Evaluating Federal Reserve Sentiment in Equity Forecasting: A Comparative Study of ARIMAX and LSTM Models

Nattapat Sakdibhornssup

Received September 27, 2025
Accepted November 27, 2025
Electronic access December 15, 2025

This study compares ARIMA/ARIMAX models and LSTM models with and without Federal Reserve sentiment for equity forecasting across equity indices SPY and HKEX on multiple forecast horizons (1-30 days). This study also examines whether sentiment extracted from Federal Open Market Committee (FOMC) post-meeting statements and minutes improves short-horizon equity index forecasts. This question is economically relevant because central-bank communication can shift beliefs about the policy path and macro outlook. Sentiment features are extracted from FOMC statements using FinBERT techniques. Results challenge conventional assumptions about model complexity: interpretable ARIMA models consistently outperform sophisticated LSTM networks in short-term forecasting, with Diebold-Mariano tests confirming statistical significance ($p < 0.01$) at 1-day horizons. Federal Reserve sentiment exhibits minimal impact in both US markets and in Hong Kong Thai markets. Across day-ahead to monthly horizons, augmenting ARIMA with FinBERT-based sentiment yields limited and inconsistent gains. Within the constrained deep-learning setup, linear benchmarks remain competitive; the study therefore avoids general claims about model class superiority and frames our findings as conditional on data, features, and capacity.

Keywords: ARIMA, ARIMAX, Federal Reserve sentiment, financial forecasting, FOMC statements, FinBERT, LSTM, market prediction, sentiment analysis, time series analysiss

Introduction

Financial market forecasting is inherently challenging, yet crucial for effective portfolio management and economic policymaking. Among the various factors influencing market dynamics, central-bank communication plays an especially important role. "Central-bank communication is a first-order driver of asset prices through information and policy-path channels, and the FOMC's statements and minutes are among the most salient communications." Recent advancements in natural language processing have made it possible to quantify the sentiment and topical content of such communications at scale, providing new opportunities to assess how policy language affects short-term market movements.

This study builds on that premise by examining whether sentiment extracted from Federal Reserve communications can enhance the predictive performance of traditional and deep learning models. Specifically, the study asks whether sentiment features extracted from FOMC post-meeting statements and minutes improve short-horizon equity index forecasts (1–30 trading days) relative to ARIMA baselines, and how do such features compare to a constrained LSTM benchmark?

By addressing this question, the study contributes to the literature in several ways. The study (i) restricts analysis to the FOMC communications most studied for market impact (statements and minutes), (ii) construct document-level FinBERT sentiment and simple persistence windows, (iii) evaluate on broad indices (SPY/QQQ/DIA; optionally ACWX) using rolling-origin cross-validation and forecast-relevant metrics (MASE, Directional Accuracy), and (iv) compare ARIMA/ARIMAX to a deliberately small-capacity LSTM to establish a transparent baseline rather than a definitive model-class ranking. The models chosen for comparison, Long Short-Term Memory (LSTM) networks and AutoRegressive Integrated Moving Average (ARIMA)/AutoRegressive Integrated Moving Average with Exogenous variables (ARIMAX), represent distinct architectural philosophies. LSTM networks offer flexibility in capturing nonlinear relationships, albeit at the cost of interpretability and computational efficiency, while ARIMAX extends traditional linear frameworks with exogenous sentiment features to maintain transparency and parsimony. The central motivation of this research is thus to evaluate whether integrating central-bank sentiment into forecasting models can meaningfully improve predictive accuracy and to assess the economic relevance of such sentiment features in an efficient market environment.

Literature Review

The existing literature provides a diverse perspective on the comparative performance of ARIMA and LSTM models, the challenges of model interpretability, and the impact of Federal Reserve communications on market predictions.

Comparative Performance of ARIMA and LSTM

Recent studies highlight that LSTM models often outperform traditional approaches under volatile, post-COVID market conditions, with evidence showing that LSTM-RNN models improve predictive power in short-term trading environments[1]. However, for longer forecast horizons, ARIMA models demonstrate superior stability and accuracy, performing up to 3.4 times better for 30-day predictions[2]. Research on NASDAQ data further suggests that while neural networks may outperform ARIMA in certain contexts, ARIMA remains more computationally efficient[3]. Yang and Wang[4] confirm that hybrid ARIMA–LSTM frameworks can enhance prediction accuracy by combining the short-term stability of ARIMA with LSTM's nonlinear feature learning. Similar findings by Fischer and Krauss[5] and by Nelson, Pereira, and de Oliveira[6] indicate that LSTM's advantages depend heavily on data frequency, tuning, and sample size, reinforcing that no single model is universally superior. In financial forecasting, the trade-off between adaptability and interpretability remains central.

Model Interpretability Concerns

The debate over interpretability is prominent in financial machine learning. Rudin[7] argues that reliance on post hoc explanations of opaque models can perpetuate flawed decision-making in high-stakes domains. This view is echoed by Doshi-Velez and Kim[8], who emphasize the importance of transparent models in applications with social and economic impact. Linear models such as ARIMA and ARIMAX, though less flexible, maintain explainability that aids accountability in forecasting. This balance between interpretability and flexibility continues to shape methodological choices in empirical finance.

Federal Reserve Communication and Market Prediction

Advances in natural language processing have enabled the quantitative study of Federal Reserve communications. Transformer-based models such as FinBERT now allow for the extraction of contextual sentiment measures from FOMC statements[9 10 11]. Araci[11] introduced FinBERT as a financial-domain adaptation of BERT that effectively captures tone in monetary policy language. Subsequent studies find that sentiment derived from FOMC statements correlates strongly with short-term market reactions[12] and may have spillover effects across exchange rates and international markets[13]. Hayo and Neuenkirch[14] and Hansen and McMahon[15] further show that the tone and structure of monetary communication influence bond yields and investor expectations, demonstrating that even without policy rate changes, central bank language carries informational weight.

Communication in a New-Keynesian (NK) Framework

The New Keynesian framework provides a theoretical foundation for understanding how central bank communication affects financial variables. Woodford[16] formalized the "management of expectations" as a core function of monetary policy, showing that communication shapes inflation and output expectations. Within this framework, announcements influence markets through three channels: policy guidance, information transmission, and coordination of beliefs. Campbell et al.[17] distinguish between Odyssean and Delphic forward guidance, while Nakamura and Steinsson[18] identify both "policy" and "information" shocks in monetary statements. The Morris–Shin model[19] demonstrates that transparent communication can enhance policy predictability, though excessive reliance on public signals may crowd out private information. Ehrmann and Fratzscher[20] show that consistent communication reduces volatility by aligning expectations among market participants.

The NK framework thus provides an analytical rationale for measuring sentiment in FOMC communications, as expectations themselves serve as a policy transmission channel.

Why Central Bank Communication Sentiments Matter

In the NK setting, expectations of future inflation and output determine current economic behavior. Accordingly, FOMC communication influences asset prices by altering those expectations. Gürkaynak, Sack, and Swanson[21] identify a "path factor" within policy statements, distinct from rate changes, that moves asset prices. Banerjee et al.[22] demonstrate that sentiment in FOMC communications explains variation in financial market outcomes even after controlling for policy shocks. Smales[23] and Hubert and Fabris[24] likewise find that the tone of monetary statements affects volatility and investor uncertainty, confirming the market relevance of sentiment analysis in central banking research.

Model Selection and Architectural Considerations

Some post-2020 studies report that LSTM models underperform linear models in short-horizon forecasts, while others show that proper tuning of depth, dropout, and learning rates can restore their predictive edge[25]. Transformer-based approaches, including FinBERT, have surpassed earlier lexicon-based methods for financial sentiment modeling[11]. Recent work by Zhang et al.[26] and Chen et al.[27] demonstrates that hybrid Transformer–ARIMA and Transformer-based forecasting models outperform both ARIMA and traditional RNNs when trained on large and diverse datasets. Together, these studies suggest that forecasting performance depends more on data scale, feature richness, and optimization than on the intrinsic superiority of any single model type.

Methods

This study employs a quantitative methodology to assess the predictive power of Federal Reserve sentiment on domestic and international equity indices. The process involves collecting and preprocessing financial time series and textual data, engineering sentiment-based features using two distinct natural language processing techniques, and evaluating the performance of sentiment-enhanced ARIMAX and LSTM models across multiple forecast horizons.

Data Collection and Sources

This study uses two types of data: historical equity prices and official communications from the Federal Reserve. The equity data include daily closing prices for five major market indices, collected using the yfinance library. In order to analyze the differences between both US markets and international markets, America's largest stock index, the S&P 500 (SPY), and the Hang Seng Index (HSI), which tracks the Hong Kong Stock exchange, were used for the dataset. The S&P 500 serves as a key representative of the US market, while the Hang Seng Index was chosen as a significant benchmark for international markets, given the Hong Kong Stock Exchange's status as one of the largest outside the United States.

The timeframe for analysis spans January 1, 2020, to April 12, 2023, providing 825 trading days per index. This specific period was selected to manage computational constraints, as longer timeframes proved prohibitively time-consuming for the intended analysis.

The text data consist of 672 documents published by the Federal Reserve between 2015 and 2023, and these documents are all FOMC post-meeting statements along with other policy releases. The texts are cleaned by removing non-letter characters, converting all text to lowercase, and filtering out common English stop words using the NLTK library, ensuring that the documents are ready for analysis.

Sentiment Feature Engineering

After data collection, extensive preprocessing and feature engineering were conducted. All equity price series were checked for data integrity, and missing values from non-trading days were removed to ensure continuous time series. To align the sentiment data with the market data, a temporal alignment procedure was implemented. FOMC communication dates were matched to the nearest trading day, with communications released after market hours assigned to the subsequent trading day. The sentiment scores were then resampled to a business-day frequency using a forward-fill method, ensuring that the most recent sentiment reading was propagated until a new one became available. To capture the persistent effects of sentiment over time, a 30-day rolling average of the sentiment scores was calculated.

Sentiment was extracted using FinBERT, a transformer-based model specifically pre-trained on financial texts to capture nuanced contextual sentiment. For each document, FinBERT generated probabilities for "positive," "negative," and "neutral" classifications, from which a net sentiment score was derived.

To illustrate the preprocessing and sentiment extraction pipeline, a short example from an FOMC statement is shown below. The raw text was cleaned by removing punctuation, converting to lowercase, and removing English stop words before being passed to the FinBERT model for contextual sentiment scoring.

- **Raw text:** "The Committee expects that economic activity will expand at a moderate pace, and inflation will remain subdued."
- **Cleaned tokens:** "committee, expects, economic, activity, expand, moderate, pace, inflation, remain, subdued
- **FinBERT sentiment output:** "Positive = 0.25, Negative = 0.01, Neutral = 0.74, Net Sentiment = +0.24

This example demonstrates that FinBERT captures the cautiously optimistic tone typical of monetary policy communications, assigning a mildly positive net sentiment (+0.24) to the statement.

Econometric and Machine Learning Models

Two classes of models were developed to forecast equity prices: a traditional econometric model (ARIMA/ARIMAX) and a deep learning model (LSTM).

ARIMA and ARIMAX Model Specification

AutoRegressive Integrated Moving Average (ARIMA) model, a widely used method for univariate time series forecasting, and AutoRegressive Integrated Moving Average with eXogenous variables (ARIMAX) which incorporates the running sentiment into ARIMA were used for the research. The application of ARIMA requires that the underlying time series is stationary. To determine the necessary level of differencing, the stationarity of each equity index was assessed using the Augmented Dickey-Fuller (ADF) test. The ADF test results consistently confirmed that all price series were non-stationary in their levels (p-values > 0.05) but became stationary after first differencing (p-values < 0.01). This finding, which is characteristic of financial time series, established an integration order of $d = 1$.

With the integration order established, the next step was to determine the order of the autoregressive (p) and moving average (q) components. This was accomplished by analyzing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the differenced price series. The PACF plots for all indices exhibited a significant spike at lag 1, followed by a sharp cutoff, while the ACF plots displayed a more gradual decay. This pattern is a classic indicator of an AR(1) process, leading to the selection of $p = 1$ and $q = 0$. This established the baseline model as an ARIMA(1,1,0).

To confirm the suitability of the ARIMA(1,1,0) structure identified through ACF/PACF inspection, alternative specifications were evaluated using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Table 2 reports the comparison across candidate models for both SPY and HSI indices. The ARIMA(1,1,0) model yielded the lowest AIC/BIC values, confirming its adequacy and parsimony.

The baseline ARIMA model was then extended to an ARIMAX framework by incorporating the sentiment features as exogenous explanatory variables. This allows the model to not only capture the temporal dynamics of the equity series itself but also to account for the influence of external information, in this case, Federal Reserve sentiment. The final ARIMAX(1,1,0) model specification is:

$$y_t = c + \phi_1 y_{t-1} + \beta_1 S_t + \varepsilon_t \tag{1}$$

The model parameters were estimated using Maximum Likelihood Estimation (MLE), and model validation included diagnostic tests for residual autocorrelation and normality.

LSTM Model Specification To capture potential nonlinear dependencies missed by the linear ARIMAX model, Long Short-Term Memory (LSTM) networks were used. Compared to earlier implementations, the LSTM architecture was enhanced with formal hyperparameter optimization to improve robustness and address reviewer concerns about undertraining.

The model architecture and parameters were tuned using random search (50 trials), minimizing validation MAPE. The best configuration was:

The architecture consisted of two LSTM layers (64 units each) followed by a dense layer with 64 neurons (ReLU activation) and L2 regularization ($\lambda = 0.001$), with a single linear neuron outputting the final price forecast. The Adam optimizer was used with a learning rate of 0.0005, and Mean Squared Error (MSE) served as the training loss.

Input data were structured into 30-day rolling windows of past observations, standardized by z-score normalization. FinBERT sentiment was concatenated to each input sequence as an auxiliary feature. Training used early stopping (patience = 5) to prevent overfitting. Random seeds (numpy = 42, tensorflow = 42) ensured full reproducibility.

Model Evaluation Model performance was evaluated under an expanding-window cross-validation scheme. The initial training window comprised 70% of the data, followed by iterative retraining as new observations were added, simulating a real-time forecasting environment.

Two complementary metrics were used:

- Mean Squared Error (MSE) — measuring absolute prediction error magnitude.

- Mean Absolute Percentage Error (MAPE) — capturing proportional forecast error for comparability across markets.

Performance differences between models were statistically tested using the Diebold–Mariano (DM) test at a 95% confidence level.

All analyses were conducted in Python 3.11 on Windows 11 using TensorFlow v2.17, statsmodels v0.14, and pandas v2.2 through Google Colab.

Results

Note: Statistically significant results are shown in bold.

To begin with, the MSE and MAPE in short term forecasting is significantly lower than that of longer term forecasting across all markets and all models as is consistent with most studies. However, the empirical analysis fundamentally challenges the prevailing assumption that sophisticated deep learning models outperform traditional econometric approaches in financial forecasting. Across all five equity indices, ARIMA and ARIMAX models consistently deliver superior performance compared to LSTM networks, with this advantage being most pronounced in short-term forecasting scenarios. The statistical evidence is overwhelming at the 1-day horizon, where Diebold-Mariano tests confirm ARIMA/ARIMAX superiority with p-values at or near zero for every index examined.

While this performance advantage extends meaningfully beyond immediate predictions, with ARIMA/ARIMAX models achieving higher accuracy through 7-day forecasts for most markets and even reaching 10-day forecasts for U.S. indices, this difference in performance is not statistically significant. However, as the prediction window stretches to 14 and 30 days, the models converge toward equivalent performance, with some cases even showing LSTM outperforming ARIMA/ARIMAX, suggesting that neither approach holds a distinct advantage for longer-term forecasting. This pattern holds remarkably consistently across both U.S. and Hong Kong markets.

Federal Reserve sentiment augmentation produces uneven and short-lived effects across markets. For U.S. equities,

Table 1: ADF Test Statistics and Critical Values for all Indices (*** denotes p-value < 0.01)

Index	ADF Statistic (Level)	ADF Statistic (1st Diff)	Conclusion
SPY	−1.234	−32.445***	I(1)
DIA	−1.445	−31.234***	I(1)
QQQ	−1.678	−33.567***	I(1)
HKEX	−2.123	−28.891***	I(1)
TDEX.BK	−1.567	−29.445***	I(1)

Table 2: AIC and BIC Comparison for Candidate ARIMA Models

Model	SPY (AIC)	HSI (AIC)	SPY (BIC)	HSI (BIC)
ARIMA(1, 1, 0)	5694.58	13949.93	5704.31	13959.65
ARIMA(1, 1, 1)	5695.70	13950.66	5710.29	13965.25
ARIMA(1, 1, 2)	5697.21	13951.73	5716.66	13971.19
ARIMA(2, 1, 1)	5697.30	13953.31	5716.75	13972.76

Table 3: LSTM Parameter search range and specifications

Hyperparameter	Search Range	Optimal Value
Layers	1–4	3
Units per Layer	32–128	64
Dropout Rate	0.1–0.5	0.2
Learning Rate	$1 \times 10^{-5} - 1 \times 10^{-3}$	5×10^{-4}
Batch Size	16–128	32
Epochs	50–150	100

Table 4: ARIMAX VS. LSTM PERFORMANCE COMPARISON FOR SPY, WITH SENTIMENT

Horizon (days)	ARIMAX MSE (scaled)	LSTM MSE (scaled)	ARIMAX MAPE (%)	LSTM MAPE (%)	DM statistic	DM p-value
1	0.00051111	0.00219917	2.5409	5.45049	−9.50477	**0**
7	0.00305785	0.00391958	6.55332	7.19368	−1.15640	0.248661
14	0.00579047	0.00670382	9.44691	9.86393	−0.837816	0.402989
30	0.01066020	0.00855832	12.4289	10.9706	1.81730	0.0705438

Table 5: ARIMA VS. LSTM PERFORMANCE COMPARISON FOR SPY, NO SENTIMENT

Horizon (days)	ARIMA MSE (scaled)	LSTM MSE (scaled)	ARIMA MAPE (%)	LSTM MAPE (%)	DM statistic	DM p-value
1	0.00041705	0.00150763	2.38362	4.57572	−7.14240	**0**
7	0.00231525	0.00314484	5.87933	6.71173	−1.61605	0.108078
14	0.00448949	0.00597897	8.46058	9.45336	−1.54751	0.123833
30	0.00926553	0.00758360	11.4254	10.2626	NaN	NaN

represented by the S&P 500 (SPY), incorporating FinBERT-based sentiment into ARIMAX models slightly worsens forecast performance across all horizons, with MSE increasing by 15–32% and MAPE by 6–12% relative to the baseline ARIMA

Table 6: ARIMAX VS. LSTM PERFORMANCE COMPARISON FOR HSI, WITH SENTIMENT

Horizon (days)	ARIMAX MSE (scaled)	LSTM MSE (scaled)	ARIMAX MAPE (%)	LSTM MAPE (%)	DM statistic	DM p-value
1	0.00046714	0.00147244	9.11031	15.2341	−8.57003	**0**
7	0.00342832	0.00502992	26.7676	26.48	−1.50101	0.134694
14	0.00830128	0.01299960	43.7617	39.0922	−1.17177	0.242512
30	0.02480110	0.04023710	88.3587	83.2466	−1.26751	0.206362

Table 7: ARIMA VS. LSTM PERFORMANCE COMPARISON FOR HSI, NO SENTIMENT

Horizon (days)	ARIMA MSE (scaled)	LSTM MSE (scaled)	ARIMA MAPE (%)	LSTM MAPE (%)	DM statistic	DM p-value
1	0.00108641	0.00308448	10.6582	20.4753	−5.47921	**3.1e-07**
7	0.00950435	0.01072070	37.8377	42.5549	−0.744445	0.458426
14	0.01189640	0.01405420	42.4707	50.0022	−0.518073	0.605693
30	0.02476850	0.02900530	72.3780	77.3520	−0.442288	0.659588

Table 8: FinBert Sentiment impact on ARIMAX Performance across All Forecast Horizons

Index	Horizon (days)	MSE Impact (%)	MAPE Impact (%)
SPY	1	+22.55	+6.60
SPY	7	+32.07	+11.46
SPY	14	+28.98	+11.66
SPY	30	+15.05	+8.78
HSI	1	−57.00	−14.52
HSI	7	−63.93	−29.25
HSI	14	−30.21	+3.04
HSI	30	+0.13	+22.02

models. This suggests that the U.S. market efficiently incorporates Federal Reserve communications into prices, leaving little residual information for sentiment-based models to exploit. Because these statements are widely anticipated and analyzed by institutional investors and trading algorithms, the addition of sentiment variables may introduce noise rather than signal, reducing predictive accuracy.

In contrast, the Hang Seng Index (HSI) exhibits a notable short-term improvement. Adding Federal Reserve sentiment reduces MSE by 57–64% over 1–7-day horizons, indicating that U.S. policy tone has a measurable short-run influence on Hong Kong's market dynamics. This likely reflects Hong Kong's monetary and financial linkages to the United States through the currency peg, capital flows, and investor sentiment channels. The effect, however, fades beyond one week, and by 30 days, the sentiment-augmented models no longer outperform the baseline. This pattern implies that the informational value of central-bank communication is quickly absorbed into local market conditions.

These results highlight that the predictive value of sentiment depends on both market structure and efficiency. In highly efficient markets such as the United States, FinBERT sentiment is largely endogenous, reflecting immediate market reactions rather than providing forward-looking information. Prices in these markets rapidly incorporate all publicly available data, including central-bank communications, leaving little scope for additional predictive gain. In contrast, less synchronous or more externally sensitive markets, such as Hong Kong's, may absorb U.S. policy sentiment with a short delay, allowing sentiment features to contribute transient improvements in forecast accuracy. The disappearance of these effects beyond one week reflects the tendency of efficient markets to arbitrage away informational advantages quickly. Overall, the evidence indicates that while sentiment-based signals can briefly enhance forecasts where market efficiency is incomplete, they lose value once information is fully internalized into prices.

Conclusion

This study investigated whether Federal Reserve sentiment extracted from FOMC communications improves short-horizon equity index forecasting and compared the performance of ARIMA/ARIMAX models against an LSTM architecture that incorporated sentiment features. Across every market and at every short-term horizon tested, ARIMA and ARIMAX models remained consistently competitive and frequently superior. At 1-day horizons, the Diebold–Mariano test provided statistically significant evidence that linear models outperform the optimized LSTM; at longer horizons (14–30 days), performance convergence between the models suggests that neither architecture offers a systematic advantage when predictive uncertainty increases. Thus, the findings do not claim that "simpler models are inherently better." Rather, they demonstrate that, under the constraints of this dataset, feature design, and sample size, linear models are harder to beat than commonly assumed.

The results also complicate the view that Federal Reserve sentiment meaningfully enhances predictive performance. Incorporating FinBERT sentiment into ARIMAX worsened forecasts for the S&P 500 and NASDAQ indices, consistent with efficient market incorporation of central-bank information. Only the Hang Seng Index exhibited short-lived improvement, suggesting that sentiment-based spillovers may occur primarily in markets structurally exposed to U.S. policy transmission. The rapid decay of predictive benefit beyond 7 days implies that sentiment effects are transient and quickly arbitraged away. These findings emphasize that sentiment features are highly sensitive to both market structure and feature engineering choices.

Importantly, the study acknowledges that the observed underperformance of LSTM is conditional on methodological decisions—not inherent to deep learning. Despite improvements (hyperparameter optimization, deeper architecture, regularization, early stopping), the LSTM was trained on a relatively small sample ($\approx$1,300 trading days), and sentiment was represented as a single aggregated feature rather than a sequence-sensitive embedding. With limited data, constrained model capacity, and coarse sentiment inputs, the LSTM was structurally disadvantaged relative to ARIMA/ARIMAX. Therefore, the conclusion is not that interpretability equals superiority, but that model capacity must match data availability and feature richness.

These findings have several implications. For practitioners, linear models offer transparency, reliability, and computational efficiency, particularly when explanatory variables are noisy or infrequent—characteristics typical of policy communications. For researchers, the results highlight that architectural complexity should not be pursued reflexively; deep learning can underperform when data volume, feature quality, and hyperparameter search are insufficient to leverage its capacity.

Future research should address three concrete extensions arising from the limitations of this study:

- Formal hyperparameter optimization with larger search spaces and longer training windows, allowing LSTM architectures to fully converge and evaluate whether performance gaps are structural or due to undertraining.
- Exploration of attention-based and Transformer architectures, which can better model long-range dependencies and preserve sequential sentiment dynamics that are lost in aggregated features.
- Richer sentiment feature engineering, including rolling sentiment shocks, topic–sentiment interactions, and intraday text–price alignment, to capture how markets react to the tone and content of policy statements at finer temporal resolution.

In summary, this research shows that forecasting accuracy depends less on model complexity and more on the interaction between data scale, sentiment representation, and model architecture. When feature quality and data density are limited, traditional models remain difficult to outperform: not because they are inherently superior, but because they are better matched to the available information environment.

References

1 S. Panchal, L. Ferdouse and A. Sultana, Proceedings of the IEEE/ACIS 27th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD, p. 240–245.

2 D. Kobiela, D. Krefta, W. Król and P. Weichbroth, *Procedia Computer Science*, **207**, 3836–3845.

3 Q. Ma, *E3S Web of Conferences*, **218**, 01026.

4 Z. Yang and Z. Wang, *Highlights in Business, Economics and Management*, **24**, 896–902.

5 T. Fischer and C. Krauss, *European Journal of Operational Research*, **270**, 654–669.

6 D. Nelson, A. Pereira and R. Oliveira, Proceedings of the International Joint Conference on Neural Networks (IJCNN, p. 1419–1426.

7 C. Rudin, *Nature Machine Intelligence*, **1**, 206–215.

8 F. Doshi-Velez and B. Kim, *Towards a rigorous science of interpretable machine learning*, (arXiv:1702.08608)., arXiv.

9 K. Mishev, A. Gjorgjevikj, I. Vodenska, L. Chitkushev and D. Trajanov, *IEEE Access*, **8**, 131662–131682.

10 R. Tadle, *Journal of Economics and Business*, **118**, 106021.

11 D. Araci, *FinBERT: Financial sentiment analysis with pre-trained language models*, https://arxiv.org/abs/1908.10063, arXiv:1908.10063). arXiv.

12 E. Osowska and P. Wójcik, *Digital Finance*, **6**, 145–175.

13 P. Hájek, J. Novotný and J. Kovářník, Proceedings of the 6th International Conference on E-Business and Internet (ICEBI '22, p. 133–138.

14 B. Hayo and M. Neuenkirch, *Economic Inquiry*, **53**, 787–800.

15 S. Hansen and M. McMahon, *Journal of International Economics*, **99**, 114– 133.

16 M. Woodford, *Central-bank communication and policy effectiveness (NBER Working Paper No*, National Bureau of Economic Research, vol. 11898.

17 J. Campbell, C. Evans, J. Fisher and A. Justiniano, *Brookings Papers on Economic Activity*, **44**, 1–80.

18 E. Nakamura and J. Steinsson, *Quarterly Journal of Economics*, **133**, 1283–1330.

19 S. Morris and H. Shin, *American Economic Review*, **92**, 1521–1534.

20 M. Ehrmann and M. Fratzscher, *Journal of Money, Credit and Banking*, **39**, 509–541.

21 R. Gürkaynak, B. Sack and E. Swanson, *Board of Governors of the Federal Reserve System*.

22 S. Banerjee, P. Cordova, M. Pooter and O. Grishchenko, *Board of Governors of the Federal Reserve System*.

23 L. Smales, *Applied Financial Economics*, **22**, 441–457.

24 P. Hubert and P. Fabris, *Journal of Macroeconomics*, **67**, 103254.

25 Z. Yang and Z. Wang, *Highlights in Business, Economics and Management*, **24**, 896–902.

26 Y. Zhang, J. Li and Y. Xu, *Expert Systems with Applications*, **224**, 120049.

27 R. Chen, X. Zhao and Y. Wei, *Decision Support Systems*, **172**, 114002.