International Journal of Modern Engineering and Management | IJMEM
ISSN No: 3048-8230
Volume 3 Issue 2 Feb- 2026

Divergent ESG Ratings in China: Measurement Inconsistency and Methodological Origins

Wanyi Zhao
Department of Finance,Johns Hopkins University/United States

Abstract:
I conduct the first systematic comparison of ESG ratings between China's two leading providers, CSMAR and CNRDS, using a matched panel of A-share firms from 2015 to 2020. I perform a wide-ranging empirical analysis at firm, industry, and temporal levels to investigate whether the two systems capture the same latent construct associated with corporate sustainability. My findings highlight significant and persistent divergence between rating levels, correlation structures, and reliability measures. CSMAR consistently assigns significantly higher ESG scores than CNRDS, particularly in the Environmental and Social dimensions, and cross-system correlations are extremely weak and frequently negative in the Social pillar. Measures of agreement, including correlation coefficients and concordance metrics, reveal sharp divergences in firm-level rankings between the two systems. Critically, I demonstrate that standard normalization methods, including percentile ranking, min–max scaling, and industry-adjusted standardization, fail to reconcile these differences. Even after removing scale and industry effects, rank-order disagreement remains pronounced, suggesting that the divergence reflects fundamental methodological differences rather than distributional or scaling artifacts. It is especially pronounced in industries with complex, qualitative, and disclosure-intensive ESG profiles, such as Finance and Information Technology, where greater measurement discretion amplifies methodological divergence. Collectively, the evidence indicates that CSMAR and CNRDS are non-interchangeable proxies for distinct ESG constructs. These findings have important implications for empirical research design, ESG-based investment strategies, and corporate sustainability assessment in China, highlighting the necessity of treating ESG rating choice as a core methodological decision rather than a neutral data input

Keywords: *ESG ratings, Methodological divergence, Rating consistency, Sustainable finance*

1. Introduction

The integration of Environmental, Social, and Governance (ESG) criteria into investment decisions and corporate strategy has become a central paradigm in global capital markets. ESG ratings, which aggregate firm-level sustainability information into standardized scores, are now widely used by investors, regulators, and researchers as proxies for corporate sustainability performance. However, a persistent challenge undermines their interpretability and comparability: substantial divergence across ESG rating providers. Prior studies document that even major international agencies often produce only weakly correlated scores—typically ranging from 0.3 to 0.6—reflecting differences in indicator selection, weighting schemes, and aggregation rules [1]-[2]. This divergence raises fundamental questions about whether ESG ratings can be treated as objective and interchangeable measures of firm sustainability. Although this has been well established in mature markets, the divergence in ESG ratings is far less well known in developing countries. This gap is critical in the context of China, where the practice of ESG disclosure, regulatory incentives, and data infrastructures are significantly different from mature markets in terms of information disclosure and monitoring with a wide variety of ESG data. As China's domestic ecosystem of ESG data continues to grow up rapidly, empirical research studies and investment practice have relied on two local institutions more and more such as the China Stock Market and Accounting Research (CSMAR) ESG database and the China National Research Data Services (CNRDS) ESG database. The relative strength of these two approaches in providing consistent, comparable ESG performance assessments at firm level remains comparatively limited, despite its dominance.

Importantly, institutional descriptions indicate that these two databases are found on different methodological philosophies. CSMAR relies on a bottom-up data-driven system whereas CNRDS has a top-down, normative structure as developed based on international ESG standards. In developed markets, similar methodological heterogeneity has negatively affected risk assessment and complicated empirical inference. In the Chinese context, where ESG research increasingly depends on a single domestic data source, such divergence raises questions about reproducibility and cross-

study comparability.

The current study fills this research gap by systematically comparing ESG ratings from CSMAR and CNRDS at firm level for Chinese A-share firms, published from 2015 to 2020. I examine the extent to which these two systems differ in score levels, temporal dynamics and cross-sectional structure; I assess both the consistency of absolute scores overall and ranking by firm; I examine whether techniques of common normalization mitigate these discrepancies. I elaborate on how divergence differs across ESG pillars and sectors. And in focusing explicitly on cross-system comparability rather than downstream economic outcomes, the paper sets methodological groundwork for the interpretation of evidence for ESG in China. The analysis provides a clear distinction as to whether ESG ratings from CSMAR and CNRDS should be assumed as interchangeable inputs in both empirical research and applied situations or if they represent distinct, fundamental constructs of corporate sustainability.

2. Literature Review

A. Global Evidence on ESG Rating Divergence

A growing literature documents substantial divergence among ESG rating agencies, even among leading international providers. Studies show that ESG scores from agencies such as MSCI, Sustainalytics, Refinitiv, and Bloomberg are often weakly correlated, raising concerns about their comparability and interpretability [1]-[2]. This divergence challenges the use of ESG ratings as objective and interchangeable measures of corporate sustainability.

Previous research blames cross-agency disagreement on systematic methodological differences and not noise. Berg, Koelbel, and Rigobon [1] identify three primary sources of divergence: scope (which ESG issues are included), measurement (how indicators are quantified), and weighting (how the sub-scores are aggregated). Such heterogeneity can undermine ESG-based risk pricing and counteract empirical inferences in finance research as demonstrated within related literature [3]-[4]. Taken together, global evidence indicates that normalization or rescaling alone is not sufficient to reconcile ESG ratings among providers.

B. Methodological Sources of Divergence

ESG rating divergence reflects the fact that rating agencies transform raw sustainability information into proprietary signals using distinct indicator systems, data treatments, and aggregation rules. Differences arise along three key dimensions: indicator selection within ESG pillars, data transformation and normalization methods, and aggregation or weighting schemes. These methodological decisions have economic consequences. Divergent ESG signals may induce information asymmetry among investors and lead to inconsistent assessments of firm risk and performance [5]. Empirically, the use of alternative ESG databases can produce conflicting estimates in asset pricing and corporate finance studies, even when analyzing the same sample of firms. An awareness of the structural origins of rating divergence is thus crucial for interpreting ESG-based evidence.

C. ESG Rating Divergence in Emerging Markets

ESG data in emerging markets is generally less standardized than in developed economies [6], and more subject to institutional heterogeneity. These challenges are exacerbated in China by policy-driven disclosure incentives as well as rapid regulatory evolution. But while CSMAR and CNRDS have become major domestic ESG sources, these companies use different rating philosophies, indicator systems, and industry adjustments. Most empirical studies in China primarily concentrate on the economic impact of ESG performance, including firm value and financing costs [7]-[8], while treating ESG ratings as input variables. Very few studies directly assess the reliability and consistency of the underlying rating systems, and systematic firm-level comparisons between CSMAR and CNRDS are limited. Consequently, the degree and structure of ESG rating divergence in China's domestic data ecosystem are not well understood.

D. Research Gap

While prior literature establishes that ESG rating divergence is widespread and structurally driven, evidence from China remains limited. Given the increasing reliance on CSMAR and CNRDS in academic research and investment practice, assessing their comparability is critical for empirical validity and reproducibility. This study addresses this gap by providing a systematic firm-level comparison of ESG ratings from CSMAR and CNRDS for Chinese A-share firms between 2015 and 2020. By examining rating levels, correlations, and reliability measures across ESG pillars and industries, the analysis documents the magnitude and sources of divergence within China's ESG rating landscape.The

intersection of strategic management and digital transformation represents a rapidly evolving research domain that draws from multiple theoretical traditions while generating new insights specific to digital contexts. This review synthesizes key contributions and identifies research gaps at this intersection.

3. Data and Methodology

Data and Sample Selection

This study constructs a firm-level panel of Chinese A-share listed companies from 2015 to 2020. ESG data are obtained from two major domestic providers: the CSMAR ESG module and the CNRDS ESG database. While both aim to assess firms' environmental, social, and governance performance, they differ substantially in coverage and methodology. CSMAR provides broad coverage of all A-share firms, whereas CNRDS primarily focuses on large-cap firms with relatively high disclosure quality.

To ensure comparability, the sample is restricted to firm-year observations jointly covered by both databases. Financial and industry information is obtained from CSMAR, with industries classified according to the CSRC standard. After matching and data cleaning, the final sample contains approximately 1,800 firm-year observations, covering about 280–320 firms per year across 11 primary industries, including manufacturing, finance, energy, and information technology.

ESG Variables and Standardization

For each firm i in year t, raw ESG scores from CSMAR and CNRDS are denoted as ESG_{it}^{CSMAR} and ESG_{it}^{CNRDS}, respectively, along with their Environmental (E), Social (S), and Governance (G) sub-pillars. Because the two systems differ in scale, weighting, and calibration, two standardization procedures are applied.

First, I apply industry–year min–max normalization[1] to rescale scores into the $[0,1]$ interval:

$$ESG_{it}^{MM} = \frac{ESG_{it} - \min\limits_{k,t}(ESG)}{\max\limits_{k,t}(ESG) - \min\limits_{k,t}(ESG)} \qquad (1)$$

where k indexes CSRC industries. Second, percentile-standardized scores are constructed to facilitate rank-based comparisons:

$$ESG_{it}^{pct} = \frac{rank_{k,t}(ESG_{it})}{N_{k,t}} \qquad (2)$$

where $rank_{k,t}(\cdot)$ denotes the within-industry–year rank and $N_{k,t}$ is the number of firms in that group. These transformations remove scale differences and mitigate industry composition effects.

Consistency and Reliability Tests

To assess cross-system consistency, I first conduct paired t-tests and Wilcoxon signed-rank tests comparing ESG levels assigned by CSMAR and CNRDS. Firm-level score differences are defined as:

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

Beyond mean differences, I evaluate agreement using correlation and reliability measures. Pearson correlations capture linear co-movement in score levels, while Spearman correlations assess rank-order consistency. [2]Reliability is further examined using the Intraclass Correlation Coefficient (ICC)[3] under the absolute agreement specification and Lin's Concordance Correlation Coefficient (CCC)[4], which jointly evaluates correlation and deviation from perfect agreement. All tests are conducted for overall ESG scores and sub-pillars, using both raw and standardized measures.

Robustness Checks

To examine whether observed discrepancies are driven by scale, distributional shape, or industry composition, all consistency and reliability tests are repeated using alternative standardization schemes. In particular, I compute score differences based on percentile-standardized ESG measures:

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

and replicate the analysis using year-specific min–max normalization. Results are consistent across specifications in such a way that the cross-system disagreement is not the result of scaling or industry composition but is indicative of deeper

[1] Min–max scaling preserves the relative shape of the distribution and is widely used in ESG harmonization frameworks [9].
[2] Pearson correlation captures linear dependence, while Spearman correlation evaluates monotonic ranking consistency [10].
[3] I adopt the ICC (1,1) "absolute agreement" form following McGraw and Wong [11], which evaluates whether two measurement systems provide interchangeable numerical values rather than merely correlated rankings.
[4] Lin [12] proposed the CCC specifically for assessing measurement agreement by combining precision (correlation) and accuracy (closeness to the 45 degree line).

methodological differences between the two ESG rating systems

Empirical Comparison of Cross-System Differences
System-Level Differences: Scale, Trend, and Stability
Rating Scale and Temporal Patterns

To assess whether CNRDS and CSMAR capture a common ESG construct, I examine their score dynamics from 2015 to 2020. Table I and Fig I show large and persistent differences in both level and time-series behavior across pillars. First, CSMAR assigns substantially higher ESG levels than CNRDS throughout the sample, with a stable average gap of roughly 20 points. The persistence of this level wedge suggests systematic scaling and benchmark differences rather than transitory noise, implying that raw-score–based analyses (e.g., regressions, portfolio sorts, benchmarking) can yield materially different conclusions depending on the data source. Second, the two providers display distinct temporal patterns. CNRDS exhibits smooth, monotonic increases across pillars, whereas CSMAR shows pronounced

Table I. System-Level ESG Differences between CNRDS and CSMAR (2015–2020)

Panel A. Average Scores and Differences (2015–2020)

Dimension	CNRDS	CSMAR	Difference (CNRDS – CSMAR)
ESG	28.215	48.224	–20.009
Environmental	11.489	23.201	–11.712
Social	28.386	23.138	+5.248
Governance	37.784	18.937	+18.847

Panel B. Structural Break Evidence (Pre-2020 vs. 2020)

Dimension	Mean Difference (2015–2019)	Difference in 2020
ESG	–19.83	–20.89
Environmental	–6.82	–37.17
Social	+11.07	–23.88
Governance	+24.39	–8.85

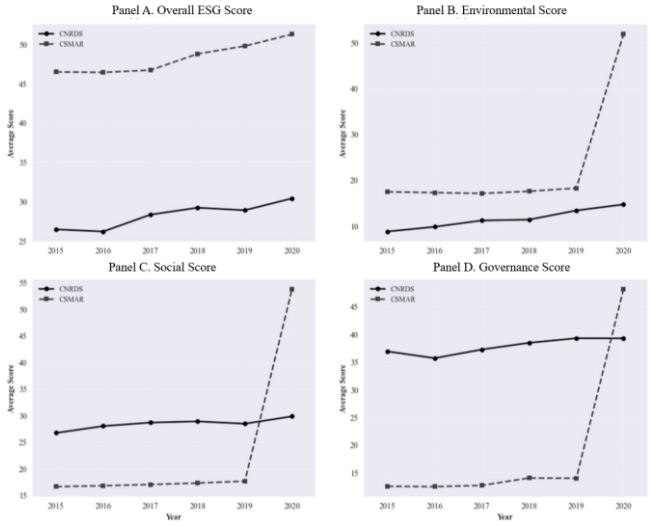

Fig I. Temporal Trends in ESG Scores: CNRDS vs. CSMAR Databases

breaks—most notably a synchronized jump in Environmental and Social scores in 2020 following relatively flat pre-2020 trajectories. Such discontinuities are more consistent with methodological recalibration than gradual changes in underlying ESG fundamentals.

Next, the Governance pillar displays a sharp regime shift: CSMAR's pre-2019 governance scores are far below CNRDS, but converge sharply by 2020, consistent with a substantial update in governance indicator definition and/or weighting. Overall, the combined evidence indicates that the two systems embed different scoring logics and differ in temporal stability, raising concerns for designs requiring longitudinal consistency unless harmonization is explicitly justified and implemented.

Industry-Level Differences in Score Levels

I next compare multi-year industry averages (Table II and Fig II). CSMAR assigns higher ESG scores in nearly all

industries, with an average gap comparable to the system-level wedge, indicating that industry composition alone cannot explain divergence. However, the magnitude of differences varies substantially across sectors: high-technology and highly regulated industries exhibit the largest gaps, consistent with greater methodological discretion when ESG signals rely on qualitative or semi-structured disclosures (e.g., data governance, compliance, operational risk). Moreover, CNRDS shows greater cross-industry dispersion, whereas CSMAR's industry distribution is more compressed, implying different discriminative power and potentially different implied industry rankings. Figure 2 further shows widening post-2018 disparities in CSMAR, consistent with system-wide recalibration rather than heterogeneous firm behavior.

Industry-Level Divergence and Agreement

Cross-Industry Divergence Patterns

To move beyond averages, I examine industry-by-pillar divergence (Table III). The direction and magnitude of gaps are systematic rather than random. CSMAR tends to rate capital-intensive and resource-extraction industries higher—especially in Environmental metrics—while several service-

Table II. Selected Industry-Level Comparison of ESG Scores: CNRDS vs. CSMAR
Panel A. Industry-Level ESG Averages and Differences

Rank	Industry	CNRDS	CSMAR	Difference
1	Construction	35.28	47.80	−12.53
2	Mining	34.83	50.23	−15.40
3	Raw Materials Manufacturing	32.21	47.89	−15.68
4	Health & Social Work	31.35	43.44	−12.09
5	Leasing & Business Services	30.63	48.25	−17.62
6	Electricity, Heat, Gas & Water Supply	30.50	50.50	−20.00
7	Agriculture, Forestry & Fishery	30.14	50.98	−20.84
8	Manufacturing	30.12	48.87	−18.75

Panel B. Cross-Industry Distribution Summary

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

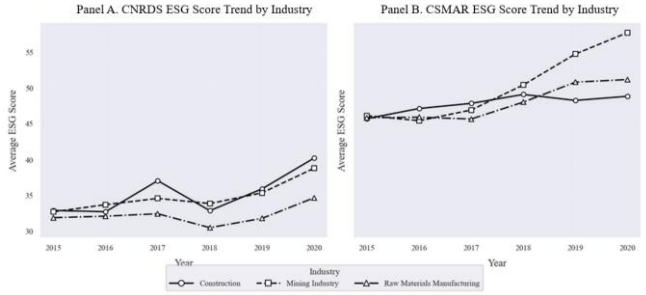

Fig II. ESG Score Trends by Industry: CNRDS vs. CSMAR

oriented industries receive lower scores relative to CNRDS. Divergence is strongest in Environmental and Social pillars, while Governance differences are smaller and more mixed, indicating pillar-specific methodological priorities.

Within-Industry Consistency and Reliability

I then test firm-level agreement within industries using regressions, ICC, and CCC (Table IV). Across pillars, the systems exhibit near-zero explanatory power ($R^2 \approx 0$) and negligible slopes, including negative slopes in the Social pillar. Absolute agreement is also poor: ICC values remain below 0.10, and CCC values are near zero, indicating that the two providers are not interchangeable measures of the same construct, even within the same industry peer group.

Table III. Selected Industry-Level ESG Rating Differences between CSMAR and CNRDS
Panel A. Overall ESG Score Differences

Industry	n	Difference	t-stat	Wilcoxon-p	Sig.
Postal Services	11	−0.547	−7.09	0.0019	***
Manufacture of Chemicals & Chemical	31	−0.395	−5.95	0.00001	***

Products					
Support Activities for Mining	11	+0.377	+4.32	0.0049	**
Smelting & Rolling of Ferrous Metals	26	+0.291	+4.03	0.00047	***

Panel B. Environmental (E) Dimension

Industry	n	Difference	t-stat	Wilcoxon-p	Sig.
Support Activities for Mining	11	+0.466	+4.95	0.0029	**
Smelting & Rolling of Ferrous Metals	26	+0.272	+3.18	0.0073	**
Processing of Agricultural & Sideline Food	20	−0.498	−7.29	0.000013	***
Postal Services	11	−0.280	−2.53	0.032	*

Panel C. Social (S) Dimension

Industry	n	Difference	t-stat	Wilcoxon-p	Sig.
Postal Services	11	−0.560	−8.85	0.00098	***
Textile, Apparel & Accessories	8	−0.375	−6.22	0.0078	***
Manufacture of Chemicals & Chemical Products	31	−0.346	−4.87	0.00020	***
Water Transport	29	+0.326	+4.64	0.00026	***

Panel D. Governance (G) Dimension

Industry	n	Difference	t-stat	Wilcoxon-p	Sig.
Postal Services	11	−0.444	−4.85	0.00195	**
Smelting & Rolling of Ferrous Metals	26	+0.388	+2.42	0.109	*
Construction	6	−0.453	−3.55	0.043	*
Postal Services	11	−0.444	−4.85	0.00195	**

Table IV. Firm-Level Agreement between CNRDS and CSMAR Ratings

Dimension	N	β (Slope)	R^2	ICC	CCC
ESG	1783	0.025	0.002	0.009	0.009
E	1783	0.154	0.017	0.093	0.093
S	1783	–0.050	0.002	0.000	–0.042
G	1783	0.060	0.003	0.031	0.031

Standardization does not resolve this disagreement: neither rescaling nor rank-based transformations materially improve correlation or agreement, implying that divergence reflects indicator choice and weighting differences rather than distributional scaling.

Why Standardization Cannot Reconcile Divergence

I evaluate common harmonization approaches—min–max scaling, percentile transformation, and industry adjustment. Three findings emerge. First, prior normalization results show that standardized scores retain systematic directional bias, with CSMAR assigning higher values—particularly in the Environmental and Social dimensions. Second, cross-system co-movement remains negligible ($|\rho| < 0.10$), and the Social pillar frequently exhibits negative association, indicating disagreement in firm ordering rather than level alone (Table V). Third, categorical alignment is weak: quintile overlap remains close to random, and chance-adjusted agreement is near zero, casting doubt on portfolio sorts and threshold-based ESG classifications that rely on a single provider (Table VI). To explain why normalization fails, I examine distributional properties (Fig III). CNRDS exhibits wider dispersion in Environmental and Social scores, while CSMAR displays upward shifted and more compressed distributions, consistent with different discriminative architectures. These differences persist across normalization schemes, reinforcing that divergence originates from rating design rather than scale choice.

Finally, divergence depends on the sector (Fig IV). Firms with qualitative, discretion-intensive ESG disclosure (e.g., Finance, IT, Real Estate) exhibit the highest gaps, while those reporting relatively objective metrics (e.g., mining-related industries) show smaller divergence, in line with variation in indicator elasticity and measurement discretion.

Table V. Cross-System Correlation of Min–Max Normalized ESG Ratings

Dimension	Pearson ρ (CSMAR–CNRDS)
ESG	0.020
E	0.047
S	–0.086
G	0.092

Table VI. Quantile-Based Agreement between CNRDS and CSMAR

Dimension	Exact Quintile Match	Cohen's κ	Rank Corr. (ρ)
ESG	0.20	–0.00	0.03
E	0.22	0.02	0.12
S	0.19	–0.01	–0.06
G	0.22	0.02	0.11

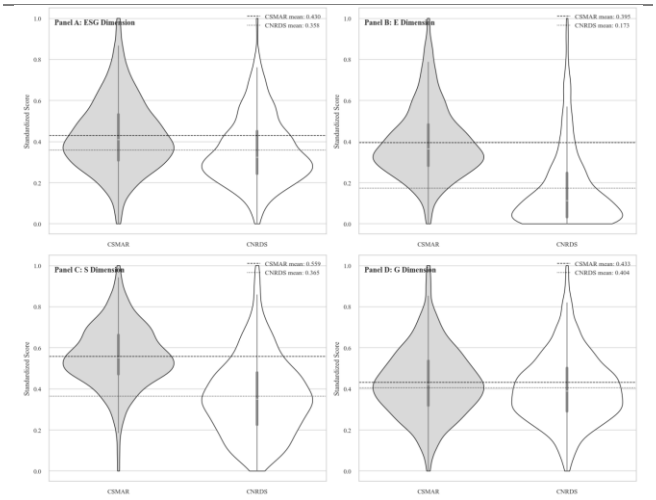

Fig III. Distribution Comparison of CSMAR and CNRDS Ratings

Section IV establishes that CNRDS–CSMAR divergence is systematic and structurally grounded across levels, trends, industries, and distributional properties, and that standard normalization fails to harmonize the two systems. Section V therefore conducts targeted robustness tests—across time, alternative statistical procedures, and aggregation levels—to confirm that these findings are not driven by specific sample partitions or methodological choices.

Robustness Tests

Temporal Stability of Correlation Patterns

To assess whether cross-system divergence is driven by specific subperiods—especially CSMAR's apparent recalibration in 2020—I examine year-by-year Pearson and Spearman correlations from 2015 to 2020 (Table VII). The evidence rejects a time-specific explanation: correlations remain consistently weak throughout the sample, with no systematic improvement in pre-2020 years. The Social pillar exhibits the weakest alignment and is frequently negative, indicating persistent disagreement in firm ordering rather than a one-off break.

Firm-Level Difference Tests

I next test whether the documented level differences are sensitive to statistical assumptions by applying paired t-tests and Wilcoxon signed-rank tests at the firm–year level (Table VIII). Both procedures deliver the same directional conclusions ($p < 0.001$): CSMAR assigns higher Environmental and overall ESG scores (≈ 11.7 and ≈ 20 points), while CNRDS assigns higher Social and Governance scores (≈ 3.2 and ≈ 15.5 points). This consistency across parametric and non-parametric tests indicates that the results are not driven by distributional features or outliers. After percentile standardization, mean gaps mechanically vanish by construction, but rank-based disagreement remains pronounced—confirming that divergence reflects conceptual weighting differences rather than scale alone.

Industry-Level Aggregation Tests

If the divergence were idiosyncratic noise at the firm level, industry aggregation should attenuate it. Instead, differences persist—and often intensify—when comparing industry means. CSMAR continues to rate Environmental performance higher, while CNRDS rates Governance higher across most industries; the largest absolute gaps (often > 20 points) occur in Finance, IT Services, and Water Production & Supply, where ESG measurement relies more on qualitative judgment

(Table IX). The magnitude of disagreement follows a clear industry gradient: it is largest

Table VII. Year-by-Year Rank Correlation between CSMAR and CNRDS

Year	ESG	E	S	G
2015	−0.002	0.050	0.034	0.018
2016	0.035	0.009	−0.042	0.075
2017	−0.004	0.070	−0.035	0.150
2018	0.011	0.117	0.024	0.194
2019	0.065	0.221	−0.038	0.196
2020	0.079	0.077	−0.084	0.007

Table VIII. Firm-Level Mean Differences between CSMAR and CNRDS

Panel A. Raw Scores

Dimension	Mean Difference (CSMAR − CNRDS)	t-test ***p***-value	Direction
ESG	+20.05	<0.001	CSMAR >CNRDS
E	+11.71	<0.001	CSMAR > CNRDS
S	−3.18	<0.001	CNRDS > CSMAR
G	−15.55	<0.001	CNRDS > CSMAR

Panel B. Percentile-Standardized Scores

Dimension	Mean Difference	***t***-test ***p***-value
ESG	≈ 0	1.000
E	≈ 0	1.000
S	≈ 0	1.000
G	≈ 0	1.000

Table IX. Industry-Level Differences in Raw ESG Scores (CSMAR − CNRDS)

Dimension	Industries with Largest Absolute Differences	Mean Difference	Direction
ESG	Finance; IT Services; Water Production & Supply	+20 to +25	CSMAR > CNRDS
E	Health; IT Services; Finance	+16 to +28	CSMAR > CNRDS
S	Construction; Mining; Utilities	−15 to −19	CNRDS > CSMAR
G	Real Estate; Construction; Finance	−17 to −29	CNRDS > CSMAR

where indicator discretion is greatest and smallest where metrics are more standardized and physical.

Comprehensive Standardization Assessment

Finally, I evaluate whether standardization can reconcile the two systems within tighter subsamples (year-by-year and year–industry adjustments). Standardization continues to fail: even when mean differences are removed within year–industry cells, cross-system rank alignment remains extremely weak (typically < 0.10). The 2020 jump in CSMAR's Social and Governance scores relative to CNRDS is consistent with methodological realignment rather than noise, but weak agreement is present in every year, not only in 2020. Overall, the robustness battery confirms that the divergence is persistent across time, aggregation levels, test procedures, and normalization schemes (Table X).

Together, Sections IV and V establish that the divergence is not a scaling artifact, a sample peculiarity, or an episodic recalibration effect. The natural next step is therefore explanatory: what in the rating architecture causes such stable disagreement? Section VI links the empirical patterns to differences in philosophy, indicator construction, and industry customization.

Discussion

Sources of Divergence between CNRDS and CSMAR

The results are best explained by differences in rating design rather than measurement error.

First, the two systems embed different methodological philosophies: CSMAR's bottom-up, data-driven approach aggregates extensive disclosure items and therefore rewards the presence and completeness of policies and management processes, while CNRDS's top-down, framework-driven

Table X. Industry-Level Differences between CNRDS and CSMAR Ratings

Dimension	2015–2019 Mean	2020 Mean	Direction
ESG	≈ +20	≈ +22	CSMAR > CNRDS
E	≈ +6	≈ +37	CSMAR > CNRDS
S	≈ −10	≈ +29	Reversal in 2020
G	≈ −21	≈ +11	Reversal in 2020

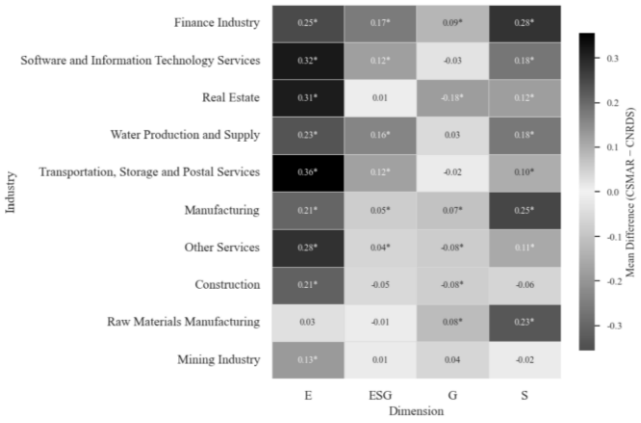

Fig IV. Industry-Level Rating Differences between CSMAR and CNRDS

design emphasizes performance outcomes, risk exposure, and alignment with international standards.

Second, indicator composition differs materially—most sharply in Social and Governance—so firms can rank highly under one construct yet poorly under the other, consistent with the near-zero (and sometimes negative) cross-system correlation, especially in S. Third, industry customization amplifies divergence: where ESG is more qualitative and discretion-intensive (e.g., Finance and IT), differences in indicator choice and weighting translate into larger cross-system gaps; where metrics are more physical and standardized, alignment improves but remains limited. These mechanisms jointly explain why standardization fails: the disagreement is not "how to scale the same signal," but "which signal is being measured."

Theoretical Implications

These findings reinforce the view that ESG ratings are constructed measures shaped by methodological priorities, not neutral readings of a single underlying "true ESG." Empirically, provider choice becomes a first-order research design decision: results and inference can shift simply because the construct differs across databases. This offers a structural explanation for mixed findings in ESG–return and ESG–risk studies and highlights a replication challenge in ESG finance when studies rely on different providers without explicit justification. The industry- and pillar-specific patterns further suggest that disagreement is systematically concentrated where qualitative judgment and discretionary weighting are most influential.

Practical Implications

For investors, the evidence implies that CNRDS and CSMAR are not interchangeable inputs; mixing or averaging them can create internally inconsistent signals. Instead, rating choice should be matched to purpose—for example, management-system and compliance orientation versus outcome- and risk-oriented assessment. For firms, the results suggest that "optimizing for all ratings" is infeasible; a more robust strategy is to focus on industry-material ESG issues and transparent improvements rather than headline score chasing. For regulators and data providers, full convergence may be unrealistic, but greater transparency on indicator definitions, weighting, and aggregation would materially improve interpretability and comparability for markets and researchers.

Conclusion

This research provides the first systematic, firm-level comparison of ESG ratings from CNRDS and CSMAR for Chinese A-share firms from 2015 to 2020. The evidence shows persistent divergence in levels, trends, distributions, and—most importantly—firm rankings: CSMAR assigns substantially higher ESG scores, while cross-system correlations and agreement metrics remain extremely weak and are often negative in the Social pillar. Standard harmonization methods, including min–max scaling, percentile ranking, and industry adjustment, fail to produce meaningful convergence, indicating that the disagreement is structural rather than a superficial scaling artifact. Robustness checks confirm that these patterns persist across time, statistical procedures, and aggregation levels.

The findings imply that CNRDS and CSMAR operationalize distinct ESG constructs rooted in different rating philosophies, indicator compositions, and industry customization strategies. For researchers, database choice is therefore a substantive methodological decision with direct consequences for inference, comparability, and replication. For investors and firms, the results caution against treating ESG scores as interchangeable labels and motivate a more purpose-aligned and industry-material interpretation of ESG information. More broadly, the Chinese ESG landscape is not converging toward a single "correct" score; maturity instead requires transparency, clarity of methodological intent, and sophistication in interpreting multiple ESG constructions.

REFERENCES

[1] 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
[2] 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
[3] Dimson, E., Marsh, P., & Staunton, M. (2020). Divergent ESG ratings and their implications for investors. Credit Suisse Global Investment Returns Yearbook 2020, 1–20.
[4] 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
[5] 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
[6] 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
[7] 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
[8] 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
[9] OECD (2020). ESG Data Harmonization Report. OECD.
[10] Kendall, M. G. (1970). Rank correlation methods. Charles Griffin.
[11] McGraw, K. O., & Wong, S. P. (1996). Forming inferences about intraclass correlation coefficients. Psychological Methods, 1(1), 30–46.
[12] Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45(1), 255–268.