ESG and Bank Financial Performance: A Critical Literature Review

Yingxuan Zhang
Beijing Normal-Hong Kong Baptist University, Zhuhai, Guangdong, China

Abstract: This article critically reviews the relationship between environmental, social, and governance (ESG) measures and bank financial performance. A targeted narrative synthesis distinguishes disclosure, performance ratings, and controversy scores, and compares accounting profitability with market valuation and intermediate outcomes. The evidence includes positive, negative, and nonlinear associations; recent meta-analyses report small positive point estimates but differ in statistical precision. These findings do not establish a uniform financial return to ESG engagement. Risk reduction, financing conditions, operating efficiency, and customer relationships offer plausible explanations, although the evidence supporting each pathway differs. In particular, an association with customer loyalty or lower risk does not by itself demonstrate an increase in bank profitability. The review contributes an interpretive framework that connects measurement choices, financial outcomes, and the strength of inference. It concludes that research should assess comparable constructs and competing mechanisms before attributing disagreement to geography or bank type. The synthesis is selective and does not estimate a new pooled effect or establish causality.

Keywords: ESG Measurement; Bank Profitability; Bank Valuation; Mediation; Literature Review

1. Introduction

Banks connect sustainability objectives with financial decisions through their operations, lending, and investment activities. Yet the financial implications of environmental, social, and governance (ESG) engagement remain difficult to generalize. Buallay reports a positive aggregate relationship between ESG disclosure and European bank performance, whereas Gutiérrez-Ponce and Wibowo report negative aggregate associations in Southeast Asia. This contrast raises a substantive question: do studies disagree about the same relationship, or do they measure different aspects of sustainability and financial performance[1,2]?

Existing reviews already address much of this debate. Salem et al. examine the scope of evidence and the need for mediation research. Mohapatra et al. estimate an average association across selected studies, while Zarea et al. use multilevel meta-analysis to examine ESG dimensions, financial indicators, rating providers, and time periods. A further review therefore requires a focused purpose beyond reporting that findings are heterogeneous[3-5].

This article examines how measurement and evidential differences change the interpretation of the ESG–performance relationship. It asks three questions: what associations are reported; which proposed mechanisms receive empirical support; and which contextual or methodological differences limit comparisons? Its contribution is an interpretive synthesis that distinguishes disclosure, ratings, and controversies; separates profitability from valuation and intermediate outcomes; and evaluates whether a study supports an association, a possible pathway, or a stronger causal claim. The aim is to make conflicting findings more intelligible while recognizing the limits of a selective narrative review.

2. Review Approach and Conceptual Distinctions

2.1 Scope and Source Selection

The review uses a targeted narrative approach. The preliminary bibliography contained ten banking studies and reviews. These records were checked through title and DOI searches, publisher pages, bibliographic metadata, and accessible original articles. Targeted searches for ESG rating divergence, bank risk and value, and CSR-related customer responses supplemented this starting set. Source verification was completed on 14 September 2026. One

Copyright @ STEMM Institute Press

overlapping empirical reference was omitted from the final synthesis, and four sources were added for specific analytical purposes.

The final bibliography contains thirteen sources: eight empirical banking studies, three banking reviews or meta-analyses, one cross-industry study of ESG ratings, and one official scoring methodology. The cross-industry study informs measurement interpretation rather than supplying banking performance evidence. The CSR customer study informs a potential stakeholder pathway rather than representing an aggregate ESG rating. Sources were retained for direct banking relevance or a clearly identified conceptual role, irrespective of the sign of their findings. This process does not constitute an exhaustive database search or a systematic review.

For each empirical study, the synthesis records the setting and period, sustainability measure, financial outcome, reported direction, and relevant interpretive limitations. Full texts were used where accessible; otherwise, claims were restricted to information in original abstracts and verified bibliographic records. Unavailable coefficients, model diagnostics, or subgroup tests were not reconstructed. Reviews and their constituent studies are discussed at different levels and are not counted as independent replications. No new pooled estimate, numerical quality score, or causal effect is calculated. As shown in Table 1, the present review complements existing banking reviews by connecting measurement choices, financial outcomes, and the strength of inference.

Table 1. Positioning This Review Against Existing Banking Reviews

Review	Approach and coverage	Contribution and remaining question
Salem et al. [3]	Scoping review; 18 studies in the final selection	Maps mixed evidence and motivates mediation research. Which pathways are actually tested?
Mohapatra et al. [4]	Random-effects meta-analysis; 16 studies included in the quantitative synthesis	Reports a small, imprecisely estimated association. What explains uncertainty and heterogeneity?
Zarea et al. [5]	Three-level meta-analysis; 28 studies and 194 effect sizes across analyses	Examines dimensions and moderators. How should measurement and inference guide the reading of individual findings?
Present review	Selective critical narrative synthesis; 13 sources with distinct roles	Connects constructs, outcomes, and evidential strength; does not claim exhaustive coverage or a new pooled effect

2.2 Sustainability Measures Are Not Interchangeable

ESG disclosure measures the information a bank reports. An ESG performance rating evaluates attributes using a provider's criteria, while a controversy score summarizes adverse events and related media information. A bank can disclose extensively without performing equally well on every underlying sustainability outcome. The official LSEG methodology also distinguishes its ESG score from the combined ESGC score, which incorporates a controversy adjustment. Consequently, disclosure–profitability and controversy–profitability estimates should not be placed on a common positive-versus-negative scale without examining what each variable means[6].

Score direction is especially important. Under the LSEG methodology examined here, companies without controversies receive a controversy score of 100. A higher controversy score therefore does not mean more controversies. Interpretation must follow the variable's coding and the authors' substantive findings rather than the sign of a regression coefficient alone. Provider versions and dates also matter: the documented scoring convention should be checked against the data vintage used by each study[6].

Measurement disagreement is not merely hypothetical. Berg et al. show that ESG ratings diverge because agencies differ in the attributes covered, how those attributes are measured, and how they are weighted. Their evidence is not restricted to banks, but it provides a reason to treat provider choice as a substantive research decision. This review accordingly distinguishes measurement differences from genuine differences in the underlying economic relationship[7].

2.3 Financial Outcomes and Strength of Inference

Return on assets (ROA) and return on equity (ROE) measure accounting profitability relative to different balance-sheet bases. ROE is also

 Copyright @ STEMM Institute Press

sensitive to leverage. Tobin's Q is a valuation measure reflecting market expectations, while net interest margin measures a component of intermediation income rather than total profitability. Risk, financing costs, cash flows, and efficiency can be outcomes in their own right or intermediate variables in a proposed pathway. Improvement in one does not imply improvement in all the others.

The synthesis uses "association" for observational estimates unless a credible identification strategy supports a stronger interpretation. Lagged variables and dynamic panel estimators can address particular problems, but they do not automatically eliminate omitted variables, reverse causality, or measurement error. Likewise, evidence consistent with a channel is weaker than a design that estimates the full indirect pathway and assesses the assumptions required for causal mediation.

3. Evidence on Financial Performance

3.1 Positive, Negative, and Nonlinear Associations

Buallay's study covers 235 European banks during 2007–2016 and reports a positive association for aggregate ESG disclosure. Its disaggregated findings differ: social disclosure is negatively related to the three financial measures, while governance disclosure has different signs for accounting performance and market valuation. Even this apparently positive study therefore does not support a uniform return to every ESG component[1].

Azmi et al. analyse 251 banks in 44 emerging economies during 2011–2017 using system GMM. They report that low levels of ESG activity are positively related to bank value, with diminishing returns as activity increases. This supports a nonlinear interpretation rather than the proposition that additional ESG engagement necessarily raises valuation. A turning point estimated in one sample should not be treated as a universal investment threshold[8].

Gutiérrez-Ponce and Wibowo study 19 banks in five Southeast Asian countries over 2010–2020. Their aggregate ESG measure is significantly negatively associated with ROA, ROE, and Tobin's Q, although individual pillars yield different findings. This evidence belongs in the negative or mixed category. The small, data-dependent sample also limits its geographical reach. Comparing it with Buallay does not isolate a regional explanation: measurement, sample composition, observation periods, and modelling choices differ simultaneously[2].

3.2 Controversies and the Interpretation of Signs

Agnese et al. examine 63 European banks during 2015–2022. Their abstract reports that banks with more ESG controversies display higher ROA, ROE, and net interest margins in the sample. This interpretation is compatible with a negative coefficient on an inversely coded controversy score. It should not be rewritten as evidence that greater controversy exposure reduces profitability. Equally, the result does not demonstrate that misconduct causes better performance or that controversy exposure is financially harmless over longer horizons. It concerns a particular indicator, set of outcomes, and research design[9].

3.3 Bank Type and Changing Periods

Smaoui et al. examine 67 Islamic and conventional banks in 13 countries during 2009–2019. Their reported positive aggregate association is particularly linked to governance; the positive governance finding for Islamic banks is not reproduced in the same way for conventional banks. This is a specific bank-type result. It does not establish that ownership, size, and every business-model characteristic independently moderate ESG outcomes. Moreover, a significant coefficient in one subgroup and a nonsignificant coefficient in another do not, by themselves, establish a statistically significant difference between subgroups[10].

Dragomir et al. study 333 banks in 53 countries across 2019–2021. They report a negative relationship between environmental performance in 2019 and ROE in 2020, and a positive relationship between social performance in 2020 and profitability in 2021. These results illustrate why dimension and timing must be reported together. Although the article invokes causality, using a one-year lag establishes temporal ordering rather than, on its own, exogenous variation. The short pandemic window limits conclusions about persistent effects[11]. Table 2 summarizes the selected empirical studies discussed in Sections 3.1–3.3, including their measures, financial outcomes, reported findings, and interpretive limitations.

Copyright @ STEMM Institute Press

Table 2. Selected Empirical Findings and Their Interpretation

Study and sample	Measure and outcome	Reported finding and reading constraint
Buallay [1]: 235 European banks, 2007–2016	ESG disclosure; ROA, ROE, Tobin's Q	Positive aggregate association; mixed pillar results. Disclosure does not directly measure all sustainability outcomes.
Gutiérrez-Ponce and Wibowo [2]: 19 banks in five countries, 2010–2020	Thomson Reuters ESG measures; ROA, ROE, Tobin's Q	Significantly negative aggregate associations; pillar differences. Limited sample coverage.
Azmi et al. [8]: 251 banks in 44 economies, 2011–2017	ESG activity; bank value	Positive at low ESG levels, with diminishing returns. A linear summary loses this qualification.
Agnese et al. [9]: 63 European banks, 2015–2022	Controversy score; ROA, ROE, net interest margin	More controversies accompany higher profitability. Check inverse score coding before interpreting coefficients.
Smaoui et al. [10]: 67 banks in 13 countries, 2009–2019	ESG and governance; profitability	Governance findings differ across Islamic and conventional banks. Do not infer untested moderators.
Dragomir et al. [11]: 333 banks in 53 countries, 2019–2021	Lagged ESG pillars; accounting and market outcomes	Signs depend on dimension and year. Lagging the predictor alone does not establish causality.

Notes: The table summarises reported findings; it is not a vote count or a ranking of study quality. Sample sizes are not equivalent to independent observations.

3.4 Reading the Meta-Analytic Evidence

Mohapatra et al. report a pooled correlation of approximately 0.10, with a 95% confidence interval from −0.11 to 0.31 and very high heterogeneity. Zarea et al. report a small statistically significant positive association, with an aggregate estimate of 0.1085. The point estimates are close. Their different significance labels therefore should not be presented as evidence of opposing economic effects. Nor does the nonsignificant estimate prove that the true relationship is zero[4,5].

The useful comparison concerns inclusion rules, outcome definitions, effect-size dependence, and uncertainty. Multiple estimates from the same banks or article do not provide the same information as independent samples. Differences in synthesis methods can affect precision, but this review does not reproduce the analyses and cannot attribute their different confidence intervals to one cause. The appropriate conclusion is that the average association is estimated to be small and positive in these syntheses, while its precision and applicability remain unsettled.

4. Mechanisms and Competing Explanations

4.1 Risk Management and Valuation

Di Tommaso and Thornton provide a useful test of the assumption that lower risk necessarily creates higher bank value. They find that higher ESG scores accompany a modest reduction in risk-taking, conditional on board characteristics, but also lower bank value. Their analysis reports a positive indirect connection through risk-taking alongside the negative valuation relationship. Thus, risk reduction can coexist with a financial trade-off. This provides more specific channel evidence than a general claim that responsible governance improves profitability[12].

An additional distinction concerns the objective being evaluated. A more stable institution may benefit depositors and the financial system without maximising short-term shareholder valuation. The review therefore treats risk reduction and profitability improvement as separate propositions. This also leaves room for competing explanations, including compliance costs, foregone risky returns, or spending that investors regard as excessive; the cited evidence does not identify every component of that trade-off.

4.2 Customer Relationships and Reputation

Pérez and Rodríguez del Bosque use a structural equation model with 648 savings-bank customers and 476 commercial-bank customers in Spain. Customer-oriented CSR perceptions are positively related to identification, satisfaction, recommendation, and repurchase responses[13]. These findings concern CSR perceptions rather than aggregate ESG scores and provide evidence of favourable customer

 Copyright @ STEMM Institute Press

responses. However, greater customer satisfaction does not guarantee profit growth; establishing a financial benefit requires additional evidence on customer retention, revenues, and the costs of the initiatives.

4.3 Financing, Efficiency, and Expenditure

Azmi et al. report positive associations of ESG activity with cash flows and efficiency, a negative association with the cost of equity, and no effect on the cost of debt. Their findings therefore do not justify the broad claim that ESG reduces all financing costs. They identify particular financial channels within that study's setting[8].

The interpretation must also allow benefits and costs to operate together. Investment in reporting systems, staff capabilities, or risk assessment uses resources, while any resulting efficiency or financing gains may arrive on a different timetable. These are possible explanations for nonlinear or mixed relationships, not findings that every reviewed paper directly tests. A convincing mechanism analysis would measure both the proposed intermediate outcome and the associated expenditure rather than infer net benefits from a favourable ESG coefficient.

5. Explaining Heterogeneity

5.1 Comparing Like Measures and Outcomes

Measurement should be examined before geography is used to explain disagreement. Researchers can first compare studies using similar disclosure or rating constructs, then separate aggregate scores from individual pillars, and finally group financial outcomes by profitability, valuation, or intermediate channel. Provider disagreement and changing scoring conventions further complicate comparisons. A score increase may reflect different information or assessment rules, so it should not automatically be interpreted as the same economic change across datasets[6,7].

Outcome alignment is equally important. A lower risk measure, a higher equity valuation, and a higher accounting return answer different questions. If ESG is associated with lower risk but lower valuation, reporting only the risk result produces an incomplete account. If governance relates differently to ROA and Tobin's Q, averaging the signs conceals the distinction between current earnings and market expectations. These comparisons should guide interpretation before studies are grouped as supportive or unsupportive.

5.2 Context and Research Design

Country institutions, bank type, and sample periods may influence both ESG activities and financial outcomes. However, cross-study contrasts do not isolate these influences when measurement and research design change at the same time. Bank-type interactions, comparisons within consistent data systems, and explicit temporal tests are more informative than broad claims about developed versus emerging economies. The subgroup and pandemic evidence reviewed above should be interpreted within these limits[10,11].

Methodological differences also matter. Better-performing banks may have more resources to disclose information or invest in ESG, producing reverse causality. Unobserved management quality may influence both variables. Panel controls, lag structures, and GMM instruments address selected concerns only under their assumptions. Empirical studies should therefore report instrument choices, serial-correlation and specification diagnostics, robustness to alternative measures, and the economic magnitude of estimates. Where these details are inaccessible, a review should narrow its claims rather than infer methodological strength from the estimator's name.

6. Discussion and Research Priorities

The synthesis supports a conditional answer to the question of whether ESG "pays" in banking. Positive associations occur, but their meaning depends on the sustainability construct, financial objective, and study design. Negative or nonlinear findings are not peripheral exceptions to be removed from the argument. They help identify situations in which costs, risk adjustments, or measurement differences change the financial interpretation.

A useful assessment therefore begins with three linked checks. First, what does the ESG variable actually measure, and which direction indicates improvement? Second, is the outcome profitability, market value, risk, or an intermediate variable? Third, what inference does the design support? Applying these checks corrects apparent contradictions without assuming that all remaining differences have been explained. It also prevents a mechanism suggested by theory from acquiring the status of

Copyright @ STEMM Institute Press

an established empirical result through repetition.

For bank managers, the implication is to evaluate specific initiatives against explicit objectives rather than rely on an aggregate score as a profitability guarantee. An initiative intended to reduce downside risk should be assessed with risk and loss measures alongside its cost; a customer-focused initiative requires evidence about customer behaviour and its financial consequences. For researchers, the central priority is comparability and identification rather than accumulating additional regressions with different ESG proxies.

Future studies could strengthen this literature by recording provider versions and scoring direction, reporting effect sizes with confidence intervals, separating short-run expenditure from longer-run outcomes, and testing competing indirect pathways. Designs using regulatory changes or other plausibly external variation should explain why their identifying assumptions hold. Future syntheses should document database coverage and exclusions, examine overlapping bank samples, and evaluate publication bias and within-study dependence. These are priorities for further research, not procedures completed in the present review.

The review has its own limitations. Its targeted selection may omit relevant studies, including evidence unavailable in accessible sources. Some findings were verified from abstracts rather than complete model results, limiting assessment of specification and robustness. The source set is small and includes overlapping geographical and temporal coverage. No pooled estimate is calculated, so the review cannot determine the overall magnitude of the relationship or explain all differences between meta-analyses. Its contribution is to improve interpretation of selected evidence and identify what a more comprehensive review would need to establish.

7. Conclusion

The relationship between ESG and bank financial performance cannot be reduced to a single positive or negative effect. The reviewed evidence includes favourable aggregate associations, negative results, nonlinear patterns, and differences across pillars and financial indicators. Recent meta-analyses suggest small positive average point estimates, but they do not remove uncertainty about particular banks, measures, or settings.

The most consequential distinction is between what a study measures and what its findings can support. Disclosure, performance ratings, and controversy scores capture different constructs; profitability, valuation, risk, and customer responses capture different outcomes. Risk management, financing conditions, efficiency, and stakeholder relationships offer relevant explanations, but the strength of mechanism evidence varies. Observational associations should therefore not be translated into universal causal or managerial claims.

A more defensible research agenda combines transparent measurement, explicit comparisons of financial outcomes, and stronger tests of competing pathways. Within its selective scope, this review provides a framework for making those distinctions and reading apparently conflicting findings with greater precision.

References

[1] Buallay, A. (2019). Is sustainability reporting (ESG) associated with performance? Evidence from the European banking sector. *Management of Environmental Quality: An International Journal, 30*(1), 98–115. https://doi.org/10.1108/MEQ-12-2017-0149

[2] Gutiérrez-Ponce, H., & Wibowo, S. A. (2024). Do sustainability practices contribute to the financial performance of banks? An analysis of banks in Southeast Asia. *Corporate Social Responsibility and Environmental Management, 31*(2), 1418–1432. https://doi.org/10.1002/csr.2641

[3] Salem, M. R. M., Shahimi, S., & Alma'amun, S. (2024). Does mediation matter in explaining the relationship between ESG and bank financial performance? A scoping review. *Journal of Risk and Financial Management, 17*(8), Article 350. https://doi.org/10.3390/jrfm17080350

[4] Mohapatra, A. K., Das, S., Nayak, Y., Sahoo, A. P., & Matta, R. (2025). Do ESG practices affect the financial performance of banks? A meta-analysis perspective. *Banks and Bank Systems, 20*(3), 117–128. https://doi.org/10.21511/bbs.20(3).2025.09

[5] Zarea, M., Prior, D., & Hegazy, M. (2026). ESG and banking financial performance: A multi-level meta-analysis review. *The Journal of Economic Asymmetries, 33*, Article e00467.

 Copyright @ STEMM Institute Press

https://doi.org/10.1016/j.jeca.2026.e00467

[6] LSEG. (2024, October). *Environmental, social and governance scores from LSEG*. Official methodology

[7] Berg, F., Kölbel, 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

[8] Azmi, W., Hassan, M. K., Houston, R., & Karim, M. S. (2021). ESG activities and banking performance: International evidence from emerging economies. *Journal of International Financial Markets, Institutions and Money, 70*, Article 101277. https://doi.org/10.1016/j.intfin.2020.101277

[9] Agnese, P., Cerciello, M., Oriani, R., & Taddeo, S. (2024). ESG controversies and profitability in the European banking sector. *Finance Research Letters, 61*, Article 105042. https://doi.org/10.1016/j.frl.2024.105042

[10] Smaoui, H., Aassouli, D., & Elakhdar, Y. (2025). Does ESG influence bank profitability? A comparison between Islamic and conventional banks. *Corporate Social Responsibility and Environmental Management, 32*(5), 6048–6065. https://doi.org/10.1002/csr.70010

[11] Dragomir, V. D., Bătae, O. M., Ionescu, B. Ș., & Ionescu-Feleagă, L. (2022). The influence of ESG factors on financial performance in the banking sector during the COVID-19 pandemic. *Economic Computation and Economic Cybernetics Studies and Research, 56*(4), 71–88. https://doi.org/10.24818/18423264/56.4.22.05

[12] Di Tommaso, C., & Thornton, J. (2020). Do ESG scores effect bank risk taking and value? Evidence from European banks. *Corporate Social Responsibility and Environmental Management, 27*(5), 2286–2298. https://doi.org/10.1002/csr.1964

[13] Pérez, A., & Rodríguez del Bosque, I. (2015). Corporate social responsibility and customer loyalty: Exploring the role of identification, satisfaction and type of company. *Journal of Services Marketing, 29*(1), 15–25. https://doi.org/10.1108/JSM-10-2013-0272

Copyright @ STEMM Institute Press