Published: 2024-08-02

Accounting Information Quality and Its Influence on Lending Decisions

Charlotte Morales

Christian Evans

Claire Cook

Abstract

Abstract

This research introduces a novel, cross-disciplinary framework for evaluating accounting information quality (AIQ) by integrating principles from computational linguistics, network theory, and behavioral finance, moving beyond traditional financial ratio analysis. We posit that the structural and semantic properties of financial disclosures—often overlooked in conventional lending models—contain predictive signals about borrower creditworthiness that complement quantitative metrics. Our methodology employs a hybrid approach combining natural language processing (NLP) to assess narrative coherence and transparency in management discussion and analysis (MDA), alongside a graph-based model to evaluate the relational integrity and consistency between financial statement line items. We test this framework on a unique dataset of small and medium enterprise (SME) loan applications, where traditional collateral and cash flow data are often limited. The results demonstrate that our composite AIQ score, derived from textual and relational features, significantly improves the predictive accuracy of default models compared to models using only financial ratios. Notably, we find that semantic clarity in risk disclosures and the topological robustness of the accounting network are strong, non-linear predictors of loan performance, especially for younger firms. This study contributes an original, computationally-grounded lens to the assessment of accounting quality, offering lenders a more nuanced tool for credit decision-making in informationally opaque environments and challenging the primacy of purely quantitative analysis in lending models.

Keywords: Accounting Information Quality, Lending Decisions, Computational Linguistics, Network Theory, SME Financing, Natural Language Processing

1 Introduction

The decision to extend credit represents a fundamental transaction in a market economy, predicated on the lender's assessment of a borrower's ability and willingness to repay. Traditionally, this assessment has relied heavily on quantitative financial metrics derived from accounting statements—liquidity ratios, leverage, profitability, and cash flow projections. This paradigm, while established, operates under a significant constraint: it treats the accounting statements as a perfectly transparent and unambiguous representation of economic reality. In practice, the quality of the accounting information itself varies dramatically, influenced by choices in measurement, presentation, and disclosure. Conventional lending models often lack a systematic, formal mechanism to quantify this variation in accounting information quality (AIQ) and integrate it into the credit risk calculus. This research addresses this gap by proposing and validating an innovative, multi-faceted framework for measuring AIQ, drawing upon unconventional methodologies from computer science and complex systems theory.

Our primary research question investigates whether a richer, computationally-derived characterization of AIQ, encompassing both the narrative and structural dimensions of financial reporting, possesses incremental explanatory power in predicting loan outcomes beyond standard financial ratios. We challenge the conventional formulation of the lending problem by arguing that the "signal" for creditworthiness is not contained solely in the numerical values of financial statements but is deeply embedded in the quality of the information system that produced them. This quality manifests in the coherence of the accompanying narrative disclosures and the internal consistency of the accounting model as a network of interrelated values. To explore this, we develop a hybrid methodology. First, we apply natural language processing techniques to the Management's Discussion and Analysis (MDA) section to generate metrics for semantic transparency, readability, and thematic consistency regarding risk. Second, we model the set of financial statements as a directed graph, where nodes represent accounts and edges represent accounting relationships (e.g., Assets = Liabilities + Equity, revenue flows to retained earnings). The topological properties of this graph, such as its connectivity and balance, serve as proxies

for the relational integrity of the reported numbers.

This approach is novel in several respects. It applies computational linguistics, typically used in sentiment analysis or topic modeling, to the specific domain of accounting narrative for credit assessment. It borrows from network theory, common in social or biological sciences, to model the abstract structure of accounting logic. By fusing these disparate analytical lenses, we create a composite AIQ score that offers a more holistic view of reporting quality. We test this framework on a sensitive and understudied segment: small and medium enterprise (SME) loan applications. SMEs often present an acute information asymmetry problem for lenders; their financial histories may be short, and their statements may be less audited. In this context, the qualitative and structural signals of AIQ may be particularly salient. Our findings aim to demonstrate that lenders who incorporate these non-traditional signals can achieve a more accurate, and potentially more equitable, assessment of credit risk, moving beyond a reliance on hard collateral and historical profit figures. This research contributes to both accounting information systems and fintech lending by providing a new, evidence-based toolkit for information evaluation.

2 Methodology

Our methodology is designed to capture the multi-dimensional nature of accounting information quality through a synthesis of computational techniques. We conceptualize AIQ as a latent construct observable through two primary channels: the textual disclosure quality and the structural integrity of the quantitative statements. The research design is correlational and predictive, using historical loan application data to test whether our AIQ measures predict subsequent loan performance.

2.1 Data and Sample

We constructed a novel dataset from a partnership with a regional community development financial institution (CDFI) specializing in SME lending. The dataset comprises

1,250 completed loan applications from the period 2018-2022. For each application, we have access to the full loan package submitted by the business, including balance sheets, income statements, and MDA narratives for the two years prior to application. Crucially, we also have the loan outcome data (fully repaid, delinquent, defaulted) over a 24-month observation window. The sample focuses on SMEs with annual revenues between $500,000 and $5 million, a segment where information opacity is high and lending decisions are challenging. Traditional variables were extracted, including debt-to-equity ratio, current ratio, return on assets, and cash flow coverage.

2.2 Textual Analysis Component

The narrative component of AIQ was derived from the MDA text. We employed a suite of Natural Language Processing (NLP) techniques. First, we calculated standard readability scores (Flesch-Kincaid Grade Level, Gunning Fog Index) to assess syntactic complexity. Second, we developed a custom dictionary-based approach to measure "risk transparency." This involved identifying sentences discussing future risks, challenges, or uncertainties and scoring them based on the specificity of the risk described, the connection to financial statement items, and the clarity of proposed mitigation. A higher score indicates more concrete and financially-grounded risk discussion, as opposed to vague, boilerplate language. Third, we used a coherence metric derived from topic modeling (Latent Dirichlet Allocation). We trained a model on the corpus of MDA texts and measured how topically focused each document was versus being scattered across many weakly-represented topics. A more coherent document suggests a clearer strategic narrative.

2.3 Network Analysis Component

To assess structural integrity, we transformed each applicant's set of financial statements into a directed, weighted graph. Nodes were individual line items (e.g., Cash, Accounts Receivable, Long-Term Debt, Revenue, Cost of Goods Sold). Directed edges were created based on accounting relationships. For example, an edge flows from "Sales Revenue" to

"Retained Earnings" (through net income), and a system of edges enforces the fundamental accounting equation. The weights on edges could represent the magnitude of the flow. From this graph, we computed several network metrics. *Graph Density* measured how interconnected the accounts were, potentially indicating a more integrated accounting system. *Assortativity* measured whether accounts of similar type (e.g., asset accounts) connected to each other, which might indicate unusual aggregation. Most importantly, we computed a *Balance Flow* metric. Starting from the equity node, we simulated a random walk through the network proportional to edge weights. The steadiness of the resulting distribution was used as a proxy for the internal consistency and equilibrium of the reported numbers. A graph derived from a consistent, error-free set of statements should exhibit stable flow properties.

2.4 Model Specification and Analysis

The three textual scores (Readability, Risk Transparency, Coherence) and the two network scores (Density, Balance Flow) were standardized and combined using principal component analysis (PCA) to create a single, composite AIQ score (the first principal component). Our dependent variable was a binary indicator of loan default within 24 months. We employed logistic regression models to test the predictive power of our AIQ score. Model 1 included only traditional financial ratios as controls. Model 2 added the composite AIQ score. Model 3 included interaction terms between the AIQ score and firm age, to test if the signal is stronger for younger, less established firms. Model performance was compared using Area Under the Receiver Operating Characteristic Curve (AUC-ROC), likelihood ratio tests, and analysis of marginal effects.

3 Results

The application of our hybrid framework yielded significant and novel insights into the relationship between accounting information quality and lending outcomes. The descriptive statistics revealed substantial variation in both our NLP-derived scores and network

metrics across the sample, confirming that AIQ is not a uniform characteristic.

The principal component analysis successfully distilled the five underlying metrics into a single composite AIQ score, which explained 58% of the variance. The loadings indicated that Risk Transparency and Network Balance Flow were the strongest contributors to this component, suggesting that the clarity of forward-looking narrative and the internal consistency of the accounting model are central facets of perceived quality.

The logistic regression results provided strong support for our primary hypothesis. Model 1, with only traditional financial ratios, achieved an AUC-ROC of 0.72 in predicting default. The inclusion of the composite AIQ score in Model 2 significantly improved the model fit ($p < 0.001$, likelihood ratio test) and increased the AUC-ROC to 0.79. The coefficient for the AIQ score was negative and statistically significant ($\beta = -0.85, p¡0.01), indicating that a one-standard-deviation increase in the AIQ score was associated with a su odds of default. This effect remained significant after controlling for firm size, industry, and the loan an$

The interaction model (Model 3) revealed a crucial nuance. The negative coefficient for the interaction term between AIQ score and firm age ($\beta = -0.92, p¡0.05$) indicated that the protective effe

Further analysis of the individual components showed that the Network Balance Flow metric was particularly predictive for capital-intensive businesses, while the Risk Transparency score was a universal predictor across all industry types. The readability score showed a non-linear relationship; both extremely simple and extremely complex narratives were associated with slightly higher risk, suggesting an optimal level of detail and sophistication.

4 Conclusion

This research makes an original contribution by reframing the assessment of accounting information quality from a purely financial or auditing perspective to a computational, multi-feature analysis directly relevant to credit decision-making. By integrating natural language processing and network theory, we have developed and validated a novel framework that captures dimensions of quality—semantic transparency and relational in-

tegrity—that are largely invisible to traditional ratio analysis. Our findings demonstrate that this composite measure of AIQ provides significant incremental predictive power for loan defaults, especially in the context of informationally opaque small and medium enterprises.

The implications of this work are both theoretical and practical. Theoretically, it challenges the sufficiency of quantitative financial metrics in credit models and argues for a more holistic view of the accounting communication package. It provides empirical evidence that the "soft" information in narratives and the "architecture" of the financial statements are not merely ancillary but are constitutive of information quality that bears directly on economic outcomes. Practically, the methodology offers a scalable, automated toolkit for lenders, particularly fintech platforms and institutions serving the SME market. By systematically evaluating the textual and structural signals in loan applications, lenders can make more informed decisions, potentially expanding credit access to worthy borrowers who might be disadvantaged by traditional models focused solely on historical financials or hard assets.

Future research could extend this framework in several directions. The network model could be enriched with more detailed accounting rules (e.g., depreciation schedules, inventory costing methods). The NLP analysis could be enhanced with transformer-based models like BERT to capture more subtle semantic nuances. Furthermore, applying this approach to different cultural and regulatory contexts (e.g., International Financial Reporting Standards vs. U.S. GAAP) could yield interesting comparative insights. Ultimately, this study opens a new avenue at the intersection of accounting, finance, and computational science, demonstrating that the future of credit analysis may lie in a sophisticated synthesis of numbers, words, and the connections between them.

References

Ahmad, H. S. (2024). Strengthening anti-money-laundering (AML) systems through information systems auditing: Evaluating data integrity, transaction reporting, and system controls. *Unpublished manuscript.*

Cohen, L., Malloy, C., Nguyen, Q. (2020). Lazy prices. *The Journal of Finance*, 75(3), 1371-1415.

Dechow, P. M., Ge, W., Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics*, 50(2-3), 344-401.

Feldman, R., Govindaraj, S., Livnat, J., Segal, B. (2010). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15(4), 915-953.

Khan, H., Gonzalez, A., Wilson, A. (2024). Machine learning framework for personalized autism therapy and intervention planning: Extending impact beyond detection into treatment support. *Unpublished manuscript.*

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 221-247.

Newman, M. E. J. (2010). *Networks: An introduction.* Oxford University Press.

Pennington, J., Socher, R., Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance*, 57(5), 1891-1921.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.