

Forecasting Monetary Policy with Central Bank Sentiment

Group 10: Gaia lori, Michael Ladaa, Luca Milani, Matteo Roda, Sofia Villa

Research question

Can sentiment analysis of central bank communications improve the forecasting accuracy of monetary policy decisions?



Dataset

4,818 speeches from 1999 to 2025:
53% ECB and 47% Fed

Speeches
1999-2024

Speeches
2025

Interest
rates



Hugging Face



Classification Models

Tasks

- Binary (cut/rise)
- 5-class classification

Evaluation

- Train/validate split: 14/15 (train), 1/15 (test)
- Randomized vs. time-aware splits

Models

RoBERTa:

- max length 128
- learning rate 1e-5
- dropout 0.2
- weight decay 0.05
- early stopping

TF-IDF:

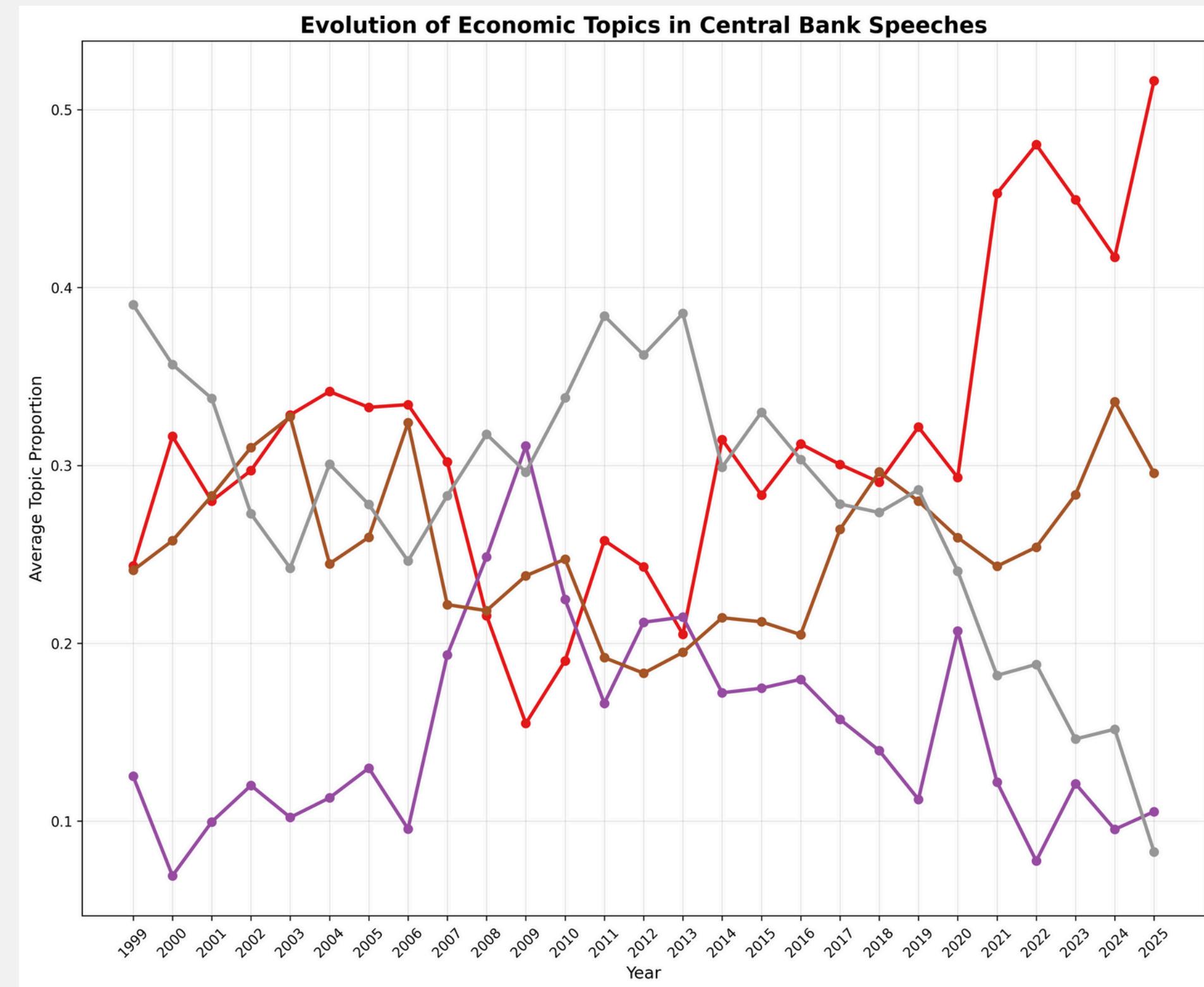
- char n-grams (2–6)
- 30k features
- class balancing

Classification Models

Macro F1 Score – Model Performance by Task & Split Type

Configuration	TF-IDF	RoBERTa
Random Split – Binary	> 0.70	0.95
Random Split – 5-Class	> 0.30	0.86
Time-Aware – Binary	≈ 0.14	≈ 0.14
Time-Aware – 5-Class	≈ 0.14	< 0.10

Aspect Based Sentiment Analysis



1. LDA

- Topic 1: Inflation
- Topic 2: Financial stability
- Topic 3: Microeconomic focus
- Topic 4: Post-crisis reforms

Aspect Based Sentiment Analysis

2. ABSA

- Extract topic-relevant text segments from speeches.
- Use **FinBERT** to analyze sentiment per topic (positive/negative/neutral)

3. Classifier

- Binary classifier (rise/cut) using Logistic Regression and Random Forest
- Variables: Sentiment scores per topic, topic distribution, rate level, days to decision, speech length, dominant topic
- Time-aware Cross-Validation → Random Forest **F1 score: 0.554**

Model Selection

I RoBERTa vs. HULK

RoBERTa outperformed TF-IDF.
Binary macro F1:

- RoBERTa 0.95
- HULK 0.71

II Binary vs. Multiclass

Binary more robust and interpretable.
Multi-class were arbitrarily balanced.

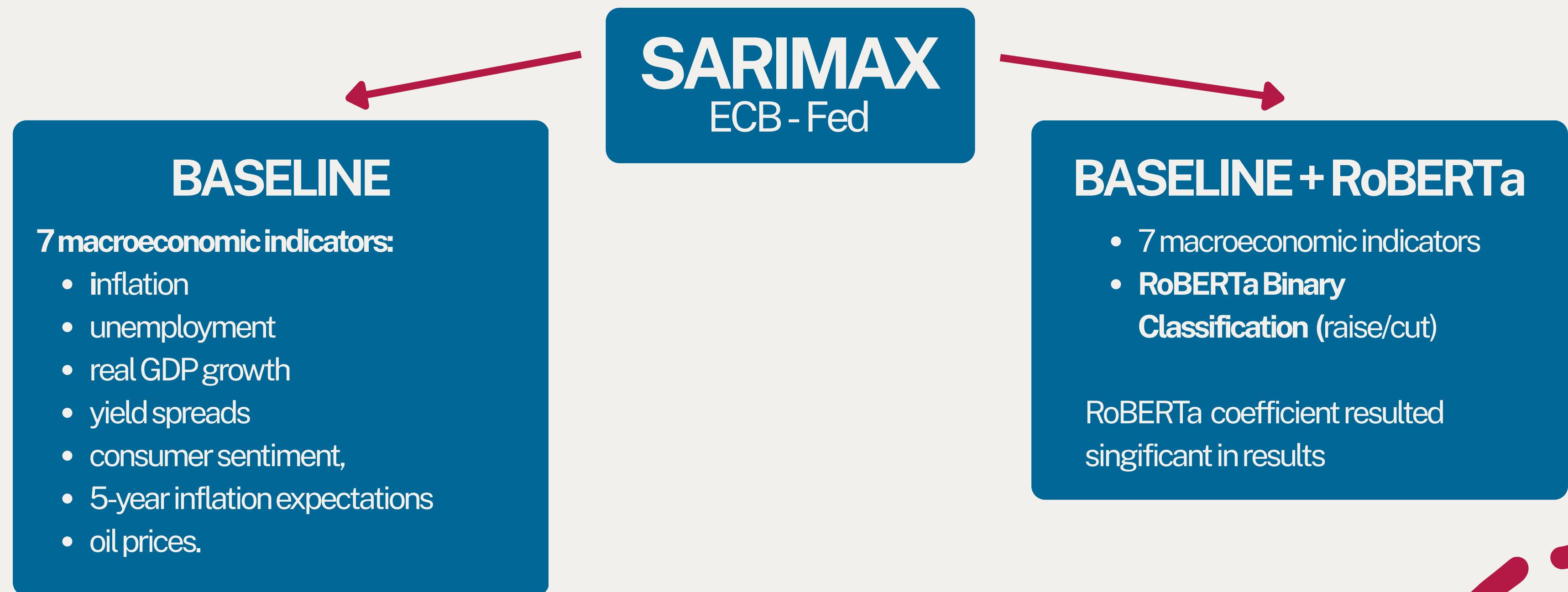
III Random vs. Time-aware

Random splits leaked future info.
Time-aware splits more realistic, lower performance.
Language drift over time.

IV RoBERTa vs. ABSA

ABSA had economically meaningful topics, but lower F1 score (0.554) compared to RoBERTA.

Interest rate forecasting



Results

ECB

- RoBERTa enhanced the ecb-version SARIMAX and improved the overall scores
- RMSE from 8.53 to 5.64
- The strongest predictor is RoBERTa, followed by GDP and sentiment.

FED

- RoBERTa integration did not change much results of SARIMAX
- RMSE from 13.9 to 14.2
- The strongest predictor is yield spread, followed by sentiment, inflation, and finally RoBERTa.

Interpretation:

FED speeches tend to be more voluntarily vague, whilst ECB speeches are more clear and defined¹

1. Chortareas, G., Papailias, F. and Shuku, L. (2025), Does Central Bank Talk Matter for Forecasting? Evidence From Speeches of the BoE, ECB, and Fed. Int J Fin Econ. <https://doi.org/10.1002/ijfe.3136>



Conclusion

Key takeaways:

- NLP analysis can be effective in improving the forecasting accuracy
- Transformer models outperform Aspect-Based Sentiment Analysis for monetary policy prediction

Limitations

- Temporal instability
- Speech transparency

Future improvements:

- Extend the analysis to other CB
- Explore other models: BERTopic, ML over SARIMAX



thank you for the
attention

Group 10: Gaia lori, Michael Ladaa, Luca Milani, Matteo Roda, Sofia Villa