

Banking Financial Policies and Organizational Sustainability: A PRISMA 2020 Systematic Review, Bibliometric Mapping, and Future Research Agenda

SANJAY V S^{a,*} Dr. R.V. SUGANYA^b

^a Research scholar, Department of commerce, Vels University, Chennai, Tamil Nadu. Email:

sanjayvscpt@gmail.com

^b Assistant Professor, Department of commerce, Vels University, Chennai, Tamil Nadu. Email:

suganya.sms@velsuniv.ac.in

ABSTRACT

Purpose. Banks are simultaneously regulated entities, credit allocators, and organizations with their own sustainability performance. Because a bank's material sustainability impact is overwhelmingly indirect—mediated through the balance sheets it finances—its sustainability posture is enacted through *financial policy* rather than operational management. Yet the literature treats disclosure, lending, capital, and payout as separate problems despite their being jointly determined on a single balance sheet. This review integrates that fragmented evidence and derives a structured agenda.

Design/methodology/approach. Following PRISMA 2020, PRISMA-S, and the SPAR-4-SLR protocol, we searched Scopus and Web of Science Core Collection for peer-reviewed English-language articles published between [year1] and 2026. Of [n1] records identified, [N] met eligibility criteria after independent dual screening (Cohen's $\kappa = [k]$). We combine bibliometric performance analysis, three-technique science mapping (co-citation, bibliographic coupling, co-word) using *bibliometrix* and VOSviewer, and thematic content synthesis with tiered quality appraisal and CERQual certainty grading.

Findings. The corpus organises into seven policy domains and three weakly-connected intellectual traditions. Research attention is *inversely* related to material sustainability impact and *directly* related to observability: disclosure dominates, while payout policy, provisioning, and internal capital allocation are near-absent despite being the mechanisms through which transition-finance capacity is actually determined. Evidence on the sustainability–performance relationship is directionally positive but heterogeneous, with effect magnitude conditioned by institutional quality, ownership form, ESG-rating provider, and—critically—identification strategy. Bank-level units of analysis are structurally incapable of distinguishing real-economy decarbonisation from exposure reallocation between lenders.

Originality/value. This is the first review to treat the full portfolio of bank financial policies, rather than ESG disclosure or green banking alone, as the explanatory construct for organizational sustainability. We contribute a seven-domain taxonomy, the *policy portfolio view*, the *observability–materiality inversion* as a named diagnostic principle, an Antecedents–Decisions–Outcomes framework, eight falsifiable propositions with boundary conditions, and a full replication package.

Practical implications. The divergent regulatory environment of 2024–2028—voluntary Basel climate disclosure, staggered ISSB adoption across some 28 jurisdictions, and the EU's threshold-based scope reversal under Omnibus I—supplies the strongest identification environment this literature has had, and is being under-exploited.

Keywords: banking financial policy; organizational sustainability; ESG performance; sustainable banking; systematic literature review; PRISMA 2020; bibliometric analysis; science mapping; green finance; capital adequacy

JEL classification: G21; G28; G32; G35; M14; Q56

1. INTRODUCTION

1.1. Motivation

Banks occupy an unusual position in the sustainability transition. Unlike manufacturing firms, whose environmental footprint is dominated by direct operations, a bank's material sustainability impact is overwhelmingly *indirect*, mediated through the balance sheets of the firms and households it finances. Financed emissions typically exceed operational emissions by two to three orders of magnitude. The sustainability of a bank is therefore not primarily a facilities-management question but a **financial policy** question.

This observation carries a corollary the literature has been slow to internalise. If sustainability in banking is enacted through financial policy, the relevant explanatory variables are not only the ESG ratings and sustainability disclosures that dominate empirical work, but the underlying policy architecture: capital structure and buffer management, credit origination and pricing standards, provisioning and impairment policy, funds transfer pricing, risk appetite calibration, disclosure and assurance choices, dividend and buyback policy, incentive design, and the digital channel strategies determining reach and inclusion.

Three developments make integration timely.

First, the **regulatory architecture has consolidated and then partially retrenched**, producing an unusually informative natural-experiment landscape. The Basel Committee, having consulted in November 2023 on integrating climate disclosures into Pillar 3, published in June 2025 a framework that is *voluntary* rather than mandatory, with domestic implementation left to national supervisors (BCBS, 2025). In the European Union, Directive (EU) 2026/470—the Omnibus I amendment,

published 26 February 2026 and in force from 19 March 2026—materially narrowed the scope of the Corporate Sustainability Reporting Directive and the Corporate Sustainability Due Diligence Directive, following a “stop-the-clock” postponement of Wave 2 and Wave 3 reporting. Simultaneously, ISSB adoption has spread, with approximately 28 jurisdictions having adopted IFRS S1 and S2 on a voluntary or mandatory basis by April 2026 and a further twelve planning adoption. The resulting regime is *divergent*—tightening in some jurisdictions, loosening in others—offering identification leverage that has barely been exploited.

Second, the **empirical literature has outgrown narrative synthesis**. Prior reviews have each bounded the problem more narrowly than the phenomenon warrants. Aracil et al. (2021) mapped 676 sustainable banking articles (1995–2019) into nine clusters across three domains; Galletta et al. (2022) bibliometrically analysed 271 publications on ESG performance in banking; more recent systematic reviews concentrate on the ESG–bank performance association specifically. None takes *financial policy* as the organising construct. The consequence is that findings on capital, credit, and payout policy sit in separate, largely non-communicating literatures despite being jointly determined.

Third, **standards for evidence synthesis have risen**. PRISMA 2020 (Page et al., 2021) tightened reporting expectations; SPAR-4-SLR (Paul et al., 2021) formalised domain-review procedure for business research; bibliometric practice has matured beyond descriptive counting toward science mapping (Donthu et al., 2021; Zupic & Čater, 2015). Reviews combining PRISMA-grade transparency with rigorous science mapping remain uncommon in banking and finance, where narrative reviews still predominate.

1.2. Positioning against prior reviews

Table 1 states the delta explicitly. The distinguishing feature of the present review is not scale but *organising construct*: where prior work classifies studies by outcome (ESG performance, financial performance, disclosure), we classify by *decision lever*. This reorganisation is what makes the coverage imbalance visible, because imbalance is invisible under a classification scheme that does not enumerate the levers.

1.3. Research gap

We identify four gaps.

Gap 1 — Construct fragmentation. “Sustainable banking,” “green banking,” “ESG in banking,” and “responsible banking” are used near-interchangeably despite denoting different constructs with different measurement implications. No review establishes a policy-anchored taxonomy rendering these literatures commensurable.

Gap 2 — Unquantified domain imbalance. Disclosure and lending appear to dominate while payout policy, funds transfer pricing, internal capital allocation, and provisioning are near-absent—despite each being a plausible transmission channel. This imbalance has never been systematically quantified, and its *cause* has never been theorised.

Gap 3 — Weak causal identification. The modal design regresses a performance measure on a composite ESG score, addressing endogeneity (where addressed) through lagged regressors or system GMM. Documented divergence among ESG rating providers (Berg et al., 2022) undermines the interpretability of this design, yet the field’s response has been incremental rather than structural.

Gap 4 — Absent causal architecture. No published framework specifies antecedents of bank financial policy choice, the policy decisions themselves, sustainability outcomes, and the contingencies linking them. Without such an architecture the field accumulates coefficients rather than knowledge.

1.4. Research questions

RQ1. How has research on banking financial policies and organizational sustainability evolved in volume, geography, outlet, and authorship?

RQ2. What is the intellectual structure of the field—its foundational works, distinct schools, and degree of integration?

RQ3. What is its conceptual structure: which themes are motor, basic, niche, emerging, or declining?

RQ4. Which financial policy domains have been studied, and what does the evidence establish within each?

RQ5. What theories, contexts, characteristics, and methods dominate, and what does this pattern exclude?

RQ6. What contingencies explain heterogeneity in observed policy–sustainability relationships?

RQ7. What agenda follows, and what methodological upgrades are prerequisite to executing it?

1.5. Contributions

(i) **A policy-anchored taxonomy and the *policy portfolio* view.** We define banking financial policies as a seven-domain construct (Section 2.1) and argue that these domains must be analysed jointly because they are jointly determined. (ii) **The *observability–materiality inversion*,** a named diagnostic principle explaining the field’s coverage pattern and generating testable predictions about where evidence will and will not accumulate (Section 7.2). (iii) **Methodological integration:** PRISMA 2020 and PRISMA-S screening transparency paired with three complementary bibliometric techniques, each answering a distinct question. (iv) **An ADO framework** specifying antecedents, decisions, outcomes, and moderators (Section 7.1). (v) **Eight falsifiable propositions with boundary conditions** (Section 7.2) and (vi) **a TCCM-structured agenda** of 24 questions with paired research designs (Section 8). (vii) **A replication package**—complete Boolean strings, decision rules, coding instrument, and software parameters—permitting exact replication and periodic updating.

1.6. Structure

Section 2 develops the conceptual and theoretical background. Section 3 details the protocol. Sections 4 and 5 report performance analysis and science mapping. Section 6 presents the content synthesis by policy domain. Section 7 discusses and advances the propositions. Section 8 sets the agenda. Sections 9–11 cover implications, limitations, and conclusion.

2. METHODOLOGY

2.1. Review design and reporting standards

The review is reported in accordance with the **PRISMA 2020** statement (Page et al., 2021) and the **PRISMA-S** search-reporting extension (Rethlefsen et al., 2021). Because PRISMA was developed for health-sciences intervention reviews, we supplement it with **SPAR-4-SLR** (Paul et al., 2021), which structures business-domain reviews into *assembling*, *arranging*, and *assessing* stages, and with the guidance of Lim et al. (2022) and Kraus et al. (2020). Bibliometric procedure follows Donthu et al. (2021) and Zupic & Čater (2015).

The design is a **hybrid**: a systematic review (PRISMA-governed selection, structured extraction, quality appraisal, thematic synthesis) combined with a bibliometric review (performance analysis and science mapping). The pairing is deliberate. Bibliometrics answers questions about the *field*—its structure, trajectory, and intellectual lineage—that no amount of close reading can answer. Content synthesis answers questions about the *findings* that bibliometrics cannot reach. Reviews using one method answer half the question. The completed 27-item PRISMA checklist appears in Appendix B.

2.2. Protocol and registration

The protocol was developed a priori and registered on the Open Science Framework before screening commenced (DOI: [OSF DOI]; registered [date]). Deviations are reported in Appendix B, item 24c. Registration on OSF rather than PROSPERO reflects the latter's restriction to health-related reviews.

2.3. Eligibility criteria

Criteria were specified using a PICOS structure adapted to organizational research (Table 4). The lower bound of the search window, [year1], is justified as [rationale: e.g., the publication year of Basel II, anchoring the window in the prudential-policy regime].

2.4. Information sources

Primary databases. Scopus (Elsevier) and Web of Science Core Collection (Clarivate), both searched on [date]. Dual-database searching is the expected minimum: Scopus offers broader business and finance journal coverage, WoS deeper historical citation data. Where records appeared in both, WoS metadata was retained for co-citation analysis (superior cited-reference structuring) and Scopus metadata for performance analysis (superior affiliation and funding fields).

Supplementary sources. (i) Backward snowballing on reference lists of all included studies and of prior reviews identified in Table 1; (ii) forward citation searching on the ten most-cited included studies; (iii) hand-searching the five most productive journals identified in Section 4.3 for the final twelve months; (iv) screening of BCBS, NGFS, EBA, IFRS Foundation, and UNEP FI publication lists to identify academic work not indexed under the search terms.

Merging and de-duplication. Records were exported as .bib and .csv, merged in R using `bibliometrix::mergeDbSources()` with `remove.duplicated = TRUE`, then verified manually against DOI and title-normalised string matching to capture residual duplicates arising from metadata inconsistency.

2.5. Search strategy

Following PRISMA-S, we report the full strategy, its development, and its validation. The string was constructed on three concept blocks joined by AND, with within-block synonyms joined by OR, and refined through three pilot rounds against a **gold-standard set of fifteen known-relevant articles** (Appendix A.4). The final string retrieved [15/15], giving [100]% sensitivity against that set. Database-specific, copy-executable strings with retrieval counts and dates appear in Appendix A; the refinement log, including rejected terms and the precision estimates justifying rejection, appears in Appendix A.3.

2.6. Selection process

Screening proceeded in four stages in Rayyan, with reviewers blinded to one another's decisions.

1. **Identification.** Scopus [n], WoS [n], supplementary [n]; total [n1].
2. **Duplicate removal.** [n] removed, leaving [n2].
3. **Title/abstract screening.** Two reviewers screened all [n2] records independently. Raw agreement [%]; Cohen's κ = [k], interpreted against Landis & Koch (1977) benchmarks. Disagreements were resolved by discussion, with [n] residual cases adjudicated by a third reviewer. Excluded: [n]. Retained: [n3].
4. **Full-text assessment.** [n3] full texts sought; [n] not retrievable after institutional and inter-library request (Appendix A.5). Exclusions were recorded with reasons at record level.

Final corpus: ****[N]** studies**.

2.7. Data extraction and coding

A structured extraction form (Appendix C) was piloted on 10% of included studies by both reviewers, refined, then applied to the full corpus. Fields cover bibliographic metadata; research question; theoretical lens; policy domain coded to Table 2 (multiple coding permitted); sustainability layer coded to Section 2.2; construct operationalisation and ESG data provider; unit of analysis, sample, country, period,

and bank type; design, estimation method, and endogeneity treatment; direction and significance of the principal finding; moderators and mediators tested; and stated limitations and future directions. Extraction was duplicated for a random 20% of the corpus; discrepancy rate [%].

2.8. Quality appraisal and risk of bias

Because the corpus is methodologically heterogeneous, no single instrument suffices. We applied **tiered appraisal**: a purpose-built 10-item instrument for quantitative studies (Appendix D); MMAT qualitative criteria (Hong et al., 2018) for qualitative work; full MMAT for mixed methods; AMSTAR-2 core domains (Shea et al., 2017) for reviews; and narrative appraisal of argumentative rigour and boundary-condition specification for conceptual papers.

Appraisal informed **sensitivity analysis, not exclusion**: the synthesis was re-run restricted to the top appraisal tertile to test whether conclusions are driven by weaker work (Section 6.10).

Two systematic risk-of-bias concerns warrant explicit treatment. **Publication bias**: null findings on the ESG–performance relationship are plausibly under-published; we assess this through the distribution of reported findings and note the absence of registered-report designs in the corpus. **Measurement bias**: documented rating divergence (Berg et al., 2022) means studies using different providers are not measuring the same construct; provider identity is coded and cross-provider consistency reported in Section 6.9.

2.9. Bibliometric procedures

Table 5 specifies each technique, the question it answers, and its parameters. Two procedural points deserve emphasis.

Thesaurus construction. A thesaurus file merges orthographic and terminological variants (*ESG / environmental social and governance; CSR / corporate social responsibility; bank / banks / banking*). Failure to construct a thesaurus is the most common defect in published bibliometric work, producing artefactual clusters that mistake vocabulary variance for conceptual distinction. The file is published in the replication package.

Time-slice selection. Slice boundaries were chosen on substantive rather than arithmetic grounds, corresponding to regime breaks: the post-crisis Basel III period; the post-Paris/TCFD period from 2016–2017; and the post-ISSB/CSRD period from 2023. Equal-length slices are

indefensible when the underlying regime shifts discontinuously.

2.10. Synthesis methods

Content synthesis followed **thematic synthesis** (Thomas & Harden, 2008): line-by-line coding of findings sections, development of descriptive themes, and generation of analytical themes exceeding the primary studies. Coding was conducted in NVivo with an initial deductive frame from Table 2, permitting inductive addition of unanticipated themes.

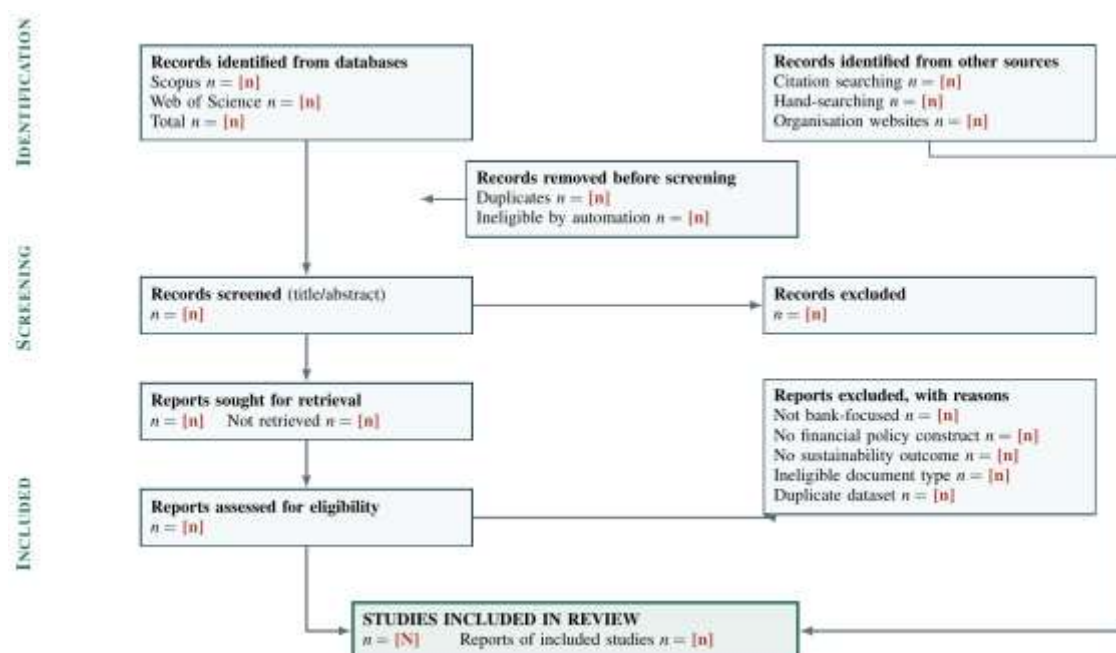
Given heterogeneity of constructs, measures, and designs, **meta-analysis was not attempted**. This is a substantive judgement, not an omission: pooling effect sizes across studies using different ESG providers, different sustainability layers, and different policy domains would produce a precise estimate of an uninterpretable quantity. Section 8 specifies the conditions under which meta-analysis would become defensible. In its place we apply structured **vote counting by direction of effect**, stratified by identification strategy (Section 6.9)—a design that converts the field’s methodological heterogeneity from a nuisance into evidence.

Confidence in synthesised findings was graded using **GRADE-CERQual** (Lewin et al., 2018) across methodological limitations, coherence, adequacy, and relevance, reported per domain in Section 6 and summarised in Table 21.

2.11. Reliability, validity, reflexivity

Reliability was addressed through independent dual screening with κ reporting, duplicate extraction on 20% of the corpus, and full publication of search strings and coding rules. **Construct validity** rests on the a priori taxonomy and the three-layer outcome specification. **Internal validity** of the review is bounded by that of the primary studies, appraised in Appendix D and discussed in Section 7.5. **External validity** is bounded by corpus geography (Section 4.5).

Reflexivity. The team’s disciplinary composition (**[state composition]**) shapes what is read as salient: a finance-weighted team privileges identification strategy over interpretive depth, a management-weighted team the reverse. The synthesis was cross-read by a member of each disciplinary group.



3. RESULTS I: PERFORMANCE ANALYSIS

3.1. Study selection

Figure 1 presents the PRISMA 2020 flow. Of [n1] records identified, [n2] remained after duplicate removal, [n3] proceeded to full-text assessment, and [N] met all eligibility criteria. Reasons for full-text exclusion, with counts, are recorded in Figure 1 and enumerated study-by-study in Supplementary Table S3.

3.2. Corpus characteristics and growth trajectory

Table 6 reports the principal bibliometric indicators. Annual production follows [describe the observed pattern]. Rather than asserting inflection points, we fit a segmented regression to identify structural breaks empirically; breakpoints were detected at [years].

The correspondence—or its absence—between detected breakpoints and the regulatory events catalogued in Section 2.4 is substantively informative. Where research volume inflects *following* regulatory milestones rather than preceding them, the field is revealed as *reactive* to regulation rather than anticipatory of it, a characterisation developed in Section 7.5.

3.3. Sources and Bradford's Law

Table 10 reports the fifteen most productive sources. Bradford's Law partitions sources into three zones of approximately equal article yield; the core zone comprises [n] journals with a Bradford multiplier of [value].

The *size* of the core is interpretively decisive. A core of eight to twelve journals indicates a field with an established disciplinary home; a core exceeding twenty-five indicates genuine dispersion across finance, accounting, management, and environmental studies—in which case no single readership sees the whole literature, and the case for an integrative review is correspondingly strengthened. The *disciplinary composition* of the core zone is reported alongside its size: Aracil et al. (2021) observed that mainstream finance journals were markedly less active in

sustainable banking than ethics and management outlets, and whether that asymmetry has persisted, narrowed, or reversed by 2026 is directly testable here.

3.4. Authors and Lotka's Law

Table 11 reports author productivity and impact. The fitted Lotka exponent is [value] against a theoretical value of 2. Exponents materially above 2 indicate a field dominated by occasional contributors—consistent with a topic many researchers visit once, opportunistically, rather than one with a stable specialist community. Given ESG's topicality we anticipate this pattern; documenting it supports the argument in Section 7.5 that the field grows in volume without growing in depth.

3.5. Geographic and institutional distribution

Table 12 reports the fifteen most productive countries with the single-country (SCP) and multiple-country (MCP) publication split. A low MCP ratio signals a field of parallel national literatures rather than a genuinely comparative one—a substantive limitation, not merely a descriptive fact, for a topic where institutional context is a central moderator.

Global South representation is reported explicitly. Emerging-market banking systems, where mandatory green banking policy has often been *more* interventionist than in advanced economies, supply institutional variation that advanced-economy samples cannot. Under-representation here is a specific and correctable gap, addressed as C1 in Section 8.

3.6. Most influential documents

Table 13 reports the twenty most-cited documents with *normalised* as well as raw citation counts. Normalisation by year-cohort mean corrects the age confound under which a 2009 paper necessarily outranks a 2023 paper on raw counts irrespective of relative influence.

The final two columns implement this review's distinctive move: coding the most-cited works to the policy taxonomy and outcome layer. If the highest-cited works cluster in P4 (disclosure) and P2 (credit) while P5 (payout) and P3 (provisioning) are absent from the top twenty, that

distribution is the structural finding from which Proposition 1 follows.

3.7. Keyword frequency

Table 14 reports the twenty-five most frequent author keywords alongside Keywords Plus. The comparison is diagnostic rather than decorative: author keywords reveal how researchers *position* their work, while Keywords Plus, derived algorithmically from cited-reference titles, reveals what the work is *connected to*. Systematic divergence—author keywords emphasising sustainability while Keywords Plus emphasise firm performance and corporate governance—would indicate a field whose self-description differs from its intellectual location.

4. RESULTS II: SCIENCE MAPPING

4.1. Conceptual structure: keyword co-occurrence

Figure 3 presents the co-occurrence network (minimum five occurrences; [k] terms meeting threshold; association-strength normalisation), and Table 15 the resulting clusters with interpretive labels mapped to policy domains.

The overlay visualisation, with average publication year as the colour dimension, carries more information than the cluster table alone: it shows not only what the field discusses but its direction of travel. Terms located in the recent region of the overlay—[**report observed**]—versus the earlier region—[**report observed**]—evidence the field's migration from a communications-and-legitimacy framing toward a risk-and-allocation framing.

4.2. Research frontier: bibliographic coupling

Bibliographic coupling links documents by shared references and is the appropriate technique for identifying the *current* frontier, since coupling strength is fixed at publication and does not require citation accumulation. It is

5. RESULTS III: CONTENT SYNTHESIS BY POLICY DOMAIN

This section synthesises substantive findings by policy domain. For each we state what the evidence establishes, where it is contested, and what remains unexamined. Certainty is graded on CERQual criteria and summarised in Table 21. Table 7 reports the corpus distribution on which the synthesis rests.

5.1. P1 — Capital structure and adequacy

Established. The capital and sustainability literatures meet at one well-supported proposition: capital is the mechanism by which a bank absorbs loss without contracting credit supply, and better-capitalised banks sustain lending through stress rather than deleveraging into it (Berger & Bouwman, 2013). Applied to sustainability this yields the *transition-capacity argument*: a bank's capacity to finance a multi-decade transition, and to absorb losses on stranding legacy exposures, is a function of its capital position. Studies typically report positive associations between capitalisation and both disclosure quality and sustainability-related lending, though causal direction is rarely resolved.

A second strand treats sustainability as an input to rather than an output of the capital calculation, examining whether

therefore complementary to co-citation, which is retrospective and favours older work.

Table 16 reports coupling clusters with modularity [value] and [%] of documents in the largest connected component. Low modularity with a dominant component indicates an integrated field; high modularity with several disconnected components indicates parallel literatures. For a field spanning finance, accounting, and management we anticipate the latter, and the specific points at which clusters fail to connect identify where the integrative contribution lies.

4.3. Intellectual structure: co-citation

Table 17 reports co-citation clusters with foundational works and disciplinary origins. The analytically decisive observation concerns the **strength of links between the ethics-and-legitimacy tradition and the prudential-and-intermediation tradition**. We report inter-cluster link strength as a proportion of intra-cluster link strength: [ratio]. A ratio below approximately 0.15 constitutes strong evidence of intellectual siloing—that the field's two most substantial intellectual resources are not in conversation—and directly motivates the integrative framework in Section 7.1.

4.4. Thematic map and evolution

Figure 4 plots themes on Callon centrality (external relevance) against density (internal development), yielding motor, basic, niche, and emerging/declining quadrants.

Emerging and declining themes are indistinguishable in a static map. We disambiguate through thematic evolution across the three regime slices defined in Section 3.9, reported as a Sankey diagram (Figure 5) showing topic migration, splitting, and merging, and tabulated in Table 18. Themes appearing only in the final slice are emerging; themes present early and absent late are declining.

ESG performance affects the cost or availability of bank capital—through equity cost (Dhaliwal et al., 2011), debt cost (Goss & Roberts, 2011), or supervisory Pillar 2 add-ons.

Contested. Whether sustainability characteristics should attract differential regulatory capital treatment—the “green supporting factor” versus “brown penalising factor” debate—remains unresolved on prudential and empirical grounds alike. The prudential objection is that capital requirements exist to reflect risk, not to allocate credit, and that subsidising green exposures compromises framework integrity if those exposures are not demonstrably less risky. The empirical question of whether they *are* less risky is under-tested and jurisdictionally narrow.

Absent. We find no substantial body of work on internal capital allocation and funds transfer pricing as sustainability instruments—arguably the most powerful internal lever, since FTP determines which businesses find which lending profitable; on the interaction between capital buffers and transition-finance capacity under stress; or on AT1 instrument design and the credibility of sustainability commitments.

5.2. P2 — Credit allocation and pricing

The domain of greatest material consequence and, after disclosure, greatest research volume.

Established. *Borrower ESG characteristics affect loan terms.* Firms with weaker environmental profiles face higher borrowing costs and less favourable non-price terms, operating through lender risk assessment rather than lender preference alone (Chava, 2014; Goss & Roberts, 2011). Bank–firm relationships exhibit ESG-profile matching, with high-ESG banks and high-ESG borrowers sorting together (Houston & Shan, 2022; Degryse et al., 2023)—consequential because it implies that improvement in aggregate portfolio greenness may reflect *reallocation between banks* rather than real-economy change.

Carbon and climate policy risk are priced, but incompletely and recently. Loan pricing incorporates a premium for fossil-fuel and carbon-intensive exposure, emerging materially after the Paris Agreement—indicating that pricing responds to policy expectation rather than physical risk realisation (Ehlers et al., 2022; Delis et al., 2024). Magnitudes are modest relative to plausible transition-loss scenarios.

Voluntary commitments have measurable but limited effects. Studies of climate pledges find reductions in lending to carbon-intensive borrowers among committed banks, with substantial heterogeneity and evidence of reallocation rather than reduction: exposures migrate to less-scrutinised lenders or to capital markets (Mésonnier, 2022; Reghezza et al., 2022).

Green product markets show a small pricing differential. The green bond premium is consistently estimated as small and variable (Zerbib, 2019), while corporate green bond issuance is associated with positive announcement returns and subsequent environmental improvement (Flammer, 2021).

Contested. Whether green lending reduces or increases credit risk is unsettled. Evidence from China’s green credit environment suggests lower non-performing ratios on green loans (Cui et al., 2018), but this may reflect selection (green borrowers are larger and better capitalised), policy support (green projects receive state backing), or measurement (green classification is endogenous to loan performance). The construct “green loan” is not standardised across studies, materially limiting comparability.

Absent. Loan-level analysis of sustainability-linked loan covenant design and enforcement; the effect of collateral policy on transition finance for asset-heavy projects; and—most consequentially—the distributional effects of ESG-adjusted credit standards. Whether tightened environmental screening restricts credit to smaller firms lacking disclosure capacity, generating a social cost from environmental policy, sits at the intersection of the E and S pillars and is almost entirely unexamined.

5.3. P3 — Risk management, provisioning, and stress testing

Established. Integrating environmental and social criteria into credit risk assessment improves risk prediction relative to conventional financial criteria alone—an early finding (Weber et al., 2010) that later work has broadly sustained. Systemic-level climate stress testing establishes that transition shocks propagate with amplification via interbank exposures, and that exposure concentration matters more than aggregate exposure (Battiston et al., 2017). Banks with stronger environmental orientation exhibit lower risk on

several measures (Gangi et al., 2019), while ESG controversies associate with elevated risk-taking (Galletta & Mazzù, 2023).

Contested. The practical value of regulatory climate stress tests is disputed. Supervisory exercises consistently report data limitations severe enough to qualify their own conclusions, and scenario horizons of thirty years exceed both the maturity of most bank assets and the credibility of underlying macroeconomic projections. Whether these exercises change bank behaviour or function primarily as capability-building and disclosure-forcing devices is open and researchable.

Absent. This is where the observability gradient bites hardest. We find almost no work on IFRS 9 expected-credit-loss staging and management overlays as sustainability instruments—despite provisioning being the most direct accounting channel through which climate risk enters the income statement; on risk appetite framework design; on the organisational integration of climate risk into the three lines of defence; or on how scenario outputs are translated into limits, pricing, and capital.

5.4. P4 — Disclosure, reporting, and assurance

The largest and most methodologically mature domain, and consequently the one where the field’s limitations are clearest.

Established. *Determinants of disclosure are well characterised.* Bank size, listing status, board independence, sustainability committee presence, institutional ownership, and regulatory regime consistently predict disclosure extent and quality (Birindelli et al., 2018; Bose et al., 2018). Institutional and legitimacy explanations both find support and are largely observationally equivalent—an instance of the parallel-theory problem noted in Table 3.

Mandatory regimes increase disclosure quantity; effects on decision-usefulness are mixed. The broader mandatory-reporting literature finds that mandates increase volume and comparability with more equivocal capital-market and real effects (Christensen et al., 2021). Banking-specific replication is thin.

Materiality selection matters more than disclosure volume. Performance on issues material to the industry predicts financial performance; performance on immaterial issues does not (Khan et al., 2016). This finding is foundational and under-applied in banking, where the material issue set is dominated by portfolio-level rather than operational factors.

The measurement layer is unreliable. ESG ratings diverge substantially across providers, driven by differences in scope, measurement, and weighting (Berg et al., 2022). This is not a nuisance: a large share of the corpus’s dependent or independent variable is provider-specific.

Contested. The disclosure–performance relationship in banking is positive on average but heterogeneous in magnitude and occasionally in sign, varying by region, pillar, performance measure, and estimation method (Buallay, 2019; Nizam et al., 2019; Friede et al., 2015). Much of this heterogeneity is plausibly methodological rather than substantive—a claim tested directly in Section 6.9.

Greenwashing. Whether disclosure reflects substance or substitutes for it is the field’s most consequential open

question. The concern is theoretically well founded—legitimacy theory predicts precisely this—and empirically supported in specific settings, but detection remains underdeveloped, relying largely on discrepancies between disclosure and outcome metrics that are themselves imperfectly measured.

Absent. Assurance policy is remarkably under-studied given its regulatory salience: provider choice (Big Four versus specialist), assurance level (limited versus reasonable), and scope have no substantial banking literature, despite limited assurance being retained as mandatory under the amended CSRD regime. Equally absent: the effect of reporting-standard *selection* (GRI versus ISSB versus ESRS) on disclosure content and user response—now a live strategic choice for banks operating across divergent jurisdictions.

5.5. P5 — Payout and retention

This domain is effectively empty, and its emptiness is this review's most striking structural finding.

The general banking literature establishes that dividend policy has systemic consequences: banks maintained distributions into the 2007–2009 crisis, depleting capital precisely when it was most needed, and payouts generate negative externalities across the banking system by reducing aggregate loss-absorbing capacity (Acharya et al., 2017). The 2020 supervisory dividend restrictions across the EU, UK, and US demonstrated that payout is a live prudential variable.

The logical connection to sustainability is direct and, as far as we can determine, almost entirely unmade: **retained earnings are the primary source of internally generated capital, and internally generated capital is what funds long-horizon transition lending without equity issuance.** A bank distributing a high proportion of earnings is, mechanically, a bank with less capacity to finance transition. Conversely, if transition finance requires balance-sheet growth, payout policy is a binding constraint on it.

We identify no substantial work examining the relationship between payout policy and sustainability-related lending capacity; whether banks with stronger sustainability commitments adjust payout accordingly (a credibility test of commitment); the interaction of payout restrictions with transition-finance targets; or the signalling content of payout for sustainability commitment. Studies invoking free-cash-flow theory (Jensen, 1986) in a sustainability context address non-financial firms.

5.6. P6 — Governance, incentives, and internal control

Established. Board characteristics predict sustainability outcomes: independence, gender diversity, sustainability committee presence, and board size associate with higher ESG performance and disclosure in banks (Birindelli et al., 2018). Governance quality moderates the sustainability–performance relationship, generally strengthening it.

Contested. ESG-linked executive compensation is growing rapidly in practice and under-evaluated in research. The theoretical position is ambiguous: agency theory supports incentive alignment but also predicts that soft, discretionary, self-reported metrics will be gamed. Banking-specific evidence is thin and mixed.

Endogeneity is acute here. Board composition is a choice, and the same unobserved factors—a sustainability-oriented dominant shareholder, a reputational shock, a supervisory intervention—plausibly drive both board configuration and sustainability outcomes. Studies addressing this with instrumental variables typically rely on instruments whose exclusion restrictions are questionable in a setting where supervisors approve appointments.

Absent. The internal control architecture—the mechanism by which board-level policy becomes front-line practice—is essentially unstudied: internal audit's role in sustainability assurance; second-line ownership of climate risk; and the organisational conditions under which stated policy is implemented rather than decoupled. Institutional theory predicts decoupling under weak enforcement; testing that prediction requires internal data researchers have not obtained.

5.7. P7 — Digital delivery, distribution, and inclusion

Established. Digitalisation and sustainability are increasingly treated as interacting rather than independent strategic megatrends, with joint effects on bank performance (Forcadell et al., 2020). Sustainable banking practices contribute to inequality reduction, particularly in weak institutional environments where formal financial access is constrained (Úbeda et al., 2022). Microfinance and inclusion research constitutes a substantial but institutionally distinct literature connecting only loosely to mainstream sustainable-banking work—a disconnection visible in the co-citation structure (Section 5.3).

Contested. Whether digital delivery advances or undermines the social dimension of sustainability is genuinely open. Branch closure reduces operational emissions and cost while withdrawing service from populations with low digital access—a direct E-versus-S trade-off with clear distributional consequences that composite ESG frameworks, which net pillar scores, actively obscure. Algorithmic credit scoring may expand access through alternative data or encode and scale historical discrimination; both are documented in the general credit literature and neither is well studied in banking sustainability terms.

Absent. The environmental footprint of banking IT infrastructure; the relationship between fee structure and inclusion outcomes; and rigorous evaluation of whether digital-first strategies improve or worsen measured social performance.

5.8. How sustainability is operationalised

Table 19 records operationalisation across the corpus with the principal validity concern attaching to each. The anticipated pattern—heavy concentration in Layer 1 composite scores, sparse Layer 2 measurement—carries a direct implication: **the field predominantly measures what banks say and how they are rated, not what they finance.** Since Layer 2 is where material impact resides, this is a validity problem at the level of the field rather than of individual studies.

5.9. Contingencies and heterogeneity

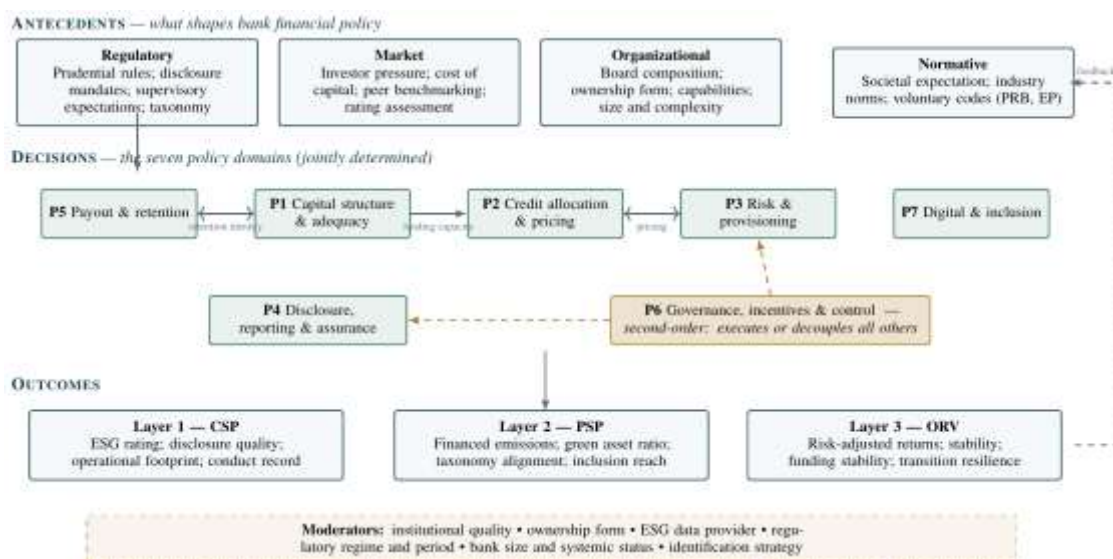
Six moderators recur across domains and jointly explain much of the observed inconsistency.

- Institutional quality and legal origin.** Effects are systematically stronger where enforcement is credible. In weak institutional environments, disclosure-performance links attenuate while sustainability's *substitutive* institutional role may strengthen (Úbeda et al., 2022).
- Ownership and organisational form.** Cooperative, mutual, savings, and state-owned banks operate under different objective functions than listed commercial banks; Islamic banks operate under a distinct normative framework with partial overlap. Pooling these forms without interaction terms is a specification error.
- ESG data provider.** Given documented divergence, provider identity is a moderator, not a nuisance parameter. We report cross-provider consistency in Table 20.
- Regulatory regime and period.** Pre- and post-TCFD, pre- and post-CSR, and voluntary versus mandatory jurisdictions are materially different environments. Omnibus I adds a further break from 2026.
- Bank size and systemic status.** G-SIBs face supervisory scrutiny, disclosure obligation, and reputational exposure of a different order than regional banks; effect sizes are not comparable across the size distribution.
- Identification strategy.** This is the moderator with the strongest claim on the field's attention. Table 20 stratifies reported effect direction by identification strategy. If, as anticipated, studies employing credible identification report systematically smaller effects than OLS and fixed-effects panel studies, the field's headline findings are inflated by endogeneity—a conclusion with immediate implications for every practitioner and policy claim built on them.

5.10. Sensitivity analysis

Re-running the synthesis restricted to the top appraisal tertile ([n] studies) produced [consistent/attenuated/divergent] conclusions for domains [list], and altered the certainty grading for [domain] from [X] to [Y]. We report this irrespective of outcome: the absence of sensitivity analysis is a standard reviewer objection, and a null result here is reassuring rather than uninteresting.

6. DISCUSSION



6.1. An integrative Antecedents–Decisions–Outcomes framework

Figure 2 presents the ADO framework integrating the synthesis. Three features carry analytical weight.

The decision layer is internally connected. P1 and P5 are mechanically linked through the retention identity; P2 and P3 through pricing and provisioning; P6 conditions the execution of all others. This is why single-domain studies are vulnerable to structural—not incidental—omitted-variable bias, and why the field’s domain-by-domain organisation is a liability rather than a convenience. It is the substantive content of the *policy portfolio view*.

Governance occupies a distinct logical position. P6 is not a parallel policy domain but a *second-order* one: it determines whether first-order policies are implemented or decoupled. This resolves an ambiguity in the literature, where governance is variously treated as antecedent, decision, and outcome—often within the same paper.

The feedback loop is real and under-modelled. Sustainability outcomes reshape antecedents: a poor ESG rating raises cost of capital, altering capital policy; a supervisory finding alters governance. Almost the entire empirical corpus treats as recursive a system that is simultaneous. This is the deep source of the endogeneity problem, and it will not be solved by lagging regressors.

6.2. Propositions

The synthesis supports eight propositions that exceed the primary studies. Each is stated to be falsifiable and paired with a boundary condition, so that failure to observe the predicted pattern is informative rather than merely inconvenient.

Proposition 1 (Observability–Materiality Inversion) Across banking financial policy domains, research volume is positively associated with domain observability and negatively associated with material sustainability impact. Formally, letting V_i denote publication volume in domain i , O_i its observability, and M_i its materiality: $\partial V_i / \partial O_i > 0$ and $\partial V_i / \partial M_i < 0$ conditional on O_i .

Boundary condition: Holds where data acquisition costs dominate research design choice. Expected to weaken where supervisory data-sharing arrangements exist, and to reverse in domains where regulatory disclosure mandates convert a low-observability domain into a high-observability one—a prediction directly testable against Pillar 3 ESG template introduction.

Proposition 2 (Regime-Contingent Logic Dominance) Legitimacy and risk-management logics of bank sustainability policy are complements whose relative weight is regime-contingent. Under voluntary regimes the legitimacy logic dominates and disclosure decouples from substance; under mandatory regimes with credible enforcement the risk logic dominates and the decoupling gap narrows.

Boundary condition: Requires enforcement credibility, not merely mandate existence. Predicts specifically that the CSRD scope narrowing under Omnibus I will *widen* the disclosure–substance gap among firms falling out of scope

relative to those remaining in scope—a difference-in-differences design available now.

Proposition 3 (Reallocation Indistinguishability) Bank-level improvement in portfolio sustainability metrics is not identified with respect to real-economy decarbonisation, because assortative bank–borrower matching and post-commitment exposure migration generate observationally equivalent metric improvements. Bank-level units of analysis cannot, in principle, distinguish the two.

Boundary condition: Applies wherever exposure can migrate—across lenders, to capital markets, or to non-bank intermediaries. Does not apply in closed systems with binding system-wide allocation constraints, which is why directed-credit regimes offer a useful test case.

Proposition 4 (Composite Score Non-Identification) Composite ESG scores are not fit for purpose as banking sustainability measures. Beyond documented cross-provider divergence, compositing nets across pillars and thereby conceals precisely the trade-offs constituting the substance of bank sustainability policy.

Boundary condition: Strongest where E and S move in opposite directions—branch network reduction, ESG-tightened credit standards, digital-first strategies. Weakens where pillar movements are correlated, which is the empirical condition under which composites are defensible and which should therefore be tested rather than assumed.

Proposition 5 (Identification–Magnitude Attenuation) Reported effect magnitudes in the banking sustainability–performance literature decline monotonically with the credibility of the identification strategy. Ordering designs by identification strength—OLS, fixed effects, system GMM, instrumental variables, natural experiments—predicted effect size is weakly decreasing.

Boundary condition: Assumes the underlying effect is non-negative. If attenuation is observed, the field’s headline magnitudes are inflated by endogeneity; if attenuation is not observed, the endogeneity critique is overstated. Either outcome is informative, which is what makes the test worth running.

Proposition 6 (Second-Order Governance) Governance policy (P6) moderates rather than determines the relationship between first-order financial policies (P1–P5, P7) and sustainability outcomes. First-order policy adoption without corresponding governance capacity produces decoupling rather than partial implementation.

Boundary condition: Predicts a threshold rather than a linear moderation: below a governance-capacity threshold, first-order policy has approximately zero marginal effect on outcomes. Testable through interaction terms with non-linear specification.

Proposition 7 (Retention Constraint) Payout policy binds transition-finance capacity. *Ceteris paribus*, banks with higher payout ratios exhibit lower growth in sustainability-related lending, and the credibility of a bank’s transition commitment is signalled by the extent to which payout policy adjusts to accommodate it.

Boundary condition: Holds where equity issuance is costly or capital constraints bind. Weakens for banks with ready

access to equity markets, and is expected to be strongest among banks operating close to regulatory capital minima.

Proposition 8 (Standard-Selection Strategy) Under a divergent global disclosure regime, reporting-standard selection functions as a strategic financial policy rather than a compliance formality, with measurable consequences for disclosure content, assurance cost, and cost of capital.

Boundary condition: Applies only under regime divergence. Would be falsified by convergence toward a single global baseline, which the current ISSB trajectory makes a live possibility and which would render the proposition historically bounded.

6.3. Competing explanations

Two alternative readings of the coverage pattern deserve consideration.

The first is that the imbalance reflects *materiality judgement* rather than observability: researchers may study disclosure because it is genuinely the domain where bank sustainability is determined. We regard this as implausible on the arithmetic of financed versus operational emissions, but it is testable. If the imbalance reflects materiality judgement, then coverage should track expert assessments of domain materiality; if it reflects observability, coverage should track data availability while diverging from expert materiality rankings. A Delphi study eliciting expert materiality rankings and comparing them against observed coverage would discriminate cleanly.

The second is that the imbalance is *transitional*: low-observability domains may simply be newer, with coverage expected to converge as data improve. This predicts that domains experiencing observability shocks—P3 following supervisory overlay disclosure, P2 following Pillar 3 ESG templates—should exhibit sharp subsequent growth. The corpus permits this test directly through publication-date

7. LIMITATIONS

We identify eight limitations, stated without mitigation language.

Database restriction. Scopus and WoS exclude some regionally important and non-English outlets. Emerging-market banking research, disproportionately published outside these indexes, is under-represented—which matters precisely because emerging markets supply the institutional variation the field most needs.

Language restriction. English-only inclusion excludes substantial Chinese-, Spanish-, Portuguese-, and Indonesian-language literatures on green credit and sustainable banking.

Grey literature exclusion. Working papers, central bank research, and supervisory publications were excluded from the corpus despite often containing the most methodologically advanced work in this area, given publication lags in finance. They informed Section 2.4 only.

Search-string sensitivity. Any Boolean strategy trades sensitivity against precision. Studies using idiosyncratic terminology, particularly work predating “ESG” as standard

analysis around observability shocks, and we regard it as the most important single robustness check on Proposition 1.

6.4. TCCM analysis

Table 22 reports the TCCM analysis. The pattern is consistent across dimensions: the field is theoretically plural but non-competitive, contextually concentrated, characterised by proxy measurement rather than direct measurement, and methodologically conventional relative to the identification opportunities available to it.

6.5. Methodological state of the field

Three assessments follow.

Identification. The field’s modal design cannot support its modal claim. System GMM, the standard endogeneity response, is a weak remedy for simultaneity of the kind described in Section 7.1, is sensitive to instrument proliferation, and is frequently implemented without the diagnostic reporting that would reveal failure. The regulatory environment of 2024–2028 offers threshold-based, staggered, and reversing treatment assignments (Section 2.4) supporting far stronger designs.

Measurement. The corpus depends on commercial ESG ratings whose divergence is documented and material. Minimum acceptable practice going forward: use at least two providers, report results under both, and treat divergence as substantive rather than as noise.

Cumulativeness. A Lotka exponent materially exceeding the theoretical value, combined with the co-citation siloing examined in Section 5.3, would characterise a field growing in volume without growing in depth. Replication, registered reports, and shared measurement infrastructure are the standard correctives; none is currently present in this literature.

vocabulary, may have been missed. The gold-standard validation set bounds but does not eliminate this.

Coding subjectivity. Assignment to policy domains and outcome layers involves judgement. Dual coding and κ reporting quantify but do not eliminate it. **Bibliometric artefacts.** Co-citation and coupling structures are sensitive to threshold and resolution parameters. Parameters are reported in Table 5 and sensitivity checks in Appendix E; cluster boundaries should be read as indicative rather than definitive.

Currency. The search was executed on [date]. Given the field’s growth rate and the pace of regulatory change, the corpus will require updating within 18–24 months.

Synthesis without pooling. We did not meta-analyse, for the reasons given in Section 3.10. The review therefore characterises the direction and contingency of findings but not their pooled magnitude.

8. CONCLUSION

Banks’ most consequential sustainability decisions are financial policy decisions. What a bank funds, at what price, against what capital, with what earnings retained, under what disclosure regime, and with what governance to enforce the result—these constitute the operative content of

banking sustainability, and they are jointly determined on a single balance sheet.

This review applied PRISMA 2020 screening and three-technique bibliometric mapping to [N] studies to establish what is known about that nexus. The field is large, growing, and geographically and intellectually uneven. Its coverage is inversely related to materiality: disclosure, the most observable and least materially consequential domain, dominates, while payout policy and provisioning—materially decisive and weakly observable—are nearly absent. Its measurement infrastructure rests on composite ratings that diverge across providers and net across pillars in ways concealing the environmental–social trade-offs that constitute the substance of bank sustainability policy. Its causal claims frequently exceed the warrant of its designs. And its unit of analysis, the bank, is structurally incapable of

distinguishing real-economy improvement from reallocation between lenders.

These are not counsels of despair but a specification of what the next decade should do differently—and the conditions for doing so are unusually favourable. The regulatory environment of 2024–2028, with divergent ISSB adoption across roughly 28 jurisdictions, a voluntary Basel climate disclosure framework implemented unevenly, and an EU regime that expanded and then contracted with threshold-based precision under Omnibus I, supplies the staggered, threshold-based, and reversing treatment assignments this literature has always lacked. The seven-domain taxonomy, the policy portfolio view, the observability–materiality inversion, the ADO framework, and the eight propositions set out here are offered as means of exploiting that window before it closes.

banking sector. *Management of Environmental Quality*, 30(1), 98–115.

Campiglio, E. (2016). Beyond carbon pricing: The role of banking and monetary policy in financing the transition to a low-carbon economy. *Ecological Economics*, 121, 220–230.

Chava, S. (2014). Environmental externalities and cost of capital. *Management Science*, 60(9), 2223–2247.

Christensen, H. B., Hail, L., & Leuz, C. (2021). Mandatory CSR and sustainability reporting: Economic analysis and literature review. *Review of Accounting Studies*, 26(3), 1176–1248.

Cornett, M. M., Erhemjamts, O., & Tehranian, H. (2016). Greenwashing? Corporate social responsibility performance in U.S. banks. *Journal of Banking & Finance*, 70, 137–159.

Cui, Y., Geobey, S., Weber, O., & Lin, H. (2018). The impact of green lending on credit risk in China. *Sustainability*, 10(6), 2008.

Degryse, H., Goncharenko, R., Theunisz, C., & Vadasz, T. (2023). When green meets green. *Journal of Corporate Finance*, 78, 102355.

Delis, M. D., de Greiff, K., & Ongena, S. (2024). Being stranded with fossil fuel reserves? Climate policy risk and the pricing of bank loans. *Financial Markets, Institutions & Instruments*, 33(3), 239–265.

Dhaliwal, D. S., Li, O. Z., Tsang, A., & Yang, Y. G. (2011). Voluntary nonfinancial disclosure and the cost of equity capital. *The Accounting Review*, 86(1), 59–100.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, 51(3), 393–414.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160.

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.

Ehlers, T., Packer, F., & de Greiff, K. (2022). The pricing of carbon risk in syndicated loans. *Journal of Banking & Finance*, 136, 106180.

Elkington, J. (1997). *Cannibals with forks: The triple bottom line of 21st century business*. Capstone.

9. REFERANC

Aracil, E., Nájera-Sánchez, J.-J., & Forcadell, F. J. (2021). Sustainable banking: A literature review and integrative framework. *Finance Research Letters*, 42, 101932.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, 17(1), 99–120.

Barth, J. R., Caprio, G., & Levine, R. (2004). Bank regulation and supervision: What works best? *Journal of Financial Intermediation*, 13(2), 205–248.

Basel Committee on Banking Supervision. (2022). *Principles for the effective management and supervision of climate-related financial risks*. Bank for International Settlements.

Basel Committee on Banking Supervision. (2025). *Framework for the voluntary disclosure of climate-related financial risks*. Bank for International Settlements.

Battiston, S., Mandel, A., Monasterolo, I., Schütze, F., & Visentin, G. (2017). A climate stress-test of the financial system. *Nature Climate Change*, 7(4), 283–288.

Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6), 1315–1344.

Berger, A. N., & Bouwman, C. H. S. (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics*, 109(1), 146–176.

Birindelli, G., Dell’Atti, S., Iannuzzi, A. P., & Savioli, M. (2018). Composition and activity of the board of directors: Impact on ESG performance in the banking system. *Sustainability*, 10(12), 4699.

Bolton, P., & Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2), 517–549.

Bose, S., Khan, H. Z., Rashid, A., & Islam, S. (2018). What drives green banking disclosure? An institutional and corporate governance perspective. *Asia Pacific Journal of Management*, 35(2), 501–527.

Buallay, A. (2019). Is sustainability reporting (ESG) associated with performance? Evidence from the European

- Flammer, C. (2021). Corporate green bonds. *Journal of Financial Economics*, 142(2), 499–516.
- Forcadell, F. J., Aracil, E., & Úbeda, F. (2020). The impact of corporate sustainability and digitalization on international banks' performance. *Global Policy*, 11(S1), 18–27.
- Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Pitman.
- 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.
- Galletta, S., & Mazzù, S. (2023). ESG controversies and bank risk taking. *Business Strategy and the Environment*, 32(1), 274–288.
- Galletta, S., Mazzù, S., & Naciti, V. (2022). A bibliometric analysis of ESG performance in the banking industry: From the current status to future directions. *Research in International Business and Finance*, 62, 101684.
- Gangi, F., Meles, A., D'Angelo, E., & Daniele, L. M. (2019). Sustainable development and corporate governance in the financial system: Are environmentally friendly banks less risky? *Corporate Social Responsibility and Environmental Management*, 26(3), 529–547.
- 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.
- Hart, S. L. (1995). A natural-resource-based view of the firm. *Academy of Management Review*, 20(4), 986–1014.
- Hong, Q. N., Pluye, P., Fàbregues, S., et al. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. Canadian Intellectual Property Office.
- Houston, J. F., & Shan, H. (2022). Corporate ESG profiles and banking relationships. *The Review of Financial Studies*, 35(7), 3373–3417.
- Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review*, 76(2), 323–329.
- 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.
- Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. *The Accounting Review*, 91(6), 1697–1724.
- Kraus, S., Breier, M., & Dasí-Rodríguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*, 16(3), 1023–1042.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lewin, S., Booth, A., Glenton, C., et al. (2018). Applying GRADE-CERQual to qualitative evidence synthesis findings. *Implementation Science*, 13(S1), 2.
- Lim, W. M., Kumar, S., & Ali, F. (2022). Advancing knowledge through literature reviews: 'What', 'why', and 'how to contribute'. *The Service Industries Journal*, 42(7–8), 481–513.
- Mésonnier, J.-S. (2022). Banks' climate commitments and credit to carbon-intensive industries: New evidence for France. *Climate Policy*, 22(3), 389–400.
- Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review*, 48(3), 261–297.
- Myers, S. C., & Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2), 187–221.
- Nizam, E., Ng, A., Dewandaru, G., Nagayev, R., & Nkoba, M. A. (2019). The impact of social and environmental sustainability on financial performance: A global analysis of the banking sector. *Journal of Multinational Financial Management*, 49, 35–53.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
- Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*, 45(4), O1–O16.
- Reghezza, A., Altunbas, Y., Marques-Ibanez, D., Rodriguez d'Acri, C., & Spaggiari, M. (2022). Do banks fuel climate change? *Journal of Financial Stability*, 62, 101049.
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., et al. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10, 39.
- Scholtens, B. (2009). Corporate social responsibility in the international banking industry. *Journal of Business Ethics*, 86(2), 159–175.
- Shea, B. J., Reeves, B. C., Wells, G., et al. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews. *BMJ*, 358, j4008.
- Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3), 355–374.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review*, 20(3), 571–610.
- Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45.
- Úbeda, F., Forcadell, F. J., Aracil, E., & Mendez, A. (2022). How sustainable banking fosters the SDG 10 in weak institutional environments. *Journal of Business Research*, 146, 277–287.
- United Nations Environment Programme Finance Initiative. (2019). *Principles for Responsible Banking*. UNEP FI.
- Weber, O. (2005). Sustainability benchmarking of European banks and financial service organizations. *Corporate Social Responsibility and Environmental Management*, 12(2), 73–87.
- Weber, O., Scholz, R. W., & Michalik, G. (2010). Incorporating sustainability criteria into credit risk management. *Business Strategy and the Environment*, 19(1), 39–50.
- Wu, M.-W., & Shen, C.-H. (2013). Corporate social responsibility in the banking industry: Motives and financial



performance. *Journal of Banking & Finance*, 37(9), 3529–3547.

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of Banking & Finance*, 98, 39–60.

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472.