

Inflation Forecasting with Monetary Policy Narratives: Evidence from an Emerging Economy

Antoaneta AMZA*

Bucharest University of Economic Studies, Bucharest, Romania

**Corresponding author, antoaneta.amza@csie.ase.ro*

Vlad BOLOVĂNEANU

Bucharest University of Economic Studies, Bucharest, Romania

vlad.bolovaneanu@ase.ro

Daniel Traian PELE

Bucharest University of Economic Studies, Bucharest, Romania

IDA Institute Digital Assets

Institute for Economic Forecasting, Romanian Academy, Bucharest, Romania

danpele@ase.ro

Abstract. *In recent years, central bank communication has become an important part of modern monetary policy. Beyond setting interest rates, central banks have started to use communication to signal their views on the future economic path.*

A growing number of studies suggest that the tone extracted from central bank reports or policy statements can improve forecasts for inflation and other macroeconomic variables. However, most of the evidence comes from developed countries or the euro area while emerging countries remain relatively underexplored. Our paper addresses this gap by examining whether narratives contained in official monetary policy documents improve inflation forecasting accuracy in Romania.

We used quarterly data from the first quarter of 2007 to the fourth quarter of 2024 to construct a sentiment indicator derived from the content of monetary policy minutes and complementary press releases published by the National Bank of Romania. The sentiment measure is determined by using a transformer-based language model trained on financial text and it is then included in a multivariate Bayesian time series model along with inflation. To evaluate its usefulness, we generated forecasts in a recursive out-of-sample framework at horizons of one to four quarters ahead which were then evaluated relative to a univariate autoregressive benchmark.

The results showed that the inclusion of the sentiment index extracted from the monetary policy documents published by the central bank improves forecast accuracy, especially in calm periods. These findings suggest that central bank narratives can serve as a complementary source of information for inflation forecasting in an emerging economy context.

Keywords: Inflation forecasting, Central bank communication, Monetary policy minutes, Sentiment analysis, Text as data, Bayesian vector autoregression, Emerging economies.

Introduction

He main mandate of most central banks is related to price stability, which makes inflation forecasting a key input into monetary policy decisions. Traditionally, forecasting models rely mainly on structured historical macroeconomic data that are compiled from official statistics. However, recent advances in natural language processing have made it possible to incorporate unstructured textual information into econometric or machine learning forecasting models. In this context, since central bank communications serve to guide expectations and contain information

about how policymakers assess inflationary pressures, risks and economic developments, it is interesting to analyze whether they contain forward-looking information that is not fully captured by historical macroeconomic data.

Therefore, this paper examines if narratives found in monetary policy minutes add useful predictive information for inflation. The issue is particularly relevant for Central and Eastern European (CEE) economies. In these countries inflation is often influenced by external shocks and communication plays an important role in guiding expectations.

More specifically, the research question of the paper is whether sentiment extracted from monetary policy minutes improves the accuracy of inflation forecasting relative to standard time series benchmarks. As a starting point, the analysis is based on two main assumptions. First, sentiment derived from minutes contains relevant information about future inflation dynamics and improves forecast accuracy when it is incorporated into an econometric model. Second, the predictive contribution of sentiment is likely to be stronger at medium-term horizons, reflecting the forward-looking nature of monetary policy discussions.

Our paper contributes to the literature in two ways. First, it provides such predictive evidence for Romania, which is part of the CEE region that remains relatively underexplored in macroeconomic forecasting using textual data. Second, it constructs a sentiment indicator from central bank communication derived using FinBERT, a transformer model trained on financial text.

From an empirical perspective, this paper evaluates the forecasting value of monetary policy minutes-based sentiment within a Bayesian vector autoregression, which is estimated recursively in an expanding window. For this purpose, forecasts are generated for horizons from one to four quarters and are compared with a standard autoregressive benchmark. The results indicate that adding sentiment to the Bayesian VAR consistently improves the accuracy of inflation forecasts, with gains increasing with the forecast horizon and under calm economic regimes. This supports the idea that the policy narratives embedded in minutes contain useful information for inflation forecasting at medium-term horizons, especially for periods when the economic environment is not affected by strong external shocks.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 describes the data, construction of the sentiment indicator and the econometric methodology. Section 4 presents and discusses the results, while Section 5 concludes.

Literature review

Recent advances in natural language processing (NLP) have made it increasingly feasible for economists to use unstructured text data as quantitative input in empirical analyses. According to the text as data paradigm, formalized by Gentzkow et al. (2019), language is transformed into a numerical representation that can be included in econometric and machine learning models. The increasing availability of digital text has opened new possibilities for macroeconomic research. The value-added is even greater for analysis where key variables such as inflation, gross domestic product, or unemployment are observed at low frequencies and released with substantial delays. In this context, textual data can complement traditional time series, both by helping to forecast future developments and by providing timely signals about current macroeconomic conditions, a practice often referred to as nowcasting (Banbura et al., 2013).

Early studies in this area relied primarily on sentiment measures based on dictionaries and bag-of-words methods. Although they have the advantage of providing transparency, the disadvantages are related to the fact that these methods have a limited ability to capture semantic context and changes in language over time. Subsequent research has introduced more flexible tools, such as topic modeling and word embeddings, that allow researchers to identify underlying themes

and semantic relationships in large textual datasets (Blei et al., 2003; Mikolov et al., 2013). More recently, transformer models and large language models (LLMs) have further extended the scope of text analysis, producing representations that take context into account. Also, survey papers such as those by Marcucci (2024) document the growing use of these tools in macroeconomic forecasting. They present evidence that forecasts using LLMs for inflation and unemployment can outperform traditional benchmarks, including professional forecasters, in certain contexts.

Central bank communication is a particularly useful source of textual data in macroeconomics. Official publications such as inflation reports, monetary policy statements, meeting minutes or speeches provide insights into how policymakers interpret economic conditions, inflationary risks and future policy directions. A growing body of literature shows that these communications contain information beyond that conveyed by policy actions alone. For example, Boukus and Rosenberg (2006) demonstrate that FOMC minutes provide information relative to policy rate decisions, while Hansen and McMahon (2016) show that changes in tone of central bank communication generate “language shocks” with effects on financial markets and macroeconomic outcomes. Importantly, narrative aspects contain information that is not fully captured in quantitative forecasts. An important contribution is provided by Hansen et al. (2019) who find that narratives from Bank of England inflation reports about risks and uncertainty contain predictive information not fully captured by published numerical projections.

These results have stimulated a growing area of research that incorporates textual indicators directly in macroeconomic forecasting models. Ferreira (2021) proposes econometric models that augment vector autoregressions and dynamic factor models with latent textual factors extracted from central bank communications, finding improvements in forecasting accuracy for key macroeconomic variables. Also, Clements and Reade (2020) show that sentiment extracted from Bank of England reports improves macroeconomic forecasts when incorporated into econometric models.

A more specific literature looks at the importance of central bank communication in forecasting inflation. Early studies typically relied on dictionary approach or dovish - hawkish measures derived from policy statements or monetary policy minutes. Recent studies are using data-driven representations. Araujo et al. (2025) show that textual embeddings can yield predictors that improve inflation forecasts, particularly at medium-term horizons, even after controlling for the central bank’s own projections.

Also recently, LLMs have been used both as forecasting tools and as instruments for extracting structured information from policy narratives. Allard et al. (2024) introduce a model base on BERT architecture that was fine-tuned to capture sentiment related to inflation from news and demonstrate that incorporating the resulting sentiment index into a standard nowcasting framework improves inflation predictions during periods of elevated volatility. In a related manner, Faria-e-Castro and Leibovici (2024) show, in retrospective exercises, that inflation forecasts generated by LLMs can outperform the Survey of Professional Forecasters, while Zarifhonarvar (2025) provide evidence that LLM-generated inflation expectations exhibit biases similar to those observed in household surveys. These suggest that such models can be used as experimental tools to evaluate the effects of communication strategies on expectations.

Despite these advances in research, empirical results for emerging economies are still relatively scarce. Most studies tend to focus on advanced economies, leaving open the question of how well these approaches translate to different institutional and macroeconomic conditions. In the context of emerging countries, Simionescu (2025) shows that incorporating sentiment extracted from the National Bank of Romania’s (NBR) quarterly Inflation Reports improves inflation forecasting performance. Another contribution is provided by Nițoi et al. (2023), who fine-tuned a

transformer model on central bank communication, for the purpose of predicting the monetary policy rate. Similarly, de Jesus and da Nóbrega Besarria (2025) analyze public documents from the Central Bank of Brazil and find that sentiment indicators constructed using a time-varying improve forecasts of inflation and GDP growth.

Methodology

From a methodological perspective, this paper draws on the intuitive forecasting comparison proposed by Araujo et al. (2025), who evaluate the marginal predictive contribution of European Central Bank communication by comparing simple autoregressive benchmarks with vector autoregressions augmented with textual information. We adopt this general framework, but we have several key differences. First, we focus on monetary policy minutes and press releases rather than on short excerpts from ECB press conference statements. Second, we extract sentiment with a transformer specifically fine-tuned for economic and financial text instead of employing a dictionary-based approach. Finally, the analysis is carried out in the context of an emerging economy, where the results may differ from those observed in advanced economies.

Therefore, our focus is on enhancing traditional inflation forecasting models with inputs from official documents related to monetary policy meetings of the National Bank of Romania Board. These documents summarize policy deliberations and include forward-looking assessments of macroeconomic developments, including the inflation projections.

Monetary policy minutes are available from Q4 2016 to Q4 2024 and represent the main source of textual information used in the analysis. However, to extend the sample period of the analysis and ensure a sufficiently long estimation sample, we supplemented it with sentiment derived from NBR press releases starting from Q1 2007. By applying this strategy, we enlarge our series without a meaningful loss of informational content. One argument in favor of our approach is the Pearson correlation between the sentiment indicators computed from minutes and from press releases on the overlapping period, 0.71, indicating strong linear co-movement between the two sources. Thus, we construct a unified quarterly sentiment indicator from Q1 2007 to Q4 2024.

All monetary policy documents are obtained from the official website of the National Bank of Romania. Year-on-year quarterly inflation data for Romania were gathered for the period Q1 2007 to Q4 2024 from Eurostat.

Before extracting sentiment, documents were cleaned to remove institutional content that does not contain information about macroeconomic conditions. This meant in particular to exclude document headers as well as standard sections listing the members of the Board present at each meeting and their official titles.

After preprocessing, the remaining narrative content was divided into individual sentences, which serve as the unit of analysis for the sentiment classification.

The next step was to extract the sentiment indicator using FinBERT (Araci, 2019), a publicly available BERT encoder fine-tuned on financial classification datasets. The model has three probabilities as outputs for each sentence: positive, negative, and neutral sentiment.

Let $p_{d,s}^+$, $p_{d,s}^-$, and $p_{d,s}^0$ denote the positive, negative, and neutral probabilities assigned by FinBERT to sentence s in document d . We define the sentence-level net sentiment as

$$Sent_{d,s} = p_{d,s}^+ - p_{d,s}^- \quad (1)$$

The document-level sentiment indicator is obtained by averaging across all S_d sentences in document d :

$$Sent_d = \frac{1}{S_d} \sum_{s=1}^{S_d} Sent_{d,s} \quad (2)$$

where S_d denotes the total number of sentences in document d , $s = 1, \dots, S_d$, and d indexes individual monetary policy documents.

We assume an end-of-quarter forecasting scheme, which allows us to use report(s) in the quarter of interest before inflation is computed. To construct a quarterly sentiment series, in quarters in which multiple monetary policy minutes were published, the corresponding sentiment indicators are averaged, resulting in a single quarterly observation.

As for the econometric approach, past inflation values and sentiment data were included in a Hierarchical Bayesian VAR (BVAR) model as defined in (Giannone et al., 2015). The model can be specified as:

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + \epsilon_t \quad (3)$$

$$\epsilon_t \sim N(0, \Sigma)$$

where y_t is a vector of endogenous variables including inflation and the sentiment indicator, p denotes the number of lags, and ϵ_t is a vector of structural shocks. The coefficient matrices a_i and the covariance matrix Σ are estimated using a Minnesota-type prior. Similarly to (Araujo et al., 2025), we set $p = 4$.

Model estimation and forecasting are carried out in a recursive out-of-sample setting. The initial estimation window includes 20 quarters, which corresponds to five years of data. At each forecast point in time, the model was recursively re-estimated as new observations became available, and inflation forecasts are generated for horizons of one to four quarters ahead. This setup allows us to evaluate the model's performance in a way that closely replicates real-time expert forecasting. In total, there are $W = 49$ windows. The MCMC procedure uses 1000 burn-in draws and 4000 retained draws, from which the median is derived and used as a point forecast for inflation. Finally, we compare the BVAR forecasts with those obtained from a simple but widely used autoregressive (AR) benchmark model.

The lag order of the AR benchmark is set equal to the lag length used in the BVAR specification. This ensures comparability between models and avoids differences in forecast performance from alternative lag selections.

Forecast accuracy is evaluated using the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) computed per horizon:

$$RMSE_h(y, \hat{y}) = \sqrt{\frac{1}{W} \sum_{w=1}^W (y_{w,h} - \hat{y}_{w,h})^2} \quad (4)$$

$$MAE_h(y, \hat{y}) = \frac{1}{W} \sum_{w=1}^W |y_{w,h} - \hat{y}_{w,h}| \quad (5)$$

where $y_{w,h}$ and $\hat{y}_{w,h}$ represent the actual and predicted value for window w and horizon h , respectively.

Finally, we rely on the Diebold–Mariano test to formally compare the predictive accuracy of competing models and determine whether the improvements associated with sentiment including are statistically significant.

Results and discussions

Before comparing the forecasting performance of the models, it is helpful to first take a closer look at the behavior of the sentiment indicator over time. Figure 1 shows the evolution of the quarterly monetary policy sentiment index between 2007 and 2024.

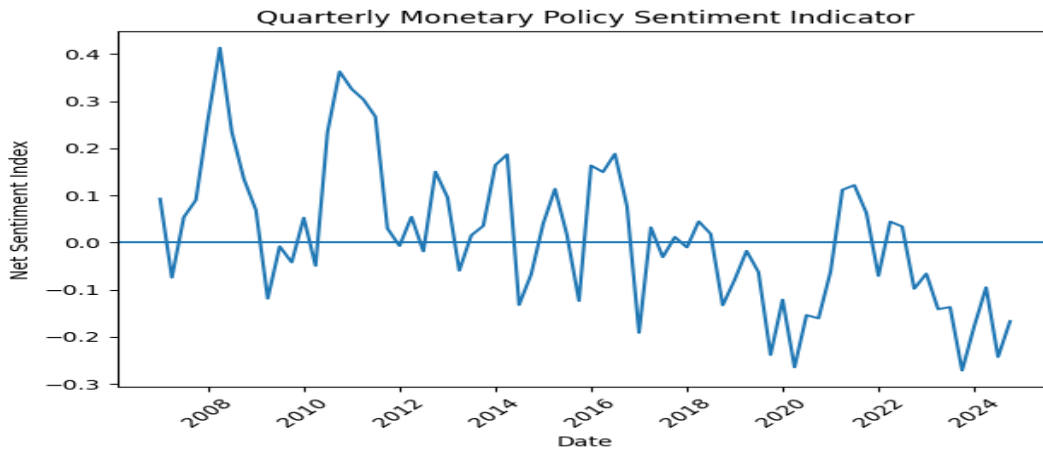


Figure 1. Quarterly monetary policy sentiment indicator (2007–2024)

The sentiment indicator shows notable variation over time, with pronounced declines during the global financial crisis, the COVID-19 pandemic, and the increase in inflation after 2021. The dynamics suggest that the index captures shifts in the tone of monetary policy communication consistent with major macroeconomic tensions.

Table 1 presents the out-of-sample forecast accuracy for the Bayesian VAR augmented with sentiment extracted from minutes versus the autoregressive benchmark, evaluated using RMSE and MAE at horizons from one to four quarters ahead. Overall, the results show that the sentiment-augmented model performs better than the benchmark across all horizons. These results indicate that incorporating information extracted from monetary policy minutes improves inflation forecasting accuracy relative to a model based solely on past inflation dynamics.

The magnitude of the improvement increases with the forecast horizon. At the one-quarter-ahead horizon, the reduction in forecast errors is modest, while at horizons of three and four quarters ahead the gains become more pronounced. This pattern is consistent with the hypothesis that policy narratives contained in monetary policy minutes convey forward-looking information that is more relevant for medium-term inflation dynamics than for short-term fluctuations. In contrast, the autoregressive benchmark relies exclusively on historical inflation and therefore lacks access to such narrative signals.

Table 1. Out-of-sample RMSE and MAE for the BVAR and AR (benchmark) models for 1 to 4 quarters-ahead forecasting. Best values are highlighted in bold.

<i>Horizon</i>	<i>RMSE_{BVAR}</i>	<i>RMSE_{AR}</i>	<i>MAE_{BVAR}</i>	<i>MAE_{AR}</i>
h1	1.40	1.49	1.11	1.17
h2	2.18	2.44	1.64	1.93
h3	2.94	3.32	2.26	2.67
h4	3.63	4.17	2.82	3.36

Source: Authors' own research.

Additional insight is provided by Figure 2, which compares actual inflation with forecasts generated by the BVAR and AR models at selected horizons. One-quarter-ahead predictions on the left track the actual inflation rate closely, with minimal differences. For farther horizons, on the other hand, variability increases. On the right, we can see that BVAR forecasts track actual inflation more closely during calm periods of relative macroeconomic stability (Q1 2012 – Q4 2019), while in the turbulent period of Q1 2020 – Q4 2024, both models fail to predict the inflation spike in time.

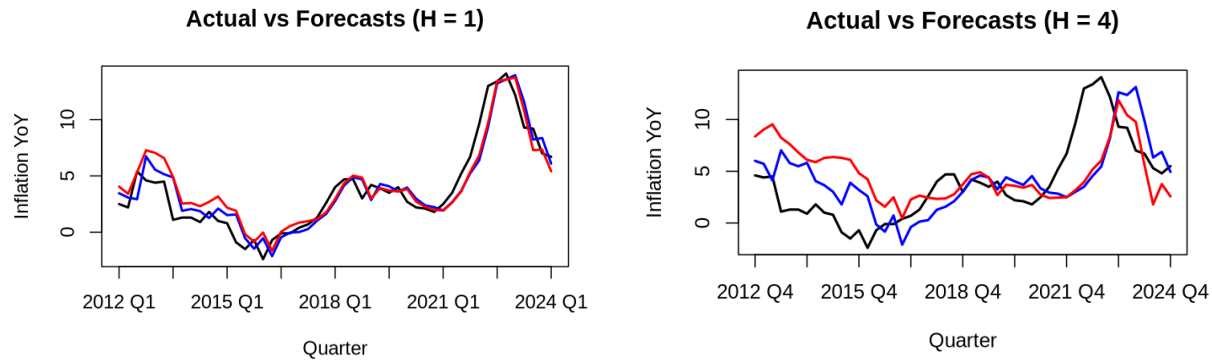


Figure 2. Actual, BVAR, AR forecasts for 1 and 4 prediction horizons

To expand on the previous observations, we formally evaluate whether the differences in forecasting accuracy between the sentiment-augmented BVAR model and the autoregressive benchmark are statistically significant in the two specified regimes. The Diebold–Mariano (DM) test for equal predictive accuracy is applied to each horizon and regime.

Table 2 presents the DM test statistics for each forecast horizon across the two regimes. Negative sign means that BVAR outperforms the AR benchmark. During the calm period, the BVAR model is shown to outperform the AR benchmark in a statistically significant manner. However, in the turbulent regime, the AR benchmark performs slightly better, but without the DM test rejecting the null hypothesis. Our results suggest that sentiment extracted from monetary policy communication provides statistically significant improvements in inflation forecasting during periods of relative macroeconomic stability, while periods characterized by large and unexpected shocks show a diminished contribution. One possible explanation is that during episodes of heightened uncertainty, inflation dynamics are driven more strongly by exogenous shocks, such as geopolitical tensions, oil crises, or pandemics, than by signals embedded in central bank communication.

Table 2. Diebold–Mariano test statistics for BVAR vs. AR forecasts

Horizon	Calm	Turbulent
h1	-1.8*	1.16
h2	-2.37*	1.50
h3	-2.02*	1.81
h4	-1.68**	1.66

Note: * significant at 5%, ** significant at 10%. Negative values indicate lower forecast errors for the BVAR model.

Source: Authors' own research.

In line with recent findings in the literature on macroeconomic forecasting using textual data (Araujo et al., 2025), our evidence suggests that narrative indicators are particularly valuable over longer horizons, where expectations, credibility, and assessments of persistence play a more important role. These findings are especially relevant in the context of an emerging economy where such communication plays a key role in shaping expectations and signaling policy decision-making, particularly for internal shocks.

Conclusion

This paper investigates whether the narrative contained in monetary policy minutes can improve inflation forecasts when included in a standard econometric model. By extracting a sentiment index from documents published by the National Bank of Romania, using a transformer model tailored to economic and financial language, and incorporating it into a Bayesian vector autoregression, we

obtain improvements in inflation forecasts, especially in calm periods. The gains in forecast accuracy are observed for several forecast horizons and become more pronounced in the medium term, indicating the forward-looking nature of monetary policy communication.

While most of the studies have focused on advanced economies, our findings contribute to the growing literature on text-based macroeconomic forecasting by providing new evidence in the context of an emerging economy.

The approach adopted in this paper has practical implications. In particular, sentiment measures derived from official communications may serve as useful complementary sources of information for macroeconomic forecasting models. More broadly, the methodology can be extended to other central banks in the Central and Eastern European region, allowing for cross-country comparisons of communication effectiveness and its role in shaping inflation expectations.

However, our analysis also has some limitations and future research could provide additional empirical evidence on this topic. First, the sentiment indicator is constructed using a single NLP model and aggregation scheme, and alternative representations of textual information may produce different results. Second, for the forecasting exercise we focused only on two variables and a single benchmark model, leaving scope for further comparisons with richer multivariate or nonlinear models. Third, our analysis assumes a specific timing structure for the availability of policy documents. While consistent with the forecasting design, it may not fully capture real-time information constraints faced by central bank experts.

There are also several natural extensions. Future studies could compare different approaches to sentiment measurement, expand the information set with additional macroeconomic indicators, or apply the same methodological framework to other Central and Eastern European economies. A more ambitious direction would be to go beyond aggregate sentiment and extract more granular signals from text, assessing their contribution to forecast performance.

Acknowledgements

This paper is supported through the project "IDA Institute for Digital Assets", CF166/15.11.2022, contract number CN760046/ 23.05.2023; the project "AI for Energy Finance (AI4EFin)", CF162/15.11.2022, contract number CN760048/23.05.2023, financed under the Romania's National Recovery and Resilience Plan, Apel nr. PNRR-III-C9-2022-I8; and the Marie Skłodowska-Curie Actions under the European Union's Horizon Europe research and innovation program for the Industrial Doctoral Network on Digital Finance, acronym DIGITAL, Project No. 101119635.

This paper was co-financed by the Bucharest University of Economic Studies during the PhD program.

During the preparation of this work, the authors utilized ChatGPT 5.5 to enhance readability and language clarity. Following the use of this tool, the authors meticulously reviewed and revised the content as necessary and assume full responsibility for the final content of the publication.

References

- Allard, M. A., Teiletche, P., & Zinebi, A. (2024). Enhancing inflation nowcasting with LLM: Sentiment analysis on news. *arXiv preprint arXiv:2410.20198*.
- Araujo, D., Borovicka, J., & Gopinath, G. (2025). *Word2Prices: Embedding central bank communication for inflation prediction* (BIS Working Papers No. 1253). Bank for International Settlements.

- Araci, D., (2019). *Finbert: Financial sentiment analysis with pre-trained language models*. arXiv preprint arXiv:1908.10063.
- Bañbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In *Handbook of economic forecasting* (Vol. 2, pp. 195-237). Elsevier.
- Bholat, D., Hansen, S., Santos, P., & Schonhardt-Bailey, C. (2015). Text mining for central banks. *Centre for Central Banking Studies Handbook No. 33*. Bank of England.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*.
- Boukus, E., & Rosenberg, J. V. (2006). *The information content of FOMC minutes* (Staff Report No. 278). Federal Reserve Bank of New York.
- Clements, M. P., & Reade, J. J. (2020). Forecasting and forecast narratives: The Bank of England inflation reports. *International Journal of Forecasting*, 36(4), 1488-1500.
- de Jesus, D. P., & da Nóbrega Besarria, C. (2025). Central Bank Narratives and Macroeconomic Forecasting: Using Textual Analysis from Machine Learning: D. Pitta de Jesus et al. *IMF Economic Review*, 1-56.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
- Faria-e-Castro, M., & Leibovici, F. (2024). Artificial Intelligence and Inflation Forecasts. *Federal Reserve Bank of St. Louis Review*, 106(12), 1-14.
- Ferreira, L. N. (2021). Forecasting with VAR-teXt and DFM-teXt Models: exploring the predictive power of central bank communication. Banco Central do Brasil.
- Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535–574.
- Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2), 436-451.
- Hansen, S., & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, 99, S114-S133.
- Hansen, S., McMahon, M., & Tong, M. (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics*, 109, 1–19.
- Marcucci, J. (2024). Macroeconomic forecasting with text-based data. *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, 425-468.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Nițoi, M., Pochea, M.-M., & Radu, Ș.-C. (2023). Unveiling the sentiment behind central bank narratives: A novel deep learning index. *Journal of Behavioral and Experimental Finance*, 38, 100809. <https://doi.org/10.1016/j.jbef.2023.100809>
- Simionescu, M. (2025). Machine learning vs. econometric models to forecast inflation rate in Romania? The role of sentiment analysis. *Mathematics*, 13(1), 168.
- Wang, R., Sun, M., & Wang, L. (2025). From news to trends: A financial time series forecasting framework with LLM-driven news sentiment analysis and selective state spaces. *Journal of Intelligent Information Systems*, 63(6), 1955–1980. <https://doi.org/10.1007/s10844-025-00971-3>
- Zarifhonarvar, A. (2025). Generating inflation expectations with large language models. *Journal of Monetary Economics*, 103859.