

ESG Rating Divergence and Corporate Financial Performance: A Review of the Measurement Problem

Siqiang An*, Mengyun Wei†

Farabi International Business School, Al-Farabi Kazakh National University, Almaty 050040, Republic of Kazakhstan

*Corresponding email: siqiangan@gmail.com, †Corresponding email: wmy210327@gmail.com

Abstract

One of the best-tested hypotheses in sustainable finance concerns the relation between environmental, social, and governance (ESG) performance and corporate financial performance (CFP). There are well over 2,000 published papers in the last decade, but the results are wildly conflicting, ranging from positive to null to negative, and these findings can arise within the same data sample and methodology. A weak theory cannot explain the discrepancy because three channels (cost of capital, operating efficiency, and valuation premium) lead clearly in the same direction, namely, a positive association. The more likely reason is the poor quality of the ESG measurement. Prior studies document that the pairwise correlations among the six major rating providers covering the same set of companies range between 0.38 and 0.71 with a mean of 0.54, compared to the correlation of 0.99 for Moody's and S&P credit ratings on the same issuers. Such measurement errors can bias the estimated effect towards zero and make empirical results dependent on which rating provider is used by the authors. Further evidence shows that greater ESG disclosures can lead to wider, rather than narrower, rating divergence, challenging the premise of the recent mandated disclosures. This paper reviews the theoretical rationale for why ESG performance affects financial performance and provides an econometric analysis of how the divergence of ESG ratings occurs and its consequences. It also evaluates the limits of recent developments in regulation, including the International Financial Reporting Standards (IFRS) Sustainability Disclosure Standards (IFRS SDS) and the EU's Corporate Sustainability Reporting Directive (CSRD). A section is devoted to the Hong Kong market, where both international and mainland Chinese rating agencies rate H-share issuers under different methodologies, producing institutionally meaningful divergence with portfolio relevance. The paper's central claim is that the next advance in ESG-CFP research will come not from more elaborate regression specifications but from resolving the reliability of the measurement instrument itself.

Keywords

ESG ratings, Rating divergence, Corporate financial performance, Disclosure paradox, Sustainable finance, Hong Kong market

Introduction

Over the past twenty years, the question of whether ESG performance improves corporate financial performance has been examined extensively. Early empirical results pointed to a positive average correlation across more than fifty instances. Later, a larger set of over 2,200 observations indicated that roughly 90 per cent showed a non-negative relationship. More recent data covering the period up to 2020 confirmed this pattern. Given such consistency across three independent collections of evidence over nearly two decades, one might have expected the debate to be resolved. Yet disagreement has

actually intensified over the past five years.

That combination-near-consensus in meta-analytic summaries alongside persistent and sharpening controversy-points to a possibility worth taking seriously. This debate is arguably less about theory or statistics, and more about the quality of the primary independent variable, that is, ESG ratings. Ratings of the same companies do not converge across the main corporate social performance rating systems. Pairwise correlations of 0.38-0.71 among rating agencies that rate the same firms have been reported, versus nearly 0.99 for

corporate credit ratings by Moody's and Standard & Poor's. Put differently, a single company could hold drastically different ranks according to different ESG rating systems. The presence of measurement error in an independent variable of this size leads to attenuation bias in estimations based on it and renders comparisons across different rating systems difficult [1].

Christensen et al. added another wrinkle to the narrative, finding that higher ESG disclosure correlated with increased, not reduced, divergence among rating agencies [2]. This challenges the assumption underlying current regulation that increasing disclosure reduces information asymmetry and thus reduces the gap between rating agencies.

In the following sections, the potential channels through which ESG performance affects financial performance are discussed, providing a baseline to evaluate empirical outcomes. The source and magnitude of rating divergence and its implications are then examined. In particular, measurement is treated not as a side issue, but as the central concern.

It should be noted at the outset that major regulations like those from the International Sustainability Standards Board (ISSB) and the European Union's Corporate Sustainability Reporting Directive (CSRD) have been developed to solve for information asymmetry. They are not designed to tackle issues around ESG rating methods. The Hong Kong market, however, offers a concrete setting in which the consequences of that gap can be observed directly.

The paper begins by discussing the theoretical background, then moves on to explore the issue of rating divergence and its impact on the literature. Next, it examines the paradoxical effect of ESG disclosure on rating divergence, followed by an analysis of regulatory responses and the Hong Kong market. Finally, it concludes by outlining a research agenda.

Theoretical mechanisms linking ESG to financial performance

Theory's own predictions deserve a brief initial statement before the discussion turns to empirical measurement. Three central mechanisms frequently recur across the relevant literature.

Cost of capital

The first channel is that ESG performance can reduce a

firm's cost of capital by reducing perceived firm risk and increasing the number of investors willing to own a firm's securities. U.S. firms with higher corporate social responsibility scores receive a significantly lower cost of equity capital. Most of the reduction occurs in the areas of employee relations, environmental management, and product quality. Good environmental practices mitigate the likelihood of environmental fines and lawsuits. Sound labor practices reduce the risk of public scandal. Solid corporate governance curbs the ability of insiders to steal from or deceive shareholders. At the same time, ESG-oriented capital worldwide now exceeds USD 35 trillion, and the systematic inflow of this capital compresses the financing premium demanded of well-rated firms. Green lending programs at several banks have already translated into interest-rate discounts of several dozen basis points.

Operating efficiency

The second channel runs through operating efficiency. Resource conservation, more effective use of human capital, and better-aligned management incentives are the usual mechanisms cited, and each has a plausible route to the bottom line. This channel can be traced to internal organizational processes.

Firms that voluntarily adopted sustainability policies in the early 1990s subsequently developed more effective stakeholder engagement and more long-term-oriented governance than their peers, and outperformed them financially over an eighteen-year horizon.

Environmental measures such as emissions reduction and energy efficiency cut costs directly. Social measures such as employee welfare and workforce diversity reduce turnover and raise output per employee. Governance measures such as board independence and compensation transparency redirect managerial attention from short-term earnings toward long-term value creation. These channels should ultimately register as improvements in return on assets and return on equity.

Valuation

The third channel is one of valuation. Markets may place higher valuation multiples on ESG-performing firms because of expected future regulatory compliance, reduced reputational risk, and enhanced stakeholder relationships. These are discounted to present value, rather than solely on current cash flows. Bolton and

Kacperczyk showed strong evidence for such a mechanism in the context of carbon pricing [3]. A carbon premium exists in equity returns, where firms with higher direct carbon emissions tend to have higher subsequent stock returns. Investors require compensation for exposure to transition-related risk, as they expect higher returns for bearing unpriced risk associated with a shift to a lower-carbon economy. Firms with fewer carbon emissions have a lower tail risk on future cash flows and therefore should have a lower required rate of return and a higher valuation. All three channels imply a positive relationship between ESG and financial performance, yet empirical evidence regarding the association between the two has proven mixed and inconsistent. The primary reason for this gap lies in the limitations of how ESG is measured rather than in the theoretical understanding of ESG-finance links.

Rating divergence: The core measurement problem

The magnitude of divergence

Ratings issued by six leading providers - Kinder, Lydenberg, Domini & Co. (KLD), Sustainalytics, Moody's ESG, Refinitiv, Morgan Stanley Capital International (MSCI), and S&P Global - were compared across the same sample of listed firms. Their pairwise correlations ranged from 0.38 to 0.71, with a mean of approximately 0.54. Billio et al., examining a largely similar set of providers, report an equally broad spread. These providers also demonstrate that portfolios constructed on ostensibly the same ESG logic can vary considerably in both their holdings and their returns simply by switching the data source [4].

Credit ratings offer a natural yardstick: When Moody's and Standard & Poor's rate the same issuer, their assessments typically correlate at about 0.99. That striking contrast is not merely an arcane methodological footnote. It directly affects the credibility and generalizability of any study that treats an ESG score as an independent variable. Because the provider is typically selected on practical grounds rather than theoretical ones, that choice alone can leave a measurable imprint on the findings.

Decomposing the sources of divergence

The aggregate divergence can be traced to three distinct sources, an exercise useful for locating the root of the problem.

Measurement divergence makes up the largest share, at 56 percent. It emerges when agencies define the same conceptual indicator differently in practice. A useful illustration is carbon emissions: Providers take opposing positions on including scope 3, on using self-reported figures versus third-party estimates, and on standardizing by absolute output rather than intensity. These operational choices can produce substantially different scores even when the firms' underlying footprints are identical.

Scope divergence accounts for 38 percent and is more foundational: Agencies simply do not assess the same basket of attributes. One provider may work with slightly over 100 indicators, another with more than 300. The relative emphasis placed on environmental, social, and governance pillars also varies across the board. When the evaluation sets differ so fundamentally, a gap of some sort is practically guaranteed.

The remaining 6 percent, weight divergence, stems from unequal weighting applied to an otherwise identical set of indicators.

The decomposition has a practical bearing on remedies. Disagreement is driven chiefly by decisions over what gets measured and how, rather than by how individual indicators are weighted. That implies adjustments limited to weights, or adding more indicators, will not by themselves close the gap.

Consequences for ESG-CFP research

Standard results on measurement error warn that random noise in an explanatory variable pulls regression coefficients toward zero. ESG ratings are exactly that type of regressor. Consequently, the weak and often insignificant coefficients could reflect nothing more than an attenuating bias that compresses the estimated relationship, not the genuine absence of one.

Using different providers can add another layer of distortion. An article relying on MSCI data can reach conclusions that seemingly contradict one based on Sustainalytics data, even when the two are probing the same underlying relationship. To a meaningful degree, then, the fragmented state of the empirical evidence on ESG and CFP reflects a systematic measurement artifact rather than true variability in the economic effect.

The fallout is not limited to academic regressions. Gibson Brandon et al. reported that stocks with the

greatest disagreement across ESG raters subsequently deliver higher returns, a pattern consistent with the idea that investors demand extra compensation for bearing rating-related uncertainty [5]. Avramov et al. reached a complementary finding: Uncertainty about ratings blunts, and in some plausible cases reverses, the portfolio gains investors expected from tilting toward high-ESG firms [6]. The point is pushed further, questioning on governance grounds whether the observed dispersion leaves aggregate ESG ratings with any reliable informational content for capital allocation.

None of this means ESG ratings carry no information at all. The noise-versus-signal question has been addressed by arguing that a large part of the observed divergence behaves like classical measurement noise rather than like irreconcilable disagreement over what firms are really worth. Using one provider's score as an instrumental variable for another can undo much of the attenuation: The corrected effects of ESG on stock returns are several times larger than the uncorrected ones. That view is consistent with the argument developed here, though it points to a different terrain. An instrumental-variable approach assumes that providers are noisy readings of one underlying truth. It has little to say when agencies are actually measuring different constructs, which is precisely what the earlier decomposition suggests. The correction restores a lost signal. It does not, however, supply the missing coordination in how ESG is defined and evaluated in the first place. Thus the two positions complement each other rather than conflict.

The disclosure paradox: Testing the transparency assumption

The intuitive policy logic and its limits

If rating divergence is attributed to insufficient information, the natural policy response is to require firms to disclose more ESG information and to standardize how they do so. This logic underlies the mandatory-disclosure rules recently adopted by the ISSB, the CSRD, and several stock exchanges. The implicit reasoning runs as follows: Greater disclosure reduces information asymmetry, which brings the information basis for ratings closer together, and this convergence in turn narrows divergence.

Contrary evidence

This assumption was tested directly. The results show

that higher levels of ESG disclosure are associated with a statistically significant widening, rather than a narrowing, of rating divergence. Kimbrough et al. reach a closely related conclusion using a different research design and sample, finding that voluntary ESG reporting fails to resolve, and in some specifications amplifies, disagreement among rating agencies [7]. This is evidence that the disclosure paradox is not an artifact of a single methodology.

This counterintuitive result stems from the qualitative character of ESG information. Financial accounting information is constrained by standardized recognition and measurement rules under frameworks such as Generally Accepted Accounting Principles (GAAP) or IFRS. Most ESG indicators lack any comparable hard constraint. What counts as sound environmental management or adequate social responsibility depends inherently on the evaluator's values and stakeholder orientation. When a firm releases more data, rating agencies gain more material for independent interpretation rather than a common informational basis that would compel convergence. Additional disclosure expands, rather than compresses, the space available for interpretive disagreement.

Implications for regulatory design

The disclosure paradox amounts to an empirical correction to regulatory strategies built primarily around mandatory disclosure. It does not imply that disclosure is without value; disclosure remains foundational to investor decision-making and public accountability. It does suggest, however, that regulators should not expect disclosure alone to resolve rating divergence. Meaningfully narrowing divergence will require deeper coordination on methodology and on the definition of core indicators, not merely more reporting. This is arguably the harder and more consequential fix, precisely because it requires coordination among competitors rather than compliance with a single rulebook.

Regulatory frameworks and the Hong Kong market

The contribution and limits of the ISSB and the CSRD

The ISSB issued IFRS S1 (General Requirements for Disclosure of Sustainability-related Financial Information) and IFRS S2 (Climate-related Disclosures) on 26th June 2023, establishing the first global baseline

for investor-oriented sustainability disclosure. The EU's CSRD, operationalized through the European Sustainability Reporting Standards (ESRS), further specified disclosure indicators. Directive (EU) 2026/470 of 24th February 2026 - the Omnibus amending directive - was published in the Official Journal on 26th February 2026 and entered into force on 18th March 2026. It subsequently simplified elements of that framework, principally by narrowing its scope and reporting thresholds.

These frameworks concentrate on standardizing disclosure: specifying what firms must report and in what format. Standardized disclosure is not equivalent to standardized evaluation. Even if every firm reported under an identical ESRS template, rating agencies reading the same report could still assign different scores because of differences in evaluation methodology. Neither the ISSB nor the CSRD has yet addressed the coordination of rating-agency methodology itself.

The distinctive position of the Hong Kong market

Hong Kong, as the principal offshore listing venue for mainland Chinese H-share issuers and a gateway for international capital entering the Chinese market, sits at the intersection of two distinct ESG evaluation systems. Zhu et al. documented exactly this pattern among Chinese listed firms, finding that ratings assigned by domestic Chinese agencies and by major international rating agencies frequently diverge and, for a nontrivial share of firms, move in opposite directions over time [8]. Zhang and Geng extended this finding to firm valuation, showing that the gap between domestic and foreign ESG ratings is itself associated with measurable differences in how the market prices the affected firms [9]. International rating agencies, among them MSCI, Sustainalytics, and S&P Global, and mainland Chinese rating agencies frequently produce systematically divergent assessments of the same H-share issuer.

This divergence is especially pronounced along the governance dimension. International rating frameworks emphasize board independence, minority shareholder protection, and executive compensation transparency; domestic frameworks place greater weight on state-ownership status, policy compliance, and the discharge of social responsibility. Identical governance arrangements can be read in materially different ways

depending on which framework is applied. This is not a technical error but a consequence of the different institutional expectations and stakeholder structures embedded in the two evaluation systems.

For institutional investors active in the Hong Kong market, this divergence carries direct allocation consequences. Whether a portfolio manager is justified in reducing exposure to a stock that scores poorly under international ratings but well under domestic ratings remains an open question. A 2024 consultation on climate-related disclosures introduced mandatory reporting requirements under the Main Board Listing Rules on a phased basis from 1st January 2025. This continues the trend of expanding disclosure. On the logic of the disclosure paradox, however, the additional disclosure this reform generates is unlikely to narrow the gap between the two rating systems. It may instead give both systems more material for systematically divergent interpretation of the same public data.

The limited role of artificial intelligence

Some researchers and technology developers have proposed using natural language processing and machine learning to automate the extraction and standardization of ESG data, reducing the divergence introduced by manual processing. A systematic review of this literature covering forty-five studies published between 2020 and 2025, documents rapid growth in such applications, ranging from text-mining tools that parse disclosure documents to machine-learning models that generate ESG scores directly from unstructured data [10]. Such tools can assist with data collection and cleaning, but they do not reach the deeper source of divergence: the value-laden nature of the judgment involved in defining good ESG performance in the first place. An algorithm still reproduces the evaluative logic built into it by its designers; it cannot substitute for substantive coordination among evaluation frameworks [11].

Conclusion

This paper's central claim is that the empirical controversy over the ESG-CFP relationship is, to a substantial degree, a measurement problem rather than a theoretical or methodological one. An average pairwise correlation of only 0.54 among six major rating agencies is sufficient, on its own, to place any empirical study relying on a single rating source on a fragile

identificational footing. Measurement error in the independent variable attenuates estimated coefficients, and findings shift as the rating source changes, a mechanism sufficient to account for much of the weak association and inconsistent conclusions reported across this literature.

The disclosure paradox further shows that policies relying on mandatory disclosure to narrow rating divergence lack empirical support. Disclosure addresses the availability of information, but it does not address the unification of evaluation frameworks.

Four adjustments to current practice follow from this diagnosis. Empirical analyses should draw on multiple rating sources or construct composite indices, and test how sensitive their results are to that choice. Rating divergence itself deserves treatment as an object of analysis in its own right. Examining how the degree of divergence affects the predictive power of ESG measures for financial performance may offer more incremental value than repeating tests of whether ESG matters at all. Treating disagreement itself as an object of asset-pricing analysis, rather than as noise to be averaged away, aligns with emerging research directions. Methodological reporting should also be more transparent about which rating source is used, with explicit discussion of how conclusions might change under an alternative source. The Hong Kong market merits treatment as an independent research setting in which to examine how divergence between international and domestic ratings affects capital flows, valuation, and corporate behavior in practice.

The regulatory priority should extend beyond expanding disclosure requirements toward substantive coordination among rating agencies on core methodology and key indicator definitions. Recent evidence on rating-agency competition offers a cautionary note: Competitive pressure among ESG raters does not straightforwardly improve rating quality and can, under some conditions, degrade it. This suggests that market forces alone are unlikely to resolve the coordination problem this paper identifies. Even where full unification proves unattainable, disclosing the structure and sources of inter-agency divergence to the market would allow investors to interpret rating information on a better-informed basis.

A genuine positive relationship between ESG performance and financial performance may well exist; theory and a substantial body of non-experimental evidence point in that direction. Confirming it, however, requires a reliable measurement instrument. The next stage of progress in this field lies not in more elaborate model specifications or estimation techniques, but in the foundational work of fixing the measurement itself.

Funding

This work was not supported by any funds.

Acknowledgements

The authors would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] Berg, F., Kölbel, J. F., Rigobon, R. (2022) Aggregate confusion: the divergence of ESG ratings. *Review of Finance*, 26(6), 1315-1344.
- [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.
- [3] Bolton, P., Kacperczyk, M. (2021) Do investors care about carbon risk? *Journal of Financial Economics*, 142(2), 517-549.
- [4] Billio, M., Costola, M., Hristova, I., Latino, C., Pelizzon, L. (2021) Inside the ESG ratings: (Dis)agreement and performance. *Corporate Social Responsibility and Environmental Management*, 28(5), 1426-1445.
- [5] Gibson Brandon, R., Krueger, P., Schmidt, P. S. (2021) ESG rating disagreement and stock returns. *Financial Analysts Journal*, 77(4), 104-127.
- [6] Avramov, D., Cheng, S., Lioui, A., Tarelli, A. (2022) Sustainable investing with ESG rating uncertainty. *Journal of Financial Economics*, 145(2), 642-664.
- [7] Kimbrough, M. D., Wang, X., Wei, S., Zhang, J. (2024) Does voluntary ESG reporting resolve disagreement among ESG rating agencies? *European Accounting Review*, 33(1), 15-47.
- [8] Zhu, Y., Yang, H., Zhong, M. (2023) Do ESG

- ratings of Chinese firms converge or diverge? A comparative analysis based on multiple domestic and international ratings. *Sustainability*, 15(16), 12573.
- [9] Zhang, Z., Geng, L. (2026) Does divergence in ESG ratings between domestic and foreign agencies impact firm valuation in China? *Asia-Pacific Journal of Accounting & Economics*, 33(3), 481-497.
- [10] Vilchez Olivares, P. A., Astorga De La Cruz, B. J. (2026) AI and data analytics in sustainable financial reporting and ESG disclosure: a systematic literature review. *Sustainability*, 18(11), 5393.
- [11] Chen, C., Dube, S., Froymovich, S. (2025) ESG rating competition and rating quality. *Journal of Accounting Research*, 63(5), 1995-2037.