

Growth Without Integration: A Bibliometric Audit of the ESG–Financial Performance Literature

Kanchan Rani¹, Satnam Singh², Prakhar Wadhwa^{3*}

¹Department of Commerce, University of Delhi, Delhi, India

²Department of Commerce, University of Delhi, Delhi, India

³Ramanujan College, University of Delhi, Delhi, India

Corresponding author: Prakhar Wadhwa, Email: prakhar.wadhwa@gmail.com

Abstract: The relationship between environmental, social and governance (ESG) performance and financial performance has generated one of the largest empirical literatures in management research, and a steady supply of reviews summarising it. This study audits that literature rather than summarising it. Search strings are treated as measurement instruments: calibrated in stages, diagnosed for precision and recall, and reported with their failures intact. The resulting corpus of 2,923 documents is mapped along four dimensions: conceptual, thematic, intellectual and social structure.

Three findings converge from independent evidence. The conceptual core occupies the basic quadrant of the thematic map: connected to everything, internally coherent with almost nothing. The citation base has bifurcated, with a contemporary ESG measurement school of 22 works forming separately from the corporate social performance tradition rather than absorbing it. The social structure fragments into 221 components, the largest holding nine of 6,600 authors, with productivity departing sharply from Lotka's law. The literature grows without integrating. Its vocabulary has also turned over almost completely, and a direct test shows this was displacement rather than relabelling: the co-word neighbourhood of ESG performance today overlaps that of corporate social responsibility before 2015 by only 11 per cent.

Two methodological findings follow. Keyword screening reproduced manual inclusion decisions at 95.4 per cent raw agreement but a Cohen's kappa of 0.297, because unit of analysis is a property of research design rather than of vocabulary. And because Scopus uses one delimiter between references and between authors, naive parsing silently returns the final author of each cited work.

Keywords: ESG; corporate social responsibility; bibliometric analysis; co-citation analysis; search validation

1. INTRODUCTION

Whether corporate environmental, social and governance performance pays has been asked continuously for five decades and answered inconsistently throughout. The literature is now large enough that the more tractable question is no longer what the relationship is, but what the accumulated body of work has established, and whether its structure is capable of establishing anything.

Reviews of this literature are plentiful. Meta-analyses have pooled it repeatedly, most influentially Orlitzky, Schmidt and Rynes (2003) and, two decades later, Friede, Busch and Bassen (2015). Bibliometric mappings appear at a rate of several per year. What these share is a treatment of the search itself as preliminary: a string is reported, a count is given, and analysis proceeds. The string is rarely calibrated, its precision is almost never measured, and its recall is essentially never tested against works the corpus could not credibly omit.

This matters because the conclusions of a mapping study are downstream of its corpus in a way that the conclusions of a primary study are not. A search that systematically under-retrieves one period, one vocabulary or one



region does not merely lose observations; it changes the shape of the map. When the object of study is the structure of a literature, the instrument that defines that literature is not a preliminary step. It is the measurement.

This study therefore audits rather than summarises. Three commitments follow. First, the search is calibrated in stages with counts recorded after every clause, and both precision and recall are assessed against standards fixed before the searches ran. Second, where those diagnostics fail, the revision is reported with its cause rather than presented as the original design. Third, failures that survive revision are stated in the results rather than buried in the limitations.

“Integration” is used here in a specific sense, and the paper stands or falls on it. A field integrates when its central themes become internally coherent as well as widely connected, when new work extends existing citation communities rather than founding parallel ones, and when its authors form a connected collaboration structure capable of transmitting standards. Each of these is measurable, and each is measured below: theme density against centrality, community formation in the co-citation network, and component structure in the co-authorship network. A field can grow rapidly on all three counts without integrating on any, and the argument of this paper is that this one has.

Three questions follow. RQ1: does the ESG–financial performance literature show evidence of conceptual consolidation, or of extensive growth without integration? RQ2: does the shift from corporate social responsibility to ESG represent conceptual development, relabelling of an existing construct, or displacement by a distinct one? RQ3: what does treating the search string as a measured instrument, rather than a reporting formality, reveal about the reliability of bibliometric evidence in this field?

The audit produces one substantive finding and two methodological ones. Substantively, four independent analyses converge on a literature that grows without integrating: its central constructs connect to everything and cohere with nothing, its citation communities have split rather than accumulated, and its authors work in hundreds of small disconnected groups. Methodologically, automated screening is shown to fail on this literature for a structural rather than a remediable reason, and a widely used reference format is shown to corrupt co-citation analysis silently.

2. LITERATURE REVIEW

Two bodies of work bear on this study. The first is the empirical literature on ESG and financial performance, whose structure is the object of analysis. The second is the practice of bibliometric review itself, whose methods this study both uses and interrogates.

2.1 Five decades of an unresolved question

The empirical question predates the vocabulary now used to ask it. Early studies through the 1970s and 1980s asked whether socially responsible firms were more or less profitable, and returned answers in every direction. Aupperle, Carroll and Hatfield (1985) found no relationship between corporate social responsibility and profitability; McGuire, Sundgren and Schneeweis (1988) found that prior financial performance predicted subsequent social performance better than the reverse, an early and under-heeded warning about causal direction. Cochran and Wood (1984) and Ullmann (1985) had already attributed the inconsistency to measurement rather than to substance.

Waddock and Graves (1997) gave the field its durable formulation. Using the KLD ratings, they reported a positive association running in both directions and named it the “virtuous circle”, in which slack resources permit social investment and social investment in turn improves performance. The paper supplied a construct label of corporate social performance, together with a data source and a research design, and the literature adopted all three. It remains the most co-cited empirical work in the corpus assembled here.

The meta-analytic syntheses of the early 2000s were the field’s first attempt to resolve the inconsistency by aggregation. Orlitzky, Schmidt and Rynes (2003) pooled 52 studies and reported a positive mean correlation, concluding that the relationship was reliable and that measurement error explained much of the observed variance. Margolis and Walsh (2003) reviewed 127 studies and reached a more guarded verdict, arguing that the accumulated

evidence was too heterogeneous in construct and design to license a general claim. That the two syntheses of the same literature diverged in confidence rather than in sign is itself informative.

A theoretical reframing followed. McWilliams and Siegel (2001) modelled social investment as a firm-level optimisation problem with an interior optimum determined by the demand for social attributes, implying that no uniform relationship should be expected. Porter and Kramer (2006) recast the question strategically. Barnett (2007) and Barnett and Salomon (2012) developed the contingency position furthest, arguing that returns to social investment depend on a firm's stakeholder influence capacity and are therefore curvilinear. Under these accounts, the inconsistent findings of three decades are not a failure of measurement but a correct reading of a heterogeneous phenomenon.

Since roughly 2010, financial economics has approached the question through the cost of capital rather than through profitability. Dhaliwal and colleagues (2011) linked voluntary non-financial disclosure to reductions in the cost of equity; El Ghoul and colleagues (2011) and Goss and Roberts (2011) reported comparable effects for equity and bank debt; Cheng, Ioannou and Serafeim (2014) connected CSR performance to access to finance through reduced agency costs and information asymmetry. Lins, Servaes and Tamayo (2017) showed that firms with high social capital outperformed during the financial crisis, framing social performance as insurance rather than investment. This strand asks how markets price ESG, which is a different question from whether ESG pays, and the co-citation results below show that it has formed a correspondingly distinct community.

Friede, Busch and Bassen (2015) aggregated more than 2,000 primary studies through existing reviews and reported that roughly 90 per cent found a non-negative relationship. The paper is the second most co-cited work in this corpus and anchors the contemporary ESG literature. Its method, however, pools across construct definitions, outcome measures and research designs that the field has never demonstrated to be commensurable, an issue this study returns to.

The most recent quantitative synthesis is Hirsch and colleagues (2023), whose meta-regression of the CSR–profitability literature reports a small positive mean effect that attenuates substantially once publication selection is modelled. Their finding that the estimated effect depends heavily on how the underlying studies are specified is consistent with the structural argument developed here.

The most recent development is scepticism about the ratings themselves. Berg, Kölbel and Rigobon (2022) documented substantial divergence between major ESG raters and traced most of it to differences in measurement rather than scope or weighting. If the same firm receives materially different scores from different providers, then studies using different providers are not estimating the same quantity, and the accumulated evidence is less comparable than its volume suggests.

2.2 Reviews of the literature, and what they leave unexamined

Reviews of this literature are numerous and take three forms. Narrative reviews trace conceptual development. Meta-analyses pool effect sizes, most recently Hirsch and colleagues (2023), whose meta-regression of CSR and profitability is the current state of the art. Bibliometric mappings describe the field's structure using citation and keyword data, and have proliferated since software such as bibliometrix (Aria & Cuccurullo, 2017) and VOSviewer (Van Eck & Waltman, 2010) made the analysis accessible; guidance for conducting them is now well established (Zupic & Čater, 2015; Donthu et al., 2021).

The ESG–performance literature has attracted several such mappings. Khan (2022) combines bibliometric and meta-analytic treatment of ESG disclosure and firm performance. Barauskaite and Streimikiene (2021) survey the concepts, definitions and measures used across the CSR–financial performance literature and conclude that measurement heterogeneity is its central problem. Cardillo and Basso (2025) revisit the field with attention to moderating variables. Each reports a search string and a resulting count.

What none of them reports is a diagnostic on that string. Precision is not assessed against a screened sample; recall is not tested against works the corpus could not credibly omit; and intermediate counts showing what each

clause contributed are not given. This is not a criticism of these studies specifically, since it is the convention of the genre, and this study follows the same convention up to the point where the diagnostics are run. The observation is simply that a search cannot be shown to have found a literature by counting what it found.

Screening is increasingly automated or semi-automated, and agreement between automated and manual decisions is seldom reported. Where inter-coder agreement appears at all it is usually raw percentage agreement, which is close to uninformative in a corpus where most records are includable, a statistical artefact documented by Feinstein and Cicchetti (1990) and quantified as the prevalence paradox by Byrt, Bishop and Carlin (1993). Section 3.4 demonstrates the point rather than asserting it.

Finally, mappings tend to report structure without triangulating it. Thematic maps are presented quadrant by quadrant; co-citation clusters are named; co-authorship networks are visualised. What is rarely attempted is asking whether the conceptual, intellectual and social structures tell a consistent story about how the field is developing. The convergence reported in Section 6 is only visible when the three are read together.

2.3 The gap this study addresses

The substantive gap is that no existing review asks whether the ESG framing represents conceptual development or vocabulary substitution, and none is designed to answer it. That question requires comparing the structural position of the ESG literature against the tradition it displaced, in keyword space, in citation space and over time, which in turn requires a corpus whose coverage of the earlier period can be demonstrated rather than assumed.

The methodological gap is that the instrument defining these corpora is not itself measured. This study treats the search string as an instrument with calibration, precision and recall, reports what those diagnostics found including where they failed, and demonstrates two specific ways in which standard practice produces results that are wrong in undetectable ways.

3. METHOD

3.1 Design

The corpus was constructed once, diagnosed, revised once in response to those diagnostics, and then frozen before any mapping was performed. No analysis informed any subsequent search decision. This ordering is stated because the alternative, namely adjusting a search after seeing what it does to results, is difficult to detect in a finished paper and impossible to correct for.

3.2 Databases and search construction

Scopus and Web of Science were searched on 10 August 2026. Both were queried on the same day, since both re-index continuously and their counts drift.

The query was built as two blocks. A construct block captures ESG, corporate social responsibility and sustainability terminology; an outcome block captures financial performance, firm value and cost of capital. Blocks were run separately before being combined so that the contribution of each was observable.

Calibration proceeded in stages, with counts recorded after every clause. The initial combined Scopus set returned 6,081 records, falling to 3,835 after restricting to journal articles, to the Business, Economics and Decision Sciences subject areas, and to English. This exceeded the pre-specified target band, and the pre-specified remedy was applied: the outcome block was restricted to the title field while the construct block remained across title, abstract and keywords. This is the change most likely to cost recall, which is why the recall diagnostic was run immediately afterwards. Table 1 records the sequence. Final strings are given in Appendix A.

3.3 Search diagnostics

Two diagnostics were specified before the searches were run.

Precision was assessed on a random sample of 30 records, classified as in scope, borderline or out of scope, with a pass threshold of 70 per cent for in-scope plus borderline. The first calibration returned exactly 70.0 per cent. Inspection showed that seven of the nine failures shared a single cause: the term “non-financial performance”, which in the management accounting literature denotes performance measures that are not financial rather than anything to do with ESG. It retrieved 175 records uniquely.

Recall was assessed against a seed list of works the corpus could not credibly omit. The list was constructed from the most-cited references of the retrieved literature rather than from the search itself, so that it was independent of the instrument being tested. The first calibration retrieved 11 of 22 seeds, or 50.0 per cent, against a standard requiring every seed to be found.

Both diagnostics triggered revision. The construct block gained “corporate social responsibility”, which had been present only in the narrower form “corporate social responsibility performance”; the outcome block gained the wealth-based vocabulary of shareholder value, shareholder wealth, financial returns, access to finance and profitability; and “non-financial performance” was removed. Re-run diagnostics returned precision of 93.3 per cent and recall of 72.7 per cent. Precision passed comfortably. Recall improved by 22.7 points and did not reach the specified standard; Section 4.3 states what follows.

3.4 Screening

The merged corpus was screened at title and abstract against rules settled before screening began. Reviews and bibliometric studies are retained. The unit of analysis is the firm, excluding portfolio and fund-level studies. A construct qualifies if it measures an environmental, social or governance dimension of firm conduct, whether by vendor rating, disclosure index or direct indicator.

The construct rule required the most judgement. Studies of workplace safety, disability inclusion, employee satisfaction, pollution-reduction expenditure and regulatory penalties measure genuine ESG dimensions without using ESG vocabulary; they are included. Studies whose predictor is a business practice with no such dimension, such as advertising expenditure, brand equity, working capital management or trade credit, are not. The rule further turns on whose conduct is measured: a predictor describing what investors do about a firm, such as ownership by socially responsible funds or membership of an ESG index, is not a measure of the firm’s own conduct.

Three procedures were used, with different error properties, and all three are reported. 740 records were read individually in priority order. A keyword rule then identified a further 254 as exclusion candidates, and all 254 were read individually. The remaining 1,929 were accepted as included, with a random sample of 60 read to measure the error rate. No record was excluded without being read: the rule determined what to read, never what to conclude.

That distinction is load-bearing. The rule was tested against the 740 manual decisions and reached 95.4 per cent raw agreement but a Cohen’s kappa of 0.297, which is “fair” on the Landis and Koch (1977) benchmarks, catching 8 of 28 exclusions. The gap between the two statistics is itself instructive: where one category dominates, chance agreement is high and kappa is suppressed even when raw agreement is excellent, the prevalence paradox described by Feinstein and Cicchetti (1990) and Byrt and colleagues (1993). Reporting raw agreement alone would have made this rule look adequate. The failure is structural rather than remediable. A record titled “Data element utilization and firm value” contains governance and performance vocabulary somewhere in its abstract, so no keyword test can establish that its predictor is not an ESG construct; and unit of analysis is a property of research design, not of the words used to describe it. Automated screening is not adequate for this literature, and reporting that is more useful than concealing it.

The audit of the unread stratum found 58 correct inclusions and 2 errors, a false-inclusion rate of 3.3 per cent, implying approximately 64 residual errors in that stratum. This is a bounded and stated error rather than an unmeasured one.

Inter-coder reliability was assessed separately. A second coder independently screened 100 records drawn at random, returning 91 per cent raw agreement with the first coder. Applying the same prevalence correction as above, the corresponding kappa is approximately 0.36, again “fair” rather than substantial, and again for the reason that inclusion decisions in this corpus are heavily imbalanced. Disagreements were concentrated in the exclusion calls, which is where the screening rules do their work, and were reconciled to consensus before the corpus was frozen.

3.5 Analytical procedures

Mapping follows the performance-and-science-mapping distinction set out by Zupic and Čater (2015) and Donthu and colleagues (2021), addressing conceptual structure through author-keyword co-occurrence, intellectual structure through co-citation (Small, 1973), social structure through co-authorship (Newman, 2004), and thematic evolution across periods (Cobo et al., 2011). Analysis was performed in Python; results were cross-checked against bibliometrix (Aria & Cuccurullo, 2017).

Keyword analysis uses author keywords. Index keywords are present for only 326 of 2,923 records, a Scopus indexing lag on recent material, and using them would bias the map towards older work. Synonyms were merged before mapping, pairing “csr” with “corporate social responsibility”, spelled-out ESG variants with “esg”, and “firm performance” with “financial performance”, because leaving them separate manufactures a divide that reflects labelling convention rather than intellectual structure. JEL classification codes and phrase fragments arising from authors entering the ESG expansion as three separate keywords were removed; unremoved, the latter generate a spurious high-density cluster of the three pillars.

Networks are weighted by association strength (Van Eck & Waltman, 2010; Waltman et al., 2010) and partitioned by the Louvain method (Blondel et al., 2008). Themes are positioned on the centrality–density plane of Callon and colleagues (1991), in which centrality measures a cluster’s external connectedness and density its internal cohesion, and the resulting quadrants distinguish motor, niche, basic and emerging or declining themes. Resolution was selected by requiring between four and nine communities with none exceeding 45 per cent of nodes; lower resolutions left a single component holding two-thirds of the network. Journal concentration is assessed against Bradford’s law (Bradford, 1934) and author productivity against Lotka’s (1926). All parameters are reported so that they can be varied.

Co-citation analysis required a correction that generalises beyond this study. Scopus uses the semicolon as a delimiter both between references and between authors within a reference. Splitting on it shreds every multi-author reference, orphaning the first author and attaching the last author to the title. The uncorrected analysis returned Graves (1997), Rynes (2003) and Bassen (2015) as the field’s most-cited works, in each case the final author of a well-known paper. References were reconstructed by accumulating fragments until one terminated with a four-digit year in parentheses. The error is invisible without domain knowledge of the citations involved.

Two databases were searched because their journal coverage differs materially (Mongeon & Paul-Hus, 2016), and because retrieval quality varies across search systems in ways that bear directly on review validity (Gusenbauer & Haddaway, 2020). Screening and reporting follow PRISMA 2020 (Page et al., 2021) where applicable to a bibliometric corpus. Inter-rater agreement is reported as Cohen’s kappa (Cohen, 1960) alongside raw agreement, with the caveats of Feinstein and Cicchetti (1990) and Byrt and colleagues (1993) noted in Section 3.4; benchmark interpretations follow Landis and Koch (1977).

Two sensitivity checks were run rather than promised. For the co-word network, edge thresholds of 2, 3 and 5 co-occurrences were crossed with Louvain resolutions of 1.6, 2.0 and 2.5. Across eight of the nine specifications the basic quadrant holds between 79 and 87 per cent of keyword occurrences; only the sparsest network at the highest resolution departs from this. For the co-citation network, node sets of the top 70, 90 and 110 references were crossed with edge thresholds of 20, 25 and 35 co-citations, returning between 8 and 10 communities in every specification with the largest holding 19 to 27 per cent of nodes. Both central findings, an undeveloped core and a bifurcated citation base, are properties of the data rather than of the parameters.

Author disambiguation uses Scopus Author IDs. Name-based matching merges distinct researchers: surname-plus-initial matching returned one identity with 28 documents that resolves to several people. Analysis of production over time excludes 2026, since the export was taken on 10 August and a partial count plots as a false collapse.

4. THE CORPUS

4.1 Construction

Stage	Scopus	Web of Science
Construct block alone	35,914	12,662
Outcome block alone	124,965	71,690
Combined	6,081	3,321
Journal articles only	4,917	1,850
Subject area restriction	3,883	1,210
English only	3,835	1,205

Table 1. Calibration sequence, pre-revision. Counts recorded after each clause.

Following revision, the final searches returned 2,651 Scopus records and 1,205 Web of Science records, the latter reducing to 1,193 after excluding composite document types. Deduplication matched on normalised DOI first and on normalised title where a DOI was absent or malformed; 921 records overlapped. The merged corpus is 2,923 documents spanning 1989 to 2026. Screening excluded 222 records, leaving 2,701 eligible. Figure 1 summarises the sequence and Table 2 gives the exclusion reasons.

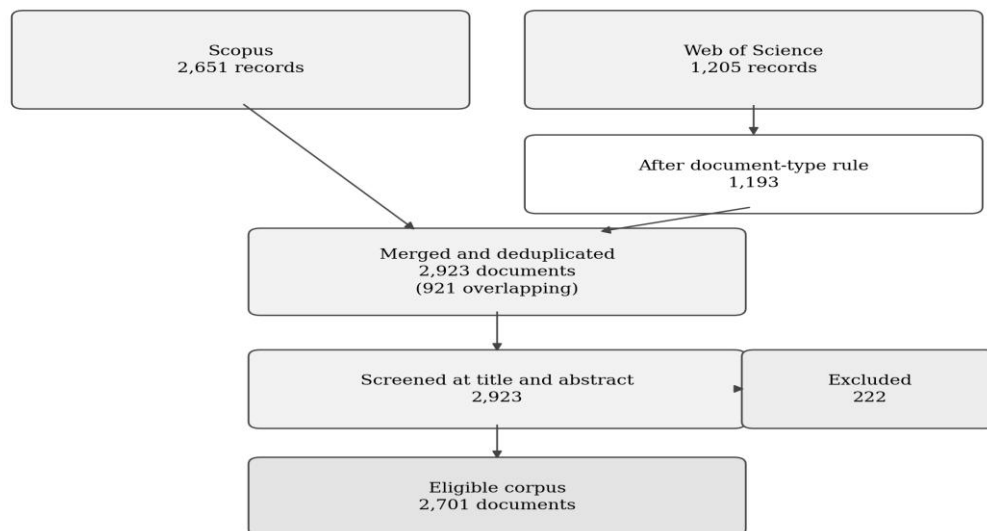


Figure 1. Corpus construction and screening, following PRISMA 2020 reporting conventions (Page et al., 2021).

Exclusion reason	n
Survey reporting no firm-level financial data	72
Review, meta-analysis or bibliometric study	69
No ESG or CSR construct in the predictor	60

Portfolio or fund level rather than firm level	16
Outcome is not financial performance	4
Withdrawn publication	1
Total	222

Table 2. Screening exclusions.

4.2 Database overlap

76.4 per cent of the Web of Science set appears in Scopus; 34.8 per cent of the Scopus set appears in Web of Science. Web of Science contributes 272 unique records, 9.3 per cent of the merged corpus.

This asymmetry could in principle be an artefact of query construction, since the two strings are not word-for-word identical, because Web of Science permits truncation where Scopus requires singular and plural forms enumerated separately. It is not. Only 8 Scopus records are retrieved solely by terms absent from the Web of Science block, and 1,435 of the 1,464 records in the pre-revision Scopus set contained at least one term present in both. Fewer than 2 per cent of the gap is attributable to the strings; the rest is indexing coverage, consistent with the comparative journal analysis of Mongeon and Paul-Hus (2016).

The practical consequence is asymmetric, and it matters because most reviews of this literature search Scopus alone. A Scopus-only corpus would have missed 272 records, 9.3 per cent of the total. A Web of Science-only corpus would have missed 1,729, or 59 per cent. Single-database reviews of this field are therefore defensible in one direction and not the other, and the direction is rarely stated. Reporting the overlap in both directions costs nothing and tells a reader how much the choice of database constrained what follows.

4.3 Coverage limitations

Three constraints bound what this corpus can support, and they are stated here rather than in the limitations because they qualify the results directly.

The Emerging Sources Citation Index returned no records across three independent exports; the institutional subscription does not appear to include it. Emerging-market research indexed only in ESCI cannot enter the corpus.

Pre-2005 coverage is thin: 16 Scopus and 9 Web of Science records. Adding “corporate social responsibility”, the dominant construct term of that period, moved the corpus from 1,464 to 2,651 records but barely moved the pre-2005 count. Two seed papers whose titles contain operative search terms were not retrieved despite qualifying on document type, language and subject area, and the cause was not established.

Recall stands at 72.7 per cent against a standard requiring complete retrieval. Of six unretrieved seeds, two are reviews excluded by design and one is unretrievable without admitting the bare term “performance” to a title-restricted outcome block. Three represent genuine gaps. The honest statement is that this search achieved high precision and incomplete recall, with under-representation of pre-2005 work.

5. RESULTS

5.1 Performance analysis

The corpus comprises 2,923 documents in 701 journals, attracting 185,390 citations, a mean of 63.4 per document. Annual production grew at 17.4 per cent between 1997 and 2025, from 6 documents to 541, with output roughly doubling between 2022 and 2025 (Figure 2). The growth rate is anchored on a base year holding 6 records and is reported as provisional; the pre-2005 sparsity discussed above constrains its precision.

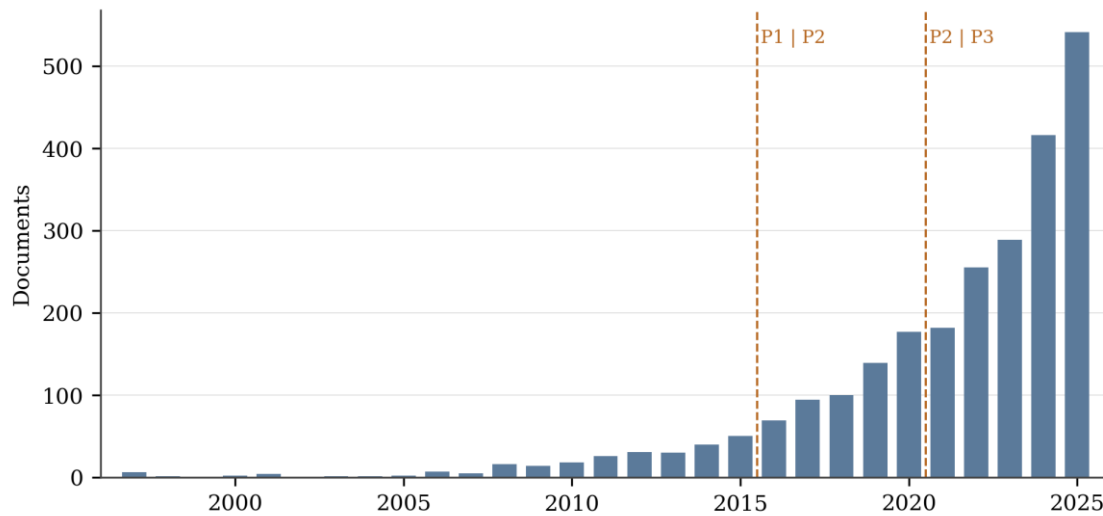


Figure 2. Annual production, 1997–2025. Dashed lines mark the period cuts used in Section 5.3. 2026 excluded as a partial year.

Sources are moderately concentrated: Bradford’s core comprises 23 journals accounting for 33 per cent of the corpus, and Table 4 lists the leading outlets. Corporate Social Responsibility and Environmental Management leads on volume with 175 documents, followed by Business Strategy and the Environment (99) and the Journal of Business Ethics (73).

Source	Documents	Citations	Per document
Corporate Social Responsibility and Environmental Management	175	10,165	58
Business Strategy and the Environment	99	9,579	97
Journal of Business Ethics	73	15,819	217
Finance Research Letters	64	3,094	48
Cogent Business and Management	54	1,676	31
Journal of Risk and Financial Management	54	767	14
Social Responsibility Journal	41	2,730	67
Journal of Cleaner Production	40	5,420	136
Journal of Business Research	37	4,719	128
Business Strategy and Development	35	693	20

Strategic Management Journal	19	20,421	1,074
Management Science	8	6,156	769

Table 4. Leading sources by document count, with Strategic Management Journal and Management Science added to show the volume–influence divergence.

Volume and influence diverge by two orders of magnitude. Strategic Management Journal contributes 19 documents at 1,074 citations each and Management Science 8 documents at 769; Journal of Risk and Financial Management contributes 54 documents at 14. The journals that publish most of this literature are not the journals that shape it. Any ranking by document count alone, which is the convention in most mappings of this field, misrepresents where its influence sits, and the divergence is itself consistent with the integration argument: a field whose highest-impact work appears in outlets that publish it rarely is not one whose central venues are consolidating knowledge.

Geographically, China (492 documents), the United States (397), India (289), the United Kingdom (269) and South Korea (259) lead. India in third place, with Indonesia and Malaysia in the top seven, indicates substantial emerging-market participation despite the ESCI gap.

The most-cited documents are the seed papers used in the recall diagnostic: Waddock and Graves at 4,958 citations, Friede, Busch and Bassen at 3,217, Dhaliwal and colleagues at 3,167. The corpus contains the field’s anchors rather than only its periphery.

5.2 Conceptual structure

Author-keyword co-occurrence across 2,805 documents carrying two or more keywords produced a network of 119 nodes and 442 edges, partitioned into fifteen themes and plotted in Figure 3.

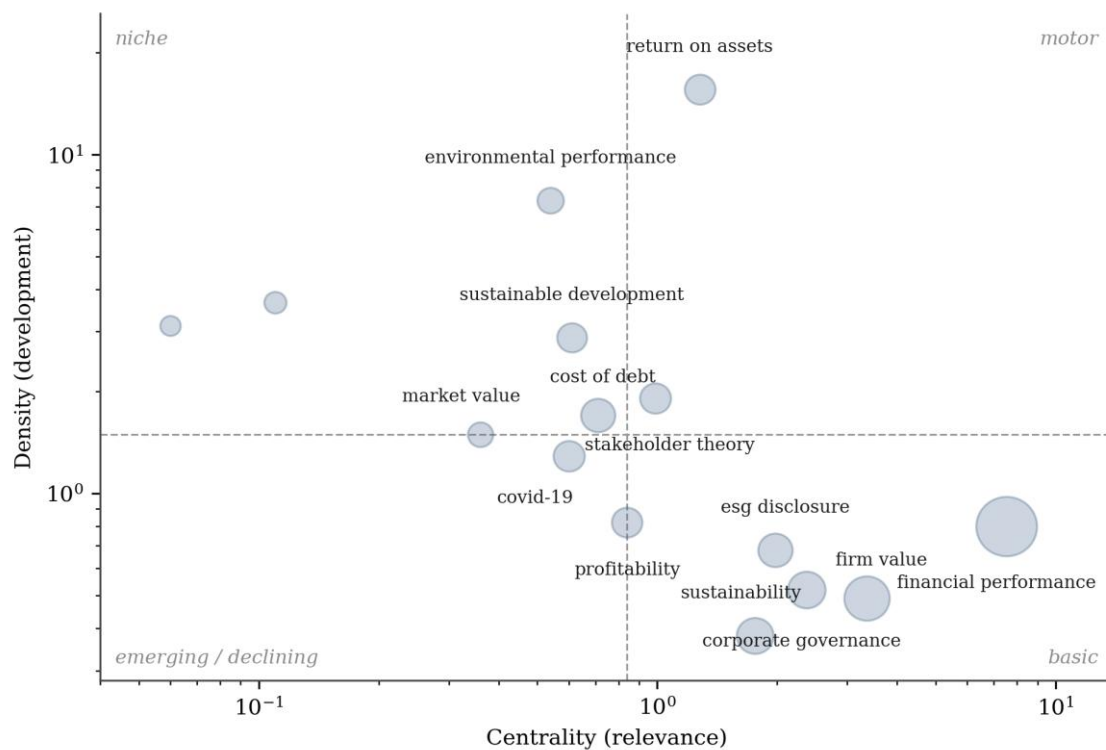


Figure 3. Thematic map. Bubble area is proportional to keyword occurrences; dashed lines mark median centrality and density.

The field's core occupies the basic quadrant, meaning high centrality and low density. Financial performance (2,785 occurrences, centrality 7.52, density 0.80), firm value (869, 3.36, 0.49), corporate governance (409), sustainability (376), ESG disclosure (285) and profitability (177) connect to everything and cohere internally with almost nothing. The financial performance cluster alone absorbs 32 terms including innovation, machine learning, content analysis and the sustainable development goals.

This is the study's first substantive finding. After five decades the ESG–performance question remains conceptually unconsolidated: papers attach core constructs to whatever contextual variable is current without building cumulative sub-literatures. The field's growth has been extensive rather than intensive.

Only one genuine motor theme exists, namely stakeholder theory with agency theory, legitimacy theory, institutional investors and the resource-based view (189 occurrences, centrality 0.99, density 1.91). The theoretical apparatus is well connected and internally coherent, but small against 2,785 occurrences in the core. The other high-density cluster, return on assets with Tobin's Q and return on equity (density 15.57), is a measurement convention crystallising rather than a research theme; those three measures co-occur almost deterministically.

Cost of debt, cost of capital, cost of equity, sustainable finance and information asymmetry form a coherent niche of 281 occurrences: methodologically self-contained and weakly connected to the accounting-performance mainstream, which mirrors the separate finance school identified in the co-citation analysis below.

5.3 Thematic evolution

Three periods were cut at 2015 and 2020 from the production curve. Prevalence is reported as a percentage of documents within each period, since the periods differ sevenfold in size, and the leading trajectories are plotted in Figure 4.

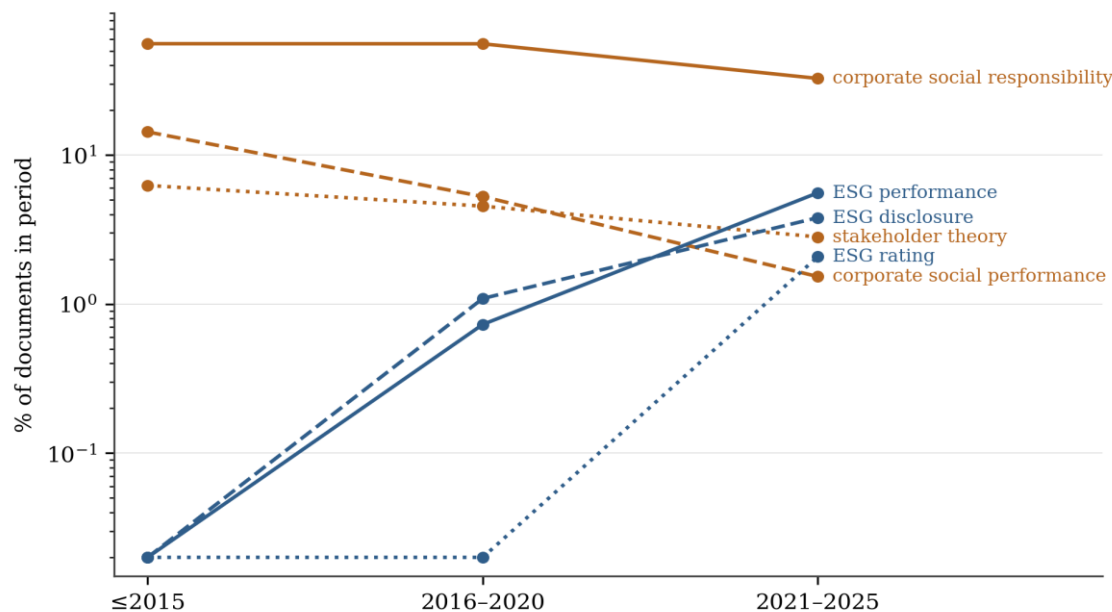


Figure 4. Vocabulary turnover. Orange series are corporate social responsibility vocabulary; blue series are ESG vocabulary. Log scale; ESG series plotted at 0.02 in periods where prevalence is zero.

Keyword	≤2015	2016–20	2021–25
corporate social responsibility	55.80	55.72	32.62

corporate social performance	14.29	5.26	1.53
stakeholder theory	6.25	4.54	2.82
esg performance	0	0.73	5.58
esg disclosure	0	1.09	3.80
esg rating	0	0	2.08
esg controversies	0	0	1.35
return on assets	0	1.81	2.51
covid-19	0	0	2.64

Table 3. Keyword prevalence by period, per cent of documents in that period.

The field's vocabulary turned over almost completely. Corporate social responsibility fell from 55.8 per cent of documents before 2015 to 32.6 per cent after 2021, while an entire ESG vocabulary appeared from nothing. None of the ESG terms in Table 3 existed as author keywords before 2016.

Prevalence alone cannot distinguish relabelling from displacement, so the two were tested directly by comparing co-word neighbourhoods. For each focal term the fifteen most strongly associated keywords were extracted by association strength and the overlap measured by Jaccard index. If ESG were a new name for the same construct, the neighbourhood of ESG performance today should closely resemble that of corporate social responsibility before 2015. Table 5 reports the result.

Comparison	Jaccard
CSR ($\leq$ 2015) vs ESG performance (2021–25)	0.111
CSR ($\leq$ 2015) vs CSR (2021–25)	0.200
CSR (2021–25) vs ESG performance (2021–25)	0.304

Table 5. Co-word neighbourhood overlap, fifteen strongest associates per term.

It does not. The overlap is 0.111. Pre-2015 corporate social responsibility sits among stakeholders, stakeholder theory, socially responsible investment and social reporting; ESG performance today sits among ESG controversies, ESG ratings, board gender diversity, greenwashing and the banking sector. These are different conceptual constellations, not one label exchanged for another.

The second row is equally informative: corporate social responsibility overlaps its own former neighbourhood by only 0.200. The term survived while the concepts around it were replaced. And the third row shows that the two present-day terms resemble each other (0.304) more than either resembles the pre-2015 literature, which is what a wholesale generational turnover looks like rather than a substitution of one construct for another.

The appropriate description is therefore displacement without consolidation. The ESG framing did not inherit the corporate social performance tradition's conceptual position; it established a different one while that tradition receded. Whether the two measure a common underlying construct is precisely the question the field has not answered, and Section 6 argues it is the question that most needs answering.

The corporate social performance tradition has retired. Its fall from 14.3 to 1.5 per cent is the sharpest single movement in the data, and its apparatus thinned in step: stakeholders 3.6 to 0.7, stakeholder management 1.8 to 0.3. This is the vocabulary of the meta-analytic syntheses of the early 2000s, and its disappearance means that studies from that era measure a different operationalisation under a different name.

Outcome vocabulary arrives late. Return on assets, return on equity, profitability and market value are absent from pre-2015 author keywords and material afterwards. Authors did not index by outcome measure before roughly 2016, which partly explains why a title-restricted outcome block under-retrieves early work, connecting this finding directly to the recall limitation in Section 4.3.

COVID-19 is emergence rather than decline. A static thematic map cannot distinguish an emerging theme from a dying one, since both occupy the same quadrant. The period analysis resolves it: covid-19 moves from zero to 2.64 per cent and attaches to the event-study tradition within the 2021–2025 sustainability cluster.

Method has become an identity. Panel data, machine learning, endogeneity and quantile regression enter the keyword space only from 2016, consistent with a field professionalising empirically.

The pre-2015 period cannot be thematically mapped. Its 224 keyword-bearing documents produced a co-occurrence network that fragmented into twelve single-term components with uninterpretable centrality and density. Term trajectories are robust to this sparsity and are reported for all three periods; cluster structure is not, and is reported only from 2016. Any claim about how themes evolved from the early period rests on those 224 documents.

5.4 Intellectual structure

Co-citation analysis across 3,830 citing documents, averaging 67.7 references each, identified 63,367 distinct cited works. The network of the 90 most-cited references partitions into eight schools (Table 6); Figure 5 plots the denser core of that network.

School	Works	Co-citations	Anchor
Finance and capital markets	22	5,793	Jensen & Meckling (1976)
CSP–CFP empirical synthesis	12	5,058	Waddock & Graves (1997)
Contemporary ESG measurement	22	4,991	Friede, Busch & Bassen (2015)
Strategic CSR and theory of the firm	12	3,410	McWilliams & Siegel (2001)
Resource-based and stakeholder foundations	11	3,119	Barney (1991)
Foundational normative debate	3	1,719	Freeman (1984)
Contingency and curvilinearity	5	1,200	Brammer & Millington (2008)
Dynamic panel econometrics	3	827	Arellano & Bond (1991)

Table 6. Co-citation schools. Louvain resolution 1.3, association strength weighting, edges retained at 25 or more co-citations.

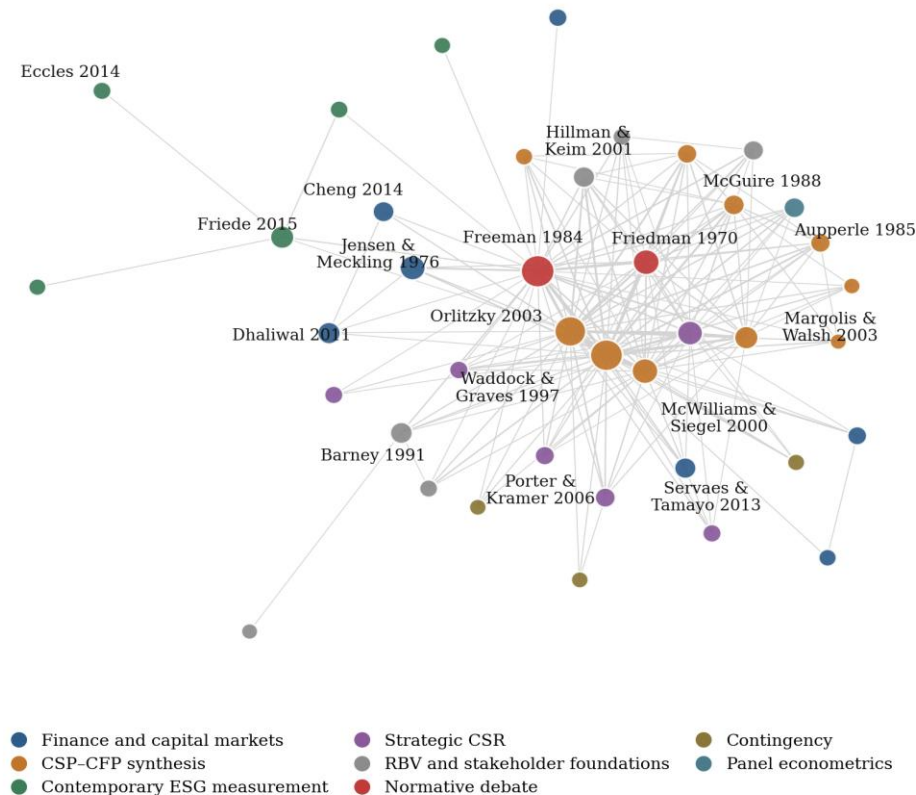


Figure 5. Co-citation network of the 46 most-cited works, edges at 100 or more co-citations. Node area is proportional to citation frequency; colour denotes school.

The displacement reading of Section 5.3 is corroborated from the citation side. If ESG were merely a new label on corporate social responsibility, the ESG literature would co-cite the corporate social performance canon and dissolve into that community. It does not. Contemporary ESG measurement forms a distinct school of 22 works, almost entirely post-2014, anchored on Friede rather than Waddock. The keyword evidence and the citation evidence therefore agree: a different constellation of concepts, cited by a different community.

Recent ESG work cites recent ESG work. It has grown alongside the corporate social performance tradition rather than absorbing it, with limited traffic between them. This is what the conceptual-structure finding looks like from the citation side: a field expanding without integrating, and doing so by adding communities rather than by consolidating existing ones.

The largest school by weight is finance and capital markets, and it asks a different question from the others: not whether ESG pays, but how markets price it. Berg, Kölbl and Rigobon (2022), on ESG rating divergence, sits inside the contemporary measurement school, indicating that the field has internalised measurement disagreement as a live problem rather than treating ratings as interchangeable.

5.5 Social structure

2,649 records carry parseable Scopus Author IDs, yielding 6,600 distinct authors, a mean of 2.98 per paper and 286 single-authored papers. The most productive are Cemil Kuzey and Ali Uyar with 13 and 12 documents, Seoki Lee with 12, and Santi Gopal Maji and Abdullah Karaman with 10 each.

Author productivity departs sharply from Lotka's law. Lotka (1926) predicts that the proportion of authors producing n papers falls as n raised to about -2 , implying roughly 60 per cent single-paper authors. This corpus returns a fitted exponent of 3.72 and 86.6 per cent single-paper authors (Figure 6a). The distribution is far steeper than the informetric norm: the field has an unusually large transient population and an unusually thin core of sustained contributors. This supplies the benchmark that a bare fragmentation count cannot, since the structure is not merely dispersed, it is dispersed well beyond what productivity distributions in established fields lead one to expect.

The field has no invisible college. Restricting to authors with two or more joint papers gives a network of 559 authors and 436 ties, fragmenting into 221 components of which the largest holds nine authors, or 1.6 per cent of the network, at a density of 0.0028. The largest groups are small and national: a Malaysian–Yemeni group of nine, a Taiwanese group of seven, an Indian group of six, and the Kuzey–Uyar–Karaman group of five accounting for 39 documents.

International co-authorship stands at 27.0 per cent of records with identifiable countries, with the strongest ties China–UK (34 joint publications), China–USA (28), China–Hong Kong (26) and UK–USA (25).

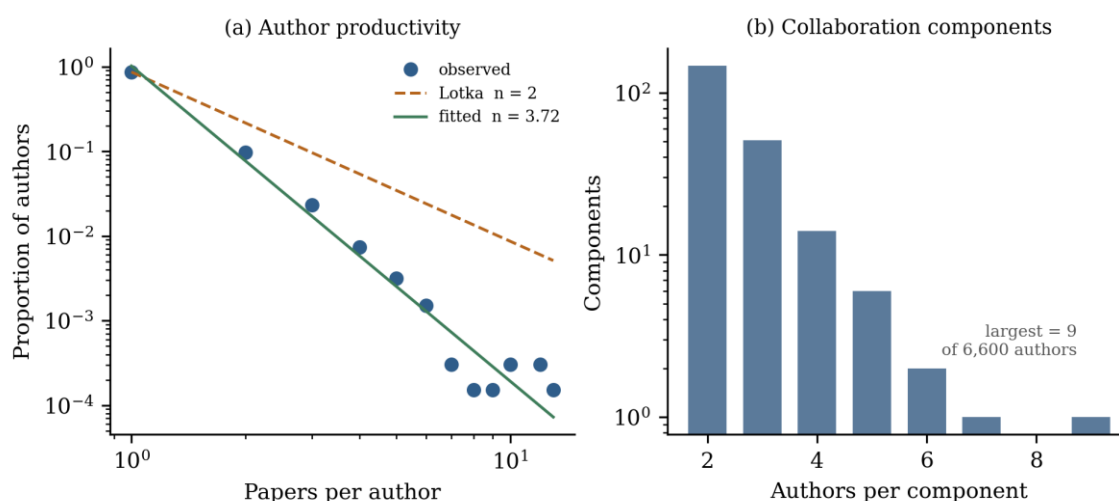


Figure 6. Social structure. (a) Author productivity against Lotka's law. (b) Size distribution of collaboration components among authors with two or more joint papers.

For a literature of this size and growth rate the fragmentation is notable, and it converges with the conceptual-structure finding. A community organised into hundreds of disconnected small groups produces breadth without cumulative depth, which is what an undeveloped conceptual core looks like from the social side.

6. DISCUSSION

Three findings converge from independent evidence, and the convergence is the argument. The conceptual core sits in the basic quadrant, connected to everything and coherent with nothing (5.2). The citation base has bifurcated between a retiring corporate social performance tradition and a self-referential ESG literature (5.4). The social structure fragments into 221 disconnected groups with no component exceeding nine authors (5.5). Three methods, three data sources, one conclusion: the field grows without integrating.

The vocabulary turnover is displacement rather than relabelling, and the distinction matters. A direct test of co-word neighbourhoods (5.3) returns an overlap of 0.111 between corporate social responsibility before 2015 and ESG

performance today, which is too low to sustain the common claim that ESG is simply a new name for an old construct. The ESG framing established its own conceptual position while the corporate social performance tradition receded, and it built its own citation community rather than inheriting one (5.4).

This is a stronger problem than relabelling would have been. If ESG were merely CSR renamed, accumulated evidence would remain comparable and synthesis would be straightforward. Instead the field has replaced its central construct with a different one, retained the outcome variables, and continued to pool across the break. Nothing in the literature establishes what the two measurement traditions have in common, and the structural evidence indicates the communities that would have to do that work barely communicate.

This bears directly on the syntheses the field relies on. Orlitzky, Schmidt and Rynes (2003) pooled corporate social performance measures; Friede, Busch and Bassen (2015) pooled across the ESG era; Hirsch and colleagues (2023) provide the most recent and methodologically careful treatment. Section 5.3 shows that the constructs these syntheses aggregate are not the same operationalisation under a different name but different operationalisations under different names. Any quantitative pooling that treats CSP measures, ESG disclosure indices and vendor ratings as one construct is averaging quantities the field has not shown to be commensurable.

Hirsch and colleagues are worth dwelling on, because their result and this one point the same way from opposite directions. Their meta-regression finds a small positive mean effect that attenuates once publication selection is modelled, and reports that the estimate depends substantially on how the primary studies are specified. That is a statistical demonstration that the underlying evidence is not homogeneous. The structural evidence here supplies the reason: the studies being pooled belong to citation communities that barely intersect (Section 5.4), use construct vocabularies that replaced one another rather than accumulating (Section 5.3), and are produced by research groups that do not communicate (Section 5.5). Heterogeneity in a meta-regression is usually treated as a nuisance parameter to be modelled. Here it is the finding.

The practical implication is not that synthesis should stop, but that it should condition on construct generation. A pooled estimate across the CSP era and the ESG era answers no well-posed question until someone establishes what the two measurement traditions have in common. The presence of rating-divergence work inside the contemporary measurement school (Berg et al., 2022) indicates researchers know this; synthesis practice has not caught up.

The contingency literature identified as its own small school (Table 6) offers the alternative reading, and the structural evidence is consistent with it. If returns to social investment genuinely depend on firm and context, as Barnett and Salomon argue, then five decades of inconsistent findings are not a measurement failure to be resolved by better pooling but a correct description of a heterogeneous phenomenon. That the contingency school remains small, at five works and 1,200 co-citations, while the field continues to seek a general answer is itself a finding about how this literature allocates attention.

One caution on the fragmentation evidence. Bibliometric networks are always sparse at the author level, and a low largest-component share is not by itself pathological. What makes the figure here interpretable is the Lotka comparison (5.5): a fitted exponent of 3.72 against a predicted 2.0, and 86.6 per cent single-paper authors against a predicted 60 per cent. The field is not merely dispersed but dispersed well beyond the informetric norm, and it is that departure rather than the raw component count that supports the reading offered here.

For the audit framing, the methodological results are not incidental. Automated screening failed at a kappa of 0.297 not because the rule was poorly specified but because the properties that determine inclusion, namely what the predictor measures, whose conduct it describes and what the unit of analysis is, are properties of research design that vocabulary does not encode. Reviews reporting keyword-based screening without a manually validated agreement statistic should be read accordingly, and raw percentage agreement in a corpus where most records are includable is close to uninformative.

The reliability of these structural findings was tested rather than assumed. Varying the co-word edge threshold and Louvain resolution across nine specifications leaves the basic quadrant holding 79 to 87 per cent of keyword

occurrences in eight of them; varying the co-citation node set and edge threshold across nine specifications returns 8 to 10 communities throughout (3.5). The findings are properties of the corpus, not of the parameters chosen to display it.

The reference-parsing fault has wider reach. Scopus is among the two most-used sources for bibliometric work, its reference strings are the standard input to co-citation analysis, and the delimiter collision produces results that are wrong in a specific and undetectable way: plausible author names, plausible years, systematically the wrong author. Any co-citation analysis built on unreconstructed Scopus reference fields should be re-examined.

6.1 Implications for practice

For researchers entering this field, the practical consequence is that construct choice is not a matter of convention. A study using a vendor ESG rating is not extending the corporate social performance literature, whatever its introduction claims, and citing across that boundary without addressing measurement equivalence imports a comparability assumption that the field has not earned.

For those conducting reviews, three procedures are cheap and change conclusions: calibrate the search in stages and report intermediate counts; test recall against a seed list built independently of the search; and report kappa alongside raw agreement, since in an imbalanced corpus the latter alone will make an inadequate rule look sound.

For practitioners and policymakers drawing on this evidence base, the caution is that summary claims about whether ESG pays rest on a literature whose central construct changed identity midway and whose syntheses pool across that change. The volume of studies is not the same as the weight of evidence.

7. LIMITATIONS

The Emerging Sources Citation Index was unavailable, excluding work indexed only there and bearing most on emerging-market coverage.

Recall is 72.7 per cent against a standard requiring complete retrieval, with pre-2005 coverage limited to 16 Scopus records. Two seed papers containing operative search terms in their titles were not retrieved and the cause was not established. The pre-2015 period could not be thematically mapped, so claims about evolution from the early period rest on 224 documents.

Screening decisions for 1,929 of 2,923 records rest on a rule-plus-audit procedure with a measured false-inclusion rate of 3.3 per cent, implying roughly 64 residual errors. Index keywords cover 326 of 2,923 records, so conceptual mapping uses author keywords and reflects how authors label their work rather than how it is catalogued.

Co-citation keys are first-author surname plus year; distinct works sharing both merge, and inconsistently recorded works split. At the top of the distribution, where the analysis operates, this is unlikely to bite. Network thresholds and Louvain resolutions are analyst choices; both were varied across nine specifications each and the central findings are stable (Section 3.5), but the specific cluster labels are not invariant and should be read as one defensible partition rather than the only one.

The diagnostics reported here were administered by the authors on their own search, and that is a real limitation of the audit framing rather than a formality. The precision sample of 30 was classified once, by the same people who designed the query; a blind second classification of that sample was not obtained. The 91 per cent inter-coder figure covers screening decisions, not the precision diagnostic. The direction of any resulting bias is toward leniency, since a coder assessing their own instrument is unlikely to be harsh, so the reported precision of 93.3 per cent should be read as an upper bound. The recall diagnostic is less exposed, because the seed list was constructed from the corpus's own citation data rather than by judgement, but the decision to treat two of the six misses as excluded by design was ours.

Finally, this study maps structure rather than content. It establishes that the ESG literature cites itself and occupies the position corporate social performance formerly held; it does not establish that the two constructs measure

the same thing. That is a question for measurement research, and Section 8 argues it is the one the field most needs answered.

8. CONCLUSION

This audit finds a literature that has grown by a factor of ninety since 1997 without consolidating what it knows. Its central constructs remain undeveloped, its citation communities have split rather than accumulated, and its authors work largely in isolation from one another. The transition from corporate social responsibility to ESG, ordinarily described as maturation, appears in the structural evidence as displacement: a different conceptual constellation, citing itself, established while the previous tradition receded and without any reconciliation between the two.

The prescription follows from the diagnosis. What this field needs is not more studies of whether ESG pays, of which it has thousands, but measurement reconciliation: explicit work establishing what, if anything, is common between corporate social performance as operationalised in 1997 and ESG as rated in 2025. Until that exists, each new study adds a data point to a series whose units are unknown, and each new synthesis averages across them.

For bibliometric practice, two lessons transfer beyond this literature. A search string is an instrument and should be calibrated and diagnosed as one; the diagnostics here changed recall by 22.7 points and would have changed every downstream result. And the tooling is not neutral: a delimiter collision in the most widely used bibliographic export format silently corrupts co-citation analysis, and the corruption is invisible without domain knowledge of the works involved.

Returning to the three questions. On RQ1, the evidence is of extensive growth without integration: the central themes are widely connected but internally incoherent, the citation base has added communities rather than consolidated them, and the collaboration structure is more dispersed than Lotka's law predicts. On RQ2, the shift from corporate social responsibility to ESG is neither conceptual development nor simple relabelling but displacement, evidenced by a co-word neighbourhood overlap of 0.111 and a separate citation school, established while the earlier tradition receded and without reconciliation between them. On RQ3, treating the search as a measured instrument changed recall by 22.7 points, revealed that keyword screening reproduces manual decisions at a kappa of 0.297, and exposed a reference-parsing fault that silently misattributes the field's most-cited works. Each of these would have gone unreported under conventional practice, and each would have changed what a reader concluded.

APPENDIX A. FINAL SEARCH STRINGS

SCOPUS

TITLE-ABS-KEY ("ESG" OR "environmental social and governance" OR "environmental, social and governance" OR "environmental social governance" OR "ESG score" OR "ESG scores" OR "ESG rating" OR "ESG ratings" OR "ESG performance" OR "ESG disclosure" OR "sustainability performance" OR "sustainability reporting" OR "sustainability disclosure" OR "corporate social performance" OR "corporate social responsibility" OR "corporate sustainability performance" OR "non-financial disclosure" OR "nonfinancial disclosure" OR "socially responsible") AND TITLE ("firm performance" OR "financial performance" OR "corporate financial performance" OR "firm value" OR "corporate value" OR "market value" OR "shareholder value" OR "shareholder wealth" OR "Tobin" OR "return on assets" OR "return on equity" OR "return on invested capital" OR "operating performance" OR "profitability" OR "stock return" OR "stock returns" OR "share price" OR "abnormal return" OR "abnormal returns" OR "financial returns" OR "cost of equity" OR "cost of capital" OR "cost of debt" OR "access to finance") AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "DECI")) AND (LIMIT-TO (LANGUAGE , "English"))

WEB OF SCIENCE

TS = ("ESG" OR "environmental social and governance" OR "environmental, social and governance" OR "ESG score*" OR "ESG rating*" OR "ESG performance" OR "ESG disclosure" OR "sustainability performance" OR "sustainability reporting" OR "sustainability disclosure" OR "corporate social performance" OR "corporate social

responsibility" OR "corporate sustainability performance" OR "non-financial disclosure" OR "nonfinancial disclosure" OR "socially responsible") AND TI = ("firm performance" OR "financial performance" OR "corporate financial performance" OR "firm value" OR "corporate value" OR "market value" OR "shareholder value" OR "shareholder wealth" OR "Tobin* Q" OR "return on assets" OR "return on equity" OR "operating performance" OR "profitability" OR "stock return*" OR "share price" OR "abnormal return*" OR "financial return*" OR "cost of equity" OR "cost of capital" OR "cost of debt" OR "access to finance") AND DT = (Article) AND WC = ("Business" OR "Business, Finance" OR "Economics" OR "Management") AND LA = (English)

Editions: SCI-Expanded, SSCI, CPCI-S, CPCI-SSH. ESCI unavailable under the institutional subscription. The complete query log, recording every clause with its count, timestamp and justification, is provided as supplementary material.

REFERENCES

1. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
2. Barauskaite, G., & Streimikiene, D. (2021). Corporate social responsibility and financial performance of companies: The puzzle of concepts, definitions and assessment methods. *Corporate Social Responsibility and Environmental Management*, 28(1), 278–287.
3. Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5), 423–429.
4. Cardillo, M. A. D. R., & Basso, L. F. C. (2025). Revisiting knowledge on ESG/CSR and financial performance: A bibliometric and systematic review of moderating variables. *Journal of Innovation and Knowledge*, 10(1), 100648.
5. Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6), 543–549.
6. Khan, M. A. (2022). ESG disclosure and firm performance: A bibliometric and meta analysis. *Research in International Business and Finance*, 61, 101668.
7. Aupperle, K. E., Carroll, A. B., & Hatfield, J. D. (1985). An empirical examination of the relationship between corporate social responsibility and profitability. *Academy of Management Journal*, 28(2), 446–463.
8. Barnett, M. L. (2007). Stakeholder influence capacity and the variability of financial returns to corporate social responsibility. *Academy of Management Review*, 32(3), 794–816.
9. Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal*, 33(11), 1304–1320.
10. Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.
11. Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344.
12. Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
13. Bradford, S. C. (1934). Sources of information on specific subjects. *Engineering: An Illustrated Weekly Journal*, 137, 85–86.
14. Brammer, S., & Millington, A. (2008). Does it pay to be different? An analysis of the relationship between corporate social and financial performance. *Strategic Management Journal*, 29(12), 1325–1343.
15. Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research. *Scientometrics*, 22(1), 155–205.
16. Carroll, A. B. (1979). A three-dimensional conceptual model of corporate performance. *Academy of Management Review*, 4(4), 497–505.
17. Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. *Strategic Management Journal*, 35(1), 1–23.
18. Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field. *Journal of Informetrics*, 5(1), 146–166.
19. Cochran, P. L., & Wood, R. A. (1984). Corporate social responsibility and financial performance. *Academy of Management Journal*, 27(1), 42–56.
20. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
21. Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review*, 86(1), 59–100.
22. Donaldson, T., & Preston, L. E. (1995). The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review*, 20(1), 65–91.
23. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296.

24. Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, 60(11), 2835–2857.
25. El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance*, 35(9), 2388–2406.
26. Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.
27. Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. *Journal of Sustainable Finance & Investment*, 5(4), 210–233.
28. Friedman, M. (1970, September 13). The social responsibility of business is to increase its profits. *The New York Times Magazine*.
29. Goss, A., & Roberts, G. S. (2011). The impact of corporate social responsibility on the cost of bank loans. *Journal of Banking & Finance*, 35(7), 1794–1810.
30. Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181–217.
31. Hillman, A. J., & Keim, G. D. (2001). Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal*, 22(2), 125–139.
32. Hirsch, S., Petersen, T., Koppenberg, M., & Hartmann, M. (2023). CSR and firm profitability: Evidence from a meta-regression analysis. *Journal of Economic Surveys*, 37(3), 993–1032.
33. Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360.
34. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
35. Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance*, 72(4), 1785–1824.
36. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12), 317–323.
37. Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly*, 48(2), 268–305.
38. McGuire, J. B., Sundgren, A., & Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *Academy of Management Journal*, 31(4), 854–872.
39. McWilliams, A., & Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5), 603–609.
40. McWilliams, A., & Siegel, D. (2001). Corporate social responsibility: A theory of the firm perspective. *Academy of Management Review*, 26(1), 117–127.
41. Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228.
42. Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl. 1), 5200–5205.
43. Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, 24(3), 403–441.
44. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
45. Porter, M. E., & Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, 84(12), 78–92.
46. Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. *Management Science*, 59(5), 1045–1061.
47. Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
48. Ullmann, A. A. (1985). Data in search of a theory: A critical examination of the relationships among social performance, social disclosure, and economic performance of US firms. *Academy of Management Review*, 10(3), 540–557.
49. Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
50. Waddock, S. A., & Graves, S. B. (1997). The corporate social performance–financial performance link. *Strategic Management Journal*, 18(4), 303–319.
51. Waltman, L., Van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635.
52. Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organization*