Beyond the Numbers: Soft Information in Conditional Asset Pricing

By Feng, Huang, Huang, and Shi

Discussion by Frank Weikai Li (City University of HK)

Summary

- RQ: Examine the role of textual information in conditional asset pricing
- Empirical approach:
  - First, use topic modelling to extract interpretable textual factors/themes
  - Second, embed these textual factors within an IPCA model with standard firm characteristics
- Main findings:
  - **Text dramatically improves mean-variance efficiency** (out-of-sample Sharpe ratios rise from ~1.7 to over 2.6 when text is added to characteristics), but it does almost nothing for expected-return forecasts.
  - The efficiency gains operate through a **covariance channel**: conference-call topics capture time-varying return comovement that is orthogonal to usual characteristics.
  - The importance of textual factors grow over time and is concentrated in intangible-heavy, high-growth firms where hard data is least informative

Strengths and Contributions

- The core innovation is to incorporate *qualitative* information from earnings conference calls into a conditional asset pricing framework, moving beyond standard "hard" firm characteristics.
- The results are economically large and statistically robust. The authors convincingly show that text-managed portfolios cannot be explained by standard factor models
- The empirical work is thorough, with careful handling of look-ahead bias through online LDA and extensive robustness checks.
- My Main Comments:
  - Why Text Boosts Covariance but Not Expected Returns Estimation? The limitation of text encoding using topic modelling
  - Does Text Add Value Beyond Non-Conventional Quantitative Data?
  - The Nature of Information or the Timeliness of Information?

Main Comments 1: Why Text Boosts Covariance but Not Expected Returns Estimation?

- **The empirical puzzle (from the paper):**
  - Adding conference-call topics to IPCA improves mean-variance efficiency substantially (Sharpe ratio from ~1.7 to 2.6).
  - Yet it adds almost *nothing* to out-of-sample expected-return forecasts (predictive R^2 stays flat).
- **My explanation:**
  - The way textual information is encoded matters.
  - LDA topic loadings are **directionless** by construction.
  - They measure *how much* a firm talks about a theme, but not so much on *whether* the exposure to the topic is positive or negative.
  - This asymmetry explains both the first-moment non-result and the second-moment success.

Main Comments 1: Why Text Boosts Covariance but Not Expected Returns Estimation?

- *Substance vs. Direction: What Do Signals Actually Convey?*

	Firm Characteristics	**LDA Topic Loadings**
Factor identity	Clearly defined (size, value, profitability)	Thematic ("Renewable Energy," "Cloud Service")
Direction / Sign	Embedded in the value itself: high B/M → positive exposure to value premium, low size → positive SMB loading	Sign is missing: a loading of 0.8 on "Renewable Energy" just means how **prevalent** the topic is in the call
Return prediction	Can be mapped to expected returns because the sign of exposure is known (e.g., high B/M → higher expected return)	Cannot deliver a directional signal without additional directional stance info

Main Comments 1: Why Text Boosts Covariance but Not Expected Returns Estimation?

- Example: (0.8, 0.2) on (value, quality) immediately implies the stock is a low-quality value stock → directional expected prediction.
- (0.8, 0.2) on (Renewable Energy, Health Care) says the firm operates in energy sectors but does not tell whether it benefits or suffers from the green-energy transition.
- **Why this matters? IPCA models expected returns as a linear combination of the instruments**
- When the instrument vector consists of topic proportions/prevalence:
    - Two firms can have identical high loadings on "Renewable Energy" but opposite economic exposures (e.g., a green-energy leader vs. a coal-dependent utility).
    - A linear predictor sees the same number and cannot differentiate → average predicted return is uninformative.
    - The signal from text is therefore washed out across firms with varied directional exposures.

Main Comments 1: Why Text Boosts Covariance but Not Expected Returns Estimation?

- **In contrast, standard firm characteristics contain directional information:**
  - Even after rank-transformation, a high book-to-market or low size retains its monotonic mapping to the associated expected return.
  - Hence, firm characteristics predict returns, while topic loadings do not.
- **In summary:** The null expected-return result is not evidence that textual information is irrelevant for the first momentum
- It is mainly because topic modelling encodes relative prevalence, not economic stance
- In fact, the textual analysis literature consistently document textual information is useful to predict expected return, after we encode its directional information
  - Chen, Kelly and Xiu, 2019: Predicting Returns with Text Data
  - Didisheim, Kelly, Pourmohammadi and Tian (2026): The Inefficient Pricing of News Shocks

Main Comments 1: Why the Covariance Channel Survives & How to Improve It?

- **Why covariances still improve:**
  - Topic loadings act as novel **industry/theme tags** that cluster firms with similar risk exposures → enhance estimation of the conditional covariance matrix.
  - Two firms that spend much of their calls on "Renewable Energy" tend to comove (positively or negatively), irrespective of whether the firm is a winner or loser from green transition.

- **Suggestion:**
  - Augment topic loadings with **sentiment or stance information** (e.g., net positive/negative word counts per topic).
  - A *directional* loading would encode both prevalence and stance, e.g., "firm would benefit from the transition to Renewable Energy."
  - More sophisticated way of Ke, Kelly, and Xiu: learns positive/negative sentiment-charged words supervised by returns

Main comments 2: Does Text Add Value Beyond Alt Quantitative Data?

- The paper compares conference-call text against 35 traditional firm characteristics (size, value, profitability, etc.). This shows text adds value relative to standard accounting/market data.
- But alternative quantitative data are increasingly available and useful:
  - ESG ratings, patent & product market networks, etc., have also been shown to improve mean-variance efficiency (Lindsey, Pruitt and Schiller 2024) through the covariance channel

ESG and the Conditional Pricing of Risk

61 Pages • Posted: 6 Dec 2021 • Last revised: 29 Aug 2025

Laura Anne Lindsey
Arizona State University (ASU) - Finance Department

Seth Pruitt
Arizona State University (ASU) - Finance Department

Christoph Schiller
Ohio State University (OSU) - Department of Finance

Date Written: August 18, 2025

Abstract

We evaluate ESG investing through the lens of conditional asset pricing. We find that ESG characteristics do not define a new priced factor or deliver alpha, but they do explain variation in firms' exposure to existing aggregate risks---particularly with respect to value. These conditional beta effects differ across ESG data

- Alternative data keep showing up as beta signals rather than alpha signals — is there a general lesson here, and is the text result a specific instance of a broader pattern?

Main comments 3: Is it the Nature of Information or the Timeliness of Information?

- **The implicit asymmetry in the paper's setup:**
  - Firm characteristics are constructed following standard conventions: many accounting items (e.g., book equity, earnings, investment) are stale with a delay of several months *after* the fiscal quarter ends.
  - Conference-call transcripts, by contrast, are typically released within days of the announcements. The textual signal used each month therefore incorporates more updated information.
- **Implication:**
  - The superior mean-variance efficiency and covariance improvement attributed to text might partly reflect this timeliness advantage.
- **Suggestion:** Following Bowles et al. (2025 JF) to measure accounting characteristics right after actual filing/release date.
- If the text advantage shrinks, it was partly a timeliness effect; if it survives, the "nature of the information" interpretation is vindicated.

Minor comments 1: Trading Costs and After-Cost Performance

- **The MVE portfolio results are before transaction costs:**
  - However, portfolios are rebalanced monthly using signals from quarterly conference calls;
  - Text-based conditional weights likely fluctuate substantially as new calls become available.
  - The concern is trading costs (bid-ask spreads, market impact, commissions) can erode the apparent risk-adjusted outperformance.
- **Missing metrics:**
  - Change in the turnover of the text-enhanced MVE portfolio relative to characteristics-based portfolio?
  - Net Sharpe ratio under realistic cost assumptions (e.g., 10–30 bps per trade).
  - If the incremental gains from text survive after costs, the economic significance of this paper is greatly magnified.

Minor comments 2: text topics vs. granular industry classification

- Since the topics that matter most are largely industry tags (Table 12), the natural skeptic's question is whether the covariance gain is "soft information" or just better industry classification.
- How much of the text gain survives once IPCA is given a richer industry-based instrument — Hoberg-Phillips TNIC text-network industries or Search-based peer firms (Lee et al. 2015)?
- If the gain is mostly absorbed by granular industry dummies, the "soft information" interpretation narrows to the handful of genuinely non-industry topics

Conclusion

- Novel application of textual information in asset pricing
- Rigorous tests and comprehensive analyses
- **Main suggestions:**
  - Incorporate directional info when doing text encoding
  - Better and more timely measure of quantitative characteristics
  - Trading costs and other implementation issues for practitioners