement noise.

Dimension	Exact Agreement	Cohen's Kappa	Rank Correlation	N
ESG	0.1985	−0.0018	0.0321	1783
E	0.2199	0.0248	0.1188	1783
S	0.1935	−0.0081	−0.0600	1783
G	0.2182	0.0227	0.1073	1783

Table 10. Linear Regression between CNRDS and CSMAR ESG Ratings

This appendix examines the linear relationships between the CNRDS and CSMAR ESG rating systems using ordinary least squares (OLS) regressions. For each dimension—overall ESG, Environmental (E), Social (S), and Governance (G)—the CNRDS scores are regressed on the corresponding CSMAR scores to assess how well one system can explain the variation in the other. The analysis is based on firm-level min–max normalized scores (MM) across all sample years. This table reports OLS regression results of the form $\text{CNRDS}_i = \alpha + \beta\,\text{CSMAR}_i + \varepsilon_i$ where both variables are firm-level min–max standardized scores (MM). Asterisks denote significance at the p < 0.05 (), p < 0.01 (), and p < 0.001 () levels.

The results indicate that the explanatory power of cross-system regressions is extremely weak across all dimensions, with R^2 values below 1% in every case. The regression slopes are generally small, ranging between −0.106 (for the S dimension) and 0.092 (for the G dimension), implying a weak and inconsistent linear association between the two databases. The regression analyses reveal several noteworthy patterns across the four ESG dimensions. For the overall ESG composite, the slope coefficient is statistically insignificant ($p = 0.403$), suggesting that the two systems share no systematic linear association in their aggregate assessments. The Environmental (E) dimension exhibits a weak yet positive relationship ($\beta = 0.052$, $p = 0.048$), though its explanatory power remains negligible with an R^2 of only 0.002, indicating that differences in scaling and weighting likely dominate the relationship. In contrast, the Social (S) dimension displays a negative and statistically significant association ($\beta = -0.106$, $p < 0.001$), implying that firms evaluated more favorably by CSMAR often receive lower relative scores from CNRDS—a result consistent with the earlier evidence of divergent criteria in social metrics. The Governance (G) dimension produces the highest positive slope ($\beta = 0.092$, $p < 0.001$), yet even here the model explains less than one percent of the cross-system variation, underscoring the structural misalignment between the two rating frameworks.

Overall, these results confirm that CNRDS and CSMAR ratings cannot be linearly mapped onto each other, even within the same ESG dimension. The lack of explanatory power reinforces earlier robustness findings that the discrepancies between the two systems are structural, not noise-driven.

Dimension	Intercept	Slope	R^2	Adj. R^2	p-value (Slope)	N Obs.	Significance
ESG	0.3496	0.0189	0.000	−0.000	0.403	1783	Not significant
E	0.1526	0.0520	0.002	0.002	0.048	1783	*
S	0.4243	−0.1064	0.007	0.007	0.000	1783	***
G	0.3648	0.0917	0.008	0.008	0.000	1783	***

Table 11. Summary Statistics of Standardized ESG Ratings from CSMAR and CNRDS

This table reports the mean, standard deviation, and median of standardized ESG, Environmental (E), Social (S), and Governance (G) scores from the CSMAR and CNRDS databases. Across all dimensions, CSMAR consistently assigns higher average and median scores than CNRDS, confirming a systematic upward bias in its rating scale.

Specifically, the Environmental (E) pillar shows the most pronounced difference, with the mean score under CSMAR (0.395) more than twice that of CNRDS (0.173), suggesting a more generous evaluation of firms' environmental performance. The Social (S) dimension also exhibits a substantial gap (0.559 vs. 0.365), implying that CSMAR's framework places greater weight on positive social disclosures or less penalization for weak performance. In contrast, the Governance (G) and overall ESG scores display smaller differences (means of 0.433 vs. 0.405 and 0.430 vs. 0.358, respectively), yet the pattern of higher central tendency under CSMAR persists.

Across all pillars, CSMAR scores also show slightly lower dispersion (standard deviations of 0.15–0.17) than CNRDS (0.17–0.19), indicating a compressed scoring distribution and potentially reduced discriminatory power at the lower tail. Taken together, these results reinforce that the divergence between CSMAR and CNRDS ratings is structural and directional.

Dimension	System	Mean	Standard Deviation	Median
E (Environmental)	CNRDS	0.173	0.193	0.113
	CSMAR	0.395	0.173	0.365
S (Social)	CNRDS	0.365	0.192	0.351
	CSMAR	0.559	0.154	0.551
G (Governance)	CNRDS	0.405	0.169	0.389
	CSMAR	0.433	0.169	0.422
ESG (Composite)	CNRDS	0.358	0.167	0.325
	CSMAR	0.430	0.175	0.409

Table 12. Industry-Level Differences between CSMAR and CNRDS ESG Ratings

This table provides a comprehensive comparison of industry-level ESG rating discrepancies between the CSMAR and CNRDS databases, highlighting systematic biases and sector-specific divergences in the two rating systems.

Panel A reports the sample distribution across ten major industries that contain sufficient firm-year observations ($N \geqslant 50$) for reliable statistical comparison. These sectors—ranging from traditional Manufacturing and Mining to modern Information Technology and Finance—capture a broad cross-section of the Chinese economy, ensuring that the observed rating discrepancies are not confined to a single domain but rather reflect structural patterns across industries.

Panel B presents the mean differences in ESG scores between the two databases for the overall composite rating (ESG) as well as its three subcomponents: Environmental (E), Social (S), and Governance (G). Positive differences indicate that CSMAR assigns higher ratings than CNRDS, while negative values imply more conservative evaluations by CNRDS. The results show clear and consistent upward bias in CSMAR's ratings, especially in the Environmental and Social pillars. Notably, the Finance, Information Technology, and Transportation sectors exhibit the largest cross-database gaps—each with mean differences exceeding 0.10 on the 0–1 standardized scale—suggesting that industry-specific disclosure practices and data weighting schemes may systematically inflate scores under CSMAR's methodology.

Panel C summarizes which industries exhibit statistically significant differences ($p < 0.05$) in at least one ESG dimension. The Environmental (E) dimension shows the broadest and most pronounced divergence, with significant differences in nine of ten industries, including Real Estate ($\Delta E = 0.315$), IT Services ($\Delta E = 0.321$), and Transportation ($\Delta E = 0.357$). The Social (S) dimension also reveals large positive discrepancies—particularly in Manufacturing ($\Delta S = 0.252$) and Finance ($\Delta S = 0.280$)—indicating that CSMAR systematically rewards firms with stronger self-reported social responsibility or disclosure frequency. In contrast, the Governance (G) pillar demonstrates weaker and more symmetric differences, with a few industries such as Construction and Real Estate exhibiting slightly higher CNRDS ratings, suggesting greater emphasis on board structure and compliance quality in CNRDS's evaluation criteria.

Taken together, these results underscore the heterogeneity of ESG rating construction across industries, revealing that CSMAR's upward bias is not uniform but context-dependent. The divergence is particularly pronounced in sectors where disclosure quality, data availability, or regulatory oversight varies substantially—such as in service-based and high-tech industries. This pattern implies that cross-database comparability may be especially limited in industries with evolving sustainability metrics or non-standardized environmental data. Overall, the findings highlight the need for caution when using multi-source ESG data in empirical research, as systematic inter-database discrepancies are closely linked to industry characteristics and methodological design.

Panel A. Industry Sample Distribution

Industry	N (Firm-Year Observations)
Manufacturing	593
Finance Industry	325
Raw Materials Manufacturing	166
Other Services	123
Software and Information Technology Services	118
Transportation, Storage and Postal Services	97
Real Estate	92
Mining Industry	85
Construction	74
Water Production and Supply	65

Total industries with sufficient samples ($N \geq 50$): 10

Panel B. Mean Differences in ESG Ratings by Industry

Industry	N	ESG_Mean_Diff	E_Mean_Diff	S_Mean_Diff	G_Mean_Diff
Manufacturing	593	0.0549	0.2070	0.2521	0.0705
Finance Industry	325	0.1749	0.2462	0.2800	0.0868
Raw Materials Manufacturing	166	-0.0123	0.0276	0.2259	0.0794
Other Services	123	0.0433	0.2831	0.1063	-0.0834
Software and Information Technology Services	118	0.1220	0.3205	0.1817	-0.0331
Transportation, Storage and Postal Services	97	0.1187	0.3569	0.1011	-0.0152
Real Estate	92	0.0074	0.3150	0.1241	-0.1817
Mining Industry	85	0.0108	0.1285	-0.0164	0.0395
Construction	74	-0.0547	0.2125	-0.0640	-0.0770
Water Production and Supply	65	0.1606	0.2321	0.1809	0.0338

Panel C. Industries with Statistically Significant Differences

Dimension	Industries with Significant Differences	No. of Industries
ESG	Manufacturing, Finance, Other Services, IT Services, Transportation, Water Supply	6
E	Manufacturing, Finance, Other Services, IT Services, Transportation, Real Estate, Mining, Construction, Water Supply	9
S	Manufacturing, Finance, Raw Materials, Other Services, IT Services, Transportation, Real Estate, Water Supply	8
G	Manufacturing, Finance, Raw Materials, Other Services, Real Estate, Construction	6

Table 13. In-Depth Analysis of Industries with Significant CSMAR–CNRDS Rating Differences

This table extends the cross-industry comparison by examining the ten industries that exhibit statistically significant differences between CSMAR and CNRDS ESG ratings. For each industry, we report the largest discrepancy dimension, the magnitude of the difference, relative bias, and within-industry rank-order agreement (Spearman ρ). Asterisks (***) denote statistical significance at the 1% level. Differences are defined as Mean(CSMAR) – Mean(CNRDS), where positive values indicate higher ratings under CSMAR. Agreement (ρ) represents the correlation between CSMAR and CNRDS scores within each industry.

The results demonstrate that the Environmental (E) pillar is the most divergent and structurally inconsistent dimension across industries. Eight out of ten industries—including Transportation, IT Services, Real Estate, and Water Supply—show substantial positive gaps ($\Delta E > 0.20$), with relative biases ranging from +67 % to +160 %. These findings indicate that CSMAR systematically assigns higher environmental ratings, potentially reflecting more generous scoring weights on disclosure quantity, emission-reduction targets, or qualitative narrative indicators. The Social (S) dimension also reveals pronounced differences, especially in Manufacturing ($\Delta S = 0.25$) and Finance ($\Delta S = 0.28$). This suggests that CSMAR emphasizes employee-related policies and community engagement more heavily than CNRDS, which tends to anchor its social evaluation on compliance and verifiable indicators. In contrast, the Governance (G) pillar shows mixed results with lower magnitudes and even negative differences in Construction and Real Estate. This reversal implies that CNRDS may apply stricter criteria regarding ownership concentration, board independence, and audit quality—metrics less emphasized in CSMAR's framework.

Across all sectors, agreement coefficients remain low ($\rho < 0.35$ in most cases), reinforcing that the two databases not only differ in score levels but also in the relative ranking of firms within each industry.
The divergence is especially pronounced in industries with heterogeneous disclosure standards or fast-evolving sustainability metrics (e.g., IT and Transportation). Collectively, the evidence highlights that cross-system inconsistencies are methodological rather than stochastic, driven by domain-specific weighting schemes, data sourcing practices, and ESG taxonomy definitions.

Industry	N	Largest Difference Dimension	Difference	Relative Bias (%)	Agreement (ρ)	Direction
Mining Industry	85	E	0.129	34.3	0.224 ***	CSMAR Higher
Transportation, Storage & Postal	97	E	0.357	112.2	0.186 ***	CSMAR Higher
Water Production & Supply	65	E	0.232	76.3	0.231 ***	CSMAR Higher
Real Estate	92	E	0.315	159.9	0.196 ***	CSMAR Higher
IT Services	118	E	0.321	149.3	0.636 ***	CSMAR Higher
Finance Industry	325	S	0.280	56.8	0.292 ***	CSMAR Higher
Construction	74	E	0.213	66.3	0.203 ***	CSMAR Higher
Raw Materials Manufacturing	166	S	0.226	53.0	0.193 ***	CSMAR Higher
Other Services	123	E	0.283	137.1	0.114 ***	CSMAR Higher
Manufacturing	593	S	0.252	54.8	0.211 ***	CSMAR Higher

Table 14 Industry-Level Mean Differences by ESG Dimension

This table provides a summarized visualization of the magnitude and pattern of ESG rating discrepancies between the CSMAR and CNRDS databases. Panel A reports the mean differences for the overall ESG score and each of the three sub-dimensions (E, S, and G), calculated as Mean(CSMAR) – Mean(CNRDS). Differences are defined as Mean(CSMAR) – Mean(CNRDS); positive values indicate higher CSMAR ratings. The last column reports the mean absolute difference across the four dimensions for each industry. Panel B ranks the industries by their overall mean absolute difference and identifies the dimension that contributes most to cross-database divergence.

The results show that Finance, IT Services, and Real Estate are the industries with the highest divergence ($\Delta \approx$ 0.16–0.20), driven primarily by environmental and social components. In contrast, Mining and Raw Materials Manufacturing show relatively consistent ratings, indicating closer alignment between the two data sources in resource-based sectors. Across industries, discrepancies are systematically larger in the Environmental pillar, suggesting that differences in disclosure coverage, qualitative weighting, and data collection scope contribute significantly to cross-database inconsistencies.

Panel A. Cross-Industry Mean Differences between CSMAR and CNRDS

Industry	E	S	G	ESG (Overall)	Mean Absolute Difference
Finance Industry	0.246	0.280	0.087	0.175	0.197
Software & IT Services	0.321	0.182	-0.033	0.122	0.164
Real Estate	0.315	0.124	-0.182	0.007	0.157
Water Production & Supply	0.232	0.181	0.034	0.161	0.152
Transportation, Storage & Postal	0.357	0.101	-0.015	0.119	0.148
Manufacturing	0.207	0.252	0.071	0.055	0.146
Other Services	0.283	0.106	-0.083	0.043	0.129
Construction	0.212	-0.064	-0.077	-0.055	0.102
Raw Materials Manufacturing	0.028	0.226	0.079	-0.012	0.086
Mining Industry	0.129	-0.016	0.040	0.011	0.049

Panel B. Summary of Industry Ranking by Overall Discrepancy

Rank	Industry	Average Absolute Difference (Δ)	Dominant Source of Divergence
1	Finance Industry	0.197	Social (S) and Environmental (E)
2	Software & IT Services	0.164	Environmental (E)
3	Real Estate	0.157	Environmental (E) and Governance (G, negative bias)
4	Water Production & Supply	0.152	Environmental (E)
5	Transportation & Postal	0.148	Environmental (E)
6	Manufacturing	0.146	Social (S)
7	Other Services	0.129	Environmental (E)
8	Construction	0.102	Environmental (E), Governance (G, negative bias)
9	Raw Materials Manufacturing	0.086	Social (S)
10	Mining Industry	0.049	Environmental (E)

Figure 1. Temporal Trends in ESG Scores: CNRDS vs. CSMAR Databases

Figure 1 presents the temporal patterns of average scores in each ESG dimension—Environmental, Social, and Governance—based on the CNRDS and CSMAR databases from 2015 to 2020. Across all dimensions, the two databases exhibit markedly different scale levels and temporal behaviors.

In Panel (a), the overall ESG scores show a persistent level gap of roughly 20 points, consistent with Figure A1. In Panel (b), the Environmental scores of CSMAR remain flat until 2019 and then sharply rise in 2020, suggesting a major methodology update or data reweighting, whereas CNRDS displays a steady, gradual increase. Panel (c) reveals an opposite pattern for the Social pillar—CNRDS maintains relatively stable and moderate scores, while CSMAR's series jumps dramatically in 2020. Finally, Panel (d) shows that the Governance scores in CNRDS increase steadily over time, whereas CSMAR reports very low values before 2019 and a sudden surge thereafter.

These inconsistencies indicate substantial heterogeneity in the underlying evaluation frameworks of the two ESG providers. Specifically, CNRDS applies a relatively stable scoring approach emphasizing disclosure-based measures, while CSMAR appears to have restructured its ESG model around 2019–2020, possibly to align with new data sources or international standards.

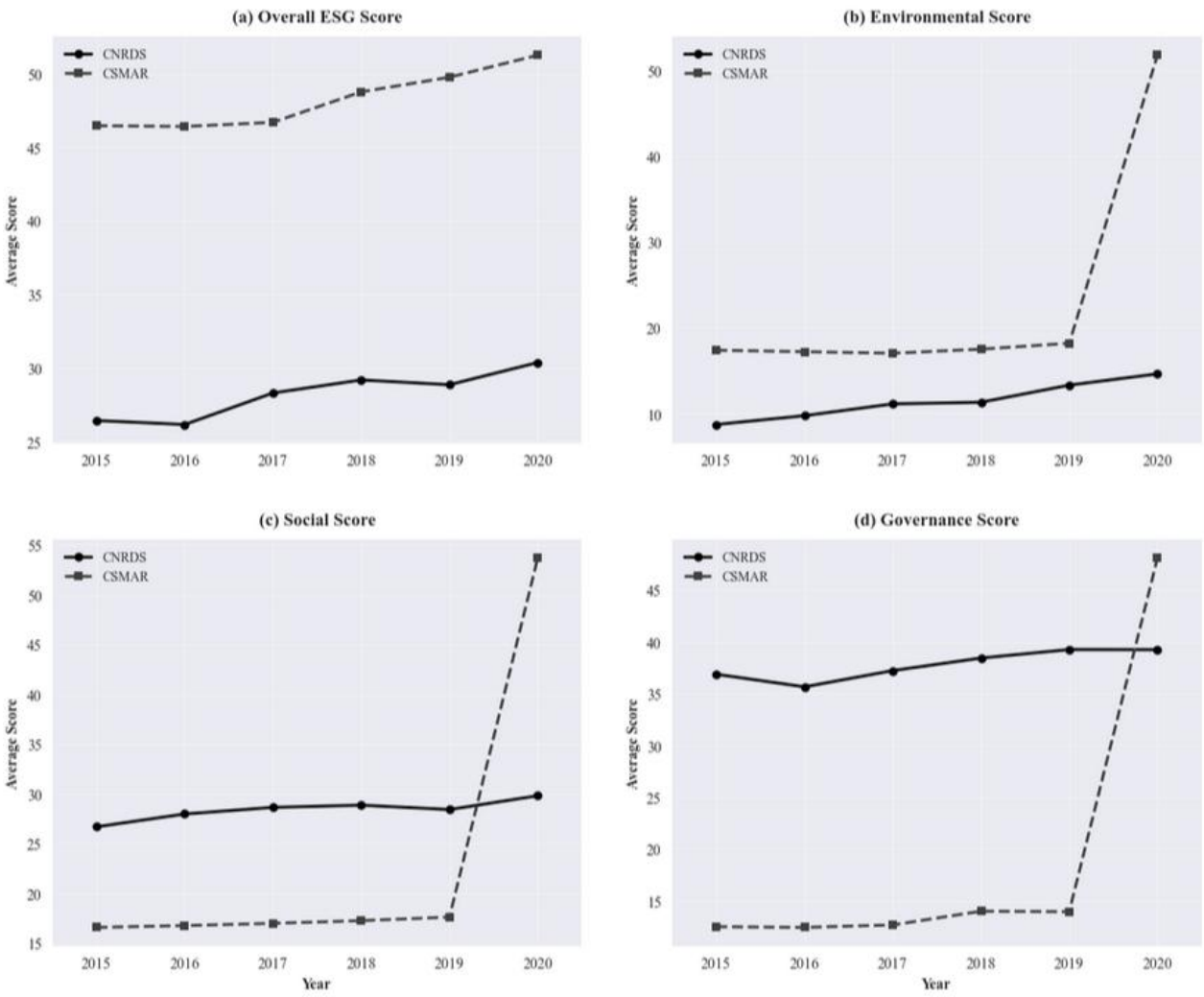

Figure 2. ESG Score Trends by Industry: CNRDS vs. CSMAR

Figure 2 plots the temporal evolution of average ESG scores across the top three industries ranked by CNRDS's overall ESG performance. The figure on the left and right correspond to the CNRDS and CSMAR datasets, respectively. Both panels show upward trajectories in ESG performance across industries over the 2015–2020 period, consistent with the broader trend of growing ESG awareness among Chinese listed firms. However, the two systems differ substantially in level and cross-industry dispersion.

In the CNRDS dataset, ESG scores increase moderately and display relatively stable industry rankings over time, suggesting a consistent evaluation framework and gradual improvement in firms' ESG practices. By contrast, the CSMAR dataset exhibits more pronounced divergence across industries after 2018, with some sectors (such as utilities and industrials) showing rapid increases while others remain nearly flat. This widening gap likely reflects the methodological adjustments introduced by CSMAR around 2019–2020—possibly involving expanded coverage or revised weighting for industry-sensitive indicators.

Overall, the comparison suggests that CNRDS provides smoother, more stable inter-industry patterns, whereas CSMAR introduces greater volatility and cross-sector differentiation in the post-2018 period, consistent with the scoring jump observed in Table A1 and Figure A2.

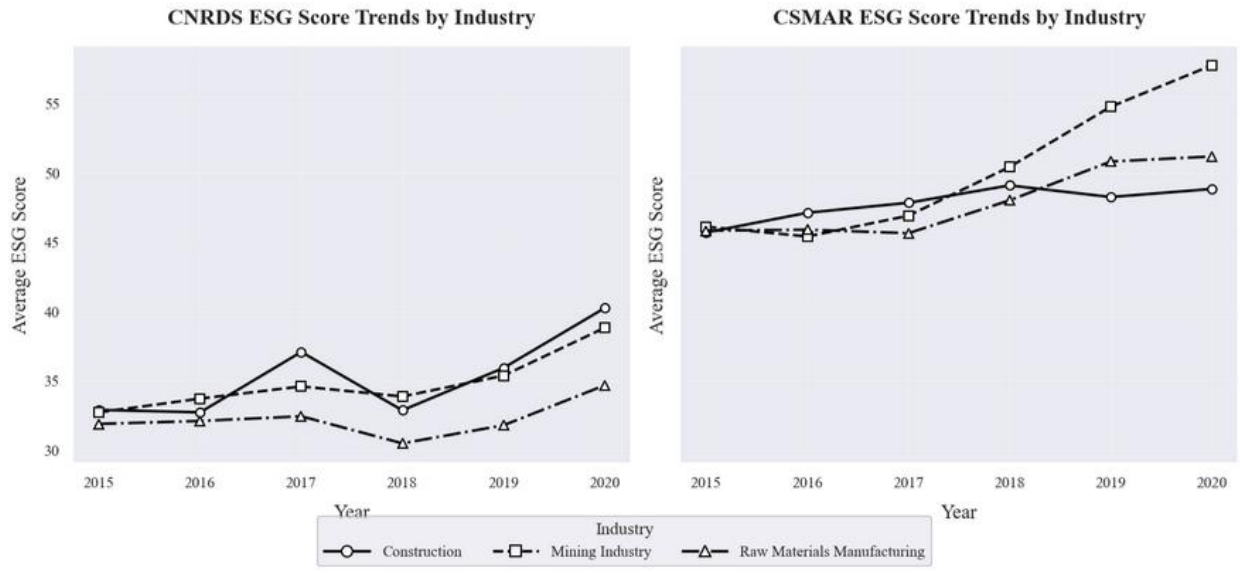

Figure 3. ESG Sub-Pillar Comparisons by Industry: CNRDS vs. CSMAR

Figure 3 compares the temporal evolution of ESG and its three sub-pillar scores (Environmental, Social, and Governance) for the top five industries ranked by CNRDS mean ESG scores—Construction, Mining, Health and Social Work, Leasing and Business Services, and Raw Materials Manufacturing. Panels (A)–(D) correspond to the overall ESG, Environmental, Social, and Governance dimensions, respectively, and present results from both the CNRDS and CSMAR systems. Across all panels, the CNRDS series exhibit smoother, more moderate increases over time, suggesting a relatively stable evaluation framework. In contrast, the CSMAR series demonstrate noticeable discontinuities in 2020 across all sub-dimensions, with especially sharp increases in the Environmental and Social components. These jumps align with the cross-sectional discrepancies observed in Table A1 and Table A2, and likely reflect a significant methodological revision or re-scaling within the CSMAR rating model. Industry-wise, Construction and Mining consistently rank at the top under both systems, though CSMAR assigns substantially higher absolute scores. Sectors such as Health and Social Work and Leasing and Business Services display the most pronounced CSMAR upswings post-2019, indicating stronger sensitivity to the expanded disclosure coverage introduced in the updated dataset. Meanwhile, the CNRDS framework captures more gradual improvements across all sectors, reflecting incremental changes in corporate ESG behavior rather than abrupt data redefinitions.

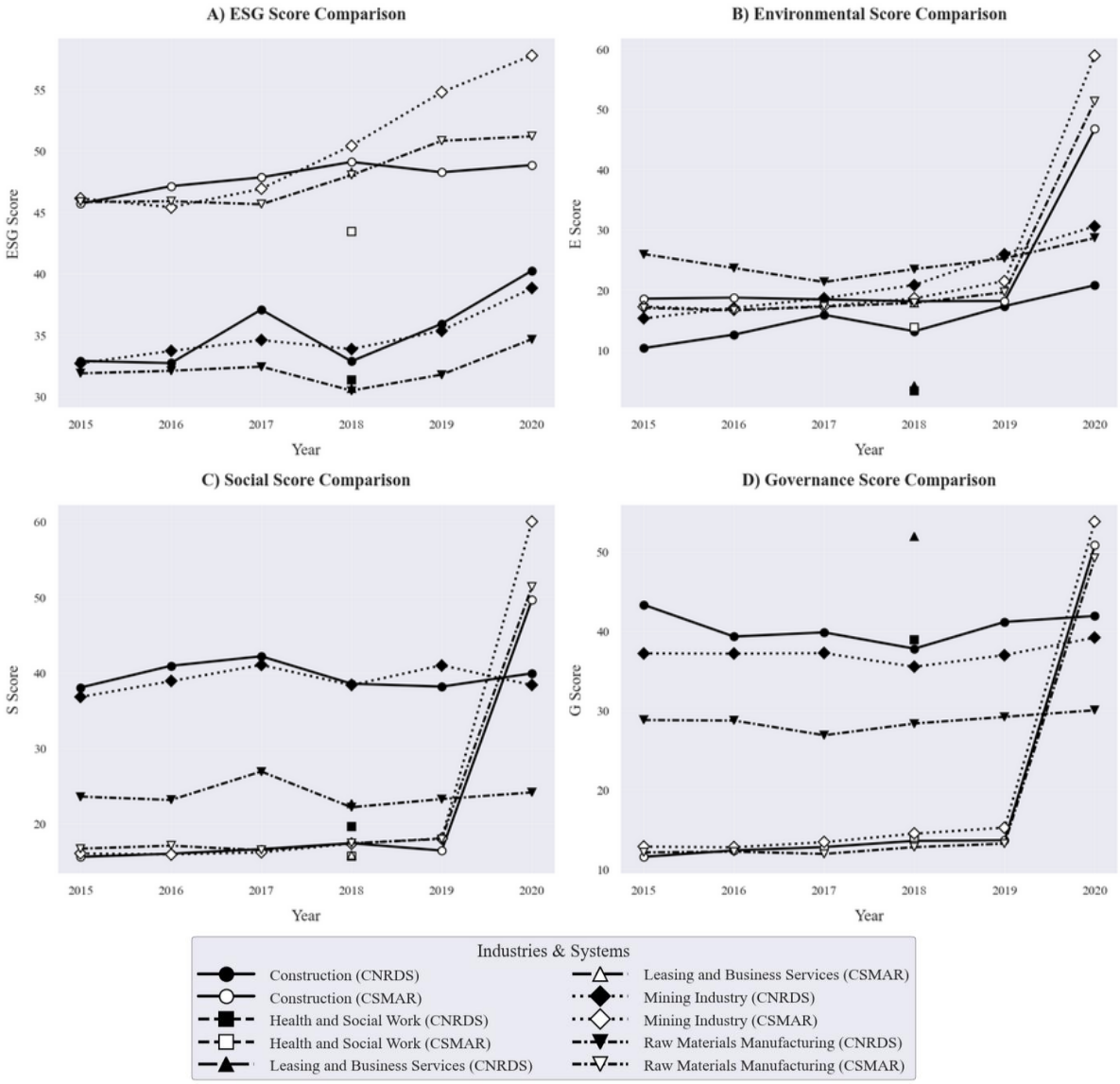

Figure 4. Yearly Spearman Correlation between CNRDS and CSMAR ESG Ratings

This figure presents the year-by-year Spearman correlation coefficients (ρ) between the CNRDS and CSMAR ESG rating systems across the four dimensions—overall ESG, Environmental (E), Social (S), and Governance (G)—for the period 2015–2020. The left panel displays correlations based on raw scores, while the right panel applies percentile-standardized scores to remove cross-database scale differences.

Across both specifications, the correlations remain weak and fluctuate within a narrow range (typically $\rho < 0.20$), indicating persistent disagreement in firm-level rankings. The Environmental and Governance pillars occasionally reach modest positive associations ($\rho \approx 0.20$ in 2019), whereas the Social dimension consistently exhibits negative or near-zero correlations, highlighting divergent evaluation criteria across the two systems. The similarity between the raw-score and standardized patterns suggests that scaling adjustments do not reconcile methodological inconsistencies. Overall, the persistently low and unstable correlations over time confirm that the divergence between CNRDS and CSMAR is structural rather than scale- or period-specific, reinforcing the robustness of the earlier findings.

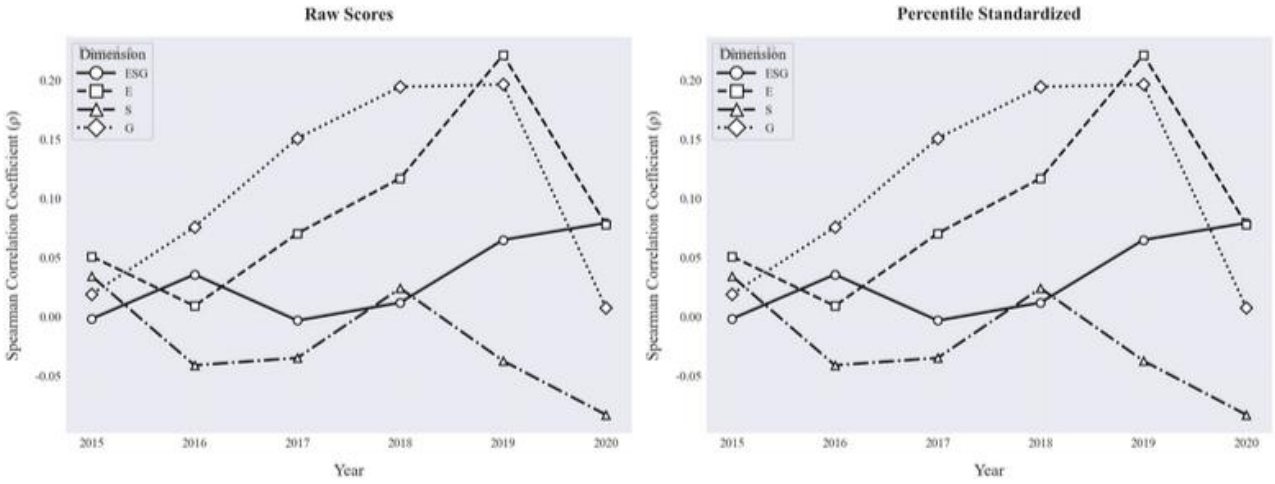

Figure 5. Visual Correlation Heatmap between CNRDS and CSMAR ESG Ratings

This figure visualizes the full pairwise Pearson correlation matrix between the CNRDS and CSMAR ESG ratings, based on firm-level min–max normalized scores across all four dimensions (ESG, E, S, and G). Lighter shades represent stronger positive correlations, while darker areas indicate weak or near-zero associations.

The heatmap reveals a clear structural pattern: within-system correlations are high and cohesive for CSMAR (e.g., $\rho \approx 0.65$–0.76 between ESG and its E and S pillars), reflecting consistent internal scaling. In contrast, cross-system correlations between the same dimensions (e.g., CSMAR_ESG_MM vs. CNRDS_ESG_MM = 0.02; CSMAR_S_MM vs. CNRDS_S_MM = −0.09) remain extremely weak. The Social pillar (S) shows the strongest disagreement—displaying negative or near-zero co-movement—while Governance (G) exhibits only minor positive association ($\rho \approx 0.09$). Overall, the heatmap visually corroborates the numerical results in Table 8: the two ESG rating systems display strong internal coherence but minimal cross-system alignment, underscoring the methodological divergence between CNRDS and CSMAR in their construction of ESG scores.

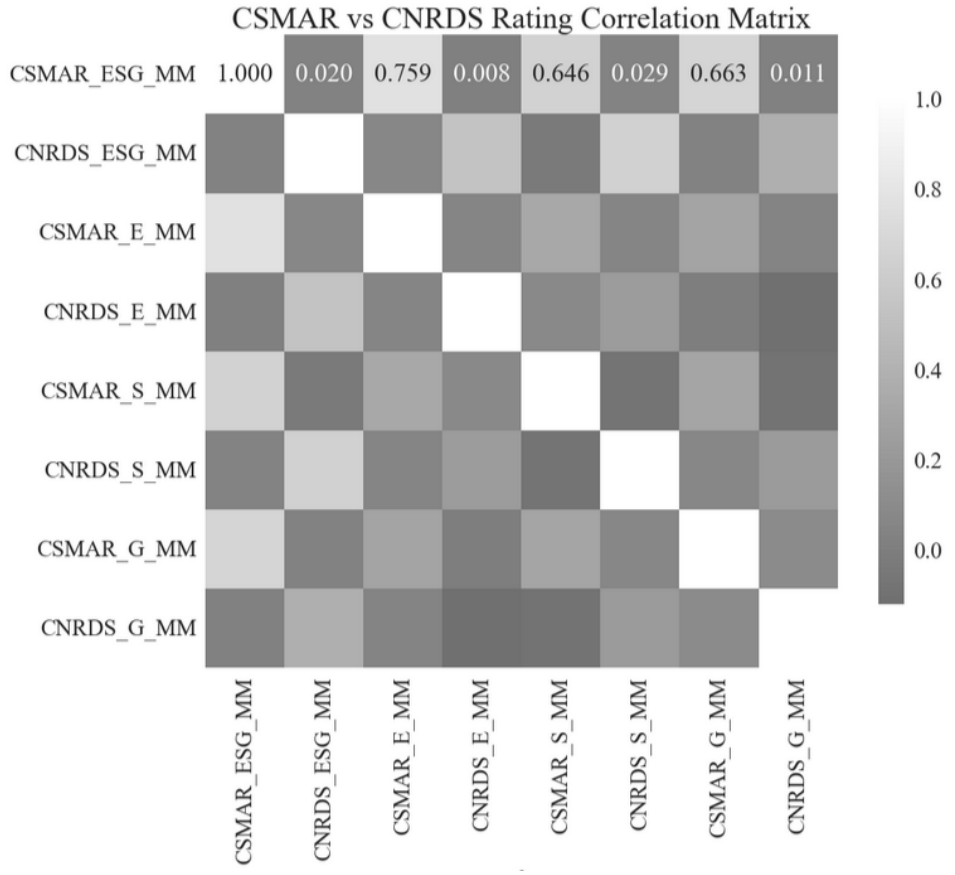

Figure 6. Cross-System Agreement Diagnostics: Bland–Altman Analysis

This appendix evaluates the measurement agreement between the CNRDS and CSMAR ESG rating systems using the Bland–Altman method, which plots the pairwise differences (CSMAR − CNRDS) against the mean of the two measures for each firm. The analysis covers four ESG dimensions—overall ESG, Environmental (E), Social (S), and Governance (G)—and three normalization variants: firm-level min–max scaling (MM), industry-adjusted normalization (ind_MM), and year-adjusted normalization (year_MM). Each subplot reports the mean difference (solid line) and the ±1.96 standard deviation limits of agreement (dashed lines), indicating the range within which 95% of score discrepancies fall.

Across all specifications, the differences between the meanings are positive and statistically meaningful, showing that CSMAR systematically assigns higher scores than CNRDS, especially in the Environmental and Social pillars (mean bias ≈ +0.22 and +0.19, respectively). Despite normalization or adjustment, the wide dispersion of points around the mean bias suggests low measurement agreement and non-interchangeability of the two datasets. Importantly, the industry-adjusted results (middle column) yield only marginally tighter agreement bands, implying that industry effects account for part—but not all—of the systematic bias. Even after controlling industry and year, large residual variability persists, confirming that the two ESG systems differ fundamentally in scaling and rating methodology rather than being subject to random noise. Together, these Bland–Altman results provide a visual diagnostic complement to the correlation and paired-test analyses. They reinforce that the divergence between CNRDS and CSMAR arises from structural design differences in their scoring frameworks rather than sampling variation, cautioning against treating the two datasets as substitutable measures of firm ESG performance.

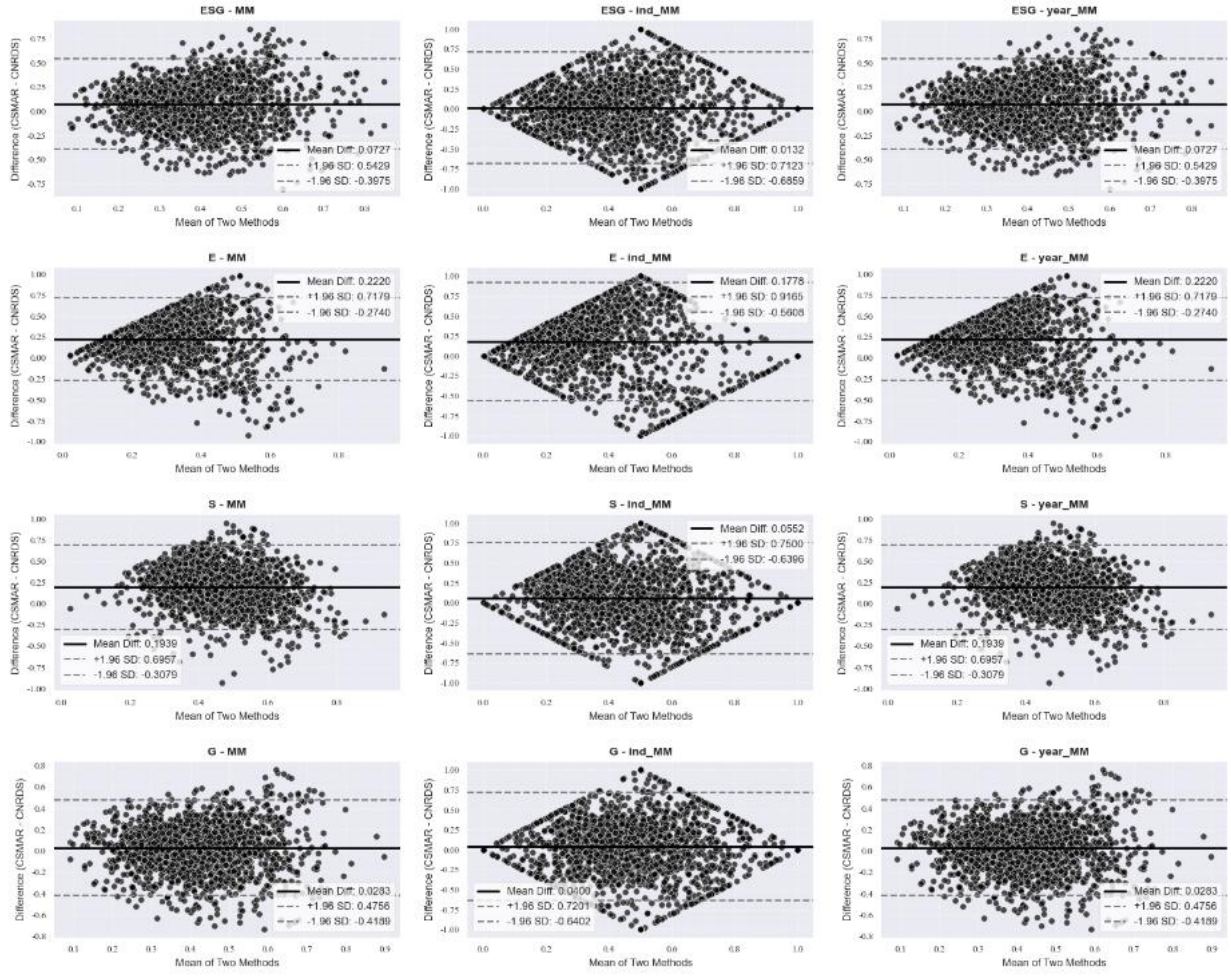

Figure 7. Time Trend of Correlation between CNRDS and CSMAR ESG Ratings

This figure illustrates the year-by-year evolution of correlation coefficients between the CNRDS and CSMAR ESG ratings over the 2015–2020 period. Four correlation series are plotted, corresponding to the overall ESG composite and the Environmental (E), Social (S), and Governance (G) pillars. The results are based on firm-level min–max normalized scores (MM), ensuring comparability across years. The time trends reveal that the overall ESG correlation remains near zero throughout, with only a modest increase after 2018. Among the sub-dimensions, the Environmental (E) and Governance (G) pillars show temporary positive correlations, peaking around 2019 ($\rho \approx 0.20$), whereas the Social (S) dimension persistently exhibits weak or negative associations across all years. This pattern indicates that the two databases do not converge over time; rather, their rating methodologies evolve independently.

Overall, the figure provides additional temporal evidence that the alignment between CNRDS and CSMAR ESG scores is both weak and unstable. Even as both datasets mature, their year-specific correlations remain low, suggesting that the disagreement between systems reflects structural differences in their rating frameworks rather than transitory noise or early-stage data inconsistency.

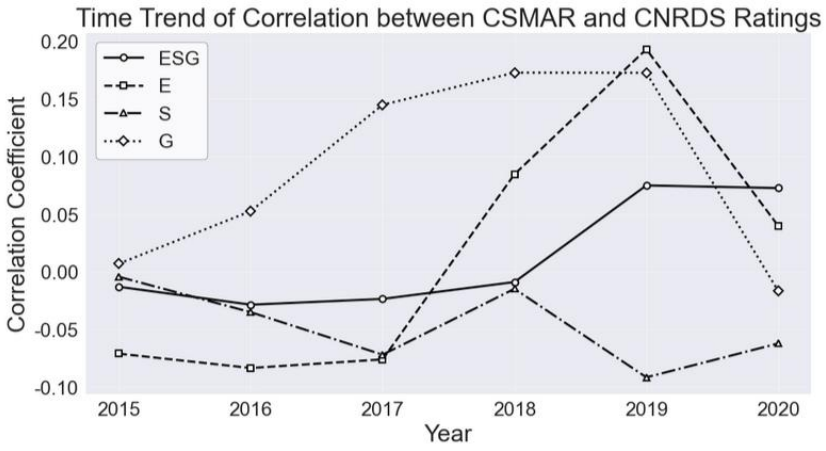

Figure 8. Scatterplots of Linear Relationships between CNRDS and CSMAR Ratings

These figures visualize the firm-level relationships between the CNRDS and CSMAR ESG scores using min–max normalized values across all sample years. Each subplot presents the fitted OLS regression line (in red), the one-to-one reference line representing perfect agreement (gray dashed), and the observed data points (black). The panel A illustrates the relationship for the overall ESG composite, where the regression line is nearly flat, confirming the absence of a systematic linear relation between the two systems ($\beta = 0.019$, $p = 0.403$, $R^2 \approx 0$). In the Environmental (E) dimension (Panel B), a slight upward slope emerges ($\beta = 0.052$, $p = 0.048$), but the explanatory power remains minimal ($R^2 = 0.002$), suggesting that scale differences rather than common evaluation logic dominate the relationship. The Social (S) panel (Panel C) shows a weak negative relationship ($\beta = -0.106$, $p < 0.001$), indicating that higher CSMAR scores are often associated with lower CNRDS scores—consistent with previously observed divergences in social metric construction. The Governance (G) panel (Panel D) exhibits the strongest positive trend ($\beta = 0.092$, $p < 0.001$), though the variance explained remains under 1%.

Taken together, these scatterplots visually confirm the regression results in Table A.25.1: despite both systems measuring similar ESG constructs, their scoring methodologies produce weak and inconsistent linear alignment across all dimensions.

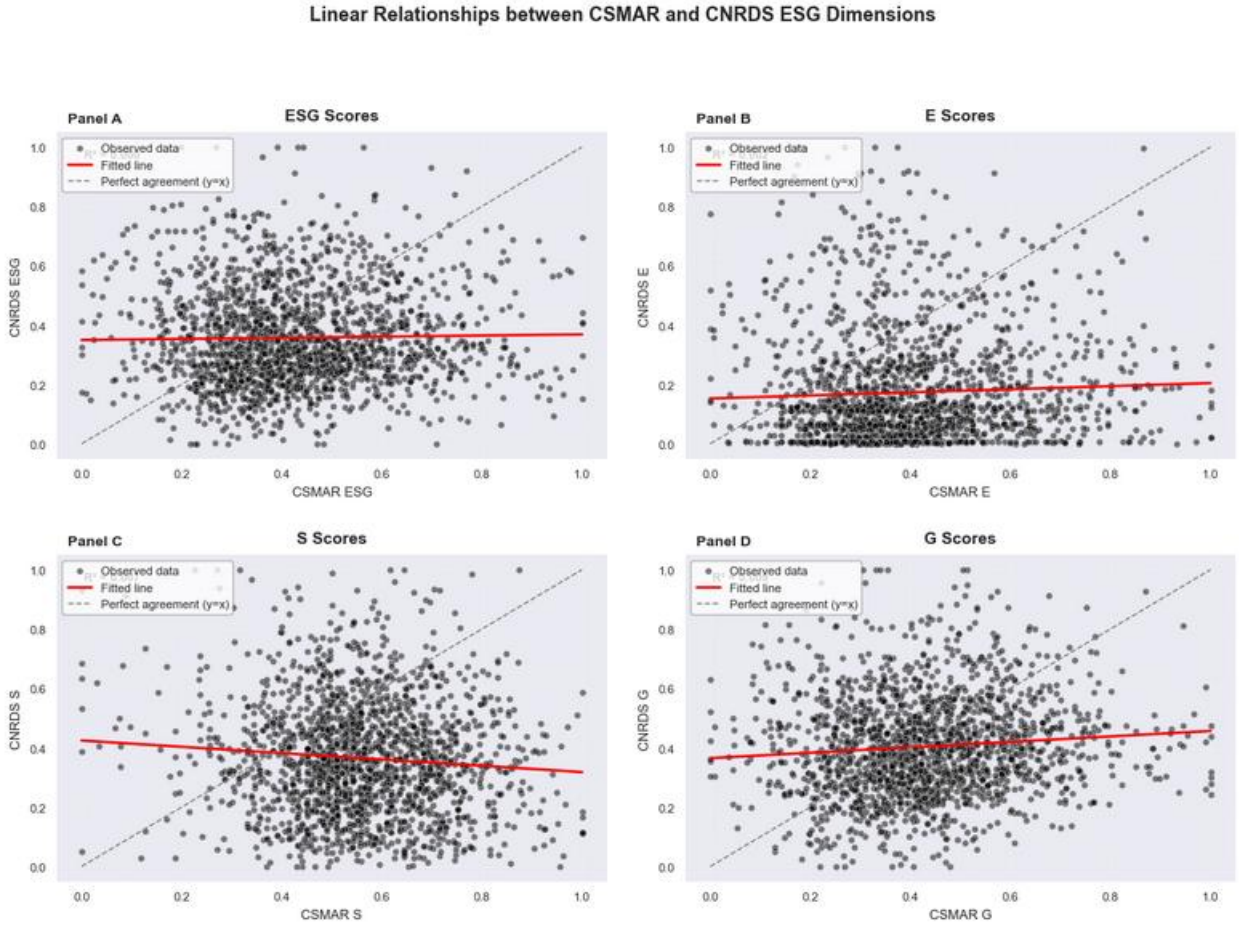

Figure 9. Distribution Comparison of CSMAR and CNRDS ESG Ratings by Dimension

This figure compares the standardized ESG ratings from the CSMAR and CNRDS databases across four dimensions—overall ESG, Environmental (E), Social (S), and Governance (G)—using violin plots that illustrate both the central tendency and the distributional spread of scores. Across all dimensions, CSMAR scores are systematically higher than their CNRDS counterparts, consistent with earlier paired-sample results. The Environmental and Social pillars exhibit the largest upward shift under CSMAR, with notably narrower distributions, indicating that the CSMAR framework tends to compress firm-level dispersion and assign relatively optimistic ratings. By contrast, CNRDS shows a broader and more left-skewed distribution, suggesting greater differentiation among low-performing firms.

Overall, the consistent upward bias and reduced variability in CSMAR scores reinforce the structural discrepancy between the two rating systems—reflecting not random noise but systematic differences in scaling, weighting, and methodological emphasis across ESG sub-dimensions.

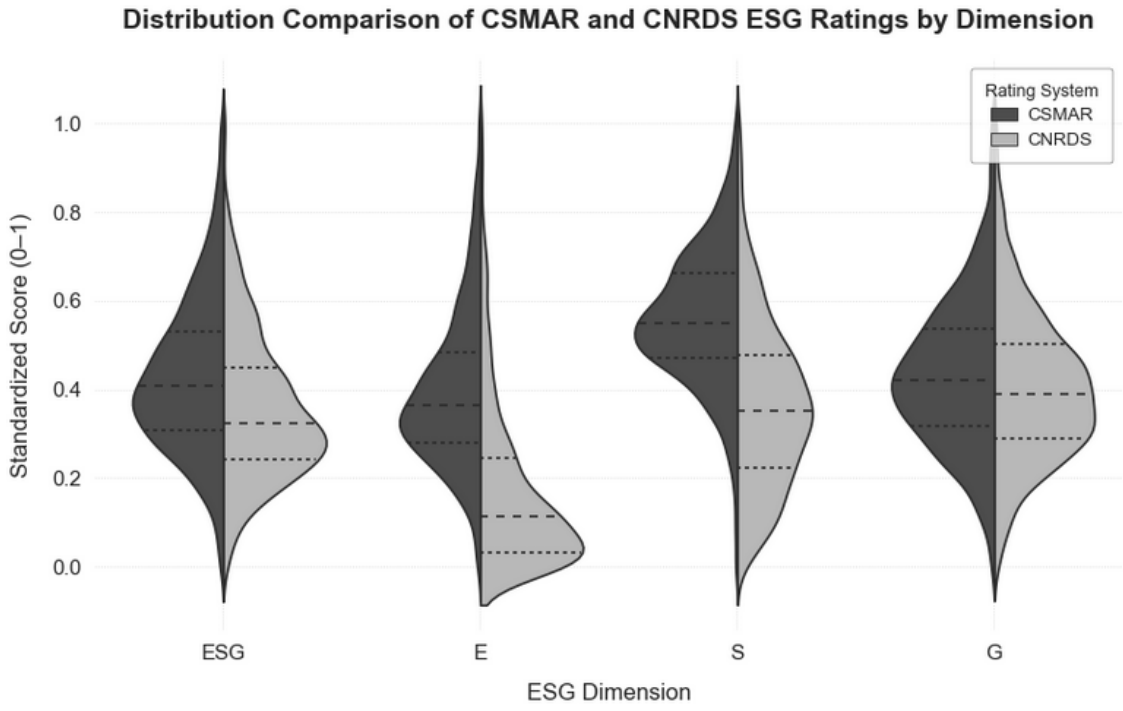

Figure 10. Distribution Comparison of CSMAR and CNRDS ESG Ratings by Dimension

This figure presents violin plots of the standardized ESG, Environmental (E), Social (S), and Governance (G) scores from the CSMAR and CNRDS databases. Each panel (A–D) shows the score distributions across the two systems, with dashed lines indicating respective mean values. Across all dimensions, CSMAR consistently assigns higher scores than CNRDS, with the largest upward bias observed in the Environmental (Panel B) and Social (Panel C) dimensions. The CSMAR distributions also exhibit smaller dispersion, reflecting a narrower range of firm-level differentiation and a tendency toward more optimistic evaluations. In contrast, CNRDS ratings are more left-skewed, suggesting stricter assessment criteria or broader variation across firms.

Overall, the figure confirms a systematic level bias between the two databases, where CSMAR's framework yields compressed and higher-centered distributions relative to CNRDS. This pattern aligns with the quantitative summary statistics reported in Appendix Table X and supports the conclusion that the divergence between the two ESG rating systems is structural rather than random.

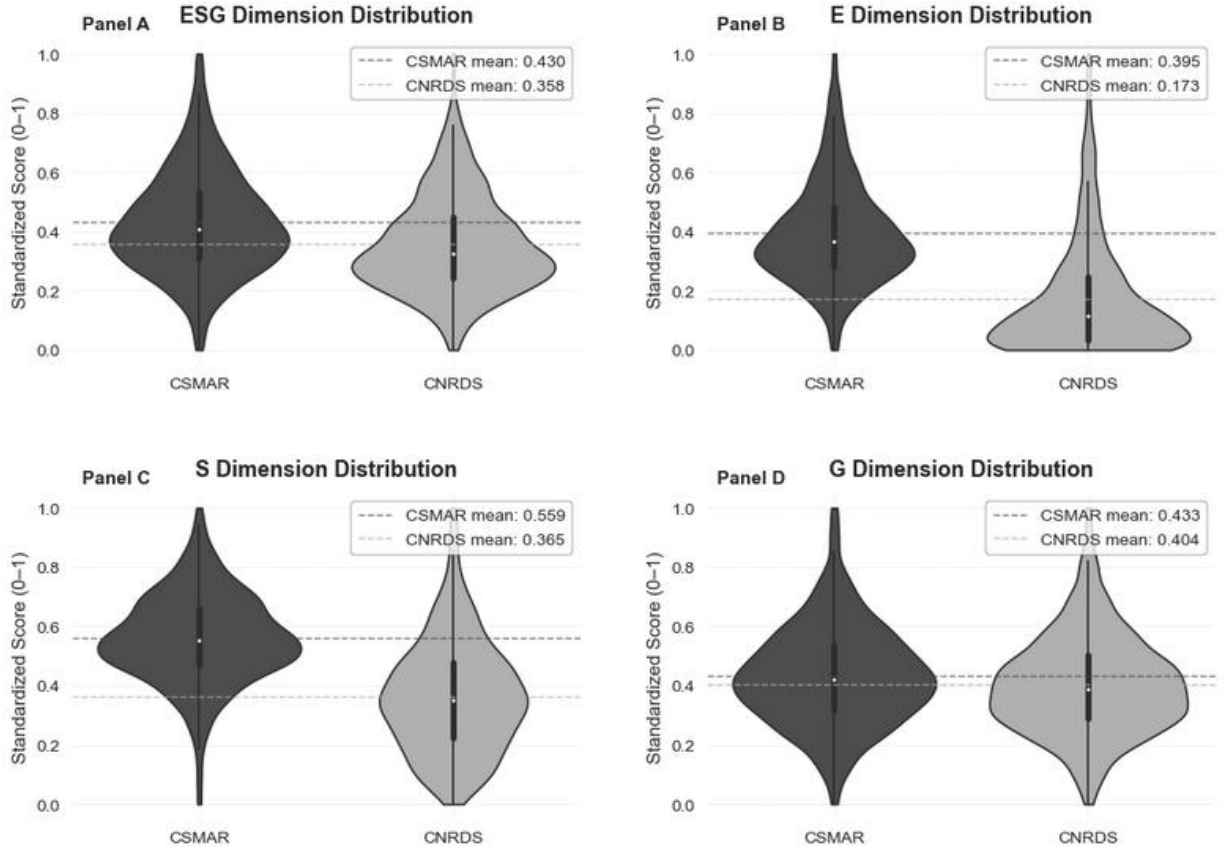

Figure 11. Distribution Overlap of Standardized ESG Ratings from CSMAR and CNRDS

This figure compares the kernel density distributions of standardized ESG, Environmental (E), Social (S), and Governance (G) scores between the CSMAR and CNRDS databases. Each panel (A–D) shows the degree of overlap between the two systems' rating distributions, with dashed vertical lines denoting mean values for each dataset. Across all dimensions, CSMAR's distributions are systematically shifted to the right, indicating higher average scores than CNRDS. The divergence is most pronounced for the Environmental (Panel B) and Social (Panel C) pillars, where CSMAR exhibits substantially higher mean values (0.395 vs. 0.173 and 0.559 vs. 0.365, respectively). By contrast, the Governance (Panel D) and overall ESG (Panel A) distributions show smaller mean gaps, though CSMAR still maintains a consistent upward bias.

Moreover, CNRDS ratings display wider dispersion and greater left-skewness, suggesting stricter or more heterogeneous scoring criteria. CSMAR, on the other hand, shows narrower peaks and lower variance, reflecting a more compressed assessment scale and reduced differentiation among firms. Together, these distributional differences confirm a systematic level bias between the two databases.

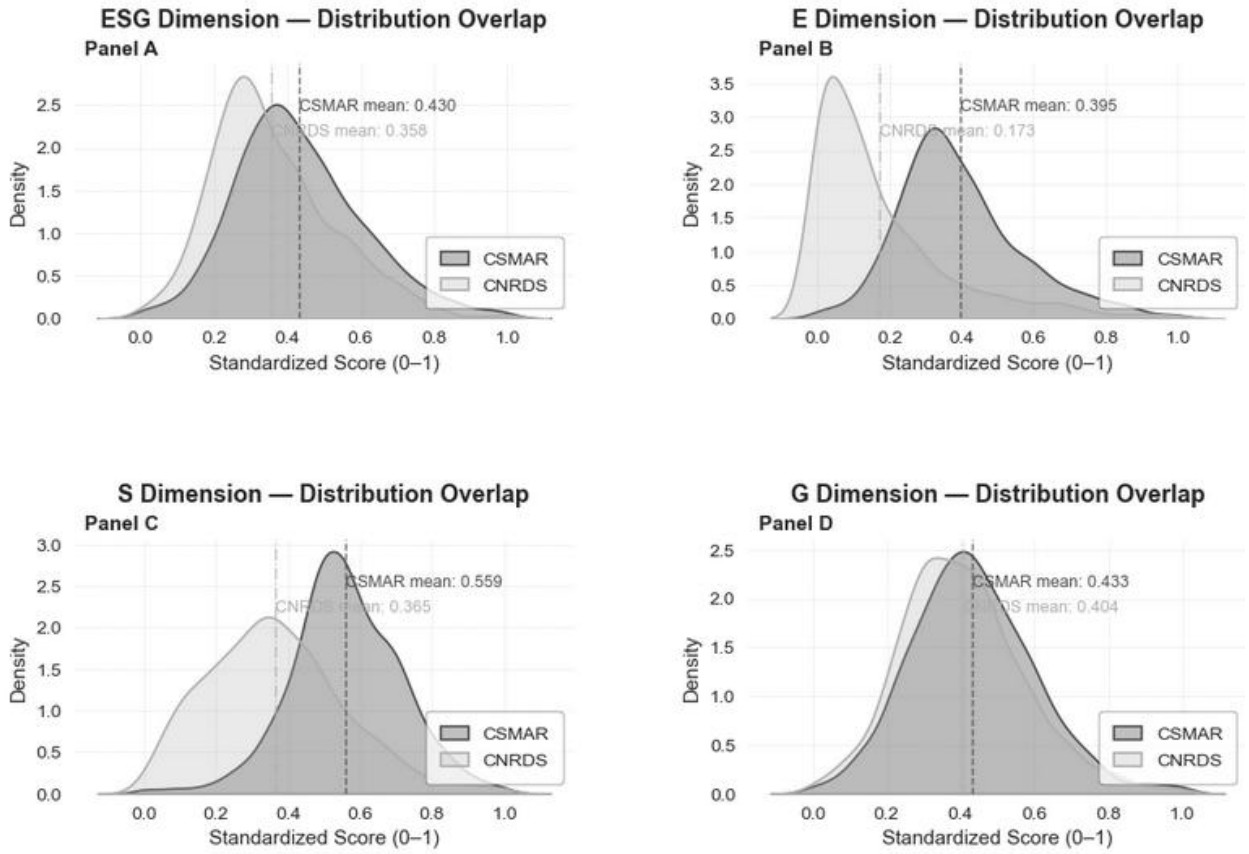

Figure 12. Boxplot Comparison of CSMAR and CNRDS ESG Ratings by Dimension

This figure presents boxplots comparing the distributions of standardized ESG, Environmental (E), Social (S), and Governance (G) scores between the CSMAR and CNRDS databases. Each box represents the interquartile range (IQR), with whiskers extending to 1.5×IQR, and dots marking outliers.

Across all four dimensions, CSMAR ratings display higher medians and narrower interquartile ranges than CNRDS, reflecting a systematic upward level bias and a more compressed scoring scale. The divergence is most pronounced for the Environmental (E) and Social (S) dimensions, where CSMAR scores cluster tightly around mid-to-high values, while CNRDS distributions are more dispersed and left-skewed, indicating a stricter evaluation framework.

These visual patterns reinforce the findings from the paired t-tests and distribution overlap plots, confirming that CSMAR consistently assigns higher and less variable ESG ratings than CNRDS, suggesting structural differences in the two systems' rating methodologies.

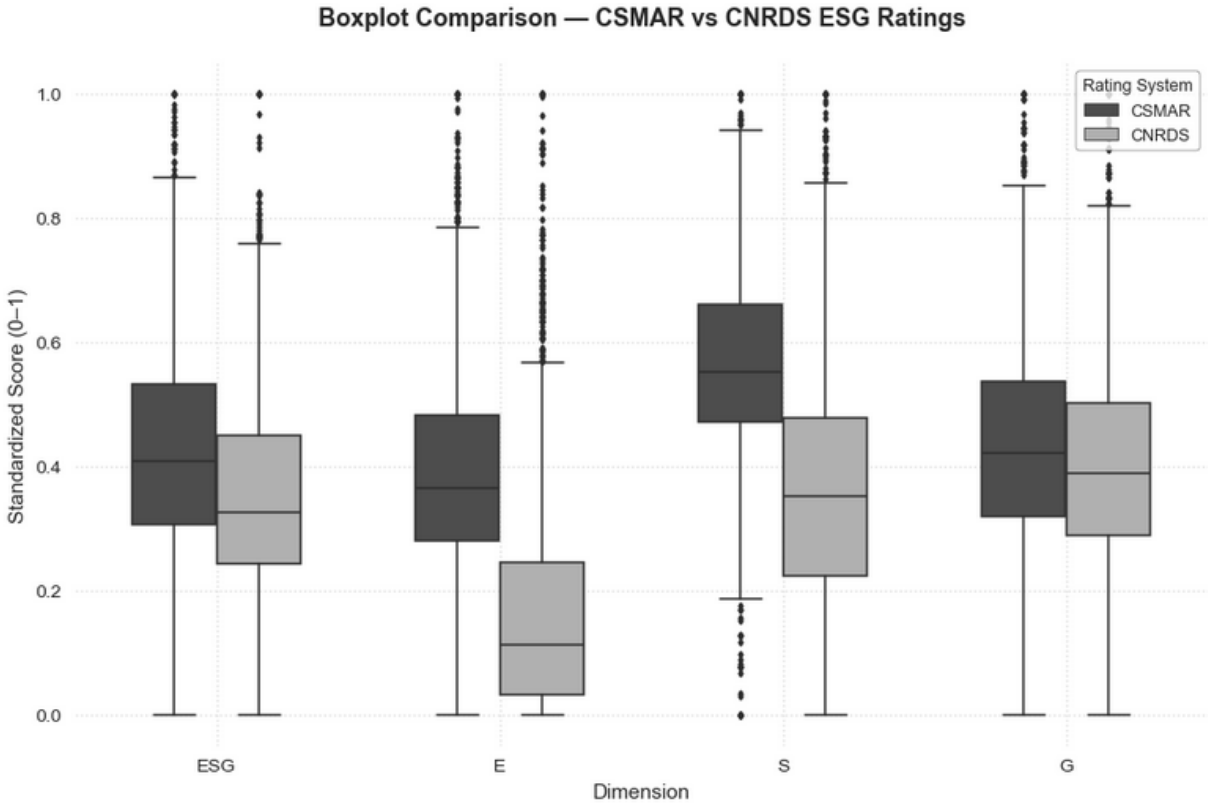

Figure 13. Industry-Level Rating Differences between CSMAR and CNRDS

This heatmap visualizes the mean rating differences between the CSMAR and CNRDS databases across ten major industries and four ESG dimensions (E, S, G, and the composite ESG). Each cell represents the mean difference in standardized scores, computed as Mean(CSMAR) – Mean(CNRDS). Darker shades indicate smaller or negative differences (where CNRDS assigns higher scores), while lighter shades indicate larger positive gaps (where CSMAR ratings are higher).

Overall, the visualization highlights substantial cross-industry heterogeneity. The Environmental (E) pillar exhibits the most pronounced divergence, with strong positive differences in sectors such as Transportation, IT Services, Real Estate, and Finance, where the mean differences exceed 0.25. These industries tend to rely more on narrative or qualitative environmental disclosures, which are weighted more heavily in CSMAR's scoring methodology. The Social (S) pillar also shows upward deviations in service-oriented and financial industries ($\Delta S > 0.20$), while Governance (G) differences appear smaller and more symmetric, occasionally negative in sectors such as Construction and Real Estate—suggesting that CNRDS enforces stricter governance standards.

Taken together, the pattern illustrates that cross-database divergence is concentrated in the Environmental and Social pillars, particularly in data-rich or disclosure-sensitive industries. This heterogeneity underscores the importance of controlling for industry-specific effects when using multi-source ESG data in empirical research.