

Sentiment-Model Complexity in Low-Frequency Environments

Andreas Stylianos Metaxas

Department of International and European Economic Studies

June 29, 2026

Motivation and Research Question

- Data frequency differences
 - Financial markets: high-frequency data, eg. VIX
 - FOMC minutes (CB communications): low-frequency
- Complexity: amount of information and computational effort a model requires to extract a "meaningful" signal from text
 - "This movie was terrible": requires little context
 - "John told Mike he was late" requires more information about surrounding text to resolve the pronoun
- Research question: Does greater NLP-model complexity improve economic usefulness of sentiment signals extracted from low-frequency data?

Contribution and Related Literature

- The literature examines CB sentiment, but does not directly examine signal–complexity tradeoff in environments with different-frequency data
- Financial econometrics & Central Bank (CB) tone
 - Apergis (2015): motivates the use of GARCH models for FOMC-related volatility analysis.
 - Tadle (2022), Hubert & Labondance (2021): CB tone contains market-relevant information
- Computational linguistics in finance
 - Araci (2019): Transformer-based models (FinBERT) outperform traditional dictionaries
- This paper
 - Does greater complexity improve economic usefulness of sentiment signals, when text is low-frequency?

Main Findings

- Some complexity is good: transformers vs simple lexicons
 - Simple lexicons (eg. Loughran-McDonald financial, LM): Assign one fixed meaning to each word
 - Contextual transformers (eg. FinBERT): Deep learning models that understand the meaning of words by considering the surrounding context
 - Increased complexity benefit: Transformers adapt a word's representation based on the words around it, enabling more accurate language understanding
- But not monotonically so: excessive domain adaptation (DA) Some
 - DA: Process of further training a pre-trained language model on text from a specific domain (eg. only FOMC), so it better understands that domain's vocabulary and writing style

- Data
 - Textual: FOMC minutes, 2000–2025, 206 documents
 - Market: 5-day change in VIX
 - Match: to the next valid trading day
- Econometric models
 - OLS and GARCH(1,1)
 - Robustness: Winsorized VIX, EGARCH

Sentiment Models

- LM lexicon: Classification of financial terms in positive, negative, neutral sentiment categories
 - Positive sentiment is simply a count of positive words
- BERT (Bidirectional Encoder Representations from Transformers): Pre-trained language model that learns deep contextual word representations by reading text in both directions
 - Can produce context-aware sentiment
 - Assigns contextual probability to positive sentiment being associated with a given word
- FinBERT: BERT trained on variety of financial documents
- Domain adaptation (DA): Here, further training BERT on FOMC-minutes corpus

Learning and DA: Baseline FinBERT

- Assigns contextual weights
 - How strongly a word like “profit” influences a positive sentiment
 - How much “loss” pushes the prediction to negative
- All (millions) model weights updated (adapted) during training (learning financial context)

Learning and DA: FinBERT + LoRA

- LoRA: Low-Rank Adaptation
- Freeze original FinBERT weights, add and train small matrices (adapter layers)
- Without LoRA: “moonshot growth” → unfamiliar → uncertain sentiment contribution
- With LoRA
 - Learns it appears in strongly positive financial context
 - Attributes strong positive sentiment signal

Learning and DA: FinBERT + MLM

- MLM: Masked-Language Modeling: Hide some words and train FinBERT to predict them
- Masked: "The startup showed moonshot [MASK] in user acquisition"
- Possible FinBERT predictions: growth (correct, most likely), expansion, increase, surge
- FinBERT learns
 - "moonshot + [MASK]" often connects to growth-related words
 - "moonshot growth" behaves like a single strong financial concept
 - The phrase appears in high-growth, positive contexts
 - It becomes associated with strong positive sentiment
- Comparison LoRA vs MLM adaptation
 - MLM: What does "moonshot growth" mean in context?
 - LoRA: Given "moonshot growth", how strongly should I predict positive sentiment?

Complexity Ordering

Model	Method	Complexity
LM Lexicon	Bag-of-words + Elastic Net filtering	Baseline (Minimal)
FinBERT (Base)	Pre-trained Contextual Transformer	High (109M params)
FinBERT + LoRA	Parameter-efficient fine-tuning	Adaptive ($\sim 0.2\%$ trained)
FinBERT + MLM	Domain-adapted pre-training	Extreme (Over-specialized)

- Models from simplest to most complex

LM Lexicon vs FinBERT: Benefits of Complexity

Sentiment Metric	LM Lexicon	FinBERT (Base)	FinBERT/Lexicon
Average	-0.02	-0.21	11
Volatility	0.01	0.22	22
Range	0.07	1.41	20
Minimum	-0.06	-0.80	13
Maximum	0.02	0.61	31

- FinBERT produces stronger average sentiment shifts
- FinBERT outputs much more variation across texts
- FinBERT spans a much wider sentiment scale
- FinBERT can detect much stronger negative and positive signals
- Benefit of complexity : Context/complexity yields sentiment variation (more nuanced sentiment)

FinBERT vs FinBERT-MLM: Limits of Complexity

Sentiment Metric	FinBERT (Base)	FinBERT (MLM)	MLM/Base
Average	-0.21	-0.02	0.10
Volatility	0.22	0.01	0.05
Minimum	-0.80	-0.05	0.06
Maximum	0.61	0.01	0.02

- With MLM, average sentiment is reduced by 90%
- MLM produces lower sentiment variation across documents → flatter sentiment signal → lower explanatory usefulness
- With MLM, extreme negative sentiment values are greatly compressed
- With MLM, extremes positive sentiment values are almost completely eliminated
- FinBERT-MLM too adapted to FOMC corpus (over-specialization)

Why Can MLM Over-Specialize?

Base FinBERT: context matters

“moderate inflation risks” → weaker risk signal

“elevated inflation risks” → stronger risk signal

- FinBERT reads bidirectionally
- Interprets adjective together with noun

MLM adaptation: wording becomes familiar

“[MASK] inflation risks”



Possible predictions: *moderate, elevated*

- MLM trains model to predict missing words
- Repeated FOMC wording → model sees very similar contexts → “moderate” and “elevated” become less distinctive → sentiment differences are compressed → flatter signal

Standardized FOMC language ⇒ repeated exposure to similar wording

Complexity and Forecasting: An Application

- FOMC sentiment of different complexity levels and changes in uncertainty (VIX)
- GARCH(1,1) Specification: Accounts for volatility clustering (Apergis 15)

$$\text{Conditional Mean: } \Delta VIX_{t-1,t+5} = \mu + \delta S_{i,t} + \epsilon_t$$

$$\text{Conditional Variance: } \sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

GARCH(1,1) Results

Metric	LM Lexicon	FinBERT (Base)	LoRA	MLM
R^2	0.01	0.04	0.04	0.01
δ	48.54	3.49***	3.49***	-92.07
α	0.26	0.30	0.30	0.36
AIC	854.7	849.6	849.6	849.7
Root MSE	3.26	3.29	3.29	3.48
ARCH-LM test	p=0.97	p=0.94	p=0.94	p=0.82

- From LM ("naive" sentiment) to FinBERT baseline (complex, contextual), sentiment coefficient δ becomes significant and of magnitude consistent with literature, while errors roughly unchanged
- From FinBERT Base to MLM (overfitted sentiment), sentiment signal lost again
- FinBERT+LoRA preserves sentiment signal while increasing efficiency

Positive Relationship between S and ΔVIX

- Significant positive coefficient between FOMC FinBERT, LoRA, LM lexicon sentiment and ΔVIX
- Why do increases in Fed optimism increase the VIX?
- Consistent with literature (Hubert & Labondance (2021), Bomfim (2003))
 - Optimistic FOMC language may be interpreted by markets as signal of future tightening
 - Anticipation of higher rates $\rightarrow$ investors hedge against downside equity risk $\rightarrow$ implied volatility spikes

① Winsorized VIX (1st & 99th percentiles):

- Extreme VIX spikes do not artificially drive the relationship
- FinBERT remains highly significant at the 1% level under GARCH

② EGARCH Specification:

- Asymmetric shock responses (leverage effect) provides marginal information criterion gains
- Does not alter ranking of NLP models.

Conclusions

- FinBERT provides strongest in-sample signal
 - Contextual sentiment captures variation in FOMC tone better than simple LM lexicon
- Excessive domain adaptation backfires
 - MLM adaptation reduces sentiment variation
 - Over-specialization can "normalize" economically meaningful central-bank language
- Simplicity remains valuable
 - LM lexicon remains competitive out-of-sample and interpretable as a benchmark

Thank You

Limitations and Avenues for Future Research

Acknowledged

- The 5-day event window allows time for market digestion, but it may also include non-FOMC macroeconomic shocks.
- The data sample is constrained to exactly 206 scheduled events, limiting deep learning scalability.

Avenues for Future Research

- Application of the LoRA-FinBERT extraction framework to high-frequency central bank communications (e.g., intra-day press conference transcripts).
- Investigating why GARCH persistence is near zero in the baseline specification, and whether this differs across monetary-policy regimes.

OLS Results

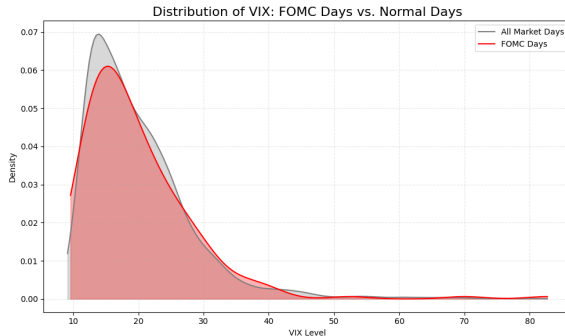
Ordinary Least Squares Specification:

$$\Delta VIX_{t-1,t+5} = \beta_0 + \beta_1 S_{i,t} + \epsilon_t$$

Metric	FinBERT (S_{Base})	LoRA (S_{LoRA})	MLM (S_{MLM})	LM (S_{LM})
R^2	0.04	0.04	0.02	0.02
β_1	2.82**	2.82**	-53.27	34.88
Kurtosis	6.54	6.54	7.12	7.05
RMSE	3.28	3.28	3.38	3.25
AIC	859.57	859.57	863.22	863.35
ARCH-LM test	p=0.0091	p=0.0091	p=0.0204	p=0.0141

- Residual diagnostics reveal non-normality (Kurtosis > 6.5) and conditional heteroskedasticity (ARCH-LM $p < 0.05$)

Data: VIX on FOMC Release Days



- VIX levels on FOMC release days (red) and normal trading days (grey) overlap
- Suggests FOMC release days are not "special" re. VIX levels → useful testing environment

Computational Trade-offs

Model	Active Parameters	Time/Doc	Memory	Trade-off
LM Lexicon	0	0.0008s	32.7 MB	Baseline
FinBERT (Base)	109,484,547	4.91s	426.0 MB	High memory
FinBERT + LoRA	Rank-8 adapters	6.89s	70.1 MB	Slower / lighter

- LoRA preserves the FinBERT signal while reducing memory use.
- Slightly higher processing time per document, but substantial computation saves