

Extracting the Structure of Press Releases for Predicting Earnings Announcement Returns

Yuntao Wu, Ege Mert Akin, Charles Martineau, Vincent Grégoire, Andreas Veneris



Link to Paper

Introduction & Contribution

We examine how textual features in earnings press releases predict stock returns on earnings announcement days. Using over 138,000 press releases (2005-2023), we find that press release content (soft information) is as informative as earnings surprise (hard information). Stock prices fully reflect the content of press releases at market open. If press releases are leaked, it offers predictive advantage. Topic analysis reveals self-serving bias in managerial narratives.

Highlights:

- **Online learning** minimizes look-ahead bias and adapts to new data
- Combines **bad-of-words**, **BERT** and **LLMs** for enhanced explainability
- Demonstrates **market efficiency** in processing soft information
- **Identifies predictive signals** from leaked press releases and earnings surprises

Methodology

We analyze **over 138,000 overnight** earnings press releases (**2005-2023**), covering about **2,500 U.S. stocks** of all market capitalizations each year.

Embedding techniques:

- BKMx: 180 topics from clusters of 3,592 words/phrases in WSJ (1984–2017), covering economics and politics
- oLDA: 50 dynamic topics from online learning on earnings press releases
- BERT family: contextual embeddings from BERT, MPNet, and FinBERT
- GPT-assisted labeling: identifies oLDA topic categories and maps top BERT tokens to metatopics

Apply Lasso regression to link press release embeddings with stock returns, generating return scores Soft^k and identifying key topics/tokens:

$$\arg \min_w \left\| X_{c,\tau} w - \text{Ret}_{c,\tau} \right\|^2 + \lambda \|w\|_1$$

- $X_{c,\tau}$: embedding of firm c 's press release on day τ
- $\text{Ret}_{c,\tau}$: adjusted announcement-day return from CRSP
- $\lambda = 10^{-5}$: fixed regularization parameter
- Rolling window: 1-year training, 1-year out-of-sample prediction
- Evaluation Period: 2006-2023

Analyze **metatopics** by **total topic weight** and **share of total variance**:

- Lag topic weights by one year, consistent with the return score generation procedure

$$w(M) = \sum_{k \in M} w_k, \quad h(M) = \frac{\sum_{i \in M} \sum_{j \in M} w_i w_j \text{Cov}(f_i, f_j)}{\sum_M \sum_{i \in M} \sum_{j \in M} w_i w_j \text{Cov}(f_i, f_j)} \times 100$$

Main Results

Prediction analysis with earnings surprise as control:

$$\text{Ret}_{c,\tau} = \alpha + \beta_0 \text{Surprise}_{c,\tau} + \beta_1 \text{Soft}_{c,\tau}^k + \epsilon$$

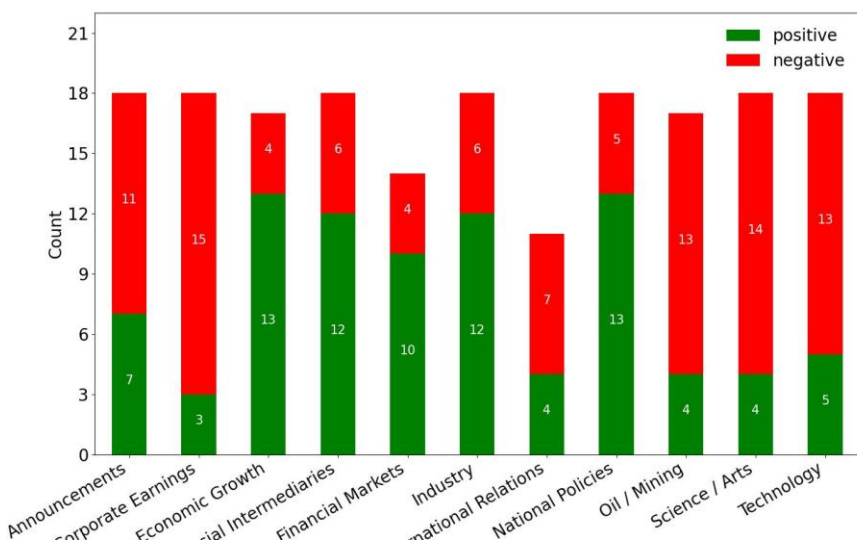
- Surprise: excess reported EPS relative to consensus analyst forecasts
- Soft: return scores derived from press release embeddings
- Soft and Surprise contribute about equally

Model	R^2	SHAP	Notes
BKMx	3.4%	19.2%	Enhances topic interpretability
oLDA	3.3%	14.6%	Uses customized, time-evolving topic categories
BERT	4.0%	45.8%	Captures rich context; less interpretable than bag-of-words
MPNET	3.5%	33.5%	Learns strong contextual relations but underperforms on this task
FinBERT	4.4%	52.1%	Fine-tuned for financial text; achieves best performance
Mean	4.1%	48.4%	Aggregated prediction used for downstream evaluation

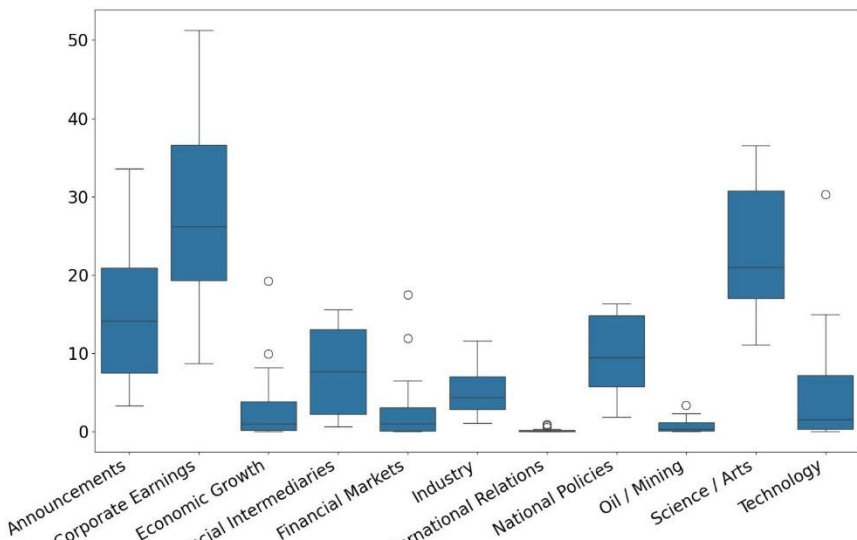
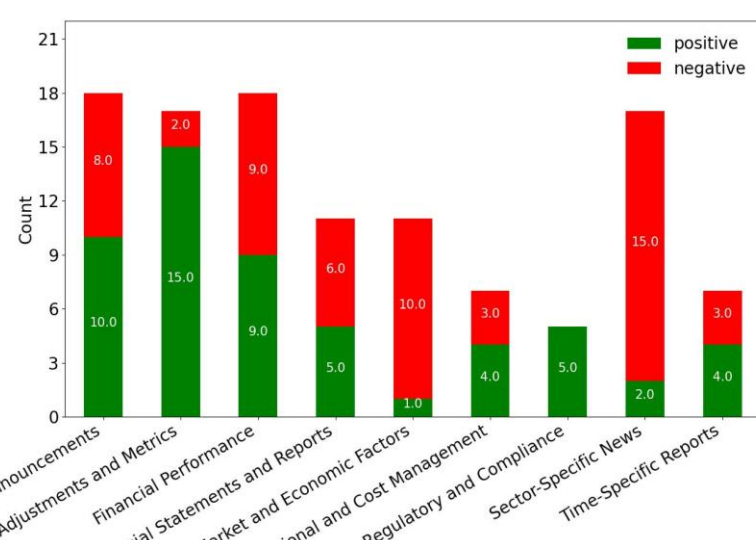
Cross-sectional regression summary

Topic analysis:

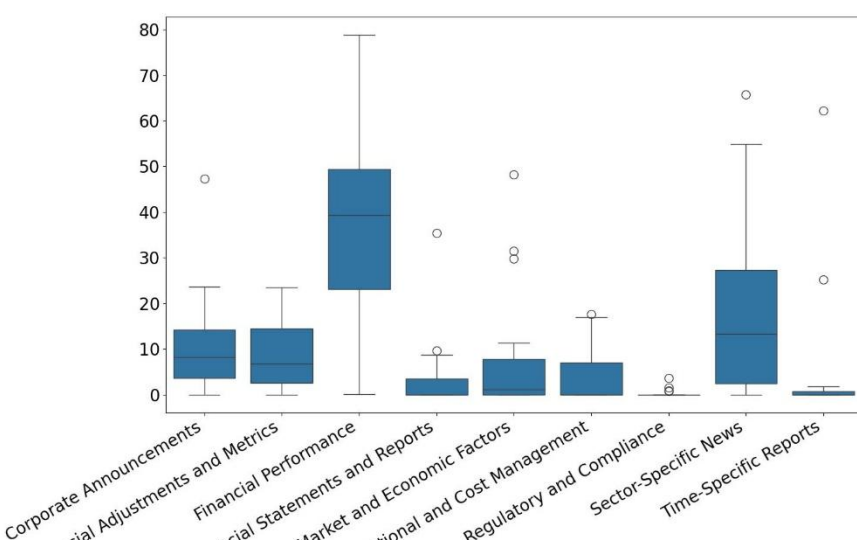
- BKMx: most influential metatopics include *Announcements*, *Corporate Earnings*, and *Science/Arts* (covering *challenges* and *marketing* themes)
- oLDA: reveals **self-serving** bias in managerial narratives
- BERT-based models: interpretable through oLDA topics, showing consistent conclusions across methods



Metatopic weight polarity for BKMx (left) and oLDA (right)



Metatopic explained variance for BKMx (left) and oLDA (right)



Model	Sign	Corp. Ann.	Fin. Metrics	Fin. Perf.	Fin. Reports	Market / Econ. Factors	Op. / Cost Mgmt	Regulatory	Sector News	Time-Spec. Reports
BERT	Pos	0.58%	5.75%	7.96%	11.78%	6.36%	11.75%	1.29%	7.95%	6.76%
	Neg	0.24%	2.34%	6.54%	3.28%	5.47%	10.11%	0.79%	5.83%	5.22%
MPNET	Pos	0.90%	4.63%	5.53%	4.41%	3.37%	6.12%	0.43%	4.82%	17.48%
	Neg	0.59%	6.75%	7.87%	2.62%	5.18%	9.53%	1.01%	9.13%	9.63%
FinBERT	Pos	1.46%	11.09%	16.07%	4.43%	5.22%	5.11%	1.49%	6.82%	7.73%
	Neg	1.85%	2.83%	6.80%	2.07%	7.94%	4.50%	1.22%	6.27%	7.11%

Top positive and negative tokens identified by BERT that contribute most to return scores



UNIVERSITY OF
TORONTO

HEC MONTRÉAL



Average returns sorted by surprise (hard information) and soft

Trading Strategy Analysis

- Trade window: enter at 9:45 AM, exit at 4:00 PM
- Filter: exclude illiquid stocks (bid–ask spread > 20% of midquote)
- Long stocks with both positive surprise and positive Soft^{Mean} and short stocks with both negative surprise and negative Soft^{Mean}
- Returns: compute long–short portfolio returns, **weighted by market capitalization**

Highlights:

- Post-announcement: strategies mainly incur transaction costs, generating **no sustained profits** (Strategy $\alpha \approx -0.5\%$)
- Pre-announcement: combining Surprise and Soft improves Precision@k in identifying top positive/negative return stocks — **from ~35% to ~52%**

		P@1	P@2	P@3	P@5	P@10
Top Positive	Surprise	9.1%/12.4%	13.5%/20.2%	17.1%/25.9%	23.6%/34.6%	34.7%/51.7%
	Soft^{Mean}	3.7%/7.0%	6.1%/12.0%	8.5%/16.9%	13.4%/25.7%	24.5%/44.1%
Top Negative	Surprise	9.4%/14.1%	14.0%/21.1%	17.5%/27.1%	23.8%/36.1%	35.8%/51.9%
	Soft^{Mean}	5.3%/9.8%	8.6%/16.4%	11.5%/21.6%	16.3%/29.8%	27.1%/46.6%

Precision@k for predicting announcement-day returns, before and after filtering by surprise and soft agreement

Acknowledgements

This research was carried out at Rotman School of Management, FinHub Financial Innovation Lab, University of Toronto. We gratefully acknowledge financial support from the Social Sciences and Humanities Research Council (SSHRC) of Canada grant number 435-2022-0745.

References

- Pat Akey, Vincent Grégoire, and Charles Martineau. 2022. Price revelation from insider trading: Evidence from hacked earnings news. *Journal of Financial Economics* 143, 3 (2022), 1162–1184.
- Dogu Araci. 2019. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. arXiv:1908.10063 [cs.CL] <https://arxiv.org/abs/1908.10063>
- Leland Bybee, Bryan Kelly, Asaf Manela, and Dacheng Xiu. 2024. Business News and Business Cycles. *The Journal of Finance*, 79: 3105–3147. <https://doi.org/10.1111/jofi.13377>
- Vincent Gregoire and Charles Martineau. 2022. How is earnings news transmitted to stock prices? *Journal of Accounting Research* 60, 1 (2022), 261–297.