Working Paper: What Matters, When

A Cognitively Structured Model of Investor Response to News

Anirudh Krovi

PhD, Northwestern University · MBA, NYU Stern · Formerly at McKinsey & Company

anirudh.krovi@stern.nyu.edu

July 11, 2025

Abstract

This paper introduces a cognitively motivated, theme-aware framework for extracting return-predictive signals from financial news. While transformer-based models like FinBERT offer general-purpose sentiment classification, they often underperform in settings where market-relevant information is implicit, structurally complex, or behaviorally delayed. We design a structured scoring system that classifies headlines by economic theme and evaluates them along four interpretable axes: actor clarity, lever relevance, magnitude, and novelty. Applied to Apple Inc. (AAPL) news from 2019–2024, our model outperforms FinBERT in explaining T+1 through T+4 returns—especially in themes such as chip supply and competitive threats. These findings suggest that return prediction from news requires more than polarity detection: it requires models that reflect how investors interpret narrative structure and act across time. The framework contributes a transparent and extensible tool for structured financial NLP, with implications for behavioral signal modeling and real-time market analysis.

1. Introduction

Financial markets react not just to numbers, but to narratives. Headlines shape expectations, and expectations move prices. As a result, understanding how news influences investor behavior is increasingly important in asset pricing—especially in an era of real-time information flow and algorithmic execution.

Recent advances in machine learning and natural language processing (NLP) have opened new frontiers in extracting predictive signals from unstructured data [Gu et al., 2020, Kelly et al., 2024]. These models—ranging from elastic nets to deep neural networks and transformer-based architectures—have demonstrated strong empirical performance across high-dimensional financial datasets. Early work by Tetlock [2007] highlighted the connection between textual pessimism in media columns and market returns. Later research [Boudoukh et al., 2012, Larsen and Thorsrud, 2019, Engle et al., 2020, Bybee et al., 2024] has explored how news volume and sentiment across macroeconomic topics influence returns and broader economic indicators such as production and employment growth.

A shared thread across many of these studies is the emphasis on *sentiment prevalence*—that is, the frequency and tone (positive/negative) of news coverage—rather than the *granular informational value* of individual headlines. However, in financial contexts, not all news is equal. For instance, an upbeat article on a supplier of iPhone components may not affect Apple Inc.'s (AAPL) stock price as strongly as a report on its quarterly sales. Similarly, two positive (negative) headlines can differ substantially in intensity, relevance, and novelty. These differences may matter—but are often obscured in traditional sentiment analyses.

One widely used tool in this space is **FinBERT** [Araci, 2019], a transformer model fine-tuned on financial text. FinBERT excels at classifying firm-level headlines with clear sentiment, such as earnings surprises or product launches, making it a popular baseline for short-term return modeling. It assigns probabilities of positivity or negativity, offering a quick and standardized way to extract polarity from financial text.

However, many economically meaningful headlines are subtle in tone and require contextual inference. News involving supply chain stress, regulatory ambiguity, or competitive threats often lacks overt sentiment cues but holds significant predictive value. In such cases, FinBERT tends to misclassify or underweight the signal, leading to false positives (e.g., corporate puff pieces) and false negatives (e.g., neutral-sounding headlines about strategic risks). Moreover, FinBERT does not distinguish between levels of *relevance*—it treats distant supply chain updates and core firm disclosures with similar weight.

These limitations motivate our central question:

> *Can a structured, domain-aware scoring system outperform generic sentiment models like FinBERT—especially in identifying return-predictive news where meaning is implicit and market reactions are delayed?*

We propose a cognitively motivated framework that mirrors how human investors parse news.

Each article is tagged with a theme and scored along four dimensions: *Actor Clarity*, *Lever Relevance*, *Lever Magnitude*, and *Novelty*. These dimensions reflect intuitive features such as proximity to the firm, salience of the trigger, and behavioral response likelihood—factors often invisible to black-box NLP systems. The resulting compound score offers an interpretable, modular view of how impactful a news article is likely to be.

To evaluate our framework, we apply it to **Apple Inc. (AAPL)**—a stock with high liquidity, broad media coverage, and heterogeneous news content. We compare our approach against FinBERT, analyzing their respective performance across multiple economic themes and time horizons. Empirically, we find that while FinBERT performs well on sentiment-heavy news with immediate implications (T+0, T+1), its performance weakens on more complex themes such as supply chain dynamics and macro spillovers. In contrast, our model captures both short-term and delayed effects (T+1 to T+4), particularly in areas where behavioral inference is critical.

Beyond performance, our method integrates smoothly into quantitative pipelines. By producing time-sensitive, theme-tagged article scores, it enables construction of interpretable factors for return forecasting, signal stacking, or macro-aware portfolio overlays. Its modular structure supports cross-sectional aggregation, orthogonalization, and hybrid use with econometric and machine learning tools. Most importantly, it appears to capture *delayed market reactions* that generic sentiment models miss—offering potential alpha in strategies reliant on fundamental or event-driven signals. While our empirical focus is AAPL, the framework generalizes easily across firms, sectors, and asset classes.

The remainder of the paper is organized as follows. Section 2 introduces the FinBERT baseline and its empirical limitations. Section 3 outlines our dictionary-based framework and its implementation. Section 4 offers an interpretation of the model's success through behavioral and structural lenses. Section 5 concludes with takeaways and future directions.

2. FinBERT as Baseline: Performance and Limitations

FinBERT [Araci, 2019], a transformer-based model trained on financial texts, has become a widely used tool for sentiment classification in financial NLP. It assigns positive, negative, and neutral scores at the sentence level and is often used in asset pricing as a plug-and-play sentiment engine. In this section, we evaluate FinBERT's performance on Apple Inc. (AAPL) headlines and discuss its strengths and limitations.

2.1. Data Collection and Preprocessing

We collected over 350 news headlines related to Apple Inc. between 2019 and 2024 using the SERP API, which aggregates search engine results from multiple sources. Given the high prevalence of duplicative, promotional, or irrelevant content, we performed manual curation to filter out headlines with little to no economic or strategic content. We retained only those articles that plausibly convey information an investor might react to.

To improve the performance of our scoring model, we also cleaned and simplified the headlines to reduce literary subtlety, wordplay, and non-informational stylistics. For example, phrases such as *"Apple takes a bite out of supply woes"* were rewritten into more sentiment-stable forms like *"Apple reports supply improvement"*. This step ensured consistency and minimized semantic confusion in the scoring pipeline, especially in magnitude and sentiment recognition.

2.2. Methodology

We use the same cleaned and curated set of AAPL-related headlines described above. Each headline is passed through FinBERT to extract a sentiment score, computed as the difference between the model's positive and negative probabilities:

$$\text{FinBERT Score} = P(\text{Positive}) - P(\text{Negative}) \tag{1}$$

Each headline is also manually tagged with one or more themes (e.g., *Company*, *Chip Supply*), enabling us to disaggregate results by theme. For each trading day, we compute the average FinBERT score across all relevant headlines. We then regress AAPL daily returns on these average scores with lags of 0, 1, 2, 3, and 4 trading days. This setup allows us to assess the predictive power of FinBERT-sourced sentiment over time and across economic themes.

2.3. Results

FinBERT performs well on news that is directly about Apple and emotionally clear in tone. In particular, themes such as *Company*—covering earnings, product launches, and executive statements—exhibit statistically significant predictive power at T+1, with a coefficient of -0.0103 ($p = 0.032$, $R^2 = 0.177$), suggesting that negative sentiment in firm-level news corresponds to next-day underperformance. However, this relationship weakens quickly: the T+2 coefficient becomes insignificant and positive, and although the T+3 and T+4 lags show modest signal ($p = 0.24$ and $p = 0.13$, respectively), the effect is not statistically robust.

By contrast, FinBERT shows issues in extracting meaningful signal from more nuanced themes. For instance, in the *Chip Supply* category, T+1 and T+2 coefficients are small and statistically insignificant, with p-values of 0.20 and 0.42, respectively. Even at longer lags (T+3 and T+4), while the coefficients remain directionally negative, they do not cross typical significance thresholds ($p = 0.13$ at T+3; $R^2 = 0.07$). Themes like *Competition* similarly fail to show consistent predictive power across lags, with low R^2 values throughout.These results are summarized in Tables 1 and 2.

To provide a complete picture, we also evaluate FinBERT's explanatory power at the same-day (T+0) horizon (Table 3). Results suggest limited immediate signal across themes, indicating that FinBERT's polarity-based model may not capture real-time return impact in many contexts.

Table 1: FinBERT Regression Results: T+1 and T+2 Returns

Theme	Horizon	n	Coef.	t-stat	p-value	R^2
Analyst	1d	31	-0.0047	-0.99	0.331	0.033
Analyst	2d	31	-0.0025	-0.71	0.485	0.017
Chip Supply	1d	42	-0.0055	-1.32	0.196	0.041
Chip Supply	2d	42	-0.0035	-0.82	0.417	0.017
Company	1d	26	**-0.0103**	**-2.27**	**0.032**	**0.177**
Company	2d	26	0.0027	0.39	0.699	0.006
Competition	1d	166	0.0018	0.91	0.362	0.005
Competition	2d	166	0.0039	1.77	0.079	0.019

Table 2: FinBERT Regression Results: T+3 and T+4 Returns

Theme	Horizon	n	Coef.	t-stat	p-value	R^2
Analyst	3d	36	-0.0023	-0.67	0.510	0.013
Analyst	4d	36	-0.0004	-0.08	0.933	0.000
Chip Supply	3d	35	-0.0072	-1.57	0.125	0.070
Chip Supply	4d	35	-0.0019	-0.49	0.631	0.007
Company	3d	27	0.0077	1.20	0.242	0.054
Company	4d	27	0.0101	1.58	0.126	0.091
Competition	3d	166	0.0028	1.43	0.154	0.012
Competition	4d	166	-0.0003	-0.13	0.893	0.000

2.4. Discussion

FinBERT's primary limitation lies in its architecture and training objective. It is trained to detect sentence-level polarity with limited consideration of the economic structure or behavioral context of the information. As a result, it misses signals when headlines:

1. Contain emotionally neutral language but economically significant content
2. Refer to entities outside the focal firm (e.g., suppliers, competitors, regulators)
3. Require time-delayed inference rather than immediate reaction

While FinBERT captures short-term sentiment shocks in direct company news, its low explanatory power in upstream and macro themes—even at longer horizons—illustrates the difficulty of extracting structured signal from generic NLP tools. These gaps underscore the need for a theme-aware [1], interpretable framework that aligns more closely with investor behavior and firm relevance.

In the next section, we introduce our cognitively motivated scoring system, which aims to address precisely these limitations by embedding economic structure and behavioral salience into the scoring process.

[1]By theme aware, we mean the capability to have news tagged by standard themes (e.g., chip supply challenges, competitive pressures) for a technology company such as Apple Inc

Table 3: FinBERT Regression Results: T+0 Returns

Theme	n	Coef.	t-stat	p-value	R^2
Analyst	37	-0.0029	-0.78	0.439	0.017
Company	27	0.0026	0.50	0.623	0.010
Competition	178	0.0017	0.86	0.389	0.004
Chip Supply	47	0.0013	0.24	0.809	0.001

3. A Cognitive Framework for News Scoring

Our methodology consists of four components: (1) data collection and cleaning; (2) a domain-specific dictionary for categorizing headlines into interpretable economic themes; (3) a cognitively motivated scoring model based on four investor-relevant axes; and (4) a regression framework to evaluate the model's predictive power with respect to Apple Inc. (AAPL) returns.

3.1. Dictionary Construction

We develop a theme-specific dictionary that maps financial terms and phrases to economically interpretable categories. Each headline is tagged with one or more themes such as *Chip Supply*, *Analyst*, *Company*, and *Competition*. These themes are chosen to reflect recurring topics that investors consider material to firm valuation. For example, *Chip Supply* captures bottlenecks in the semiconductor value chain that affect Apple's production cycle, while *Analyst* encompasses broker research and institutional forecasts from firms such as Morgan Stanley or Credit Suisse.

Each headline is then scored along four cognitive dimensions, designed to reflect how investors evaluate financial news:

- **Actor Clarity:** Identifies who is acting and the salience of their role. Statements from Apple executives are weighted more than comments from third parties. Competitor actions receive negative weights by design, reflecting likely negative valuation impact.
- **Lever Relevance:** Indicates which business lever is affected—e.g., demand drivers, pricing power, product innovation. Lever classification is based on keyword matching and business context.
- **Lever Magnitude:** Captures the implied intensity of the news. Words like *"soar," "plummet,"* score higher[2] than softer terms like *"grew"* or *"decreased."*
- **Novelty:** Measures how new or surprising the information is. Rather than using similarity metrics, we rely on novelty cues (e.g., *"unexpected," "first time," "unprecedented"*) that often trigger repricing.

Actor scores range from 0.5 to 1 in increments of 0.25, based on structured business intuition. Novelty scores range from 0.5 to 2 in 0.25 increments. Lever and Magnitude scores range from 0

[2] In terms of absolute value; sign (positive or negative) may reverse final directional impact.

to 1. These weights are not optimized or machine-learned in the current version, but are selected to reflect plausible economic salience. We view them as tunable parameters for future iterations. If a given dimension has no article support, default values are used (actor = 1, novelty = 0.5, lever = magnitude = 0.25).

Signs are flipped where appropriate. For example, a headline like *"Samsung launches rival smartphone ahead of Apple event"* receives a high actor score (prominent firm), a high lever relevance score (product strategy impact), strong magnitude (event timing), and high novelty. However, since the actor is a competitor, the final contribution is negatively signed. These sign adjustments are embedded into the dictionaries themselves.

To improve linguistic robustness, we algorithmically expand each dictionary entry using natural language processing tools that account for:

- Tense variations (e.g., "increase" vs. "increased"),
- Singular/plural forms (e.g., "deal" vs. "deals"),
- Synonym families (e.g., "surge," "spike," "jump").

This expansion ensures broader coverage and consistency in matching across diverse headline phrasings, while preserving semantic intent. The final dictionaries thus combine curated domain expertise with NLP-based augmentation.

3.2. Scoring Model

We compute the final article-level score as the product of the four component scores:

$$\text{Final Score} = \text{Actor} \times \text{Lever} \times \text{Magnitude} \times \text{Novelty} \tag{2}$$

Each dimension is computed as follows:

- **Actor:** Product of all actor-tagged keyword scores. Competitors or negative influences apply a negative multiplier.
- **Lever:** Computed as a softmax over matched levers within the headline, allowing the dominant lever to influence the score while preserving secondary effects.
- **Magnitude:** Computed as a softmax over intensity descriptors for a given headline, amplifying headlines with stronger language.
- **Novelty:** Assigned as the maximum novelty score among the headline tokens that match our novelty dictionary. Each matching token contributes a score, and the highest among them is taken to reflect the most surprising element in the headline.

This structure ensures that news which is actor-relevant, lever-salient, high in magnitude, and novel receives a higher score—while vague, repetitive, or low-impact headlines yield lower weights.

3.3. Regression Setup

To evaluate the predictive power of our model, we run daily return regressions on AAPL using the computed compound scores.

We obtain adjusted daily close price data for AAPL from Yahoo Finance, covering the period January 1, 2019 to January 1, 2024. Returns are calculated as the log difference in adjusted closing prices. Each trading day is associated with the average compound score of all news headlines published that day. If no relevant headlines exist, the day is excluded.

Although asset pricing models often use excess returns, we use raw AAPL returns to avoid filtering out macro-sensitive information that may itself be influenced by firm-specific or supply-chain-related news. This choice reflects the fact that Apple frequently acts as a bellwether for broader economic trends.

We regress AAPL returns on average daily news scores with lags of T+0 through T+4, capturing both immediate and delayed market responses. We report regression coefficients, t-statistics, and adjusted R^2 to assess explanatory power across themes and time horizons.

Table 4: Regression Results: Our Model (T+0 Returns)

Theme	n	Coef.	t-stat	p-value	R^2
Competition	178	-0.0051	-0.98	0.330	0.005
Analyst	37	0.0011	0.20	0.841	0.001
Company	27	0.0012	0.10	0.922	0.000
Chip Supply	47	-0.0002	-0.02	0.984	0.000

Table 5: Regression Results: Our Model (T+1 and T+2 Returns)

Theme	Horizon	n	Coef.	t-stat	p-value	R^2
Analyst	1d	31	0.0086	0.75	0.459	0.019
Analyst	2d	31	-0.0098	-1.17	0.250	0.045
Chip Supply	1d	42	0.0082	1.32	0.194	0.042
Chip Supply	2d	42	0.0064	1.02	0.312	0.026
Company	1d	26	-0.0037	-0.34	0.739	0.005
Company	2d	26	-0.0068	-0.45	0.660	0.008
Competition	1d	166	0.0054	0.94	0.350	0.005
Competition	2d	166	0.0004	0.07	0.947	0.000

3.4. Discussion

The regression results reveal several notable features of our scoring model's performance across themes and time horizons.

At the same-day (T+0) horizon, our model shows little to no explanatory power across themes (Table 4) similar to FinBERT. Coefficients are small, statistically insignificant, and R^2 values

Table 6: Regression Results: Our Model (T+3 and T+4 Returns)

Theme	Horizon	n	Coef.	t-stat	p-value	R^2
Analyst	3d	36	-0.0019	-0.37	0.716	0.004
Analyst	4d	36	0.0037	0.51	0.613	0.008
Chip Supply	3d	35	0.0112	1.69	0.100	0.080
Chip Supply	4d	35	-0.0029	-0.49	0.627	0.007
Company	3d	27	-0.0009	-0.06	0.956	0.000
Company	4d	27	0.0147	0.89	0.383	0.031
Competition	3d	166	**-0.0115**	**-2.02**	**0.045**	**0.024**
Competition	4d	166	-0.0048	-0.84	0.402	0.004

are near zero across all categories—including *Company* and *Chip Supply*, which often feature in real-time sentiment-based strategies. This result aligns with the design of our framework, which prioritizes interpretability and structured inference over immediate polarity detection.

Rather than capturing instantaneous reactions, the model is better suited to detecting delayed or second-order price movements—where behavioral processing, thematic context, and economic salience require time to filter into prices.

While many coefficients are not statistically significant at conventional thresholds, we observe directionally intuitive and thematically consistent effects. For example, news tagged under *Chip Supply* shows positive coefficients at lags of T+1 through T+3, with the highest R^2 (0.08) and a p-value just above the 10% threshold at T+3. This aligns with the hypothesis that supply chain information—often complex and indirect—requires time for investors to process and price in.

In contrast, *Competition* shows a negative and statistically significant coefficient at T+3 ($p = 0.045$, $R^2 = 0.024$), suggesting that competitor-related news—especially when captured using cognitive cues—has a measurable delayed impact on AAPL returns. This may reflect strategic repositioning effects or perceived loss of relative advantage.

Company and *Analyst* themes show weaker performance in our model relative to FinBERT, particularly at short horizons. This is expected: these themes tend to involve clearly framed, high-sentiment headlines where FinBERT already performs well. Our model is designed to shine where FinBERT shows limitations —namely, in inference-heavy or indirect themes—and the results reflect that distinction.

Taken together, these findings validate the intuition that a cognitively structured scoring system may surface delayed or second-order pricing effects not visible to polarity-based models. They also suggest that future refinements—such as theme-specific weights, dynamic novelty calibration, or machine-learned scoring priors—could further strengthen signal quality in complex news domains.

4. Why Our Model Works: Behavioral and Structural Advantages

FinBERT serves as a useful benchmark for capturing immediate sentiment in clearly framed firm-level news. However, as shown in Section 2, its effectiveness diminishes when headlines require economic interpretation, involve indirect actors, or produce delayed reactions. Our cognitively structured framework is explicitly designed to address these limitations by modeling how investors parse, weigh, and respond to financial information across time and theme.

4.1. Behavioral Foundations

Investor reactions are not uniform—they depend on attention, interpretation, and perceived relevance. Headlines tied to earnings or product announcements tend to generate quick responses (T+0, T+1), while news involving supply chains, policy, or competitive dynamics may take longer to price in. Our model reflects this heterogeneity by integrating dimensions such as *Actor Clarity*, *Lever Relevance*, and *Novelty*—capturing not just the tone of the news, but its behavioral salience.

For example, a Tier-2 supplier disruption may not trigger immediate price movement but can influence institutional positioning over the next several days. By requiring alignment across actors, levers, and novelty, the model filters noise and prioritizes headlines likely to result in meaningful—but possibly delayed—market responses. This approach better mirrors the way real investors integrate complex information under bounded rationality.

4.2. Structural Advantages Over FinBERT

Our model introduces several design choices that provide structural advantages over FinBERT:

1. **Theme Awareness:** Each headline is tagged with an economic theme, allowing return attribution by topic. This contrasts with FinBERT's one-size-fits-all sentiment scoring, which flattens heterogeneous content into a single dimension.

2. **Interpretability and Decomposition:** Scores are computed via modular cognitive axes—actor, lever, magnitude, novelty—each of which reflects a component of investor reasoning. This transparency enables diagnostic use and domain-informed tuning, which is not feasible with FinBERT's embedding-based architecture.

3. **Lag Structure and Temporal Design:** Rather than assuming immediate price reaction, our model is tested explicitly across lags from T+0 to T+4. This allows themes with longer absorption cycles (e.g., *Chip Supply*, *Competition*) to emerge as predictive, a feature FinBERT does not model directly.

4.3. Empirical Support

As shown in Section 3, our model offers improved explanatory power over FinBERT in themes that require contextual or inferential reasoning. For instance, in the *Chip Supply* theme, FinBERT produces flat scores and weak returns correlation at all lags. Our model, by contrast, assigns scores based on the interaction of economically grounded features and exhibits rising predictive power at T+2 and T+3.

The *Competition* theme provides an additional point of contrast: while FinBERT shows negligible signal across all lags, our model produces a statistically significant coefficient at T+3 ($p = 0.045$), suggesting that structured signals tied to strategic positioning are priced with delay.

Importantly, these improvements are not simply statistical. Because the model is constructed from interpretable, theme-aware components, it enables real-time diagnostics and supports practical use in systematic workflows. Portfolio managers or analysts can inspect which themes are producing strong signals, decompose the sources of signal strength, and evaluate the nature of investor attention in response to current news flows.

4.4. Summary

In short, our model performs because it reflects how market participants interpret and respond to news. Rather than relying purely on surface-level sentiment, it encodes structure, salience, and behavioral lag—features that align with the realities of investor cognition. This makes it not just a scoring tool, but a framework for building explainable, return-predictive signals from unstructured financial information.

5. Conclusion

This paper introduces a cognitively motivated framework for extracting return-predictive signals from financial news. Unlike polarity-based models such as FinBERT, our approach incorporates domain structure and behavioral realism through four interpretable axes: *Actor Clarity*, *Lever Relevance*, *Magnitude*, and *Novelty*. Using a theme-tagged, article-level scoring system and a lag-aware regression setup, we show that structured signals can outperform black-box models—particularly in economically complex or indirect news domains.

Key empirical findings include improvements in explanatory power for themes such as *Chip Supply* and *Competition*, especially at longer horizons (T+2 to T+4), where FinBERT's short-term sentiment model shows limitations. These gains highlight that markets do not uniformly price news: investor responses depend on who is acting, what business lever is affected, how surprising the information is, and how long it takes to absorb. Our model's alignment with these behavioral patterns—combined with its decomposable structure—makes it both explanatory and actionable.

In addition to empirical performance, our framework offers a path for scaling. Table 7 summarizes how each scoring axis maps to specific linguistic forms and extraction strategies. These components are designed to generalize across firms and asset classes, enabling automation at scale while preserving interpretability.

Looking ahead, several extensions are worth pursuing. First, applying this framework to small-cap or thinly covered stocks may uncover stronger effects, given higher information asymmetry. Second, while our current score weights are rule-based, future versions could be learned using return-grounded loss functions—via architectures such as variational autoencoders (VAEs), energy-based models (EBMs), or reinforcement-based tuning. These architectures would allow us to move beyond fixed heuristics by learning latent structures in how actors, levers, and novelty interact

Table 7: Scaling Strategy for Signal Extraction

Component	**Linguistic Form**	**Extraction Strategy**	**Tools / Heuristics**
Actor	Proper nouns, named entities	Use dictionary mismatch and NER to identify non-standard tokens	`spaCy`, `nltk`, entity types (`ORG`, `PERSON`, `GPE`)
Lever	Noun phrases	Extract as objects or attributes related to financial action	Noun chunks, dependency labels (`dobj`, `pobj`); domain-specific lexicons
Magnitude	Verbs, adverbs	Detect verbs indicating directional action and intensity modifiers	POS tags, verb impact lexicons, TF-IDF scores
Novelty	Adjectives, adverbs, rare phrases	Identify markers of surprise or deviation from norm	POS tags (`JJ`, `RB`), phrase templates, frequency rarity

to shape returns—potentially discovering nonlinear or cross-dimensional effects missed by hand-tuned scoring. Importantly, such extensions need not sacrifice interpretability: score components could be supervised or constrained using existing axis-level features, ensuring that learned signals remain decomposable and aligned with economic intuition. For example, a VAE trained on actor–lever–timing embeddings could cluster news into archetypal narratives, while an EBM could capture the nonlinear activation of market response given specific combinations of novelty and source identity. Third, FinBERT itself could be hybridized with this approach: actor and lever signals could be used as overlays to reweight sentiment scores, improving its responsiveness in structurally ambiguous cases.

Beyond immediate return prediction, our framework opens avenues for studying how attention and narrative shape investor response. Because the model isolates news by actor, lever, and novelty—while tracking signal timing—it can support investigations into *attention dynamics*: for example, how quickly supply chain disruptions linked to major vendors (e.g., TSMC vs. Foxconn) get priced, or whether institutional investors react faster to analyst downgrades than to regulatory statements. Similarly, by decomposing news into magnitude and novelty dimensions, the model enables research into *narrative framing*—such as whether crisis-oriented headlines (e.g., "Apple hit by chip shortage") drive more delayed volatility than informationally equivalent but neutrally worded articles (e.g., "Apple faces supply constraints").

Anecdotal patterns in emerging markets, such as India, further illustrate the importance of modeling actor identity and intent. For example, net flows from foreign institutional investors (FIIs) are often treated as informational events—even when their size is small relative to total market capitalization. These reactions may stem from structural asymmetries: FIIs often operate under tight discretionary caps or regional allocation models, meaning their exits or entries are interpreted as global opportunity shifts rather than firm-specific signals. This leads to capitalsource signaling, where the behavior of an actor (e.g., a large foreign fund) is perceived as more informative than the content of the news itself. Our model, by explicitly scoring actor salience and intent, provides a natural foundation for studying and quantifying such dynamics—especially in markets where identity-driven reflexivity plays a central role.

In sum, this work presents a transparent, scalable, and behaviorally grounded framework for understanding how financial news drives returns. By bridging linguistic form and economic function, the model not only enables interpretable prediction—but also opens a path for deeper insight into how narrative structure, attention allocation, and market cognition jointly shape asset pricing outcomes.

References

Didar Araci. Finbert: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063*, 2019. URL `https://arxiv.org/abs/1908.10063`.

Jacob Boudoukh, Ronen Feldman, Shimon Kogan, and Matthew Richardson. Which news moves stock prices? a textual analysis. Working Paper 18725, National Bureau of Economic Research, January 2012. URL `https://www.nber.org/papers/w18725`.

Leland Bybee, Bryan T. Kelly, Asaf Manela, and Dacheng Xiu. Business news and business cycles. *The Journal of Finance*, 79(5):3105–3147, 2024. doi: 10.1111/jofi.13377.

Robert F. Engle, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel. Hedging climate change news. *The Review of Financial Studies*, 33(3):1184–1216, 2020. doi: 10.1093/rfs/hhz072.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020. doi: 10.1093/rfs/hhaa009.

Bryan Kelly, Boris Kuznetsov, Semyon Malamud, and Teng Andrea Xu. Large (and deep) factor models, 2024. URL `https://arxiv.org/abs/2402.06635`.

Vegard H. Larsen and Leif Anders Thorsrud. The value of news for economic developments. *Journal of Econometrics*, 210(1):203–218, 2019. doi: 10.1016/j.jeconom.2018.11.013.

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: 10.1111/j.1540-6261.2007.01232.x.