D I S S E R T A T I O N P R O P O S A L

Jae Yeon (Jane) Pyo

"Accounting as Structured Information: An Information-Theoretic and Graph-Theoretic Framework for Financial Statement Analysis and Market Implications"

Wednesday, November 19, 2025
1:00pm
Tepper 4242

1. Accounting Classification Entropy

We theoretically develop and empirically apply a new information measure that quantifies the structural dimension of financial statements. Building on the entropy concept from Shannon's (1948) information theory, our approach captures firm fundamentals directly embedded in accounting numbers and formally measures the information conveyed by their classification structure. The measure is transparent, easy to implement, and adaptable across various classification-based settings. Using balance sheet data for U.S. public firms, we construct firm-quarter entropy-based measures and examine their relationship with market-based proxies that indirectly capture information content. In the first extension, we further demonstrate the measure's empirical usefulness by applying it to explain financial analysts' attention allocation decisions. In the second extension, we illustrate the framework's flexibility by incorporating richer accounting structures using the concept of mutual information and exploring additional empirical applications. Overall, our study offers a theory-based yet practical measure of financial statement classification that complements existing approaches to quantifying accounting information.

2. Accounting Graph Entropy

In this paper, we theoretically derive and empirically apply a new measure of firms' internal resource flows, built on the premise that the resource flows are keenly captured in the graph/network structure of the double-entry bookkeeping system as depicted in Ijiri (1975a) and later in Arya et al. (2000). Leveraging Graph and Information Theories, we derive a measure, labeled "Accounting Graph Entropy" (AGE), along with its associated difference measures such as K-L divergence (KLD) and Jensen-Shannon distance (JSD) to quantify news about firms' fundamentals (and their changes) at the granular level of bookkeeping graphs. The theoretical derivations proceed in two steps. First, we utilize double-entry bookkeeping laws to recover approximate weighted directed graphs/networks representing the aggregate journal entries among reported accounts on four public financial statements. Second, we develop a procedure that operates on the graph/network objects to compute the AGE measure. Built on the application of Information Theory to graph information objects, the procedure utilizes the resulting graph information functionals to extract the information embedded in the edges of the graph (representing the direction and the dollar amounts of the entries). The measure is a holistic summary measure of the multidimensional financial statement structure while remaining interpretable through edge-level or subgraph-level decomposition. In empirical applications, we construct a panel dataset of entropy-based statistics at the firm-quarter and firm-year levels using weighted directed graphs from financial statement data of US public firms. We document that the magnitude of year-to-year change in the financial statement graph structure closely mirrors expected economic patterns, including macroeconomic shocks and industry-specific transformations and, most notably, structural shifts in the banking sector during the early 2000s. Finally, by directly quantifying the structure of financial statements across firms and over time and without relying on external market-based reactions, we show that the implementation of SFAS 115 significantly improved cross-firm comparability in financial reporting.

3. Accounting Graph Entropy: Empirical Analysis

The paper extends the graph-entropy framework to a large-scale empirical setting that links accounting structure to firm performance and capital market valuation. By constructing GAAP-based financial-statement graphs for all Compustat firms from 1995 to 2024, the study measures structural change as the KLD between consecutive graph-edge distributions. In other words, the KLD measures the structural change from one quarter to the next quarter for each firm. The results reveal a consistent intertemporal performance pattern: firms undergoing large structural changes experience temporary declines in profitability but achieve higher long-term sales growth. Despite these long-term growths, markets react negatively to such changes, with persistent underperformance in cumulative abnormal returns. Option-implied volatility rises for high-divergence firms, indicating heightened investor uncertainty. Intuitively, investors' expectation of next period's earnings becomes more difficult to formulate as firms undergo larger changes, relative to firms that maintain the same structural activities. Cross-sectional Fama–MacBeth regressions further show that structural change predicts lower future returns even after controlling for established risk factors. Collectively, these findings suggest that financial-statement structure constitutes a distinct, priced but misunderstood risk factor: investors systematically discount firms that reorganize their financial architecture, even though such reorganization is associated with superior long-run fundamentals. This paper demonstrates that accounting-based information measures derived solely from the structure of financial statements can explain both real and capital-market phenomena, reinforcing the broader claim that accounting structure is itself a source of information.

4. Empirical Tests of Rational Expectations: Belief Updating to Earnings Announcements

What are the sources of investor disagreement? The competing hypotheses explaining the sources of disagreement are standard rational models with asymmetric information and differences in interpretation models. Contrary to conventional measures used to indirectly proxy for investors' disagreements (i.e., dispersion in analysts' forecasts), we present novel data that directly measure individual investors' beliefs at high frequency prior and posterior to earnings announcements. While previous literature measures sentiment using word counts or LLM models trained with sentence-level labels, investor beliefs about future stock performance are more nuanced and typically expressed through longer-form arguments that require document-level understanding. We address this by training a neural network on complete investor narratives to measure these beliefs about their expectations of future stock prices. First, we descriptively investigate the dynamics of investors' heterogeneous beliefs around earnings announcements and test the competing hypotheses. We find contradictory results: investors update their beliefs regarding the direction of new information, but overall disagreement increases after the revelation of public information. Next, we test how much of the trading volume can be explained by disagreement via the two mechanisms above. We find a positive association between disagreement and trading volume, as well as proxies for both mechanisms positively inducing trading volume. Notably, we find that differences in interpretation exhibit a stronger positive relationship with trading volume, compared to resolution in differing priors.

Proposed Committee: Pierre Liang, Gaoqing Zhang, Yucheng Liang, Bryan Routledge