Do ESG Ratings Penalize Regional Institutions? New Evidence from Original Data.

1. Motivation

ESG ratings increasingly operate as a de facto language for sustainability performance in capital markets and stakeholder governance. For financial institutions, this language matters not only reputationally but also because ESG signals feed into investor screening, stewardship priorities, and broader legitimacy dynamics. Yet the research consensus is that ESG ratings are far from a single, stable "measurement technology." Instead, they often embed distinct judgments about what to measure, how to measure it, and how to aggregate it into a score. This matters because once ESG ratings are treated as comparable, they can shape incentives and distribute attention, potentially advantaging some institutional archetypes over others.

A central empirical fact is that ESG ratings from different providers frequently disagree in economically meaningful ways. Berg et al. (2022) show that divergence is not merely noise; it is structurally produced by differences in scope (which issues enter the rating), measurement (how the same issue is quantified), and weighting (how issues are aggregated). Their decomposition is particularly important for applied work because it implies that even if two agencies appear to rate "ESG," they may effectively rate partially different objects. One implication given the latter might be that if methodology drives a substantial part of cross-provider disagreement, methodology can also generate systematic patterns across types of institutions. Disagreement is therefore not only a reliability problem; it can become a distributional one.

This paper focuses on one such distributional question in the EU context: do commercial ESG ratings systematically penalize regional financial institutions relative to non-regional institutions? "Penalize" is meant in an empirical sense: conditional on a transparent, disclosure-grounded benchmark of ESG fundamentals, do regionally anchored institutions receive systematically lower third-party ratings or systematically larger negative deviations from the benchmark? The question matters for two reasons. First, if penalization exists, it may indicate that rating methodologies disproportionately reward features correlated with global visibility, scale, or data richness rather than sustainability fundamentals. Second, in the EU, sustainability disclosure requirements have expanded and are evolving toward more structured reporting architectures, which should, in principle, improve comparability, yet the literature suggests that more disclosure does not automatically resolve rating disagreement and can even increase it (Christensen et al. 2022; Hummel and Jobst, 2024).

2. Literature review

Understanding why penalization could arise begins with the mechanics of rating divergence. Berg et al. (2022) propose a taxonomy for: (1) scope divergence, (2) measurement divergence, and (3) weight divergence. Their key insight is that measurement divergence contributes the most to disagreement, meaning providers often process the same nominal information differently, employ different estimations, and respond differently to missing data and controversy signals (Berg et al., 2022). For research design, this implies that "bias-like" outcomes can emerge without any intentional discrimination. If one provider systematically relies more on data sources

that under-cover certain institutions (for example, local-language media, locally reported indicators, or disclosures not captured by global data pipelines), then the methodology can generate systematic under-rating for those institutions. The paper's empirical strategy therefore treats penalization as an outcome consistent with methodological frictions rather than as proof of intent.

A second important insight concerns disclosure. A natural expectation is that greater disclosure should harmonize ratings by giving all agencies richer inputs. However, Christensen et al. (2022) challenge this intuition. They show that ESG rater disagreement is pervasive and that ESG disclosure often exacerbates disagreement rather than resolving it. One way to interpret this is that more disclosure expands the opportunity for differentiation, especially if agencies employ different mappings from qualitative information and heterogeneous indicators into standardized scores (Christensen et al., 2022). This is highly relevant in an EU environment that is moving toward more structured sustainability reporting architectures: even if disclosure becomes more standardized in principle (Hummel and Jobst, 2024), agencies' measurement and weighting choices may still produce divergent assessments, and those divergences may interact with institution type.

Related evidence indicates that the form and credibility of reporting can matter for convergence. Kimbrough et al. (2024) examine voluntary ESG reporting and find that disagreement among ESG rating agencies is lower for firms that voluntarily issue ESG reports, with environmental and social disclosures playing a meaningful role in reducing disagreement on those dimensions. They also discuss how third-party assurance and higher reporting standards can be associated with further reductions in disagreement (Kimbrough, Wang and Zhou, 2024). This suggests a nuanced picture: disclosure can exacerbate disagreement in general settings (Christensen et al., 2022), yet structured and credible reporting regimes can mitigate disagreement in specific ways (Kimbrough, Wang and Zhou, 2024). For this paper, the implication is that any penalization pattern for regional institutions may depend not only on the amount of disclosure but on the structure, accessibility, and "machine-readability" of that disclosure, precisely the dimensions that can differ between regionally anchored institutions and globally diversified peers.

A third strand of the literature raises the possibility that ESG ratings can reflect organizational resources and data availability rather than sustainability fundamentals. Drempetic et al. (2020) document a pronounced association between firm size and ESG scores, consistent with the idea that large organizations may be better positioned to produce, curate, and disseminate ESG-relevant information, thereby receiving higher scores even when underlying sustainability performance is not proportionally better. Their analysis highlights that ESG ratings can embed "capacity-to-report" effects, which can distort comparability, especially when stakeholders interpret ratings as pure performance measures (Drempetic et al. 2020). For the present study, this matters because regional financial institutions may differ systematically in reporting infrastructures, international investor relations, and database visibility. Even if some regional institutions are large in balance-sheet terms, they may still be "smaller" along the dimensions that matter for global ESG data capture.

Finally, recent work on aggregation directly informs the methodological choice to employ principal component analysis (PCA). Bissoondoyal-Bheenick et al. (2024) study ESG rating disagreement using multiple provider datasets and explicitly evaluate aggregation approaches, including PCA, to synthesize information across ratings. They document striking patterns, such as negative correlations involving certain rating measures in their setting, and argue that analysts should not treat ESG ratings as interchangeable, highlighting the value of systematic aggregation and robustness to methodological choices (Bissoondoyal-Bheenick et al., 2024). Their inclusion of PCA as an aggregation device supports the idea that PCA can serve as a disciplined, data-driven way to recover a latent "common" dimension from a set of correlated indicators without imposing arbitrary weights.

Taken together, these literatures motivate a design that (i) builds a transparent, disclosure-grounded measurement layer, (ii) extracts a latent Core ESG construct using PCA to limit subjective weighting, and (iii) benchmarks this construct against commercial ratings to diagnose where systematic gaps arise. The paper therefore asks: Do ESG ratings penalize regional EU financial institutions, in the sense that regionally anchored institutions exhibit systematically lower commercial ESG ratings, or systematically larger negative deviations from a disclosure-grounded Core ESG construct, compared to non-regional institutions?

3. Methodology and Data

The dataset covers about 60 financial institutions including both banks and insurers and explicitly distinguishes regionally anchored institutions (e.g., institutions with a predominantly regional mandate or ownership like the *German Landesbanken* or regionally concentrated business models) from non-regional institutions (e.g., more internationally diversified or nationally broad-based institutions). All institutions are EU-based to reduce heterogeneity in reporting environments, while still providing meaningful variation in business models and institutional structures.

For each institution, the dataset includes two main blocks. The first comprises classic finance KPIs (e.g., scale, profitability, capital and risk proxies, and other standard balance-sheet/performance indicators) used primarily for descriptive context and subgroup comparability. The second comprises structured ESG disclosure variables aligned with typical CSRD-style reporting content, coded from publicly available corporate reporting. Third-party ESG rating data are compiled for EU financial institutions from MSCI, ISS ESG, and Sustainable Fitch.

The study proceeds in two linked steps: (1) construction of a harmonized indicator set combining financial controls and structured ESG disclosure variables, and (2) PCA-based derivation of a Core ESG construct that can be benchmarked against commercial ratings.

First, ESG variables are coded to reflect typical CSRD-style sustainability disclosure content for EU financial institutions. The objective is not to replicate the full breadth of standards, but to capture a comparable set of disclosure-grounded indicators that plausibly map to rating inputs: environmental policies and performance proxies, social and workforce-related disclosures, and governance structures and controls. Indicators

are standardized and checked for comparability across institutions. Conventional financial KPIs are compiled in parallel to support stratification (banks versus insurers; regional versus non-regional) and to provide context for interpreting rating gaps.

Second, the Core ESG rating is derived using principal component analysis. PCA is chosen for three reasons. It provides a transparent, reproducible way to combine multiple correlated indicators into a smaller number of orthogonal components; it reduces reliance on subjective weights that can themselves be a source of divergence (Berg et al., 2022); and it aligns with prior work that studies ESG disagreement and evaluates PCA as a disciplined aggregation method (Bissoondoyal-Bheenick et al., 2024). Operationally, PCA is applied to the standardized ESG indicator matrix (and, where appropriate, separately within E, S, and G blocks as a robustness check). The first principal component is interpreted as the Core ESG factor when it captures a meaningful share of common variance and when loadings align with an interpretable sustainability dimension.

Because PCA can be inappropriate if the data are not factorable, sampling adequacy and factorability diagnostics are used. In particular, the Kaiser–Meyer–Olkin (KMO) measure is computed to assess whether partial correlations are sufficiently small to justify a factor-analytic approach, consistent with Kaiser's classic work on factorial structure and simplicity (Kaiser, 1974). Where KMO suggests weak adequacy, indicator selection and coding are revisited to ensure the Core ESG construct reflects a coherent common factor rather than a forced aggregation of unrelated items. This emphasis on factorability is not a cosmetic statistical step; it is conceptually aligned with the ESG measurement debate, because a low-KMO environment could indicate that the chosen indicators do not share a stable latent dimension, mirroring the broader point that ESG is not a single universally agreed construct (Berg et al., 2022; Agosto and Tanda, 2025).

Once the Core ESG factor is derived, the benchmarking step compares it to commercial ESG ratings from MSCI, ISS ESG, and Sustainable Fitch for EU financial institutions. Comparability is addressed via correlations and rank concordance between Core ESG and provider rating (whenever available and publicly disclosed) and (ii) the distribution of provider-minus-Core ESG gaps across regional versus non-regional institutions and across institutional types.

References

Agosto, A. and Tanda, A. (2025) 'Divergence and aggregation of ESG ratings: A survey', Open Research Europe, 5:28. https://doi.org/10.12688/openreseurope.19238.2

Berg, F., Kölbel, J.F. and Rigobon, R. (2022) 'Aggregate Confusion: The Divergence of ESG Ratings', Review of Finance, 26(6), pp. 1315–1344. https://doi.org/10.1093/rof/rfac033

Bissoondoyal-Bheenick, E., Bennett, S., Lehnherr, R. and Zhong, A. (2024) 'ESG rating disagreement: Implications and aggregation approaches', International Review of Economics & Finance, 96, 103532. https://doi.org/10.1016/j.iref.2024.103532

Christensen, D.M., Serafeim, G. and Sikochi, A. (2022) 'Why Is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings', The Accounting Review, 97(1), pp. 147–175. https://doi.org/10.2308/TAR-2019-0506

Drempetic, S., Klein, C. and Zwergel, B. (2020) 'The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review', Journal of Business Ethics, 167, pp. 333–360. https://doi.org/10.1007/s10551-019-04164-1

Hummel, K. and Jobst, D. (2024) 'An Overview of Corporate Sustainability Reporting Legislation in the European Union', Accounting in Europe, 21(3), pp. 320–355. https://doi.org/10.1080/17449480.2024.2312145

Kaiser, H.F. (1974) 'An index of factorial simplicity', Psychometrika, 39, pp. 31–36. https://doi.org/10.1007/BF02291575

Kimbrough, M.D., Wang, I.Y. and Zhou, X. (2024) 'Does Voluntary ESG Reporting Resolve Disagreement among ESG Rating Agencies?', European Accounting Review, 33(1), pp. 15–47. https://doi.org/10.1080/09638180.2022.2088588