

Narrative Risk in Empirical Finance From Dictionary Sentiment to LLM-Extracted Disclosure Factors

Dennis Hoffmann

Abstract

This article surveys narrative risk in empirical finance, the family of textual signals derived from corporate disclosures, news, and analyst language that move the cross-section of expected returns. The corpus consists of 23 papers across 5 thematic clusters identified by a 2026-05-02 systematic literature review over OpenAlex. The article aim is structural: to map the literature along a dictionary, firm-exposure, factor-pricing, and large-language-model arc, with chronological development handled within each thematic section. The article surveys what the systematic search surfaced rather than reporting new empirical findings, so it complements rather than competes with the underlying SLR run.

1 Introduction

Narrative risk is the price-relevance of textual signals derived from corporate disclosures, news, and analyst language. Equity returns reflect both quantitative information (earnings, cash flows, accounting accruals) and qualitative signals (tone, risk-factor language, news context, narrative structure). Quantitative inputs are well understood and have a long history of factor-pricing tests; the qualitative side is the younger and faster-moving research frontier. The literature traced in this article has progressed from dictionary methods to large-language-model extraction, with each generation surfacing new firm-level and aggregate exposures that prior methods could not detect. The earliest empirical work bridged psychology and finance by counting tonal words in newspaper columns and showing that sentiment forecasts daily index returns and trading volume (Tetlock, 2007). A finance-specific dictionary refined that approach by recognising that words like “liability” carry a different polarity in 10-K filings than in general English (Loughran and McDonald, 2011). The next decade extended these foundations into firm-level text-based exposures, news-based aggregate indices, and topic-model factor portfolios.

This article surveys the resulting corpus through the lens of a structured systematic literature review. The methodological framework that produced the corpus is an algorithmic pipeline tailored to financial narratives (Taibi and Osterrieder, 2026). A 2026-05-02 systematic search over OpenAlex screened 92,426 candidate records down to 30 scored papers and 23 selected for the final corpus, spanning 5 themes. The 5 themes are Narrative Asset Pricing and News-Based Factors (12 papers) and Text-Based Firm-Level Exposures (7

papers). They are joined by LLM-Based Disclosure and Document Extraction (7 papers), Risk-Disclosure Information Content (4 papers), and Sentiment and Dictionary Baselines (4 papers). The two smallest themes are methodologically adjacent and share Section 4 in this article, while the other three themes anchor their own sections.

The organising principle is thematic. Within each section, the prose proceeds chronologically from the oldest seed paper to the newest discovered work, so the reader sees how each theme matured even though section ordering is theme-driven rather than time-driven. This choice reflects a corpus reality: the 5 themes are unbalanced and overlap in time, and a strict chronological spine would scatter related papers across the document. It also makes the curation rule auditable. The body cites the top-13 papers ranked by the SLR pipeline’s combined relevance and quality scores, plus the methodological framework paper that produced the corpus. The 9 remaining corpus papers appear only in a corpus-at-a-glance table near the end of the document.

The rest of the article proceeds as follows. Section 2 covers narrative asset pricing and news-based factors. Section 3 covers text-based firm-level exposures. Section 4 pairs risk-disclosure information content with sentiment and dictionary baselines. Section 5 covers large-language-model-based extraction and the look-ahead-bias critique that frames it. Section 6 concludes with open questions.

2 Narrative Asset Pricing & News-Based Factors

This section surveys aggregate news-based indices and narrative factor portfolios. The label Narrative Asset Pricing & News-Based Factors is the canonical theme name used by the SLR run’s configuration. Six body-cited papers anchor the discussion in chronological order: Tetlock (2007), Baker et al. (2016), Caldara and Iacoviello (2022), Calomiris and Mamaysky (2019), Engle et al. (2019), and Bybee et al. (2023). The throughline is that aggregate textual signals, distilled from newspapers and other periodic media, can be priced as systematic risk factors once the right aggregation rule is applied.

The starting point is Tetlock (2007). The paper bridges from Section 1 because it is both the foundational sentiment study and the first quantitative demonstration that media tone forecasts market behaviour at daily frequency. The empirical design counts negative-tone words in the Wall Street Journal’s “Abreast of the Market” column over a long sample, then tests whether tone shocks Granger-cause Dow Jones Industrial Average returns and trading volume. High pessimism predicts downward pressure on prices the next day, with reversion to fundamentals over the following week. Trading volume rises when pessimism is unusually high or unusually low, consistent with media tone proxying for noise-trader sentiment. The methodology, simple word counting against a general-purpose dictionary, set the template for a generation of follow-up work.

The natural next step is to scale narrative measurement from a single newspaper column to entire newspapers and from sentiment to issue-specific risk. Baker et al. (2016) construct an Economic Policy Uncertainty (EPU) index from frequency counts of policy-uncertainty-related terms across 10 leading U.S. newspapers. The EPU index spikes around fiscal-cliff debates, debt-ceiling standoffs, and elections. The index forecasts firm-level investment declines in policy-sensitive sectors and aggregate output and employment slowdowns. This

result establishes that aggregate news-text indices can be both economically meaningful and macroeconomically predictive. The EPU template has since been replicated for over 20 countries and forms the canonical aggregate news-based index in the literature.

Caldara and Iacoviello (2022) construct a sibling index, the Geopolitical Risk (GPR) index, by counting newspaper mentions of geopolitical events such as wars, terror attacks, and diplomatic tensions across 10 newspapers. The GPR index predicts declines in investment, employment, and asset prices following geopolitical-risk shocks, with effects concentrated in firms with foreign exposure. EPU and GPR together demonstrate that a wide class of economic and political narratives can be measured by mechanical text frequency, then embedded into vector autoregressions or asset-pricing tests as exogenous shock measures.

Calomiris and Mamaysky (2019) extend the news-text-driven approach to international markets by analysing 50+ topics extracted from 7,000 Reuters articles per day across 51 countries. Topics that drive cross-country returns include macroeconomic news, monetary policy, government, and crises. The magnitudes are economically meaningful, with topic-shocks moving country-level returns by 50 to 100 basis points monthly in the worst-affected markets. Cross-country topic shocks load most strongly in markets with weaker investor protection and lower informational efficiency.

Engle et al. (2019) introduce a portfolio-construction lens. Their goal is not to measure narrative directly, but to hedge climate-related news risk using innovations in a climate-news index built from Wall Street Journal and Crimson Hexagon text. They construct climate-hedge portfolios that load on stocks whose returns covary with climate-news innovations, and show that these portfolios deliver positive Sharpe ratios over the post-2008 period. The methodology is asset-pricing-native (mimicking-portfolio construction against an exogenous text shock) and it converts a narrative measurement into a tradable hedge.

Bybee et al. (2023) bring topic-model factor pricing to a fully systematic conclusion. Their methodology applies Latent Dirichlet Allocation to 800,000 Wall Street Journal articles spanning 1984 through 2017 and extracts 180 topics. They construct mimicking portfolios for each topic and test which topic factors are priced in the cross-section. Their key finding is that a small set of news-narrative factors spans standard factor-zoo dimensions: the news-based factors absorb a substantial fraction of the alpha and explain return covariance that conventional factors miss. This closes the chapter on aggregate news-based indices by showing that narrative factors are not just predictors but priced systematic risks.

The throughline of this section is that news-based aggregate indices and narrative factors aggregate firm-level textual signals into priced risk premia. The next section turns to firm-level text-based exposures, the inputs that aggregate indices implicitly aggregate over.

3 Text-Based Firm-Level Exposures

This section covers Text-Based Firm-Level Exposures, the literature that builds firm-level measurements of textual content from 10-K filings, earnings-call transcripts, and analyst reports. Four body-cited papers anchor the methodology stack in chronological order: Hoberg and Phillips (2010), Hoberg and Phillips (2016), Hassan et al. (2019), and Sautner et al. (2023). The progression runs from cosine-similarity TNICs through bigram counts to embedding-based semantic exposures.

Hoberg and Phillips (2010) provide the foundational text-based methodology paper in this theme. The empirical setup analyses product descriptions in the business sections of 10-K filings. The authors compute pairwise cosine similarity between firms’ product descriptions and use these similarities to build a text-based product-market network. Their application is to mergers and acquisitions: bidder-target product-market overlap (measured by text similarity) predicts both deal completion probability and post-merger operating performance. The methodological insight is that firms describe themselves in their disclosures, and machine-extractable similarity in those self-descriptions reveals competitive structure that SIC codes miss.

The natural extension is to formalise text-based industries as an alternative to SIC. Hoberg and Phillips (2016) construct Text-based Network Industries (TNICs) by clustering firms based on the same cosine-similarity scores of 10-K product descriptions. TNICs are firm-specific and time-varying: each firm’s industry peers depend on its own product description in each year, so industry boundaries can shift continuously. The application is endogenous product differentiation. Firms in TNICs that face high text-based competition exhibit higher markups and lower R&D intensity, a finding that is not visible when industries are forced into static SIC categories. The TNIC dataset has become a canonical input to corporate-finance and asset-pricing tests where industry definitions matter.

Hassan et al. (2019) introduce a different text-mining methodology: bigram counting of political risk language in earnings-call transcripts. The authors construct a firm-level political-risk measure (PRisk). PRisk counts the share of bigrams in transcripts that combine political topics (such as “election” or “regulation”) with risk-synonymous words (such as “concern” or “uncertain”). PRisk varies over time within firms, increases around elections and policy events, and is concentrated in firms whose political exposures the qualitative narrative confirms. Higher PRisk predicts lower investment, lower hiring, and lower returns, and the firm-level measure aggregates to a sensible economy-wide index. The methodological contribution is that bigram-counting against carefully curated topic and risk vocabularies extracts firm-level exposures that simple sentiment-counting misses.

Sautner et al. (2023) take the firm-level-exposure literature into the embedding era with a climate-change exposure measure built by an algorithm scanning earnings-call transcripts. The methodology is a keyword-and-embedding hybrid. Starting from a small seed list of climate-related terms, the algorithm extends the vocabulary through word2vec-style embedding similarity and then counts term-frequencies in transcripts to score firms. The resulting climate-exposure measure is firm-time-specific and identifies physical, regulatory, and opportunity exposures separately. Higher climate exposure correlates with higher green-patenting activity, more climate-related capex, and (modestly) lower returns when climate news is salient. The shift from cosine-similarity (Hoberg) to bigram counts (Hassan) to seeded embeddings (Sautner) traces the methodology stack of this theme.

These firm-level exposures are the bridge between aggregate narrative indices (Section 2) and the text-derived factor models that revisit them in Section 5. Aggregate news-based indices implicitly average firm-level exposures across the cross-section; conversely, firm-level exposures can be aggregated to construct sector-wide or country-wide narrative measures. The two scales answer different questions, but they share methodological foundations.

4 Risk-Disclosure Information Content and Sentiment Baselines

This section pairs two SLR-run themes that are methodologically adjacent. They are Risk-Disclosure Information Content (papers studying mandatory Item 1A and 10-K risk-factor disclosures) and Sentiment & Dictionary Baselines (older lexical methods that newer information-content tests must beat). Four body-cited papers anchor the discussion in chronological order. The four are Tetlock (2007) for the dictionary lineage and Loughran and McDonald (2011) for the canonical financial dictionary. Campbell et al. (2013) provides the first systematic test of mandatory risk-factor disclosures, and Hope et al. (2016) adds the specificity-versus-boilerplate dimension. Pairing the two themes makes sense because Item 1A disclosure work directly inherits the dictionary methodology and benchmarks itself against the older sentiment-counting baselines.

Tetlock (2007) (revisited briefly here on the methodological dimension) used Harvard-IV-4 General Inquirer categories, a general-purpose dictionary built for psychological text analysis, to count tonal words in newspaper text. The General Inquirer’s “negative” category aggregates many words that are not negative in financial contexts (“liability”, “tax”, “cost”), which limited the precision of the resulting sentiment measure. This dictionary-mismatch problem motivated the next step in the literature.

Loughran and McDonald (2011) built the canonical financial dictionary, now universally referred to as the Loughran-McDonald (LM) dictionary, by manually classifying every word that appears in 10-K filings between 1994 and 2008. The result is a six-category dictionary (negative, positive, uncertainty, litigious, strong-modal, weak-modal) tuned to financial narrative. The empirical comparison shows that the LM negative category outperforms the Harvard-IV-4 negative category at predicting 10-K filing-day abnormal returns and abnormal trading volume. The LM measure also predicts post-filing volatility, fraud, material weakness, and unexpected earnings. The authors emphasise that the gain comes from two adjustments. The first removes finance-specific false positives (“liability”, “vice”, “mine”) from the negative category, and the second adds finance-specific terms (“adverse”, “deficit”) that the general dictionary missed. The LM dictionary is the comparator that any newer information-content test must beat.

Campbell et al. (2013) provide the first large-scale empirical test of the information content of mandatory Item 1A risk-factor disclosures, which the SEC began requiring in 10-K filings in 2005. The authors hand-classify risk-factor disclosures into 25 risk categories. They document that firms disclose more risk in categories that match their objective economic exposures: technology firms disclose technology risk, financial firms disclose financial-system risk, regulated firms disclose regulatory risk. Risk-factor disclosure quantity (word counts and risk-type-specific counts) predicts post-filing realised volatility, beta, and equity-cost-of-capital changes. This predictive content demonstrates that Item 1A disclosures contain incremental information beyond what other 10-K sections provide. The information-content claim sets the foundation for subsequent specificity tests.

Hope et al. (2016) sharpen the information-content claim by distinguishing specific from boilerplate risk-factor disclosures. The authors construct a specificity score using machine-extracted named-entity counts (more named entities indicate firm-specific text, while fewer

indicate boilerplate language). Specific disclosures predict analyst-forecast accuracy, lower analyst-forecast dispersion, and stronger market reactions to filings, while boilerplate disclosures are largely uninformative. This finding establishes that the form of risk-factor disclosure matters, not just the quantity, and motivates much of the subsequent LLM-based work that aims to extract specific severity-graded risks rather than relying on raw word counts.

The two themes tie together cleanly. Item 1A risk-disclosure work defines its incremental value over older lexical baselines, and the LM dictionary is the comparator that newer methods must beat. Section 5 picks up this thread by surveying the LLM-based extraction methods that aim to surpass dictionary baselines on both information content and specificity.

5 LLM-Based Disclosure & Document Extraction

This section covers the LLM-Based Disclosure & Document Extraction theme, the most recent and fastest-growing branch of the corpus. The theme contains 7 corpus papers, of which 2 are body-cited briefly and 5 appear only in the corpus-at-a-glance table. The body-cited revisits are Bybee et al. (2023) as the topic-model-to-LLM bridge and Taibi and Osterrieder (2026) as the article’s own methodological frame, both already discussed in earlier sections and recapped here for the LLM-specific dimension.

Bybee et al. (2023), revisited briefly, sits at the methodological transition point. Latent Dirichlet Allocation produces topic distributions over text, and topic distributions are precursors to the contextual embeddings that modern transformer-based language models produce. The Bybee, Kelly, and Su pipeline can be re-implemented with transformer embeddings rather than LDA topic loadings, and an emerging literature is doing exactly that. The methodological lesson is that the asset-pricing test stays the same (mimicking portfolios for narrative dimensions, then spanning tests against standard factors), while the upstream text-feature extractor improves.

Taibi and Osterrieder (2026) is the methodological framework paper for this article’s underlying corpus. The Algorithmic Framework for Systematic Literature Reviews adapts PRISMA-style review pipelines to financial narratives, with an emphasis on reproducibility, snowball discovery, and machine-assisted relevance scoring. The pipeline’s outputs (a frozen run directory containing query results, scored candidates, themes, and a final corpus) make the corpus selection auditable and replayable, which is the property this article relies on for its curation rule. Beyond producing the corpus, the framework is itself an LLM-augmented research workflow: large language models score relevance, extract themes, and propose seeds, all under human-checkable prefilters and quality gates. This makes the framework both a tool for the article and a member of the LLM-extraction theme it surveys.

The remaining LLM-theme corpus papers appear only in the corpus-at-a-glance table. These are Kelly, Manela, and Moreira (2021); Lu, Wang, and Zhang (2023); Hossain and Hossain (2025); Doey and de Jong (2025); Hayrapetyan and Gevorgyan (2025); and Sahu and Debata (2026). Together they span text-selection methodology, China-specific topic-driven prosperity narratives, and 10-K narrative diversification. They also cover earnings-call tone propagation to media, narrative econometrics in equity markets, and managerial sentiment effects on liquidity. Together they demonstrate that the LLM frontier is now broad enough to require its own systematic survey, which is partly the motivation for the SLR pipeline

that produced this corpus.

A central methodological hazard in the LLM frontier requires direct attention. Look-ahead bias is the dominant validity threat for LLM-extracted features applied retrospectively to historical filings. Modern LLMs are trained on data that postdates the filings they are asked to analyze (Sarkar and Vafa, 2025, arXiv:2502.21206). Notably, this benchmark paper was a seed in our systematic search but did not survive the prefilter and scoring funnel into the final corpus. That outcome is itself a methodological observation. The SLR’s relevance criteria privilege asset-pricing applications over benchmark-construction methodology, so look-ahead-bias work appears as a critique-axis here rather than as a corpus member. The Sarkar-Vafa benchmark formalises the concern by constructing a point-in-time evaluation suite. The suite holds out post-cutoff information and measures how much of an LLM’s apparent forecasting power evaporates once that information is masked. The implication for any LLM-based asset-pricing study is that knowledge-cutoff annotations and point-in-time replay are no longer optional: a paper that omits them cannot distinguish genuine textual signal from training-data leakage.

Mitigation strategies are emerging in the literature but remain ad hoc. Some studies restrict the analysis to LLMs whose training cutoff predates the earliest filing in the sample, which sacrifices model quality for temporal cleanliness. Others use embedding-only pipelines, where the LLM produces a static embedding for each document. The asset-pricing test then runs against the embeddings rather than asking the LLM to forecast directly, which limits but does not eliminate leakage. A third approach is to fine-tune small open-weight models on point-in-time data, which gives the researcher full control over the training cutoff at the cost of substantially weaker base capability. None of these is dominant, and the field is still converging on a standard practice.

The reproducibility question is the natural complement to the look-ahead-bias question. The SLR pipeline that produced this corpus illustrates one answer: every run is a frozen directory tree with the query, snowball, prefilter, scoring, and theme-classification artefacts archived together (Taibi and Osterrieder, 2026). The same principle applies to LLM-based asset-pricing studies: every prompt, every retrieved context, every model version, and every cutoff annotation should be archived alongside the eventual portfolio backtest. The reproducibility ethos and the look-ahead-bias mitigation overlap because both reduce to the same operational discipline of point-in-time provenance.

6 Conclusion

The thematic arc traced in this article runs from dictionary methods through firm-level exposures and aggregate news-based factors to large-language-model extraction. Each generation of methodology surfaced narrative dimensions that the previous generation could not see, and each is benchmarked against the older lexical baselines that came before it. The 23-paper corpus surveyed here covers all five themes that the underlying systematic literature review identified, and the corpus-at-a-glance table places every paper in its theme and tier.

Several open questions persist. Look-ahead bias is the dominant validity threat for any LLM-based study that uses modern off-the-shelf models on historical filings, and the field has not converged on a standard mitigation. Cross-market replication is uneven: U.S. evi-

Table 1: Corpus at a glance: 23 papers from the 2026-05-02 systematic literature review across 5 themes.

Author-Year	Theme	Tier	Year
Tetlock2007	Sentiment & Dictionary Baselines	Q1	2007
Hoberg2010	Text-Based Firm-Level Exposures	Q1	2010
Loughran2011	Sentiment & Dictionary Baselines	Q1	2011
Campbell2013	Risk-Disclosure Information Content	Q1	2013
Baker2016	Narrative Asset Pricing & News-Based Factors	Q1	2016
Hoberg2016	Text-Based Firm-Level Exposures	Q1	2016
Hope2016	Risk-Disclosure Information Content	Q1	2016
Calomiris2019	Narrative Asset Pricing & News-Based Factors	Q1	2019
Engle2019	Narrative Asset Pricing & News-Based Factors	Q1	2019
Hassan2019	Text-Based Firm-Level Exposures	Q1	2019
Kelly2021	LLM-Based Disclosure & Document Extraction	Q1	2021
Caldara2022	Narrative Asset Pricing & News-Based Factors	Q1	2022
Bybee2023	Narrative Asset Pricing & News-Based Factors	Q1	2023
Lu2023	LLM-Based Disclosure & Document Extraction	Q1	2023
Sautner2023	Text-Based Firm-Level Exposures	Q1	2023
Doey2025	LLM-Based Disclosure & Document Extraction	Q2	2025
Hayrapetyan2025	LLM-Based Disclosure & Document Extraction	untiered	2025
Hirshleifer2025	Narrative Asset Pricing & News-Based Factors	Q2	2025
Hossain2025	LLM-Based Disclosure & Document Extraction	Q1	2025
Li2025	Narrative Asset Pricing & News-Based Factors	Q1	2025
Han2026	Narrative Asset Pricing & News-Based Factors	Q1	2026
Sahu2026	LLM-Based Disclosure & Document Extraction	Q2	2026
Taibi2026	LLM-Based Disclosure & Document Extraction	untiered	2026

dence on news-based indices and 10-K disclosures is rich, but international replications are concentrated in a few countries and rarely test the same factor structure. Factor pricing of narrative betas is still being established, with topic-model factors competing against and partly subsuming the conventional factor zoo. Spanning tests have not yet reached the saturation that the equity-anomaly literature has achieved.

The corpus itself flags a curation question. The SLR’s relevance and quality criteria privilege asset-pricing applications, which means methodological-only papers (such as look-ahead-bias benchmarks) appear as cross-references in this article rather than as corpus members. A future SLR pass with explicit methodological-only inclusion criteria would close that gap, at the cost of a larger and noisier corpus.

This article does not propose new empirical work. The intent is structural: to provide a reader with a thematic map of what the systematic search surfaced and a reproducible record of how the body of literature was selected. The companion run directory contains every artefact (queries, scored candidates, theme classifications, exclusion logs) and the reproducibility infrastructure that the corpus rests on (Taibi and Osterrieder, 2026).

References

- Scott Baker, Nicholas Bloom, and Steven J. Davis. Measuring economic policy uncertainty*. *The Quarterly Journal of Economics*, 131(4):1593–1636, 2016. doi: <https://doi.org/10.1093/qje/qjw024>. URL <https://academic.oup.com/qje/article-pdf/131/4/1593/30636768/qjw024.pdf>.
- Leland Bybee, Bryan Kelly, and Yinan Su. Narrative asset pricing: Interpretable systematic risk factors from news text. *Review of Financial Studies*, 36(12):4759–4787, 2023. doi: <https://doi.org/10.1093/rfs/hhad042>.
- Dario Caldara and Matteo Iacoviello. Measuring geopolitical risk. *American Economic Review*, 112(4):1194–1225, 2022. doi: <https://doi.org/10.1257/aer.20191823>.
- Charles W. Calomiris and Harry Mamaysky. How news and its context drive risk and returns around the world. *Journal of Financial Economics*, 133(2):299–336, 2019. doi: <https://doi.org/10.1016/j.jfineco.2018.11.009>.
- John L. Campbell, Hsinchun Chen, Dan S. Dhaliwal, Hsin-Min Lu, and Logan B. Steele. The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19(1):396–455, 2013. doi: <https://doi.org/10.1007/s11142-013-9258-3>.
- Brandon Doey and Pieter de Jong. How negative tones in earnings calls shape media narratives. *Review of Behavioral Finance*, 17(3):406–423, 2025. doi: <https://doi.org/10.1108/rbf-08-2024-0228>.
- Robert F. Engle, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel. Hedging climate change news. *Review of Financial Studies*, 33(3):1184–1216, 2019. doi: <https://doi.org/10.1093/rfs/hhz072>.
- Qian Han, Pei-Lin Hsieh, and Lu Zhang. Commodity futures report text sentiment and returns. *Journal of Futures Markets*, 2026. doi: <https://doi.org/10.1002/fut.70090>.
- Tarek A. Hassan, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun. Firm-level political risk: Measurement and effects*. *The Quarterly Journal of Economics*, 134(4): 2135–2202, 2019. doi: <https://doi.org/10.1093/qje/qjz021>. URL https://lbsresearch.london.edu/id/eprint/1190/7/qjz021_online_appendix.pdf.
- Davit Hayrapetyan and Ruben Gevorgyan. From headlines to forecasts: Narrative econometrics in equity markets. *Journal of risk and financial management*, 18(9):524, 2025. doi: <https://doi.org/10.3390/jrfm18090524>. URL <https://doi.org/10.3390/jrfm18090524>.
- David Hirshleifer, Dat Mai, and Kuntara Pukthuanthong. War discourse and the cross section of expected stock returns. *The Journal of Finance*, 80(6):3589–3637, 2025. doi: <https://doi.org/10.1111/jofi.13482>.
- Gerard Hoberg and Gordon M. Phillips. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies*, 23(10): 3773–3811, 2010. doi: <https://doi.org/10.1093/rfs/hhq053>.

- Gerard Hoberg and Gordon M. Phillips. Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5):1423–1465, 2016. doi: <https://doi.org/10.1086/688176>.
- Ole-Kristian Hope, Danqi Hu, and Hai Lu. The benefits of specific risk-factor disclosures. *Review of Accounting Studies*, 21(4):1005–1045, 2016. doi: <https://doi.org/10.1007/s11142-016-9371-1>. URL <https://utoronto.scholaris.ca/bitstreams/cd31ffc4-af90-474d-b4ab-2b0a69f4d70e/download>.
- Md Enayet Hossain and K. S. M. Tozammel Hossain. Corporate diversification and 10-k narratives: A novel approach using large language models. *Borsa Istanbul Review*, 25(6):1391–1402, 2025. doi: <https://doi.org/10.1016/j.bir.2025.08.002>. URL <https://doi.org/10.1016/j.bir.2025.08.002>.
- Bryan Kelly, Asaf Manela, and Alan Moreira. Text selection. *Journal of Business and Economic Statistics*, 39(4):859–879, 2021. doi: <https://doi.org/10.1080/07350015.2021.1947843>.
- Xiao Li, Changhong Jin, Y. Niu, and Ruihai Dong. Two sides of the same coin: How llms reveal dual narratives in annual reports. *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 308–316, 2025. doi: <https://doi.org/10.1145/3768292.3770435>. URL <https://doi.org/10.1145/3768292.3770435>.
- Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011. doi: <https://doi.org/10.1111/j.1540-6261.2010.01625.x>. URL https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1331573_code16609.pdf?abstractid=1331573&mirid=1.
- Wanbo Lu, Yifu Wang, and Xingjian Zhang. Which news topics drive economic prosperity in china? *PLoS ONE*, 18(10):e0291862, 2023. doi: <https://doi.org/10.1371/journal.pone.0291862>. URL <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0291862&type=printable>.
- Asis Kumar Sahu and Byomakesh Debata. Optimism unleashed: managerial sentiment and stock liquidity. *Management Decision*, pages 1–24, 2026. doi: <https://doi.org/10.1108/md-05-2025-1161>.
- Suproteem K. Sarkar and Keyon Vafa. Look-ahead bench: A standardized benchmark of look-ahead bias in point-in-time llms for finance, 2025. URL <https://arxiv.org/abs/2502.21206>. External methodological cross-reference; this paper was a seed in the underlying SLR run but did not survive the prefilter+scoring funnel into the final 23-paper corpus.
- Zacharias Sautner, Laurence van Lent, Grigory Vilkov, and RUI SHEN ZHANG. Firm-level climate change exposure. *The Journal of Finance*, 78(3):1449–1498, 2023. doi: <https://doi.org/10.1111/jofi.13219>. URL <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/jofi.13219>.

Gabin Taibi and Joerg Osterrieder. An algorithmic framework for systematic literature reviews: A case study for financial narratives, 2026. URL <https://doi.org/10.48550/arxiv.2601.03794>.

Paul C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007. doi: <https://doi.org/10.1111/j.1540-6261.2007.01232.x>.