

Research Article

A Replicable Six-Dimension Rubric for Assessing IFRS 17 Disclosure Quality and Comparability in Life Insurers

Esmeralda Yryku* , Diana Lamani 

Department of Accounting, Faculty of Economy, University of Tirana, Tirana, Albania

*esmeralda.yryku@unitir.edu.al

Abstract

IFRS 17 has reshaped life insurers' reporting by replacing IFRS 4 with a stock-flow disclosure package in which users must interpret judgement-intensive constructs (CSM, risk adjustment, insurance finance effects) through audited statements and notes. Early implementation shows wide heterogeneity, yet archival research lacks a transparent issuer-year metric to test learning, determinants, and market relevance. We develop an IFRS-17-native, replicable six-dimension rubric scoring evidence strength (0–2 each) for: (D1) reconciliations/roll-forwards, (D2) actuarial assumptions, (D3) IFRS-17-specific sensitivities, (D4) method/transition transparency, (D5) audit traceability via key audit matters, and (D6) presentation/navigability. Scores aggregate to FLOW (D1–D3; 0–6), STOCK (D4–D6; 0–6), and TOTAL (0–12). Using audited annual reports for 33 life-insurance groups over FY2023–FY2024 (balanced panel, N=66), reliability is high ($\omega=0.893$; $\alpha=0.883$) and inter-coder agreement is assessed with weighted Cohen's κ on a double-coded subsample. Disclosure quality improves in year 2 (TOTAL: 9.924→10.740; paired $t(32)=4.967$; Wilcoxon $z=3.724$; $p<.001$). In pooled OLS determinants, leverage is the only significant correlate ($B=15.385$; $p<.001$). Market tests are proxy-dependent: controlling for size, STOCK is associated with tighter bid-ask spreads ($B=-0.023$; $p=.007$) while FLOW loads positively ($B=+0.016$; $p=.025$); TOTAL is positively related to log Amihud illiquidity ($B=0.245$; $p=.019$) and remains detectable with country random intercepts. Implied cost-of-equity models are not significant. The rubric provides a reusable publicly auditable measurement layer for cumulative IFRS 17 disclosure research and for practitioner benchmarking.

Keywords: IFRS 17; Life Insurance; Disclosure Quality; Comparability; Audit Traceability; Liquidity; Financial Reporting

INTRODUCTION

IFRS 17 has changed how life insurers explain performance and financial position to external users. It introduced a reporting package that users can only understand well if the insurer makes good presentation choices. It also depends on how the notes are organised and how clearly the report links figures across statements and notes. In the first mandatory

reporting years, market participants, supervisors, and practitioners have reported wide differences across issuers in how IFRS 17 disclosures appear in annual reports. This happens even when the measurement requirements are broadly the same in principle.

Measurement is often a difficult starting point in archival research due to the complexity and variability of data sources. IFRS 17 exacerbates this issue by introducing new standards that lack consistent application across firms. No issuer-year measure of disclosure quality adequately fits life insurers due to the diverse nature of their operations and reporting practices, making consistent application across firms challenging. This makes empirical work harder. Early adoption patterns are difficult to interpret with confidence.

Another challenge with the analysis of disclosure architecture is that the rubric used to categorise the report elements does not address the inconsistencies in how firms disclose information. It is clear that disclosure structures differ among firms, but it is unclear if these differences have economic consequences.

There is now a growing literature on IFRS 17 drawn from academic studies, supervisory reviews, and practitioner assessments. Some studies have been focused on the standard and implementation guidance.

Other work comes from supervisory reviews and thematic assessments. Practitioner benchmarking exercises form another part of the discussion. While these contributions aid in implementation and oversight, they highlight the need for developing a standardized issuer-year disclosure measure that archival researchers can apply directly to life insurers.

However, they rarely provide a transparent issuer year scoring protocol that researchers can reproduce from public annual reports. They also rarely set out a coding algorithm in a way that others can follow line by line. In addition, they rarely provide a measure that can be used directly as a dependent variable or an independent variable in panel econometric designs. For these reasons, early IFRS 17 research still lacks a practical issuer year measurement instrument. What is missing is a way to translate the new IFRS 17 reporting package into a measurable construct that supports hypothesis testing and independent replication.

We fill this gap by developing and applying an IFRS 17 native rubric. The rubric scores six dimensions of disclosure quality and comparability. Each dimension is scored on a harmonised scale from 0 to 2. We then aggregate the dimension scores into composite indices that can be used across issuers and years. We use a balanced panel of 33 life insurance groups observed in FY2023 and FY2024. We test four hypotheses. These analyses cover changes in reporting scores observed in the second year and examine firm-level correlates that may influence these changes. They cover associations with liquidity and illiquidity proxies. They also cover a cost of equity proxy where it is available. Because the panel is short and the setting is early post adoption, we interpret all findings as associations.

The study makes three contributions. First, it provides a replicable issuer year measurement instrument designed for the IFRS 17 life insurance reporting package. Second, it provides early panel evidence on learning and correlates in the first two mandatory years. Third, it reports market associations that speak to potential mechanisms. These use liquidity and pricing proxies, and we report model specific sample sizes in a transparent way.

The primary setting is Europe. We define Europe as the EU and EEA plus the UK and Switzerland. We also have a small set of IFRS 17 reporters outside Europe. The idea is that it stress tests portability. In the determinants model we also use a Europe indicator, EUROPE. The paper next transitions to the theory and lays out the hypotheses. The Methods section details how the rubric was constructed and scored, as well as how the sample was created and reliability established. Outcomes prove learning tests, determinants analysis, market results, and robustness checks. The paper ends with findings and limits of the study, future work and key conclusions. These include changes in scores in the second reporting year.

THEORY, PRIOR LITERATURE, AND RESEARCH GAP

Disclosure, Information Risk, and Capital-Market Outcomes

Disclosure enables users to understand how the economic phenomenon behind the numbers reflected in the statements operates, and to see what the risks involved are. However, current reporting standards do not consistently provide information that is equally useful across firms.

In general terms, better disclosure is associated with a lower cost of equity since investors face less uncertainty when they can develop a clearer picture of the firm's underlying position [1]. However, it is important not to over emphasize the association between improved disclosure and reduced cost of equity. A range of market events, changes in regulations and individual firm characteristics will affect how investors perceive the credibility and reliability of disclosure.

The difficulty is not just in the validity of the underlying economics, and whether an outside developer can draw useful conclusions from it through a firm's public filings. Noisy, incomplete, thin or expensive disclosures result in investors wanting to be compensated for the added risk with higher expected returns or adjust their trading strategies [2] (for instance increasing bid-ask spreads or decreasing trade volumes). Trading and valuation are more complicated with weak information environments.

Although IFRS 17 aims at harmonized reporting, differences in interpretation by firms themselves and selective disclosure can limit quality progress. Firms should compare the costs and benefits of compliance, allowing for consideration of preparation costs, proprietary costs (i.e. production violations), enforcement type and settings. So disclosure quality is the product of a choice under constraint, rather than just a mechanical outcome

of written requirements [3]. And this is why studies of major reporting shifts are worth the effort.

Two companies can be subject to the same standard but report their results in ways that cause investors to come to different conclusions about how healthy they are financially. Firms reporting under IFRS 17 will need to provide a significant amount of judgment and additional information for end-users to comprehend reported numbers. Evidence observed in measurement is highly subjective making users dependent upon reporting packages. Reconciliations, supporting narrative explanations, and cross-references are critical to understanding what drives performance and risk over the period [4].

Comparability as an Economic Mapping Property

Comparability in IFRS 17 life insurance reporting will not mean firms reporting 'similar' numbers. Numbers such as the CSM or the RA might be similar even if they stem from different measurement decisions. Insurers often use their own discount curves, lapse assumptions, or transition approaches, so the same responses do not necessarily imply the same economics [5].

This makes disclosure particularly important. As many IFRS 17 amounts are judgemental, users need sufficient information to understand how the numbers were calculated and what caused movements during the period. Without that context, comparison among the insurers is much more difficult.

Oversight and enforcement can promote more consistent application of the standard, which typically improves comparability [6]. However, IFRS 17 still leaves room for diversity in approaches and making individual choices — e.g. transition methods, measurement models and assumptions.

Why IFRS 17 Elevates the Role of Disclosure Architecture

Revolutionizing the reporting of insurance entities, IFRS 17 introduces new measurement models, changes presentation formats, and substantially increases disclosure requirements. Because the standard focuses on decision-usefulness, first-time applicants have easier transition options and clear guidance and examples [7, 8].

Users want detailed roll-forwards and reconciliations that connect directly to financial tables. This ensures transparency and clarity in where the figures derive from and why they change over time. They need simple explanations of why a certain change occurred during the period, not roundabout descriptions that hang tantalizingly around numbers. Rarely will an overarching description suffice. Two companies may follow the same rules for recognition and measurement, yet their reporting looks drastically different because of differences in style, cross-referencing, explanation.

These performance reporting challenges are very evident for life insurers, especially in the context of Contractual Service Margin. IFRS 17 establishes the CSM as a key element of performance reporting, but the roll-forward of the CSM is highly judgmental [4]. The same lack of information about how users should interpret these residual profits — when they

arise and how they progress over time [4, 9] makes detailed CSM roll-forwards necessary. Other areas see the same problem too; if users cannot interpret what will change the risk adjustment, then for example, this is not of very much use. The same concerns hold for reinsurance contracts owned, where stakeholders have to be shown the connections between risk levels, variations in financial metrics, employed methods, and experienced uncertainties.

Otherwise cross-firm comparison becomes thin very quickly. For it to be credible, users need clear disclosure of key inputs, sources of uncertainty, and the link between the quantitative tables and the accompanying narrative [4, 10, 11].

Earlier technical work and early empirical discussion point in much the same direction. Problems start when disclosures become boilerplate, when they are only loosely tied to actual movements, or when they are too vague to support verification and cross-firm mapping [12–15]. Practitioner material, implementation commentary, and early market-facing discussion also suggest that first-year IFRS 17 reporting was far from uniform [16–24]. Insurers differed in how they explained judgements, drivers, and uncertainty, and those differences were visible quite early in adoption [25–33].

Against this background, we conceptualise IFRS 17 disclosure quality as an issuer year disclosure architecture property. The question is whether an insurer's audited reporting package enables users to trace, interpret, and compare IFRS 17 reporting logic through disclosures that are explicit, structured, and verifiable [4, 16]. This construct motivates the issuer year rubric developed in the paper. The rubric operationalises traceability and comparability in a transparent and replicable form that is suitable for panel based archival tests.

Prior Literature and State of the Art Evidence on IFRS 17 Disclosure Practice

The evidence base on IFRS 17 reporting and disclosure practice is informative, but it is fragmented across streams with different objectives and outputs. For an archival, hypothesis testing design, the methodological issue is not whether these streams are useful. The issue is whether they provide an explicit and auditable firm year measure that can be operationalised as a dependent or independent variable in panel econometric tests.

Table 1 synthesises the state-of-the-art evidence by evidence stream and documents scope and sample, method or model, reported findings, and the specific constraints that limit direct firm year econometric use. This synthesis makes the measurement gap addressed by this study transparent.

Table 1. State-of-the-art anchored benchmarking of IFRS 17 disclosure-quality evidence and how this study advances the literature

Evidence stream	SOTA contribution	SOTA limitation	Rubric mapping (D1–D6)	This study's advancement
Disclosure, information-risk &	Defines mechanisms: disclosure	Not IFRS 17-native measurement;	Comparability-as-mapping: D1 (traceability) +	Builds an IFRS 17-native issuer-year measurement layer

comparability theory [2, 3, 5, 6, 34]	architecture affects information asymmetry, information-processing costs, and comparability-as-mapping; motivates links to liquidity and cost of capital.	lacks an auditable issuer-year “architecture quality” variable tied to IFRS 17 constructs (CSM/RA/transition) suitable for panel regressions.	D4/D6 (method clarity/navigability). Information-risk: D2 (assumptions), D3 (sensitivities), D5 (audit/KAM traceability).	(TOTAL 0–12; FLOW 0–6; STOCK 0–6) and tests falsifiable hypotheses on learning, determinants, and market associations (Tables 8–14). Reliability and replicability are documented (Table 4; Appendix A.5).
Standard-setting [4, 7, 8, 16]	Sets normative IFRS 17 disclosure requirements/objectives (CSM roll-forwards, assumptions, risk adjustment, insurance finance, transition).	Normative baseline is not a scoring system; cannot generate issuer-year variation capturing insufficient vs partial vs full evidence beyond binary compliance.	D1–D6 operationalize “degree of evidence” aligned to IFRS 17 objectives: D1 traceability; D2 assumptions; D3 sensitivities; D4 transition/method clarity; D5 audit verification; D6 connectivity/navigability.	Translates “should disclose” into auditable evidence thresholds (0–2) and issuer-year variables usable in regressions and paired tests (Appendix A). Demonstrates beyond-compliance dispersion and year-2 improvement (Tables 8 & 9).
Academic benchmarking [20]	Describes heterogeneity in year-1 IFRS 17 presentation/disclosures; highlights comparability challenges (roll-forward detail, transition transparency, connectivity).	Mainly descriptive/cross-sectional; no panel learning test; no released issuer-year scoring protocol; limited reliability documentation for replication.	Themes map to: D1 roll-forwards; D4 transition clarity; D6 connectivity; D2 assumptions; D3 sensitivities.	Extends benchmarking into a replicable scoring algorithm (Appendix A) and a balanced FY2023–FY2024 panel enabling within-issuer learning tests and econometric use (Tables 8–14).
Supervisory reviews [17–19]	Flags recurring deficiencies and “better practice” in	Thematic diagnostics are not issuer-year metrics; not econometrics-	Supervisory priorities map to: ESMA movement explanations → D1;	Converts supervisory themes into issuer-year scores and evaluates associations with liquidity/illiquidity

	first application: movement explanations, assumption transparency, connectivity/navigability, transition clarity, audit focus.	ready; cannot statistically test determinants or market relevance.	connectivity/navigability → D6; key inputs/assumptions → D2; transition clarity → D4; audit/KAM traceability → D5 (sensitivities expectations inform D3).	proxies (Tables 8–10), enabling mechanism-oriented inference via FLOW vs STOCK decomposition.
Practitioner benchmarking [23, 25, 26]	Summarizes year-2 evolution and persistent variation; prevalence-style observations on formats and disclosure practices (roll-forwards, assumptions, sensitivities, transition narratives).	Typically narrative; scoring protocol and coder reliability not disclosed; not provided as issuer-year panel dataset suitable for hypothesis testing.	Themes map to: D1 (CSM roll-forward detail), D2 (assumptions), D3 (sensitivities), D4 (transition method/bridge clarity), D6 (presentation/navigability), and D5 (audit emphasis where discussed).	Supplies measurement infrastructure for statistical inference beyond benchmarking: repeatable issuer-year scoring (Appendix A), learning and determinants tests (Tables 8 & 10), market-association models with robustness checks (Tables 8 & 14), supported by reliability evidence (Tables 5 & 6; Appendix A.5).

Table 1 delivers a straightforward message when the main evidence streams are read side by side. Standard setting materials set out what IFRS 17 disclosures are meant to accomplish. Supervisory reviews point to recurring weaknesses and describe practices that improve usefulness. Practitioner benchmarking reports the disclosure formats that insurers adopt in annual reports and tracks how those formats develop in the second year. Early academic work, based on audited annual reports, documents meaningful cross issuer heterogeneity. The limitation that remains, however, is methodological. Across these streams, it is still uncommon to find an openly specified and reusable issuer year scoring protocol that is written as a coding algorithm, can be reproduced from public annual reports, and is already arranged as a panel ready dataset suitable for determinants and market association tests.

Researchable Gap Implied by State-of-the-Art Evidence

As Table 1 synthesises, guidance and diagnostics are extensive, yet the literature still lacks a measurement layer with four practical properties. It must be constructed from audited annual reports and notes. It must be replicable through a pre specified codebook

with clear scoring rules. It must be organised as a firm year dataset that supports panel estimation. It must also be supported by reliability evidence and by tests of empirical relevance. State of the art evidence therefore shows what varies in early IFRS 17 disclosure practice, but it does not deliver a reusable firm year metric that enables econometric tests of why the variation occurs, in determinants models, and whether it is economically relevant, in market association tests.

This study supplies the missing measurement layer by providing a replicable issuer year scoring instrument anchored in disclosure and comparability theory [1–3, 5, 6, 34] and implemented using transparent content analysis principles [35, 36]. It converts the recurring themes in state of the art discussions, including traceability, transition transparency, audit traceability, and navigability, into explicit evidence thresholds and harmonised ordinal scores. The output gives a set of reusable issuer-year variables, including the total score and the flow and stock components that can be used in the next stage of the analysis. All such variables are logged according to a predetermined scoring schema, based on the IFRS 17 logic for disclosures [4, 7, 8, 16].

We then use within issuer variation in the first two mandatory reporting years, 2023 and 2024, to test learning and stabilisation patterns, to model firm level correlates using the panel dataset, and to evaluate associations with market based proxies, including liquidity and illiquidity measures, and a cost of equity proxy where available.

Positioning of the Proposed Rubric Relative to Benchmarking and Enforcement Frameworks

In Table 2, we show how the rubric differs from two other non-academic approaches: practitioner benchmarking and guidance [22–26] and supervisory review and enforcement priorities [17–19, 21]. Clearly, they differ in purpose, theoretical basis, format of output, and replication readiness. Most practitioner and supervisory materials simply reflect observed practices, priorities, and examples.

By contrast, this study provides a research-oriented measurement layer in the form of a documented issuer year scoring algorithm aligned with IFRS 17 reporting logic [4, 7, 8, 16] and grounded in disclosure and comparability as mapping mechanisms that motivate testable links to information frictions [1–3, 5, 6, 34]. Building on this positioning, Table 3 formalises the research gap structure by linking each state-of-the-art constraint, G1 through G5, to a design response and to falsifiable empirical implementation within the existing 2023 to 2024 panel dataset.

Table 2. Positioning of the proposed issuer-year rubric relative to practitioner benchmarking and regulatory/supervisory review output

Feature / Dimension	Proposed Rubric (This Study)	Practitioner Benchmarking & Guidance	Regulatory Reviews & Enforcement Priorities
Purpose	Issuer-year scoring metric for IFRS 17 disclosure	Benchmarking of observed practices and “better practice”	Thematic diagnostics and expectations in first application

	architecture/comparability (research-oriented)	guidance (implementation-oriented)	(oversight/enforcement-oriented)
Theory link	Explicitly linked to disclosure, information-risk, and comparability-as-mapping mechanisms	Practice-led; theory is usually implicit rather than formalised	Primarily policy- and compliance-oriented; limited explicit theory-building
Output format	Issuer-year scores (TOTAL plus FLOW/STOCK components) usable in regression designs	Narrative benchmarking, examples, and guidance/checklists	Thematic findings, priority areas, and illustrative examples
Coverage focus	Multi-dimension coverage of disclosure architecture and traceability within the IFRS 17 package	Scope varies by publication; emphasis on prevalent reporting approaches	Scope defined by the review mandate; emphasis on recurrent deficiencies and expectations
Market use	Used for association tests with liquidity/illiquidity and pricing proxies where data are available	Generally not framed as capital-market outcome testing	Generally not framed as capital-market outcome testing
Replication readiness	High: pre-specified scoring rules and issuer-year dataset fields enable reuse	Typically does not publish a replicable issuer-year scoring algorithm for independent reuse	Typically does not publish a replicable issuer-year scoring algorithm for independent reuse

Table 3. Gap matrix linking SOTA limitations to gaps, design responses, and falsifiable tests

SOTA limitation	Research gap	Theoretical rationale	Design response	Empirical implementation (our case study)
Qualitative diagnostics/benchmarking rarely yield a regression-ready issuer-year metric.	G1: Measurement gap	Without a firm-year observable, determinants and economic-consequence tests are infeasible in archival designs [2,3].	Six-dimension rubric (0–2 each) → TOTAL (0–12), FLOW (0–6), STOCK (0–6); explicit coding + recoding rules (Appendix A).	Rubric variables used as DV/IV across models: learning (Tables 8 & 9), determinants (Table 10), liquidity (Table 11), robustness (Tables 12 & 13), pricing proxy (Table 14).
Checklist/binary compliance misses decision-useful disclosure architecture.	G2: Construct gap	Comparability is a mapping property [5]; disclosure economics and information-risk theory emphasize processing costs and information	Ordinal evidence-strength scoring (0–2) across traceability, assumptions, sensitivities, transition clarity, audit traceability,	Component inference feasible: STOCK vs FLOW effects diverge in bid-ask model (Table 8); dispersion beyond “compliance” shown in descriptives (Table 8).

		risk [2,3]. Binary indicators cannot capture evidence strength or traceability under IFRS 17 [4].	navigability; retain FLOW/STOCK decomposition.	
Cross-sectional early evidence cannot test within-issuer learning.	G3: Panel-learning gap	Learning after year-1 feedback predicts systematic improvement; within-issuer tests net out time-invariant traits [17–19].	Balanced FY2023–FY2024 panel (33 issuers; N=66) + paired tests.	H1 supported: paired $t(32)=4.967$, $p<.001$; Wilcoxon $z=3.724$, $p<.001$ (Tables 8 & 9).
SOTA rarely decomposes disclosures along IFRS 17 stock–flow logic.	G4: Mechanism gap	Different channels may operate differently (mapping/navigability/audit assurance vs movement-detail/complexity); composites can mask offsetting effects [2,3,5].	FLOW (D1–D3) vs STOCK (D4–D6) decomposition aligned with IFRS 17 architecture; estimate models with both components.	Mechanism evidence: STOCK → lower bid–ask spreads ($B=-0.023$, $p=.007$); FLOW → higher spreads ($B=+0.016$, $p=.025$) (Table 11). Robustness via LOG_AMIHUDD + mixed effects (Tables 12 & 13).
Limited published reliability/replicability reporting for scoring instruments.	G5: Reliability gap	Without reliability, measurement error can distort inference; “black-box” constructs reduce credibility in content analysis [3,35].	Internal consistency (ω , α + CIs) + inter-coder agreement (weighted κ) + ω -if-item-dropped sensitivity; full codebook in Appendix A.	Reliability: $\omega=0.893$ (0.853–0.934), $\alpha=0.883$ (0.830–0.935) (Table 5); ω -if-dropped (Table 6); κ_w in Appendix A.

RESEARCH GAP, CONTRIBUTION, AND FALSIFIABLE HYPOTHESES

Identified Research Gaps: from SOTA Evidence to Measurement Requirements

The synthesis of the state-of-the-art evidence in Tables 1 and 2, together with the structured mapping in Table 3, indicates five gaps that still constrain hypothesis driven archival research on IFRS 17 disclosure quality and comparability.

G1 Absence of replicable issuer year disclosure metrics.

Supervisory reviews and thematic diagnostics show that early applications vary considerably and often suffer from the same recurring weaknesses [17–19]. Benchmarking work by practitioners also shows which reporting formats firms actually use and how those formats change in the following year [23–26]. The literature still lacks a clear issuer-year scoring method that researchers can replicate and use as a firm-year variable in panel analyses. Without such a measure, it is difficult to assess what motivates disclosure architecture, or determine whether divergences in disclosure are economically meaningful. Such omission is the more relevant for archival studies in that disclosure theory connects disclosure architecture to capital-market outcomes via information risk and information-processing costs [2,3]. Both mechanisms rely on variation across firms and over time.

G2 Over-reliance on binary compliance logic.

When it comes to disclosure, we have a checklist approach that only addresses whether something is disclosed or not. This approach is overly simplistic when it comes to IFRS 17. And comparability is not only the presence of information, but also involves how economic events will get transferred into accounting outcomes [5]. Understanding IFRS 17 properly requires reconciliations, clear descriptions of key drivers, transparency on assumptions and quantified uncertainty [4, 16]. The downside of this approach is that it fails to differentiate informative disclosure from standard boilerplate. So even if a report seems compliant, it may yet be a challenge to follow exactly what took place and compare one issuer with another.

G3 There is limited within issuer evidence of post adoption learning.

The early reporting of IFRS 17 implementation is helpful, but the majority only cover the first year and are largely descriptive [20]. It is not known whether the disclosures issued by issuers become better after the first year; there are studies showing that firms react to comments from auditors, supervisors, and market participants [17–19]. In the absence of tracking the same issuers over time, first-year implementation noise can be readily misconstrued as true differences in substantive reporting practice. This is significant, because the approach differs in both cases.

G4 Lack of decomposition aligned with IFRS 17 stock flow logic.

Another problem is that most existing measures do not distinguish disclosure quality according to the logic of stock-flow presented in IFRS 17. This is unfortunate as IFRS 17 reporting consists of both stock and flow elements [4] and not all of these elements are equally informative. Some disclosures provide more information about performance during the reporting period, while others provide more information about opening balances and closing balances, transition effects, or the usability of the report. In combining these attributes into a composite score, some of this distinction is lost. The measure can still capture broad differences in disclosure quality, but it is difficult to know which aspects of the reporting structure were associated with the observed association. We think this

reduces the theoretical meaning and practical utility of the findings in terms of improving disclosure design [2, 3, 5].

G5 Insufficient reliability documentation for scoring instruments.

Another issue with scores is that many of these tools are not well documented in terms of reliability. Frequently, when benchmarks are reported to practitioners, there is little discussion about coder agreement, internal consistency, or the degree to which the result may depend on another aggregation [23–26]. If the report is presented under IFRS 17, the report should be more detailed and include more information. While judgement is not an afterthought in the coding process, it is part of it. Therefore, the reliability evidence should be considered as a necessary part of the measurement rather than an optional appendix. Where measurements do not include inter-coder agreement, consistency checks, or at least some type of sensitivity analysis (assumed within the norm of transparent content analysis) [35], the score is more difficult to defend as a robust measure. In the disclosure literature, this problem is not one of measuring validity but a non-reporting error [3].

How this Study Closes the Gaps. Design Response and Contribution

Taking the gap to design as input, we respond in the same way as indicated in Table 3. We have developed a replicated issuer year rubric that addresses both G1 and G2 where IFRS 17 disclosure quality is a matter of disclosure architecture rather than compliance with the standard, and applied the rubric to a balanced set of issuers for FY2023 and FY2024 to assess if the disclosure practices change in issuers during the first reporting cycle (G3). Given the structure of IFRS 17, we separated the measure into FLOW and STOCK (G4). Replication is supported by three ways.

First, explicit scoring rules, recoding procedures, and index construction logic are documented in Appendix A. Second, reliability evidence is provided through internal consistency, inter coder agreement on a double coded subsample, and drop one dimension sensitivity checks. These are reported in Tables 5 and 6 and Appendix A, addressing G5 [35]. Third, empirical tests are implemented using the rubric variables and the firm characteristics and market proxies already contained in the study dataset, and the design does not introduce new external enforcement or supervisory intensity variables

Falsifiable Hypotheses

This study evaluates four hypotheses, in line with the stated research gaps (G1-G5; Table 3), the first two mandatory IFRS 17 reporting years (FY2023-FY2024), and posited linkages between disclosure architecture, information frictions, and price. Each hypothesis is given in falsifiable terms and linked to an empirical test that is consistent with the issuer-year panel design.

H1 Learning and Second Year Improvement.

Issuer year IFRS 17 disclosure quality, TOTAL_SCORE, is higher in FY2024 than in FY2023.

Mechanism. Initial adoption of IFRS 17 involves implementation learning. After the first reporting cycle, preparers are expected to refine note architecture, reconciliation structure, and connectivity in response to feedback from auditors, supervisors, and users [8, 17–19]. Supervisory reviews of early IFRS 17 reporting emphasise weaknesses in movement explanations, assumption transparency, and cross referencing, and these areas indicate where disclosure architecture should improve [17–19]. If preparers internalise this feedback, second year reports should display more traceable roll forwards, D1, clearer disclosure of key actuarial inputs, D2, more decision useful uncertainty evidence, D3, more transparent transition and method explanations, D4, and improved presentation and navigability, D6. Under the rubric, these improvements should appear as within-issuer score increases over time [4].

For H1, the analysis is limited to the 33 insurers observed in both FY2023 and FY2024. We then test whether reporting scores are higher in the second year. We therefore examine score changes within issuers. The main tests are paired-sample procedures. We report a paired t-test and, given the small sample and the possibility of non-normal score changes, the Wilcoxon signed-rank test.

H2 Determinants and Reporting Incentives.

For H2, we expect the TOTAL_SCORE to be higher for firms that have stronger reasons and better ability to report well. The key variables we look at are firm size (SIZE_LN_TA), analyst coverage (ANALYST_COUNT), leverage (LEVERAGE), and whether the firm is listed (LISTED). We also include controls for YEAR and whether the firm is in EUROPE.

This idea is common in the disclosure literature [3]. Larger firms generally have more established reporting systems and receive more attention, so clearer reporting has more benefits. Following them on the ladder of analysts, they will have more pressure to provide useful information for forecasting and comparing with their peers. Use can also play a role because creditors scrutinize reports to understand liabilities, performance, and financial health. For example, under IFRS 17, the disclosures about insurance liabilities and the Contractual Service Margin (CSM) can influence how people think about a firm's solvency and its ability to take on debt. Shareholders face similar pressures, because they are subjected to the eyes of the markets, and therefore, we expect these to be associated with higher scores on the reporting rubric.

We test H2 using pooled OLS regression for a fully balanced set of issuer years (N = 66) and control variables are YEAR and EUROPE. We interpret the coefficients as conditional relations and not as proof of causality. Because we have an observational sample, it is possible that some important firm characteristics are not fully captured by the model.

H3 Liquidity and Trading Frictions.

In hypothesis H3, we examine whether IFRS 17 disclosures increase trading frictions (e.g., decreased BID_ASK_SPREAD and LOG_AMIHUD). We expect more transparency to aid outside users in understanding the components of reported figures [2, 3]. With better clarity on assumptions, reconciliations, and reporting choice, the information asymmetry

between firms and market participants should diminish. Comparability should also play a role. When reports are presented in a logically connected way, the user does not have to guess how similar events are reflected in the accounting results of firms. In IFRS 17, this is likely to be stronger in the areas of movement reconciliations (D1), actuarial assumptions (D2), sensitivity disclosures (D3), transparency in methods and transitions (D4), audit traceability (D5), and navigability (D6).

But, we don't expect every section of the disclosure to have the same effect. Some sections might be useful over time, but they may slow down users at first as there's more information to process, so the effects on liquidity may differ, especially at the outset [2,3].

Clearly, the stock-flow rubric will be more appropriate for this scenario. The FLOW (D1-D3) components address tracking movements, input transparency, and uncertainty quantification, which appear to be directly associated with information risk. The STOCK (D4-D6) components are concerned with method and transitions, audit traceability, and navigation that are more important for those who want to efficiently process the report or compare different firms side-by-side [5, 6].

This structure motivates two component oriented sub hypotheses. H3a predicts that higher STOCK_SCORE is associated with lower BID_ASK_SPREAD. H3b treats the sign of the FLOW_SCORE association with BID_ASK_SPREAD as an ex ante empirical question in the early adoption window.

Tests. BID_ASK_SPREAD is regressed on FLOW_SCORE and STOCK_SCORE with SIZE_LN_TA as a control. Models are estimated using cases where bid-ask data are available. The usable number of cases N is reported in the results tables. LOG_AMIHUDD is regressed on TOTAL_SCORE. ANALYST_COUNT is used as a control. This is estimated using cases where Amihud data are available. The usable N is reported. Mixed effects specifications with country random intercepts are used as a robustness check to account for jurisdictional clustering in market microstructure. All estimates are interpreted as associations rather than causal effects.

H4 Pricing Channel and Cost of Equity

Higher IFRS 17 disclosure quality, TOTAL_SCORE, is associated with a lower implied cost of equity, CO_EQUITY.

A risk model based on information risk estimates suggests that when information is more public, but less verifiable or precise, or simply harder to process, investors should pay for the extra uncertainty. This raises the expected return of the firm and the cost of equity [1, 2].

Actually, in our context the expectation from them is that higher TOTAL_SCORE should do the reverse. If IFRS 17 disclosures are more traceable, verifiable and interpretable, investors should have less uncertainty when forming views about the firm.

That should show up as lower perceived information risk and, by extension, a lower CO_EQUITY [1, 2].

Test. We estimate this relation by regressing CO_EQUITY on TOTAL_SCORE using pooled OLS on the market-data subsample for which CO_EQUITY is available, and we report the usable N. The control variables are SIZE_LN_TA and LEVERAGE. Our implied cost of equity measure is constructed using standard valuation-based approaches that rely on market prices and analyst forecasts; the details are set out in Appendix B. Given the short panel and the fact that implied cost of equity is an indirect proxy, we treat these results with caution.

Interpretation and boundary conditions. All hypotheses are tested in a two-year post-adoption observational panel covering FY2023 and FY2024. That design is useful for examining patterns, but it does not allow strong causal claims. We therefore interpret the results as conditional associations. A number of limitations remain relevant throughout: omitted variables such as governance quality, internal controls, or supervisory intensity; the possibility that market outcomes themselves shape disclosure incentives; measurement error in the market proxies; and the limited within-issuer time-series variation available over only two years. Even so, the rubric gives us a transparent issuer-year measure of disclosure quality, and the FLOW/STOCK split makes it possible to examine mechanism-consistent associations in a more structured way than the mainly qualitative evidence currently available [2, 3, 5, 35].

METHODOLOGY

In this study, we develop a replicable issuer-year measure of IFRS 17 disclosure from documentary analysis and use panel regressions to examine its association with key outcomes. Evidence extraction is implemented as structured content analysis of audited annual reports and IFRS financial statements using a pre-specified codebook, explicit evidence thresholds, and standardized scoring rules [35, 36]. Each insurer is treated as a “case” observed in two post-adoption reporting years (FY2023 and FY2024). The design therefore integrates (1) issuer-year scoring of audited IFRS 17 reporting packages and (2) statistical analysis of within-issuer and cross-issuer variation. H1 is assessed using within-issuer paired procedures. H2–H4 are estimated using pooled OLS as the baseline, complemented by mixed-effects with a country-level random intercept to account for repeated observations and jurisdictional clustering.

Rubric Development

The rubric captures whether an issuer’s IFRS 17 annual report and notes provide sufficiently explicit and verifiable disclosure support for users to interpret judgement-intensive IFRS 17 constructs. The instrument is defined ex ante with six dimensions (D1–D6) and evidence-based scoring thresholds, consistent with protocol-driven documentary research and content-analytic design [35, 36]. Scoring evidence is restricted to the audited annual report/IFRS financial statements and notes for the relevant issuer-year. To align measurement with IFRS 17’s stock-flow reporting logic while preserving replicability, the dimensions are grouped into two channels used in mechanism-oriented tests: FLOW (D1–D3) and STOCK (D4–D6). The FLOW channel captures traceability of

period movements and interpretation of key drivers and uncertainty. The STOCK channel captures transparency of methods and transition choices, audit-linked traceability signals, and navigability that supports cross-issuer interpretation.

Inter-coder reliability (double-coding). To ensure the objectivity of the scoring process, a double-coding procedure was conducted on a subsample of the panel. Two coders independently scored $n = 13$ issuer-year observations (19.7% of the full balanced panel, $N = 66$) on the same 0–2 ordinal category definitions used in Table 4 and Appendix A. Inter-coder agreement, assessed using weighted Cohen's κ for ordinal ratings, was $\kappa_w = 0.885$, indicating almost perfect agreement. Detailed statistical parameters and confidence intervals are reported in Appendix A.5.

Table 4. Rubric dimensions (issuer-year level)

Dimension	Label	What the dimension captures (high-level evidence standard)
D1	Reconciliation	Traceable roll-forwards/bridges that allow users to follow IFRS 17 movements across the period.
D2	Assumptions	Decision-useful transparency about key assumptions and methods that shape reported stocks/flows.
D3	Sensitivities	IFRS-17-specific sensitivity evidence that helps users gauge measurement uncertainty and drivers.
D4	Method / Transition	Clarity of transition approach and method choices; transparency of bridges linking IFRS 4 to IFRS 17 positions (as applicable)
D5	Audit (KAM)	Audit traceability: whether IFRS 17 is auditable and cross-referenced through audit focus/KAM evidence.
D6	Presentation/Format	Layout and navigability that support consistent interpretation and cross-issuer comparison.

Theoretical Grounding of the Six-Dimension Structure

Each rubric dimension is linked to disclosure theory, comparability theory, and information risk theory. The purpose is to capture decision useful disclosure attributes under IFRS 17 rather than to assemble an arbitrary list of disclosure items. This subsection sets out why these six dimensions provide a theory based operationalization of IFRS 17 disclosure quality, and why a simple checklist approach would not capture the same construct.

D1: Reconciliation and roll forwards traceability and movement explanation.

Theoretical justification. Reconciliations and roll forwards make comparability observable in practice. In the comparability literature, the central requirement is the mapping property. Users can assess the performance of companies based on how an organization's accounting numbers reflect the impact of economically equivalent events and allow users to understand these events [5]. With this being said under IFRS 17, to meet this requirement, the report needs to have a direct bridge showing how the opening balance ends up as the closing balance; therefore, allowing users to see what occurred during the reporting period and what was responsible for these occurrences. This will be important when it relates to judgments-based items such as Contractual Service Margin

(CSM) [9]. If the roll forward is transparent enough, users will be able to see how the profit has developed over time by viewing such items as the development of new business, experience adjustments, assumptions updates, interest accretion and expected CSM release [9]. In addition, the underlying principle of this concept can also be seen within disclosure economics. When a company issues more complex reports, it will become more difficult and expensive to isolate useful information for users' decisions. Reconciliation is useful since reconciliation breaks down the components of the change which can be used by users [3]. If the roll forward is not traceable, then the user will perceive the disclosed information as somewhat of a "black box." While users may be able to view the opening and closing stocks, the users will not be able to determine what caused the change nor whether the change is consistent with the issuer's narrative.

D1 evidence threshold (0 to 2). D1 examines whether the issuer provides sufficient detail to support an independent assessment of the change in the major IFRS 17 stock variables during each reporting period. The major items included in D1 are Contractual Service Margin (CSM), insurance contract liabilities, the risk adjustment, and reinsurance assets/liabilities. An example of a D1 score of 2 would include a report that includes a proper quantitative bridge with adequate line item detail to enable users to identify the major causes of the change. Typically, this would involve separate disclosures related to the CSM associated with new business, the amount of CSM recognized as profit, the experience adjustments, the assumption changes and foreign currency impacts.

Score of 1 – When a roll-forward is shown but too much information is not available to follow the movement, most likely because some parts are in "other" or presented at a superficial level. Score of 0 – The movement cannot be followed because there is no reconciliation or it is more likely that the movement has been replaced by a narrative

D2: Assumptions and Actuarial Transparency: Judgement and Input Observability.

Theoretical justification. In IFRS 17, the measurement is based on assumptions, such as discount rates, lapse and mortality rates, and expense loadings. These inputs affect the figure shown in the financial statements, and must be carefully selected. This is not always easy as many insurers have only limited information and different modelling choices can lead to different results. If readers do not see these inputs or understand how they were chosen, they are more likely to misinterpret the numbers [2]. More disclosure can help reduce this misinterpretations by explaining how the figures were arrived at and why certain assumptions were made.

Assumptions are also important in the comparability of the figures [5]. For instance, two insurance companies may present the same contractual service margin (CSM), but report different discount rates or biometric assumptions. Without more information on these inputs, it is difficult to determine whether the comparison is essentially reasonable. This is important for IFRS 17 as the CSM and risk adjustment are not directly observable. Much of their reliability depends on how the inputs are described [37, 38].

D2 evidence threshold (0–2). D2 measures whether enough quantitative and qualitative information is included in the disclosure for the reader to understand the effects of the

assumptions used in the financial statements. For instance, its focus on the assumptions of discount-rate terms, mortality and lapse, and expense inflation, while also asking how the assumptions were chosen and how they have changed over time.

Score 2: The report makes numerical statements about the assumptions and describes how those assumptions were formed. Score 1: The report makes a range or general description of the methodology but does not mention enough about the method used. For example, an insurer might state that discount rates are based on market yields without explaining the term structure or the source of the data. Score 0: Provide all the standard disclosures but not enough detail about the assumptions behind the numbers.

D3 (Sensitivities): Measurement Uncertainty and Risk Communication.

Theoretical justification. Sensitivity disclosures quantify how reported outcomes change under alternative assumptions and therefore directly communicate measurement uncertainty under IFRS 17 [4]. Information-risk theory predicts that when users cannot gauge the precision of disclosed estimates, they require compensation for bearing unobservable uncertainty [2]. IFRS-17-native sensitivities (e.g., discount-rate shocks and biometric/lapse stresses) enable users to stress-test CSM and profit measures, reduce model risk, and compare uncertainty exposure across issuers. Without sensitivity evidence, users cannot distinguish between firms with stable measurement outcomes and firms whose profitability is highly sensitive to plausible assumption changes. Practitioner benchmarking highlights that sensitivity disclosures are a key differentiator in the interpretability of early IFRS 17 reporting packages [23, 24]. Sensitivities also support cross-firm comparability by clarifying whether differences in reported outcomes reflect different assumption levels or different exposure to common shocks.

Evidence threshold (0 to 2). D3 asks whether the issuer reports sensitivity analysis that is native to IFRS 17, rather than relying on generic market risk sensitivity disclosure. It also asks whether the sensitivity work quantifies how changes in key assumptions affect IFRS 17 constructs. The relevant outputs include the CSM, insurance revenue, the insurance service result, and liability balances. The score of 2 suits cases where the report is reporting quantitative sensitivity tables for several scenarios and covers those key areas of assumptions including discount rates, longevity and mortality, lapses and expenses. Where the effects are obviously material, fuller disclosure should get the full credit. For a score of 1, the sensitivities are considered only in writing, or sometimes the quantitative analysis is rather limited (for instance, a single test for discount rates). A score of 0 applies when sensitivities are completely absent, or when the sensitivity disclosures that do exist are not meaningfully tied to IFRS 17 measures.

D4: Method and Transition Transparency: Comparability Across Time and Issuers.

Theoretical justification. IFRS 17 permits the use of Full Retrospective, Modified Retrospective, and Fair Value transition approaches. These are important because they can lead to different opening positions and make comparisons in the first year across firms more problematic [8]. The main issue for users is transparency. They should know how the transition from IFRS 4 to IFRS 17 has been implemented and see how that choice of

transition affects opening CSM and its reading in subsequent performance. A clear statement of the transition method, along with numbers showing the move from IFRS 4 to IFRS 17, helps readers separate transition effects from what happened in the business after that point [6].

Users need information on whether the entity is applying PAA, VFA, or GMM, together with the reasons for that choice of method, including eligibility decisions and materiality judgements. In the absence of such information, differences across entities may appear economic when they are in fact largely due to different choices made under the standard application.

Evidence threshold (0 to 2). D4 has two questions: first, is there a clear statement of the transition method? Second, is there a quantitative bridge from IFRS 4 to IFRS 17? A score of 2 requires both a statement of method that is unambiguous and reconciliation that enables users to follow through on all major adjustments; that is, it must be detailed. A score of 1 applies where either one or the other (method or bridge) is disclosed but not fully detailed or perhaps only partially aggregated. A score of 0 applies when there is no bridge at all or if described purely narratively.

D5: Audit Traceability and Key Audit Matters: External Assurance and Verifiability Signals

Theoretical justification. Disclosure is most helpful when readers are able to understand why it is important, credible, and able to be verified. A similar concept is applicable in disclosure economics: unsubstantiated claims do little to alleviate information asymmetry, as users are unable to differentiate robust reporting from opportunistic reporting [3]. With IFRS 17, the audit report can serve as that outside signal. When IFRS 17 measurement, transition, or disclosure is identified and described as a Key Audit Matter, or similar audit focus, the readers are shown that the more judgement-heavy areas have been subjected to audit review. This can be important when the reported results rely heavily on actuary and finance assumptions. In such cases, audit-related traceability can reinforce confidence in the reported changes, methods, and uncertainty disclosures.

Evidence threshold (0 to 2). D5 asks if the audit report mentions IFRS 17 as a Key Audit Matter, or comparable to it, and whether the case provides meaningful explanation regarding the audit challenge. A score of 2 applies where IFRS 17 is mentioned and the explanation contains substantial detail on the auditor's areas of interest or specific steps. A score of 1 is given if IFRS 17 is acknowledged, but only very generally. A score of 0 is given in case IFRS 17 is ignored as an audit focus.

D6: Presentation and Navigability: Report Structure and Cross-Reference Transparency

Theoretical justification. Even with complete transparency, information is still valueless when the audience is unable to find, connect, or interpret it. In this case, the transparency is counterproductive. The ability to navigate the documents can reduce the overall costs associated with processing information. Information that is organized with the aid of a clear index, consistently named sections, and effective use of cross-references can help

readers navigate reports. Avoid the need to stitch together the analysis and tables related to the same subject [3]. The IFRS 17 standard increases the need for effective presentation for the same reason. It is important that effective cross-referencing and organization improve traceability and verifiability. This is the case as they help the audience effortlessly follow the evidence within the report.

Evidence threshold (0 to 2). D6 examines the organisation of the IFRS 17 information in the report. Do the disclosures have proper structure and labelling, and do they allow the reader to follow without too much mental effort? A score of 2 is given if organisation and cross-referencing are such that tracing and confirming information is effortless. A score of 1 is given if there is some organisation, but it is excessive or unevenly labelled, requiring users to do substantial work to integrate the different tables and narrative. A score of 0 applies when the organisation of the information have a negative effect on the traceability (e.g. absence of cross-references, inconsistently and poorly labelled table titles across different sections, shifting terminology across sections).

Integration from Theory to Testable Measurement

Across the six dimensions, disclosure quality is treated here as a multi-faceted construct, rather than as a single overall feature. The rubric is organised around three ideas from the disclosure and comparability literature.

When discussing financial reports, there are a few key things to consider. First, users may not have all the necessary information. This is where details like D2, D3, and D5 can help users understand what is behind those numbers, changes in assumptions, and verifying the information. For instance, having a better understanding of the inputs that go into reporting the numbers would be a big help. Without all the necessary pieces of information, it is not possible to develop a complete understanding of the issue. Knowing how sensitive those numbers are to changes in assumptions is very important to determine whether the disclosures are trustworthy [2]. The second concern is the cost of processing this information. D1 and D6 are particularly important because readers have to try to connect the dots between the tables, movements, and explanations throughout the report [3]. It's like trying to navigate a complex maze of parts. The third item to consider is comparability. Here, D1 and D4 provide a way for readers to understand how economic events are translated into accounting results, so they can compare things across firms and time periods [5]. It's like having a common language that everyone understands.

The split of FLOW and STOCK is determined by combining D1, D2, and D3 together. STOCK consists of D4, D5, and D6. This split is based on the stock-flow logic that IFRS 17 reporting [4] demonstrates. It's useful because we can see how different parts of the disclosure are having an effect, rather than just looking at everything as one big score.

When we talk about FLOW, we're looking at how things have changed over a certain period and how easy it is to understand those changes, including tracking movement, assumptions, and sensitivity information. STOCK is more concerned with the overall package of information being reported. How clear the methods and transitions are,

whether the audit trail is easy to follow, and whether the report is organized in a way that makes sense to those who are using it. Most of the problems with coding come naturally. Sometimes the numbers don't add up neatly, or the assumptions are described in general terms. Have you ever been trying to follow across reference and been sent on a wild goose chase through multiple notes without a clear direction? That's what the rubric is trying to measure: not just whether the information is technically there but also how easily can be understood. The goal is to make sure that the disclosure structure is clear and useful, not just to check a box and say that the information has been disclosed.

Rubric as Theoretical Extension: Formalizing Information-Risk Reduction and Comparability-as-Mapping

The objective of the six-dimensional assessment framework is to put into practice the IFRS 17 reporting structure, which has been developed for this study's theoretical framework. This framework connects issuer-year reporting structures and the capital-market frictions' by using two primary concepts: Reduced Information Risk [2] and Increased Comparability of Financial Statements [5]. The IFRS 17 standard is a judgement based standard. The significant reported values under IFRS 17 such as the Contractual Service Margin ("CSM"), the Risk Adjustment for Non-financial Risk, and insurance finance effects are dependent on internal inputs and modelling assumptions, which cannot be directly observed by users external to those who prepare financial statements [4]. Thus, the quality of disclosure will depend not only on "how much" information is available, but also on how well users are able to verify what they are being presented with, whether they can trace estimates from one statement or note to another, and if the report will facilitate making reasonable comparisons between financial statements.

The information risk caused by misaligned knowledge between insiders and outsiders is what Easley and O'Hara refer to as "information risk." As per the terms of IFRS 17 standards, there are two additional ways this information risk could escalate: 1) Input Opacity - when investors cannot see all of the variables that go into constructing the inputs, including discount rates, lapse assumptions, mortality assumptions, and expense inflation. 2) Process Opacity - when a user cannot easily identify whether reported movements are due to actual experience or discretionary estimates made by an entity. The new reporting rubric under IFRS 17 is intended to eliminate some of the opacity in regards to reporting inputs by improving how users are able to observe and verify the information provided in the financial statements. The "assumptions" and "sensitivity" dimensions provide improved visibility to investors regarding the inputs used to prepare the financial statements and also provide greater clarity regarding any uncertainty related to the preparation of those financial statements. The audit and Key Audit Matter dimension adds something slightly different, see equation (1). It brings in an outside check, which may give users more confidence in the reporting and make the disclosures easier to rely on [2, 3].

$$IR_{i,t} = \alpha + \beta \text{TOTAL_SCORE}_{i,t} + \delta' X_{i,t} + \epsilon_{i,t}, \beta < 0 \quad (1)$$

Here, $IR_{i,t}$ is the firm's information risk in year t . The key explanatory variable is $TOTAL_SCORE_{i,t}$, the disclosure score for that firm-year, measured from 0 to 12. $X_{i,t}$ contains the usual controls, including size, leverage, and analyst coverage. Because information risk cannot be observed directly, the paper looks at it through the market-friction and pricing measures used later in H3 and H4, such as bid-ask spreads, an illiquidity measure, and the implied cost of equity where that figure is available. Given the short period since IFRS 17 was introduced, the results are treated as associations only.

De Franco et al. describe comparability as a matter of seeing how similar economic events end up in accounting numbers across firms [5]. Under IFRS 17, that may not always be easy, given that firms may use different transition approaches, different measurement models, or simply set out the relevant disclosures in different ways [4, 8]. These details may also be dispersed throughout notes and cross-references. The rubric was constructed with that in mind: give more weight to those sections of the report that make this path easier to follow. D1 shows how balances change from opening to closing. D4 makes it easier to track the choices between transition and method. D6 examines whether the report can be followed without too much back-and-forth by the reader, see equation (2). That point is close to the information-processing cost argument in [3, 5].

$$C_{i,j,t} = \phi + \gamma TOTAL_SCORE_{i,t} + \gamma TOTAL_SCORE_{j,t} + \eta' Z_{i,j,t} + u_{i,j,t}, \gamma > 0 \quad (2)$$

Here, $C_{i,j,t}$ captures latent comparability between issuers i and j in year t , with the vector $Z_{i,j,t}$ encapsulating economic and institutional similarity between the pair (business mix, domicile, shared characteristics etc.). Comparability similar to information risk, is not directly observable, so the study employs proxies for liquidity and illiquidity as indirect evidence of how easily users are able to digest and assess the reporting.

Going back to the stock-flow structure of IFRS 17, and the number of composite channel is reduced from six individual dimensions to two. $FLOW_SCORE$ (D1 + D2 + D3) represents movement traceability, visibility over assumptions, and quantified uncertainty.

These features are likely to be important primarily through the information-risk channel. $STOCK_SCORE$ (D4 + D5 + D6) captures the method and transition transparency, assurance signals and navigability. These characteristics are likely to be important for the most part via like-mindedness, as mapping, and reduced handling costs.

That directly corresponds to the liquidity expectation heading into H3. If $STOCK_SCORE$ has the stronger association this would also align with the hypothesis that most users, at least in advance of any particular focus on underlying or fundamental analysis, are likely to benefit more from reporting which is easier to navigate and interpret.

$FLOW_SCORE$ may still be relevant, but its effect may show up less clearly at the start. More detail does not always help immediately. In the first years, it can simply mean that the report takes more effort to read before users become more familiar with the new system [3].

EMPIRICAL RESULTS

Measurement Reliability: Internal Consistency of the Six-Dimension Rubric

The six dimension IFRS 17 native disclosure rubric has strong internal consistency. As reported in Table 5, McDonald's omega is $\omega = 0.893$ (95% CI: 0.853–0.934) and Cronbach's alpha is $\alpha = 0.883$ (95% CI: 0.830–0.935). These results support the use of an aggregated composite index, TOTAL_SCORE on a 0 to 12 scale, while also supporting the planned decomposition into FLOW_SCORE and STOCK_SCORE. FLOW_SCORE covers D1 to D3, Reconciliation, Assumptions, and Sensitivities, on a 0 to 6 scale. STOCK_SCORE covers D4 to D6, Method and Transition, Audit and Key Audit Matter traceability, and Presentation and Navigability, on a 0 to 6 scale. Internal consistency supports aggregation, but it does not by itself establish construct validity. Construct validity is assessed in the subsequent regression based analyses.

Table 5. Internal consistency reliability

Frequentist Scale Reliability Statistics				
Coefficient	Estimate	Std. Error	95% CI	
			Lower	Upper
Coefficient ω	0.893	0.021	0.853	0.934
Coefficient α	0.883	0.027	0.830	0.935

To check whether the composite is driven mainly by one dimension, Table 6 reports ω if a dimension is removed. The values range from 0.854 to 0.902, which indicates that removing any single dimension does not materially improve reliability. Dropping D4 produces the highest omega, 0.902, compared with the full-scale omega of 0.893, but the increase is small. For that reason, all six dimensions are retained in order to preserve content validity and maintain the rubric's intended coverage of IFRS 17 disclosure architecture.

Table 6. Reliability contribution by dimension (ω if item dropped)

Frequentist Individual Item Reliability Statistics			
Item	Coefficient ω (if item dropped)		
	Estimate	Lower 95% CI	Upper 95% CI
D1_Reconciliation_02_calc	0.862	0.809	0.916
D2_Assumptions_02_calc	0.871	0.821	0.921
D3_Sensitivities_02_calc	0.854	0.797	0.910
D4_Method_02_calc	0.902	0.864	0.940
D5_KAM_02_calc	0.889	0.845	0.932
D6_Format_02_calc	0.877	0.830	0.924

Descriptive Patterns across Years (2023–2024)

Descriptive evidence indicates an upward shift in rubric-based disclosure quality from the first to the second mandatory IFRS 17 reporting year, accompanied by lower dispersion (Table 7; Figure 1). For TOTAL_SCORE_0_12_calc, the mean increases from 9.924 in 2023 (SD = 2.433, N = 33) to 10.740 in 2024 (SD = 1.653, N = 33). The minimum rises from 5 (2023)

to 7 (2024), while the maximum remains 12 in both years. Because the dataset is a balanced two-year panel of the same issuers (33 insurers observed in both 2023 and 2024), this subsection reports descriptive patterns only. Formal hypothesis testing of year-to-year change is conducted using within-issuer procedures and/or panel specifications that account for repeated measurements, and is reported in the subsequent analyses.

Table 7. Year-by-year descriptive statistics for TOTAL_SCORE_0_12_calc (2023 vs 2024)

<i>Descriptive Statistics</i>	TOTAL_SCORE_0_12_calc	
	2023	2024
Valid	33	33
Mean	9.924	10.74
Std. Deviation	2.433	1.653
Skewness	-0.845	-1.130
Std. Error of Skewness	0.409	0.409
Kurtosis	-0.847	-0.128
Std. Error of Kurtosis	0.798	0.798
Minimum	5.000	7.000
Maximum	12.00	12.00

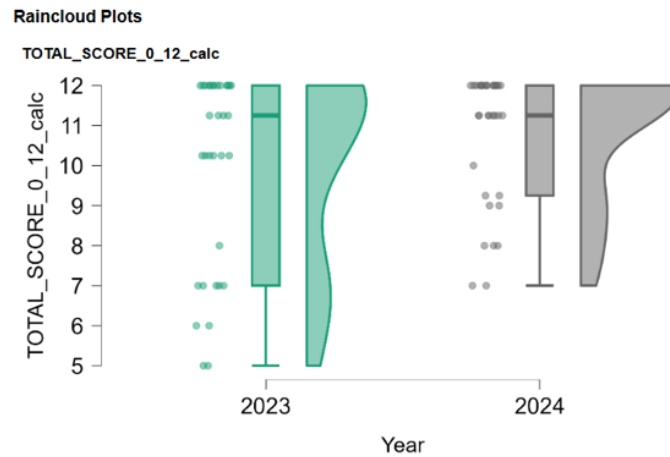


Figure 1. Year-by-year distribution of TOTAL_SCORE_0_12_calc (raincloud plot)

Hypothesis H1: Learning / Second-Year Improvement (Paired-Samples Tests)

To test H1, we compare issuer year IFRS 17 disclosure quality across the first two mandatory reporting years using the balanced panel of 33 life insurers observed in both FY2023 and FY2024. Table 8 shows that the mean composite rubric score, TOTAL_SCORE_0_12_calc, increases from 9.924 (SD 2.433) in FY2023 to 10.742 (SD 1.653) in FY2024. The mean within issuer change is 0.818 points on the 0 to 12 scale. The paired samples t test indicates that this increase is statistically significant, $t(32)$ equals 4.967, p less than 0.001. The evidence is consistent with the learning prediction that disclosure quality is higher in the second reporting year under IFRS 17.

The nonparametric Wilcoxon signed rank test yields the same conclusion, z equals 3.724, p less than 0.001, which supports the result without relying on normality.

Indeed both test results imply a potentially systemic improvement in disclosure architecture from FY2023 to FY2024 and are in line with second year stabilisation, as well as early learning on implementation. H1 is supported.

From a practical perspective, the year two increase indicates that insurers were able to enhance their IFRS 17 disclosure architecture in 2024 as compared to 2023. This trend is in line with the introduction of refinements to note presentation and also more closely aligned quantitative tables with relevant narrative description as preparers' become accustomed to the new reporting standard.

Table 8. Descriptive Statistics for TOTAL_SCORE_0_12_calc (FY2023–FY2024)

Descriptives	N	Mean	SD	SE	Coefficient of variation
TOTAL_SCORE_0_12_2024_calc	33	10.742	1.653	0.288	0.154
TOTAL_SCORE_0_12_2023_calc	33	9.924	2.433	0.424	0.245

Table 8 summarises the descriptive statistics for the composite IFRS 17 disclosure quality index, TOTAL_SCORE_0_12_calc, for the same 33 insurers observed in FY2023 and FY2024 in a balanced panel. The mean score rises from 9.924 in 2023 to 10.742 in 2024. Over the same period, the coefficient of variation falls. Taken together, these descriptive patterns indicate both higher average disclosure quality and greater consistency, which is consistent with stabilisation of disclosure architecture from the first to the second mandatory reporting year.

Table 9. Paired-Samples Tests for H1 (Learning / Second-Year Improvement)

Measure 1	Measure 2	Test	Statistic	z	df	p	VS-MPR*
TOTAL_SCORE_0_12_2024_calc	TOTAL_SCORE_0_12_2023_calc	Student	4.967		32	< .001	1565.5
		Wilcoxon	171.000	3.724		< .001	250.0

Paired-samples t -test and Wilcoxon signed-rank test comparing disclosure-quality scores between FY2023 and FY2024. Both tests confirm a statistically significant improvement ($t(32) = 4.967$, $z = 3.724$, $p < .001$), supporting H1 and indicating that disclosure quality increased in the second IFRS 17 reporting year.

Dependence and inference. The baseline regressions are estimated as pooled OLS and therefore rely on conventional independence assumptions. Because the dataset is a short balanced panel (33 issuers observed in two years), standard errors and p -values should be interpreted as conventional OLS inference and not as fully dependence-robust panel inference. To partially probe jurisdictional heterogeneity in the market-data setting, we additionally report a country random-intercept mixed-effects sensitivity model. This country-level specification is used as a coarse adjustment for cross-country baseline differences and does not, by itself, eliminate within-issuer dependence. Overall, given the

short panel and observational design, statistical evidence is interpreted as associational and exploratory rather than as causal or fully panel-robust inference.

Determinants of Disclosure Quality (pooled OLS; DV = TOTAL_SCORE_0_12_calc)

To test H2 (Determinants of disclosure quality), we estimate a pooled OLS model in which the issuer-year IFRS 17 disclosure-quality index (TOTAL_SCORE_0_12_calc) is regressed on firm-level monitoring and information-environment proxies and a regional control. Specifically, the model is expressed via equation (3):

$$\text{TOTAL_SCORE}_{\{i,t\}} = \beta_0 + \beta_1 \text{SIZE}_{\{i,t\}} + \beta_2 \text{ANALYST}_{\{i,t\}} + \beta_3 \text{LEVERAGE}_{\{i,t\}} + \beta_4 \text{LISTED}_i + \beta_5 \text{EUROPE}_i + \beta_6 \text{YEAR}_t + \varepsilon_{\{i,t\}}. \quad (3)$$

The estimation uses the full balanced rubric panel (N = 66 issuer-year observations). As reported in Table 10, the specification yields $R^2 = 0.414$, adjusted $R^2 = 0.365$, and RMSE = 1.677.

Table 10. OLS determinants model (DV = TOTAL_SCORE_0_12_calc)

Coefficients								Collinearity Statistics	
Model		Unstandardized	Standard Error	Standardized ^a	t	p	Tolerance	VIF	
M₀	(Intercept)	10.333	0.259		39.887	< .001			
M₁	(Intercept)	8.415	2.361		3.564	< .001			
	SIZE_LN_TA	-0.182	0.198	-0.135	-0.915	.364	0.449	2.229	
	ANALYST_COUNT	0.077	0.047	0.324	1.634	.107	0.249	4.022	
	LEVERAGE	15.385	3.693	0.474	4.166	< .001	0.753	1.327	
	LISTED (1)	0.283	0.774		0.366	.715	0.426	2.349	
	EUROPE (1)	-0.097	0.570		-0.170	.866	0.881	1.135	

^a Standardized coefficients can only be computed for continuous predictors.

Among the included covariates, LEVERAGE is the only statistically significant predictor of higher rubric scores (B = 15.385; SE = 3.693; t = 4.166; p < .001). The remaining covariates are not statistically significant in the multivariate specification: SIZE_LN_TA (B = -0.182; p = .364), ANALYST_COUNT (B = 0.077; p = .107), LISTED (B = 0.283; p = .715), and EUROPE (B = -0.097; p = .866). Accordingly, conditional on the included controls, we do not find evidence that listing status or the regional indicator (EUROPE) is systematically associated with IFRS 17 disclosure quality as measured by the rubric in this sample. Model fit: $R^2 = 0.414$; Adjusted $R^2 = 0.365$; RMSE = 1.677; N = 66.

Thus, H2 is partially supported: among the proposed monitoring/information-environment proxies, only leverage is robustly associated with higher rubric scores. Economic magnitude is interpreted in the units of the leverage measure used in the dataset (see Nomenclature/Appendix). Accordingly, the estimated coefficient indicates that higher leverage is associated with higher TOTAL_SCORE (0–12), holding other variables constant; any point-based translation (e.g., per 0.10-unit change) is meaningful only under the

leverage scale defined in the dataset. Collinearity diagnostics do not indicate severe multicollinearity: VIF values range from 1.135 to 4.022 (highest for ANALYST_COUNT), remaining below common cutoffs used to flag problematic inflation of standard errors. Given the design of the study, these findings are interpreted as showing associations in the IFRS 17 disclosure structures across firms at the early post-adoption period, rather than causal effects.

Economic Consequences: Liquidity Proxies and Cost of Equity

The full set of market data are not available for every firm, which means the sample is different across the models. The tables report the number of observations in each case. We study bid-ask spreads, Amihud trading frictions, and cost of equity.

Bid-ask Spread and the FLOW vs STOCK Decomposition

To see whether rubric-based disclosure quality is related to trading frictions, we estimate a pooled OLS model with BID_ASK_SPREAD as the dependent variable. The main explanatory variables are the two rubric components. FLOW_SCORE_0_12_calc covers D1 to D3, that is, reconciliation, assumptions, and sensitivities. STOCK_SCORE_0_12_calc covers D4 to D6, namely method and transition, audit and Key Audit Matter traceability, and presentation and navigability. We include firm size, SIZE_LN_TA, as a control. The model is reported in Table 11. The estimated coefficients reveal a component-specific pattern.

Controlling for firm size (SIZE_LN_TA), STOCK_SCORE_0_12_calc is negative and statistically significant ($B = -0.023$, $SE = 0.008$, $t = -2.835$, $p = .007$), indicating that stronger disclosure quality in the reporting-architecture and assurance-oriented dimensions (D4–D6) is associated with tighter bid-ask spreads (lower trading frictions). In contrast, FLOW_SCORE_0_12_calc is positive and statistically significant ($B = 0.016$, $SE = 0.007$, $t = 2.318$, $p = .025$), implying that stronger reconciliation/assumptions/sensitivities evidence (D1–D3) is associated with wider spreads in this sample. SIZE_LN_TA is included as a control and is not statistically significant ($B = 0.004$, $p = .107$).

Taken together, the evidence suggests that the liquidity relevance of IFRS 17 disclosure quality in the early post-adoption period is component-specific rather than uniform. Accordingly, support for H3 is strongest for the STOCK component and not supported for the FLOW component in the bid-ask spread tests. Consistent with the study design, these results are interpreted as associational rather than causal.

Table 11. OLS model for bid-ask spread (DV = BID_ASK_SPREAD)

Coefficients						
Model		Unstandardized	Standard Error	Standardized	t	p
M ₀	(Intercept)	0.010	0.004		2.544	.014
M ₁	(Intercept)	0.004	0.034		0.127	.900
	FLOW_SCORE_0_12_calc	0.016	0.007	0.584	2.318	.025

STOCK_SCOR	-0.023	0.008	-0.708	-2.835	.007
E_0_12_calc					
SIZE_LN_TA	0.004	0.002	0.218	1.645	.107

Amihud-based Trading Frictions (LOG_AMIHUD)

As a second liquidity-related proxy, we examine trading frictions using LOG_AMIHUD, constructed as $\text{LOG_AMIHUD} = \ln(\text{AMIHUD} + 0.001)$, where higher values indicate higher illiquidity (Appendix B). Table 12 reports the pooled OLS specification with LOG_AMIHUD as the dependent variable and includes TOTAL_SCORE_0_12_calc and ANALYST_COUNT as predictors. Model fit: $R^2 = 0.120$; Adjusted $R^2 = 0.082$; RMSE = 1.209; N = 50. Omnibus test: $F(2,47) = 3.195$, $p = .050$. Predictors: TOTAL_SCORE_0_12_calc, ANALYST_COUNT.

Table 12. Pooled OLS regression for LOG_AMIHUD (DV = LOG_AMIHUD)

Coefficients						95% CI		Collinearity Statistics		
Model		Unstand ardized	Stand ard Error	Stand ardized	T	p	Lower	Upper	Tolera nce	VIF
M ₀	(Intercept)	-6.247	0.178		-34.996	< .001	-6.605	-5.888		
M ₁	(Intercept)	-8.377	1.079		-7.762	< .001	-10.548	-6.205		
	TOTAL_S CORE_0_1 2_calc	0.245	0.100	0.345	2.438	.019	0.043	0.447	0.937	1.067
	ANALYST _COUNT	-0.034	0.027	-0.178	-1.259	.214	-0.088	0.020	0.937	1.067

The model-level omnibus test is borderline at the 5% threshold ($F(2,47) = 3.195$, $p = .050$). The coefficient on TOTAL_SCORE_0_12_calc is positive and statistically significant ($B = 0.245$, $SE = 0.100$, $t = 2.438$, $p = .019$), whereas ANALYST_COUNT is not statistically significant ($B = -0.034$, $SE = 0.027$, $t = -1.259$, $p = .214$). Given the illiquidity interpretation of LOG_AMIHUD, the positive coefficient indicates that higher rubric scores are associated with higher illiquidity in this reduced market-data subsample—i.e., the association is opposite to the hypothesized liquidity-improving direction; therefore, H3 is not supported in the Amihud tests. Consistent with the study design and the limited coverage of market variables, these findings are interpreted as associational rather than causal effects.

Robustness to Country-Level Clustering (Linear Mixed-Effects Model)

To assess whether the LOG_AMIHUD association is driven by cross-country baseline differences in market microstructure and reporting environments, we estimate a linear mixed-effects model with a random intercept for Country, using the Satterthwaite approximation for degrees of freedom and Type III sums of squares (Table 13). The model

is estimated on the market-data complete-case subsample ($N = 50$; 16 observations removed due to missing values). Because several countries in this subsample contribute only a single issuer, the country random intercept should be viewed as a coarse sensitivity adjustment that absorbs country-level baseline differences; variance-component interpretation is therefore not emphasised.

Table 13. Linear mixed-effects model (random intercept: Country)

ANOVA Summary			
Effect	df	F	p
TOTAL_SCORE_0_12_calc	1, 39.97	4.182	.047
SIZE_LN_TA	1, 44.97	1.023	.317
LEVERAGE	1, 41.93	0.906	.347
ANALYST_COUNT	1, 43.72	1.858	.180

Linear mixed-effects model with a random intercept for Country; Satterthwaite degrees of freedom; Type III sums of squares. Estimated on complete cases for the market-data specification ($N = 50$; 16 observations removed due to missingness). Several countries contribute only one issuer in this subsample; the random intercept is therefore used as a sensitivity control for country-level baseline differences rather than for detailed variance-component inference.

As reported in Table 13 (ANOVA summary), TOTAL_SCORE_0_12_calc remains statistically significant ($F(1, 39.97) = 4.182$, $p = .047$). In contrast, the additional firm characteristics included in this specification are not statistically significant (SIZE_LN_TA: $p = .317$; LEVERAGE: $p = .347$; ANALYST_COUNT: $p = .180$). Overall, this sensitivity check suggests that the association between the rubric-based disclosure score and LOG_AMIHUDD is not driven solely by cross-country baseline differences in the available market-data subsample. Consistent with the short two-year panel and observational design, the evidence is interpreted strictly as associational; moreover, the country random-intercept specification does not, by itself, remove within-issuer dependence.

Cost of Equity (CO_EQUITY)

As a broader pricing-based outcome, we test whether IFRS 17 disclosure quality is associated with the implied cost of equity (CO_EQUITY). Table 14 reports pooled OLS results for the subsample with CO_EQUITY available ($N = 50$).

Table 14. OLS model for cost of equity (DV = CO_EQUITY)

Coefficients						
Model		Unstandardized	Standard Error	Standardized	t	P
M ₀	(Intercept)	0.097	0.002		56.165	< .001
M ₁	(Intercept)	0.117	0.017		7.004	< .001
	TOTAL_SCO	-0.002	0.001	-0.238	-1.504	.139
	RE_0_12_calc					
	SIZE_LN_TA	-0.001	0.001	-0.140	-0.914	.366
	LEVERAGE	0.056	0.034	0.276	1.626	.111

The model is not statistically significant overall ($F(3,46) = 1.210, p = .317$) and explains a limited share of variation in CO_EQUITY ($R^2 = 0.073$; Adjusted $R^2 = 0.013$; RMSE = 0.012). Although the coefficient on disclosure quality is negative, it is not statistically significant (TOTAL_SCORE_0_12_calc: $B = -0.002, p = .139$). Therefore, within this specification and sample coverage, we do not find evidence of a statistically reliable association between the rubric-based disclosure measure and CO_EQUITY. Model fit: $R^2 = 0.073$; Adjusted $R^2 = 0.013$; RMSE = 0.012; $N = 50$. Omnibus test: $F(3,46) = 1.210, p = .317$.

DISCUSSION, IMPLICATIONS, AND FUTURE RESEARCH

Summary of Findings

This study develops and validates a replicable IFRS 17 native disclosure rubric and applies it to a balanced issuer year panel of 33 life insurance groups observed in FY2023 and FY2024, with N equal to 66. The design supports three questions that matter in the first two mandatory years of IFRS 17 reporting. First, does disclosure architecture improve in the second year? Second, which firm characteristics are associated with stronger rubric scores? Third, do rubric scores relate to market based proxies? The results extend the evidence base in four ways that correspond to the stated gaps G1 through G5 in Table 3. A further contribution is that supervisory and practitioner concerns about IFRS 17 disclosure architecture are translated into issuer year evidence that can be tested empirically [17–19, 23–26].

Finding 1: Measurement reliability and replicability, addressing Gap G5. The six dimension rubric shows strong internal consistency. This supports aggregation into TOTAL_SCORE while retaining the planned FLOW and STOCK decomposition. McDonald's ω equals 0.893 with a 95 percent confidence interval from 0.853 to 0.934, and Cronbach's α equals 0.883 with a 95 percent confidence interval from 0.830 to 0.935, as reported in Tables 5 and 6. Drop one-dimension checks yield ω if item dropped values between 0.854 and 0.902, which indicates that no single dimension mechanically drives the scale and that each contributes to the shared construct. Replicability is also supported by inter coder agreement in a double coded subsample using weighted Cohen's κ , reported in Appendix A, which supports use of the ordinal 0 to 2 scoring protocol.

Relative to state of the art evidence, the contribution here is methodological. Supervisory reviews and practitioner benchmarking remain useful, but they typically do not publish scoring algorithms and they rarely report reliability statistics that enable independent replication [17–19, 23–26]. In contrast, the rubric is fully specified, including explicit recoding rules in Appendix A, so it can be applied to new samples and extended in cumulative work. This addresses the black box measurement concern that arises when disclosure quality is proxied using undocumented checklists or non-replicable judgement [3].

Finding 2: Second year learning and stabilisation, addressing Gap G3. Disclosure architecture improves from FY2023 to FY2024. TOTAL_SCORE increases from FY2023,

with M equal to 9.92 and SD equal to 2.43, to FY2024, with M equal to 10.74 and SD equal to 1.65. The improvement in average scores is statistically significant. The paired-samples t-test yields $t(32) = 4.967$, $p < 0.001$ and the Wilcoxon signed-rank test yields $z = 3.724$, $p < 0.001$ (Tables 8 and 9). Issuers increased their scores by 0.82 points on average. Given the maximum score of 12, that is an improvement of 6.8%. There was less variation in scores in 2024 than in 2023. The coefficient of variation decreased from 24.5% in 2023 to 15.4% in 2024.

Firms may have learned fast once they had gone through their first mandatory IFRS 17 reporting cycle. But other things could have contributed to the improvement. Regulatory pressure and market competition are obvious examples. The fact that firms, once they had one full year under their belts, appear to have reworked portions of the reporting package also suggests that feedback – whether from auditors, supervisors, or users struggling to make sense of the notes – drove changes to layout, sequencing, and cross-referencing [8, 17-19].

The reason why we can assess changes in disclosure practice is that our panel is balanced. We follow the same issuers from year to year. This is a within-firm change. We are not comparing one set of firms to another. Our earlier benchmarking work found that disclosure in first-year IFRS 17 was heterogeneous across insurers [20]. The fact that scores are more tightly clustered in year two suggests that firms responded to the experience of year one by improving disclosure. It also counsels humility in interpreting some of the more jaundiced early supervisory commentary on IFRS 17. Weak disclosure may have reflected challenges of implementation. It may not have been entrenched [17-19]. Still, there is meaningful variation in disclosure scores in 2024. The range is 7 to 12, with a standard deviation of 1.65.

Finding 3: Firm-level determinants, particularly leverage, fill gaps in the literature related to its role in financial reporting, specifically gaps G1 and G2.

In the OLS model presented in Table 10, leverage is the only significant covariate. The coefficient is 15.385 (standard error = 3.693; $t = 4.166$; $p < 0.001$). It is positively correlated with TOTAL_SCORE. All other controls are not significantly associated: SIZE_LN_TA is -0.182 ($p = 0.364$), ANALYST_COUNT is 0.077 ($p = 0.107$), LISTED is 0.283 ($p = 0.715$), and EUROPE is -0.097 ($p = 0.866$). The models are moderately classified with $R^2 = 0.414$ and adjusted $R^2 = 0.365$. In substantive terms, a 0.10 increase in leverage is associated with an increase of approximately 1.54 points on the 0–12 TOTAL_SCORE scale. This suggests that firms with higher leverage may have more incentive for firms to disclose more information.

Although size and analyst following are often identified as important factors in the disclosure literature [3, 4], our results suggest that, in the IFRS 17 life-insurance setting, the positive association between leverage and rubric scores is consistent with a creditor-monitoring explanation. More highly leveraged insurers may face closer scrutiny from debt holders and rating agencies, and IFRS 17 increases the visibility of judgement-intensive areas such as discount rates, RA, transition choices, and the CSM. In our sample, during the early phase of IFRS 17 implementation, leverage is the only covariate associated

with higher rubric scores. This interpretation is broadly consistent with supervisory attention during rollout, particularly in relation to entity-specific disclosures, significant judgments and estimates, transition effects, and other areas where IFRS 17 requires substantial managerial judgement.

Finding 4: Market associations are channel-specific and proxy-dependent, addressing Gap G4.

Market evidence varies significantly depending on the specific disclosure component and the type of proxy used, which affects how financial data is interpreted. In Table 11, higher STOCK scores, defined as D4 plus D5 plus D6, are linked to tighter bid-ask spreads ($B = -0.023$; $SE = 0.008$; $t = -2.835$; $p = 0.007$). Higher FLOW scores (D1 plus D2 plus D3) are associated with wider bid-ask spreads ($B = 0.016$; $SE = 0.007$; $t = 2.318$; $p = 0.025$), when controlling for size.

Structure, comparability, and navigability were preferentially chosen by the market during the first years of IFRS 17 adoption, as less cognitive load and easy to interpret data was experienced. It is similar to the notion that comparability is a form of mapping, as described in the literature on information processing costs. FLOW's result is simple to interpret. In early adoption, more movement detail may increase the interpretation burden of interpreting data before users have fully integrated the new reporting system [3].

Unlike bid-ask spreads, LOG_AMIHUD is positively associated with TOTAL_SCORE ($B = 0.245$; $SE = 0.100$; $t = 2.438$; $p = 0.019$), and this relationship remains when country random intercepts are included. This suggests that the two proxies may capture different dimensions of market frictions

This difference is plausible, given the proxies capture distinct market frictions, which influence their association with market variables differently.

Bid-ask spreads are closely tied to dealer costs and adverse selection, whereas Amihud-based illiquidity is more sensitive to price impact and order-flow conditions. Those may behave differently, especially in thinner or more segmented markets, or where disclosure changes affect trading behaviour and the balance between informed and uninformed trading [2].

The sample does not support a pricing effect. In Table 14, the cost-of-equity model is not significant overall (omnibus $F(3, 46) = 1.210$; $p = 0.317$). The TOTAL_SCORE coefficient is negative, but not statistically significant ($B = -0.002$; $p = 0.139$). Given the smaller market-data subsample ($N = 50$), the noise in implied cost-of-equity proxies, and the short two-year window, a null result here is not surprising. Pricing may simply adjust more slowly than market microstructure frictions [1, 2].

Relative to existing supervisory and practitioner work, these results add more structure to repeated concerns about connectivity and navigability in IFRS 17 reporting [17–19, 23–26]. They also suggest that architecture-related features, captured by STOCK, are the disclosure features most consistently associated with tighter bid-ask spreads, whereas richer technical detail, captured by FLOW, does not produce a uniform reduction in

frictions during the early adoption period. That kind of mechanism-level inference is difficult to get from qualitative review alone without an auditable issuer-year metric [17–19].

Implications for Theory, Practice, and Policy

One of the basic difficulties in early IFRS 17 archival research is that disclosure is complicated and hard to measure in a way that can actually be tested. This study tries to deal with that problem by building an issuer-year measure that is auditable, replicable, and usable in panel regressions [17–20]. That is a methodological contribution, even if it sounds a little dry when stated so directly. Without an appropriate measure for disclosures, research about IFRS 17 could simply state that the firms provide what is reported, without providing evidence about how the reporting are useful.

The FLOW/STOCK distinction appears useful because it "divides two aspects of disclosure that are almost always confused with one another". The FLOW measures change, judgments and uncertainty, whereas STOCK tests whether the reader "can make sense of how the reporting flows together, to understand where the numbers came from and to trace the significant judgments through the disclosure". The evidence provided by the bid-ask calculation indicates that these two areas are not functioning in the same way during the early adoption years. The level of technical detail does not appear to be the key element to focus on; what seems to be of greater importance is the ability to logically follow the disclosure. The results demonstrate that in case of disclosure standards that rely on judgment, the quantity of disclosure may not be as important as the clarity and understandability. What matters is whether the disclosure is well organised, connected, and explained clearly [3].

Evidence is increasingly indicating that to the preparer and the audit committee, the structure of a report and assurance signals continue to play an important role. If the goal is to decrease user processing costs and improve the ability for a user to interpret the information, clear cross referencing between primary statements and reconciliations, assumption disclosures and risk disclosures, etc., is essential. Consistent use of indexation and signposting is equally important for the user. Users also gain from clearly defined transition bridges and explanations of methods, particularly where insurance companies have made different choices regarding adoption. In addition, audit trailability is important. IFRS 17 requires a high level of judgement, and as such, the user of the IFRS 17 report is likely to place more value on the IFRS 17 report when they can see where audit attention has focused on the individual judgements made. Of particular significance is when IFRS 17 is noted as a Key Audit Matter, or similar type of identification, and IFRS 17 is discussed in terms of procedural details.

The significance of IFRS 17 will become clear as it is identified in an audit as a Key Audit Matter among others; furthermore, when the comments are beyond the standard wording and provide insight into what was reviewed [17-19, 23-26]. The rubric's 0-to-2 rating scale provides audit committees with an easy way to differentiate between informative-only

disclosures and those that have some evidence supporting them, as well as those whose sources can be readily traced.

Through employing rubric scores, regulators and standard-setting organisations are able to better understand how disclosure patterns differ within a jurisdiction and between countries, thereby providing insight for future improvement of disclosure practices and thus ultimate transparency regarding the activities and operations of firms [17–19].

Overall, the relationship between STOCK and lower bid-ask spreads suggests some empirical support for the need to continue investigating connectivity, navigability, and traceability as part of the ongoing study of themes [17–19]. However, the results do not provide enough supportive evidence to make sweeping statements about the universal benefits of enhanced disclosures across all markets. Different proxies will capture different frictions and therefore make it impossible to establish causal relationships due to the limited time span of the panel. Therefore, any policy decisions regarding enhanced market disclosures should remain closely aligned with IFRS 17, such as decision usefulness, verifiability, and comparability [4, 7, 8]. The market data contain valuable information, but they cannot be relied on for conclusive evidence.

One limitation, which was discussed above, is the short period following the adoption of IFRS 17, specifically in the years 2023 and 2024. This gives the study some information on the first two mandatory years of IFRS 17 adoption, but not much on what may happen once reporting practices are more settled and firms have more time to adjust to the standard.

A further limitation is that the market tests are drawn from smaller subsamples. For bid-ask spreads, 52 observations are required, while for LOG_AMIHU and CO_EQUITY, 50 observations are required. As a result, the market evidence is drawn from larger and more liquid insurers, and it is not possible to exclude selection effects.

Moreover, there are also institutional limitations. Country random intercepts are useful to account for heterogeneity at a higher level (country level) but cannot reflect time-varying differences in the strictness of enforcement at the country level or specific characteristics of the local reporting environment at a more detailed level. The rubric has been purposefully limited to only cover audit annual reports and notes to IFRS financial statements, helping preserve scoring validity, yet ignoring potential additional information channels (such as investor presentations, conference calls, and interim reports).

Lastly, we are looking at observational data. We have not used falsification tests or additional checks of a more rigorous nature, and there are still concerns about missing variables or reverse causality. Because these data are observational in nature, we do not know whether the data represent associations or causal effects.

Future Research Agenda

This study opens multiple avenues for future work. The first limitation is the short time frame, which ends before the fiscal year 2025. It would be more interesting to see how firms begin to disclose in a more similar manner over time and how relevant FLOW and STOCK

stay as users become familiar with the new reporting framework as this long of a period will make that clearer.

More jurisdictions should also be included in the analysis. This is especially important for studies of enforcement alignment across comparability frameworks [6], because comparability is not just about the text of the standard. It also may depend on how consistently the standard is applied and enforced. The analysis could also be expanded beyond equity liquidity to encompass outcomes like analyst forecast properties, bond spreads, and rating changes. Such outcomes appear to be pertinent in light of the incentive patterns associated with leverage found in this study.

The rubric also merits further scrutiny on its own terms, instead of being interpreted merely through the total score. Not every component is likely to matter in equal measure in practice. Transparency at discount rates, for example, may be more informative than at biometric assumptions. Similarly, a forensic breakdown of new-business CSM roll-forwards might say more than that headline number does alone. An obvious next step will be to relate rubric-based scoring with quantitative measures of comparability, and explore whether enhanced disclosure architecture correlates with more visible comparability outcomes [5]. Future research could also consider the role of audit quality. In certain instances, heightened auditing and more robust disclosure architecture could operate in tandem. In others, they overlap on some of the same ground.

CONCLUSION

This paper develops a disclosure rubric for life insurers based on IFRS 17 and uses annual reports to construct an issuer year measure of disclosure quality and comparability. The rubric is composed of six dimensions, scored from 0 to 2, with the scores added together to produce a total score. FLOW and STOCK are separate in order to reflect the stock-flow structure of IFRS 17. These dimensions are connected to disclosure, information-risk, and comparability theory. The rubric represents the architectural structure of reporting, beyond compliance at the surface level.

The reliability tests support the use of an overall score, and the inter-coder differences suggest that the scoring could be used fairly consistently. The scores were generally higher in the second mandatory year than in the first, consistent with some early learning since the change. The companies did not move in the same direction, and there were clearly differences between issuers. The rubric was robust as a means of measuring performance.

Leverage is consistently emerges as a significant factor in all empirical tests, suggesting that it is an important determinant of firms' choices for early disclosure. One explanation is that more leveraged firms may be viewed more closely by creditors and rating agencies, which might induce firms to provide more disclosure in order to preserve credit quality. In the insurance context, however, this interpretation should be considered together with supervisory scrutiny and capital-management considerations.

NOMENCLATURE

IFRS	International Financial Reporting Standards
IASB	International Accounting Standards Board
IFRS 17	Insurance Contracts (standard)
IFRS 9	Financial Instruments (standard)
IFRS 4	Insurance Contracts (superseded standard)
CSM	Contractual Service Margin
KAM	Key Audit Matter(s)
EIOPA	European Insurance and Occupational Pensions Authority
ESMA	European Securities and Markets Authority
FRC	Financial Reporting Council
AMF	Albanian Financial Supervisory Authority
IVASS	Istituto per la Vigilanza sulle Assicurazioni
EU	European Union
OLS	Ordinary Least Squares
VIF	Variance Inflation Factor
RMSE	Root Mean Squared Error
SD	Standard Deviation
CI	Confidence Interval
df	Degrees of Freedom
N	Number of observations
ω	McDonald's omega (reliability coefficient)
α	Cronbach's alpha (reliability coefficient)
R ²	Coefficient of determination
BID_ASK_SPREAD	Bid-ask spread proxy (as defined in dataset)
AMIHUD	Amihud illiquidity measure
LOG_AMIHUD	$\ln(\text{AMIHUD} + 0.001)$
CO_EQUITY	Implied cost of equity proxy (as defined in dataset)
SIZE_LN_TA	Natural logarithm of total assets
LEVERAGE	Leverage ratio (as defined in dataset)
ANALYST_COUNT	Analyst coverage proxy
LISTED	Listing status indicator
FLOW_SCORE	D1 + D2 + D3
STOCK_SCORE	D4 + D5 + D6
TOTAL_SCORE	FLOW_SCORE + STOCK_SCORE
MRA	Modified Retrospective Approach (IFRS 17 transition)
FRA	Full Retrospective Approach (IFRS 17 transition)
FVA	Fair Value Approach (IFRS 17 transition)

AUTHORS CONTRIBUTIONS

Conceptualization, EY and DL; Methodology, EY and DL; Software, EY; Validation, EY and DL; Formal Analysis, EY; Investigation, EY; Resources, EY; Data Curation, EY; Writing – Original Draft Preparation, EY; Writing – Review & Editing, EY and DL; Visualization, EY; Supervision, DL.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interest.

REFERENCES

1. Botosan, C.A. Disclosure Level and the Cost of Equity Capital. *Acc. Rev.* **1997**, *72*(3), 323–349.
2. Easley, D., O'Hara, M. Information and the Cost of Capital. *J. Finance* **2004**, *59*(4), 1553–1583.
3. Leuz, C., Wysocki, P. D. The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *J. Account. Res.* **2016**, *54*(2), 525–622.
4. IFRS Foundation. IFRS 17 Insurance Contracts. Available from: <https://www.ifrs.org/issued-standards/list-of-standards/ifrs-17-insurance-contracts/> (Accessed on 13.10.2025).
5. De Franco, G., Kothari, S.P., Verdi, R. S. The Benefits of Financial Statement Comparability. *J. Account. Res.* **2011**, *49*(4), 895–931.
6. Prather-Kinsey, J., De Luca, F., Phan, H.-T.-P. Improving the Global Comparability of IFRS-Based Financial Reporting through Global Enforcement: A Proposed Organizational Dynamic. *Int. J. Disclosure Governance* **2022**, *19*, 330–351.
7. Vazov, R., Hristozov, Z. Specific Features of the Application of IFRS 17 – Valuation of Insurance Contracts and Profit and Loss Management. *J. Risk Financial Manag.* **2025**, *18*, 706.
8. Kamberi, A., Memeti, A., & Bexheti, A. e-UROZONE: A DSGE-Based Model for a New Financial Architecture. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, *8*(3), 734–783.
9. Yousuf, W., Stansfield, J., Malde, K., Mirin, N., Walton, R., Thorpe, B., Thorpe, J., Iftode, C., Tan, L., Dyble, R., Pelsser, A., Ghosh, A., Qin, W., Berry, T., Er, C. Contractual Service Margin under IFRS 17 from a Life Insurance Perspective. *British Actuarial Journal* **2021**, *26*, e2.
10. Signorelli, T., Campani, C.H., Neves, C. Direct Approach to Assess Risk Adjustment under IFRS 17. *Rev. Contab. Finanças* **2022**, *33*(90), e1646.
11. Zelinová, S., Blahušiaková, M., Šoltésová, T. Proposal for the Measurement of Reinsurance Contracts under IFRS 17. *Ekon. Čas.* **2022**, *70*(3), 284–304.
12. Basu, S., Grace, M. F. Insurance: In or out of the 'Too Difficult' Box? *Account. Bus. Res.* **2022**, *52*(5), 557–595.
13. Lindholm, M., Lindskog, F., Palmborg, L. Financial position and performance in IFRS 17. *Scandinavian Actuarial Journal*, **2021**, *2021*(3), 171–197.
14. Mignolet, F. Anticipating the Impact of IFRS 17 on Insurance Company Financial Statements' Transparency. Master's Thesis, Université de Liège, Liège, Belgium, 2017. <https://matheo.uliege.be/bitstream/2268.2/2782/4/Thesis%20-%20IFRS%2017%20Insurance%20Contracts%20-%20Félix%20MIGNOLET.pdf> (Accessed on 10.10.2025).

15. Nasser, A., Elsayed, M. M. S. S., Siyam, I. Analyzing the Impact of IFRS 17 on Insurance Reserves Management: A Multifaceted Assessment. *J. Financ. Commer. Res.* **2024**, 25(3), 1–20.
16. IFRS Foundation. IFRS 17 Supporting Material. <https://www.ifrs.org/supporting-implementation/supporting-materials-by-ifrs-standards/ifrs-17/> (Accessed on 11.10.2025).
17. European Insurance and Occupational Pensions Authority (EIOPA). Report on the Implementation of IFRS 17—Insurance Contracts; 15 April 2024. https://www.eiopa.europa.eu/document/download/e50ad04f-88a9-4ed4-bd76-c111bbd1f025_en?filename=EIOPA%27s+Report+on+the+implementation+of+IFRS+17+-+Insurance+contracts.pdf (Accessed on 17.10.2025).
18. European Securities and Markets Authority (ESMA). From “Black Box” to “Open Book”? Evidence from the First Application of IFRS 17 Insurance Contracts (ESMA32-1188985980-1046); 25 October 2024. <https://www.esma.europa.eu/sites/default/files/2024-10/ESMA32-1188985980-1046 - From black box to open book.pdf> (Accessed on 29.01.2026).
19. Financial Reporting Council (FRC). IFRS 17 “Insurance Contracts” Disclosures in the First Year of Application; 05 September 2024. https://media.frc.org.uk/documents/IFRS_17_Insurance_Contracts_Disclosures_in_the_First_Year_of_Application.pdf (Accessed on 15.10.2025).
20. ter Hoeven, R., Borelli, A., van Vlijmen, P. First Year’s Application of IFRS 17 in the Financial Statements of European Insurance Companies. *Maandbl. Account. Bedrijfsecon.* **2024**, 98(6), 403–422.
21. IVASS (Istituto per la Vigilanza sulle Assicurazioni). Lettera al mercato, 3 gennaio 2023: “Bilancio consolidato IAS/IFRS – Informativa sulla transizione all’IFRS 17 di cui all’Allegato 4 del Regolamento n. 7/2007, come modificato dal Provvedimento IVASS n. 121/2022”. Available from: <https://www.ivass.it/normativa/nazionale/secondaria-ivass/lettere/2023/lm-03-01/index.html> (Accessed on 15.10.2025).
22. KPMG. First Impressions: IFRS 17 Insurance Contracts (Updated for June 2020 Amendments); 2020. Available from: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/ifrg/2024/ifrs17-first-impressions-2020.pdf> (Accessed on 15.10.2025).
23. KPMG International Standards Group. Real-time IFRS 17: Insurers’ 2024 Annual Financial Statements; April 2025. Available from <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/ifrg/2025/isg-talkbook-real-time-ifrs-17-2024-annual-reporting.pdf.coredownload.inline.pdf> (Accessed on 15.10.2025).
24. EY. Reporting under IFRS 17 and IFRS 9: What We Learned from Insurers’ First Annual Reports; June 2024. Available from: <https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/industries/insurance/documents/ey-reporting-under-ifrs-17-and-ifrs-9.pdf> (Accessed on 15.10.2025).
25. EY. Observations from Year-end 2024 Financial Statements Published by Insurers: Reporting under IFRS 17 and IFRS 9; May 2025. Available from: <https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/industries/insurance/documents/ey-gl-reporting-under-ifrs-17-and-ifrs-9-05-2025.pdf> (Accessed on 14.10.2025).
26. Milliman. IFRS 17 Benchmarking: Our Key Observations on the 2024 Annual Reports; 5 June 2025. Available from: <https://www.milliman.com/en/insight/ifrs-17-benchmarking-fy-2024-report> (Accessed on 15.10.2025).

27. PwC. IFRS 17: Redefining Insurance Accounting; 2017. Available from: <https://www.pwc.com/sg/en/insurance/assets/ifrs17-redefining-insurance-accounting.pdf> (Accessed on 15.10.2025).
28. Grant Thornton International Ltd. Implementing IFRS 17 Insurance Contracts – A Practical Guide for Insurers; 2020. Available from: <https://www.grantthornton.global/globalassets/1.-member-firms/global/insights/article-pdfs/2020/gppc-ifrs-17-implementation-guide-final.pdf> (Accessed on 19.10.2025).
29. World Bank—CFRR. Challenges Applying IFRS 17 Insurance Contracts; 2023. Available <https://cfrr.worldbank.org/sites/default/files/2023-11/IFRS%2017%20Challenges%20and%20changes.pdf> (Accessed on 19.10.2025).
30. Alhawtmeh, O.M. The impact of IFRS 17 on the development of accounting measurement and disclosure, in addition to improving the quality of financial reports, considering compliance with the requirements of IFRS 4—Jordanian insurance companies—Field study. *Sustainability* **2023**, *15*(11), 8612.
31. Owais, W.O., Dahiyat, A.A. Readiness and challenges for applying IFRS 17 (Insurance Contracts): The case of Jordanian insurance companies. *Journal of Asian Finance, Economics and Business* **2021**, *8*(3), 277–286.
32. Puławska, K., Strzelczyk, W. IFRS 17 implementations: Market participants' perspective. *Central European Management Journal* **2025**, *33*(3), 455–477.
33. Hartojo, S.T., Purnamasari, V. Readiness and challenges of insurance companies in Indonesia in implementing PSAK 74 (IFRS 17). *Owner: Riset & Jurnal Akuntansi* **2023**, *7*(2), 1206–1212.
34. Beyer, A., Cohen, D.A., Lys, T.Z.; Walther, B.R. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* **2010**, *50*(2–3), 296–343.
35. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*, 4th ed.; Sage: Thousand Oaks, CA, USA, **2018**.
36. Yin, R.K. *Case Study Research: Design and Methods*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2003. Available from: https://books.google.al/books?id=BWea_9ZGOMwC&redir_esc=y (Accessed on 19.10.2025).
37. Barth, M.E.; Landsman, W.R.; Lang, M.H.; Williams, C.D. Are IFRS-based and US GAAP-based accounting amounts comparable? *Journal of Accounting and Economics* **2012**, *54*(1), 68–93.
38. Financial Reporting Council (FRC). Superseded Auditing Standards. Available from: <https://www.frc.org.uk/library/standards-codes-policy/audit-assurance-and-ethics/auditing-standards/superseded-auditing-standards/> (Accessed on 19.10.2025).

APPENDIX A — SCORING, RECODING (0–2), AND COMPOSITE INDEX CONSTRUCTION (FLOW, STOCK, TOTAL)

A.1. Unit Of Analysis and Stored Fields (Dataset-Level)

The unit of analysis is issuer–year. This appendix documents only the dataset implementation rules: (i) raw input fields as stored, (ii) deterministic recoding to a common 0–2 scale, and (iii) construction and verification of the FLOW, STOCK, and TOTAL composite indices (variable names exactly as stored).

Raw inputs stored in the dataset:

Raw in {0, 0.5, 1}:

D1_Reconciliation, D3_Sensitivities, D5_KAM, D6_Format

Raw on Likert 1–5:

D2_Assumptions

D4 stored via two components:

D4_Method coded as 1 = FV, 2 = MRA, 3 = FRA

D4_TransitionTransparency_02 coded directly on 0–2

A.2. Recoding to the common 0–2 scale (analysis fields)

A.2.1. D1, D3, D5, D6 (raw in {0, 0.5, 1})

Rule (as implemented):

$D_02_calc = \text{MIN}(2, D_raw \times 2)$

Mapping:

0 → 0

0.5 → 1

1 → 2

Stored analysis fields:

D1_Reconciliation_02_calc

D3_Sensitivities_02_calc

D5_KAM_02_calc

D6_Format_02_calc

A.2.2. D2 (Likert 1–5 Collapsed to 0–2)

Rule (as implemented):

$D2_Assumptions_02_calc = \text{IF}(D2_Assumptions \leq 2, 0, \text{IF}(D2_Assumptions = 3, 1, 2))$

Mapping:

{1, 2} → 0

3 → 1

{4, 5} → 2

Stored analysis field:

D2_Assumptions_02_calc

A.2.3. D4 (Method/Transition) Used in Composite Indices

Step 1 — method recode to 0–2:

$D4_Method_02_calc = D4_Method - 1$ (FV→0, MRA→1, FRA→2)

Step 2 — derived D4 component used in indices:

$D4_FINAL_0_2_calc = (D4_Method_02_calc + D4_TransitionTransparency_02) / 2$

Note: D4_FINAL_0_2_calc may be fractional (e.g., 0.0, 0.5, 1.0, 1.5, 2.0). Both underlying D4 components are retained for traceability.

A.3. Composite Indices (Variable Names and Formulas Exactly as Stored)

Despite the “0_12” suffix in the component names, the component indices range from 0 to 6 (sum of three 0–2 dimensions). The TOTAL index ranges from 0 to 12.

FLOW (0–6):

$$\text{FLOW_SCORE}_{0_12_calc} = \text{D1_Reconciliation}_{02_calc} + \text{D2_Assumptions}_{02_calc} + \text{D3_Sensitivities}_{02_calc}$$

STOCK (0–6):

$$\text{STOCK_SCORE}_{0_12_calc} = \text{D4_FINAL}_{0_2_calc} + \text{D5_KAM}_{02_calc} + \text{D6_Format}_{02_calc}$$

TOTAL (0–12):

$$\text{TOTAL_SCORE}_{0_12_calc} = \text{FLOW_SCORE}_{0_12_calc} + \text{STOCK_SCORE}_{0_12_calc}$$

A.4. Deterministic Integrity Checks (Replication Identities)

At issuer–year level, the following identities must hold:

Index identity:

$$\text{TOTAL_SCORE}_{0_12_calc} = \text{FLOW_SCORE}_{0_12_calc} + \text{STOCK_SCORE}_{0_12_calc}$$

Expanded identity:

$$\text{TOTAL_SCORE}_{0_12_calc} = (\text{D1_Reconciliation}_{02_calc} + \text{D2_Assumptions}_{02_calc} + \text{D3_Sensitivities}_{02_calc})$$

$$(\text{D4_FINAL}_{0_2_calc} + \text{D5_KAM}_{02_calc} + \text{D6_Format}_{02_calc})$$

Practical replication check (Excel):

$$\text{CHECK_TOTAL} = \text{IF}(\text{ABS}(\text{TOTAL_SCORE}_{0_12_calc} - (\text{FLOW_SCORE}_{0_12_calc} + \text{STOCK_SCORE}_{0_12_calc})) < 1\text{e-}9, \text{"PASS"}, \text{"FAIL"})$$

Table A1. Variable-level recoding and index construction rules

Constr uct	Raw input field	Analysis field	Rule / formula	Ran ge
D1	D1_Reconciliation	D1_Reconciliation_02 _calc	MIN(2, D1_Reconciliation × 2)	0–2
D2	D2_Assumptions	D2_Assumptions_02 _calc	IF(D2_Assumptions ≤ 2, 0, IF(D2_Ass umptions = 3, 1, 2))	0–2
D3	D3_Sensitivities	D3_Sensitivities_02_c alc	MIN(2, D3_Sensitivities × 2)	0–2
D4	D4_Method (1=FV, 2=MRA, 3=FRA)	D4_Method_02_calc	D4_Method – 1	0–2
D4	D4_TransitionTransp arency_02	D4_TransitionTransp arency_02	As coded	0–2
D4	—	D4_FINAL_0_2_calc	(D4_Method_02_calc + D4_TransitionTransparency_02)/2	0–2
D5	D5_KAM	D5_KAM_02_calc	MIN(2, D5_KAM × 2)	0–2
D6	D6_Format	D6_Format_02_calc	MIN(2, D6_Format × 2)	0–2
FLOW score	—	FLOW_SCORE_0_12 _calc	D1_02 + D2_02 + D3_02	0–6
STOC K score	—	STOCK_SCORE_0_1 2_calc	D4_FINAL + D5_02 + D6_02	0–6
TOTA L score	—	TOTAL_SCORE_0_1 2_calc	FLOW_SCORE_0_12_calc + STOCK_SCORE_0_12_calc	0–12

Appendix A.5. Inter-coder Agreement (Weighted Cohen's κ)

To assess coding replicability on the 0–2 ordinal scale, two coders independently scored a double-coded subsample of $n = 13$ issuer–year observations (i.e., 19.7% of the full balanced panel, $N = 66$).

Inter-coder agreement was evaluated using weighted Cohen's κ for ordinal ratings using the same category definitions as in Table 3 and Appendix A (0–2). The resulting agreement was $\kappa_w = 0.885$ (95% CI: 0.812 to 0.958). According to the benchmarks established by Landis and Koch (1977), this value indicates almost perfect agreement. Discrepancies were rare and were resolved via a reconcile-to-consensus session before the final pooled OLS and market-association analyses.

APPENDIX B — SAMPLE COMPOSITION, MARKET-VARIABLE IMPLEMENTATION, AND SUPPLEMENTAL OUTPUT (DESCRIPTIVES/INFLUENCE)

B.1. *Balanced Issuer–Year Rubric Panel (Fy2023–Fy2024)*

The rubric dataset is a balanced issuer–year panel of 33 life-insurance groups observed in FY2023 and FY2024 (N = 66 issuer–year observations).

Dataset fields used for descriptors/controls (as stored):

EUROPE: binary indicator (0/1)

LISTED: binary indicator (0/1)

Country: group domicile (used as grouping factor where a country random-intercept specification is estimated)

Table B1. Sample insurance groups (balanced issuer–year panel; 2023–2024)

Insurance group	Domicile (Country)	EUROPE indicator	LISTED indicator
AIA Group	Australia	No	Yes
ASR_Nederland_Life	Netherlands	Yes	Yes
Ageas	Belgium	Yes	Yes
Allianz_SE_Life	Germany	Yes	Yes
Allianz_Życie_Polska_S.A.	Poland	Yes	No
AXA_S.A.	France	Yes	Yes
Aviva_plc	United Kingdom	Yes	Yes
CNP_Assurances	France	Yes	Yes
Croatia_Osiguranje_Life	Croatia	Yes	No
Discovery_Life	South Africa	No	Yes
Fidelidade_Group	Portugal	Yes	No
GamaLife	Portugal	Yes	No
Generali_Italia_Life	Italy	Yes	No
KBC	Belgium	Yes	Yes
Legal_&General_Group	United Kingdom	Yes	Yes
MAPFRE_S.A.	Spain	Yes	Yes
Manulife	Canada	No	Yes
MetLife	Romania	Yes	No
NN_Asigurări_de_Viață_S.A.	Romania	Yes	No
NN_Group_Netherlands_Life	Netherlands	Yes	Yes
Poste_Vita_SpA	Italy	Yes	No
PZU_Życie	Poland	Yes	No
Resolution_Life_Australasia_(RLAL)	Australia	No	No
Sanlam_Life	South Africa	No	No
SEB_Life_Pension	Sweden	Yes	No
SunLife	Canada	No	Yes
Swiss_Life	Switzerland	Yes	Yes
Talanx	Germany	Yes	Yes
Triglav_Group	Croatia	Yes	No
UNIQA_Group	Austria	Yes	Yes
VIG	Austria	Yes	Yes

VidaCaixa_S.A.U.	Spain	Yes	No
Zurich_Insurance_Group	Switzerland	Yes	Yes

B.2. Market-Variable Implementation and Complete-Case Rules (Model-Specific Subsamples)

Market variables are used only where available and therefore each market specification is estimated on a complete-case subsample that matches the usable N reported in the corresponding main-text table. An observation enters a given market model only if the dependent variable and all included covariates for that model are non-missing.

BID_ASK_SPREAD is used as stored in the dataset (annual issuer-year scalar).

AMIHUDD is used as stored; the analysis uses the following transformation:

$\text{LOG_AMIHUDD} = \ln(\text{AMIHUDD} + 0.001)$ (0.001 is added to avoid undefined log values.)

CO_EQUITY is used as stored in the dataset (annual issuer-year scalar). The study does not recompute intermediate inputs inside the dataset; the regression uses the provided CO_EQUITY values on a complete-case basis. Country grouping (random intercept): where estimated, Country is used as the grouping factor and the model is estimated on complete cases for that specification.

B.3. Supplemental descriptives for regression subsamples

Table B2. Descriptives — bid-ask spread model subsample (used with Table 11)

	N	Mean	SD	SE
BID_ASK_SPREAD	52	0.010	0.028	0.004
FLOW_SCORE_0_12_calc	52	5.365	1.048	0.145
STOCK_SCORE_0_12_calc	52	5.389	0.862	0.120
SIZE_LN_TA	52	11.814	1.579	0.219

Notes: Values reproduce the “Descriptives” output for the complete-case regression subsample used in Table 11.

Table B3. Descriptives — LOG_AMIHUDD model subsample (used with Table 12)

	N	Mean	SD	SE
LOG_AMIHUDD	50	-6.247	1.262	0.178
TOTAL_SCORE_0_12_calc	50	10.865	1.778	0.251
ANALYST_COUNT	50	15.660	6.654	0.941

Notes: Values reproduce the “Descriptives” output for the complete-case regression subsample used in Table 12.

Table B4. Influence diagnostics — LOG_AMIHUDD model (Cook’s distance; “Influential Cases” output, highest values)

Case number	Std. residual	Cook’s distance
62	3.254	0.108
61	3.162	0.102
28	1.337	0.042
11	1.687	0.042
27	1.563	0.039

Notes: This table reports the largest Cook’s distance values shown in the “Influential Cases” output for the Table 12 regression (N = 50). Only the highest Cook’s distance values are shown for compact reporting; the full “Influential Cases” list is contained in the export for the same Table 12 specification.