Journal of Behavioral and Experimental Finance 38 (2023) 100809

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

Unveiling the sentiment behind central bank narratives: A novel deep learning index☆

Mihai Niţoi [a], Maria-Miruna Pochea [b,*], Ştefan-Constantin Radu [c]

[a] Institute for World Economy, Romanian Academy, Bucharest, Romania
[b] Department of Finance, Babeş-Bolyai University, Cluj-Napoca, Romania
[c] School of Advanced Studies of the Romanian Academy, National Institute of Economic Research, Bucharest, Romania

ARTICLE INFO

Article history:
Received 1 September 2022
Received in revised form 27 January 2023
Accepted 7 April 2023
Available online 11 April 2023

JEL classification:
G40
E58

Keywords:
Sentiment analysis
Natural language processing
BERT
Central bank communication
Forward guidance

ABSTRACT

This paper proposes a new framework for analyzing the sentiments of central bank narratives. Specifically, we fine-tune a pre-trained BERT model on a dataset of manually annotated sentences on monetary policy stance. We derive a deep learning domain-specific model—BERT central bank sentiment index—ready for sentiment predictions. The proposed index performs similarly to other measures in capturing financial uncertainty. Also, the sentiment index is less noisy and has the ability to forecast the future path of policy stance, augmenting the standard Taylor rule. Finally, compared to other lexicon-based sentiment indicators, our deep learning index has a higher predictive power in anticipating policy rates changes. Our framework enables future possible research in developing more accurate sentiment indicators for central banks in both advanced and emerging countries.

1. Introduction

During the global financial crisis, the European debt crisis, and, more recently, the pandemic crisis, central banks have used unprecedented tools to stabilize the economies and markets. Within their frameworks, communication has become a key tool for managing the monetary policy beliefs of all financial market participants. The role of communication has evolved over time from an instrument that supports the traditional monetary policy framework toward a tool that steers financial markets and succours central banks in achieving their macroeconomic goals. Nowadays, the tone of central banks can move the markets more than the economic fundamentals (Gardner et al., 2022) and central banks press conferences are on the spotlight of both media and academics.

The number of studies that aim to capture the sentiments behind central banks statements has grown rapidly. Initially, most papers transformed textual information coming from central banks communication into quantitative measures by using lexicon-based approaches. The main drawback of lexicon-based approaches is that they do not account for the context of the keywords. In recent years, finance researchers have embraced the use of machine learning and deep learning techniques for textual analysis. The main advantage of these approaches is that they can learn sentiment weights for words and even sentences, being able to capture the sentiment of an entire expression. Nevertheless, they require substantial labeled datasets that imply considerable time and costs to build (Shapiro et al., 2022).

The main objective of this paper is to construct a novel deep learning sentiment index for central banks narratives. We explore the following research questions: (1) Does this new central bank sentiment index vary if the language representation model is retrained on a different corpus of annotated sentences? (2) How does our index relate to other financial uncertainty measures? (3) How does our index behave compared to other central bank lexicon-based sentiment indicators? (4) Can our proposed index anticipate future policy rate changes? (5) How does the predictive power of our index perform compared to other lexicon-based sentiment indicators?

For building the sentiment index, we use a corpus of 591 central banks minutes, totalizing 50,451 sentences. Also, the empirical analysis is achieved with the help of Google's Bidirectional

☆ Funding: This work was supported by a grant of the Romanian Ministry of Research, Innovation and Digitization, CNCS - UEFISCDI, project number PN-III-P1-1.1-TE-2019-1702, within PNCDI III.

* Corresponding author.
E-mail addresses: mihai.nitoi@iem.ro (M. Niţoi), miruna.pochea@econ.ubbcluj.ro (M.-M. Pochea), stefan.radu@iem.ro (Ş.-C. Radu).

https://doi.org/10.1016/j.jbef.2023.100809
2214-6350/

Encoder Representations from Transformers (BERT), trained on a corpus of more than 3,300,000 million words, with over 110 million parameters (Devlin et al., 2018).

Our study enhances the existing literature in three keyways. First, we propose a new framework to construct a deep learning indicator for capturing the sentiment of central banks communications. By relying on a pre-trained BERT model, we fine-tune the model by re-training it on our dataset of labeled sentences on central bank monetary policy stance in order to transform its word embeddings into hawkish, neutral, and dovish. The model is then used to predict the stance of central banks minutes and to construct a BERT central bank sentiment index (hereinafter, BERT-CBSI). Afterwards, we run several tests to check the robustness of our index. We fine-tune the financial domain specific FinBERT model by re-training it on our corpus of labeled sentences on central bank monetary policy stance. Our findings reveal that BERT-CBSI accuracy is slightly higher compared to that of FinBERT. However, when both models are fine-tuned on a small labeled dataset, FinBERT performs better. We show that BERT-CBSI does not change if the model is fine-tuned on a different corpus of central bank annotated sentences. We find that BERT-CBSI performs similarly to other measures in capturing financial uncertainty. We verify if our deep index comprises valuable information for future monetary policy decisions. The results show that BERT-CBSI has the ability to forecast the future path of policy stance, augmenting the standard Taylor rule. Also, we compare the performance of our indicator with other lexicon-based indices. We find that BERT-CBSI clearly outperforms other lexicon-based measures.

Second, we extend the literature by publishing the minutes corpus that we gathered for creating BERT-CBSI, but, more importantly, the corpus of sentences manually annotated according to the monetary policy stance, i.e., hawkish, dovish, and neutral. To our knowledge, this is the first public database on labeled sentences on monetary policy stance. The minutes corpus comprises all the minutes released by Czechia, Hungary, Poland, and Romania, between January 2007 and January 2022. The sentence-by-sentence sentiment-labeled central bank dataset encompasses 1,998 sentences. Both the corpus of text and the labeled sentences dataset are available by accessing the Harvard Database.[1] By publicly and freely providing these data, we facilitate future research on measuring the sentiment of central banks communications in other emerging or developed countries and different research applications of these sentiment scores.

Third, most of the literature is focused on the sentiment expressed by the Federal Reserve (Fed) narratives (Bennani, 2020; Gardner et al., 2022; Shapiro and Wilson, 2022) and by the European Central Bank (ECB) communication (Picault and Renault, 2017; Picault et al., 2022; Parle, 2022). There are some sporadic papers focused on the materials provided by other central banks, such as Bank of Korea (Lee et al., 2019) or Bank of England (Clements and Reade, 2020). In this study, we extend the monetary policy sentiment literature by analyzing the minutes of four central banks in Central and Eastern Europe, i.e., Czechia, Hungary, Poland, and Romania. These developing countries are related through their comparable monetary policy strategies.

The rest of the paper is structured as follows. Section 2 reviews the literature. Section 3 describes the data. Section 4 presents the methodology. In Section 5 we reveal and discuss the results, together with a battery of robustness tests. Finally, Section 6 concludes the paper.

1 https://doi.org/10.7910/DVN/40JFEK

2. Related literature

There are numerous studies that investigate central bank communication. This paper lies at the crossroad of two strands of literature. The first examines the classification of central bank communication by using textual analysis and has unceasingly advanced in the literature. The seminal papers of Poole (1971), Potts and Luckett (1978), and Boschen and Mills (1995) were among the first that computed a time-varying monetary policy index for the narratives of the Federal Open Market Committee (FOMC). Generally, these studies manually assign values ranging on different scales, e.g., −1 (expansionary policy), 0 (neutral), and +1 (restrictive), to FOMC narratives. This approach has also been applied for ECB communication (Berger et al., 2011; Ehrmann and Fratzscher, 2007; Gerlach, 2007; Rosa and Verga, 2007). The main drawback of this method is that the interpreting and the labeling of central bank documents can be biased by human subjectivity. Also, reading all central bank documents can be time-consuming.

Another important body of research resorts to natural language processing (NLP) techniques to extract the tone of central bank narratives. In this area, the most common method is the lexical approach, which automatically searches and counts for specific words within a document. Jansen and De Haan (2007) were among the first who applied a word-count approach to assess the ECB communication on price stability. In an influential development, Loughran and McDonald (2011) proposed a financial specific dictionary, based on 10-Ks reports, that can be used to assess the sentiment of financial documents. Their dictionary has been widely used to measure central bank sentiment (Shapiro and Wilson, 2022; Armelius et al., 2020; Hansen and McMahon, 2016). The main drawback of the Loughran and McDonald (2011) dictionary is that it is more suited for financial communication documents. Recently, Correa et al. (2021) proposed a central bank specific dictionary aiming to capture the tone of financial stability reports. However, the main shortcoming of dictionary-based models is that they rely only on a single word search and do not take into consideration the context of the sentence. To tackle this drawback, by using a combination of methods, Apel and Blix-Grimaldi (2014), Picault and Renault (2017), Apel et al. (2021), and Gardner et al. (2022) developed specific dictionaries for the narratives of Swedish Central Bank, ECB, and FOMC, while Gonzalez and Tadle (2021) proposed particular dictionaries for 21 central banks.

The second strand of the literature assesses the effects of the central bank textual-based indicators on specific economic and financial outcomes or whether these indicators anticipate future monetary policy decisions. One of the most detailed papers in this field is that of Boschen and Mills (1995) who reveal that changes in monetary policy indices are associated with persistent changes of M2 and monetary base and with transitory changes in short-term interest rates. Afterwards, a plethora of papers investigate the impact of central bank sentiment indicators on economic and financial variables. In another extensive study, Ehrmann and Fratzscher (2007) show that the sentiment indicators of the Fed, Bank of England, and ECB are important drivers for asset prices, e.g., interest rates, equity market, and exchange rate. The effects are more pronounced for the narratives of the Fed communication tools. A similar result is emphasized by Armelius et al. (2020) who document significant spillovers induced by the tone of the Fed communication. Similarly, other studies demonstrate that central bank communications influence the interest rates (Lucca and Trebbi, 2009; Hansen and McMahon, 2016; Schmeling and Wagner, 2019), on exchange rates (Conrad and Lamla, 2010; Jansen and De Haan, 2007; Burkhard and Fischer, 2009; Rosa, 2013; Dewachter et al., 2014), and on equity markets (Sadique et al., 2013; Born et al., 2014; Schmeling and Wagner, 2019;

Picault and Renault, 2017; Cieslak and Schrimpf, 2019). Also, Heinemann and Ullrich (2007), Apel and Blix-Grimaldi (2014), Picault and Renault (2017), Lee et al. (2019), and Apel et al. (2021) show that central bank sentiment indicators are good predictors for future policy rates.

In this paper, we advance from the existent literature by proposing a new approach to measure central bank sentiments. We provide evidence that a deep learning sentiment indicator is a good predictor for future policy rate decisions. Also, the sentiment index matches other indicators in capturing financial and economic uncertainty and performs better in comparison to other lexicon-based sentiment indices.

3. Data

For building the central bank sentiment indices, we manually collected a corpus of minutes, which are published after each monetary policy meeting for four central banks, classified as inflation targeters, in the following developing countries: Czechia, Hungary, Poland, and Romania. We chose the minutes because they are one of the most important communication instruments of the central banks, covering a broad range of monetary, economic subjects, and financial subjects and have the power to form market expectations. Also, the time-varying nature of each central bank minutes is quite homogeneous. Due to these particularities, we consider that the minutes are the most appropriate tool to unveil the sentiments behind the words used by central banks.

The data corpus consists of 591 monetary policy minutes, with a total of 50,451 sentences, for the period between January 2007 and January 2022, hand-collected from the central banks websites. This timespan has been marked by significant turbulences, which had major impact on the economic activity and significant effects on central banks decisions and communications. Furthermore, we believe that sentiments were more pronounced in central banks minutes, as they tried to better anchor market expectations on future policy decisions. Therefore, this period proves to be appropriate and fruitful to investigate our research hypotheses.

Table 1 provides some descriptive statistics for each central bank corpus. The minutes corpus differs in terms of length: Hungary has the longest minute, while Czechia prefers a more concise minute. However, compared to Poland and Romania, the minutes of Czechia and Hungary have a similar readability score,[2] requiring the lowest number of years of education needed to understand the minutes. Romania records the highest readability score and the highest average words per sentence. Even if we notice some differences in terms of length and readability, Fig. A.1 shows that the topics covered by each central bank are quite similar, e.g., economic activity, growth, inflation outlook, labor market, interest rate, or monetary policy.

4. Methodology

4.1. A deep learning central bank sentiment index: BERT-CBSI

Over the past decades, a variety of methods have been used for NLP tasks on sentiment analysis. Specifically, for the financial sentiment analysis purpose, the models can be grouped in three broad categories: lexicon, machine learning, and deep learning approaches. A comprehensive literature review on the papers that employed these approaches can be found in Daudert (2021). The lexicon-based method uses a specific list of words, classified as positive, negative, or neutral, to identify and extract the sentiment of a text. The main drawback of the lexicon-based approaches is that they ignore the semantic information of a sentence. The machine-learning method uses a bag-of-words and term frequency inverse document frequency representation for a text. Afterwards, the matrix representation is passed to classifiers as Naive Bayes, Support Vector Machine or Maximum Entropy to extract sentiments. A detailed description of these approaches can be found in Tripathy et al. (2016). However, this category of models ignores the words order and the small training sets can pose modeling challenges (Shapiro et al., 2022). The deep learning approach extracts sentiments from a text by relying on similar labeled text data to learn and tune its embeddings parameters. The main drawback of these models is that they require a huge amount of human labeled text data for training and fine-tuning.

In our paper, we will employ a deep learning model. Hereinafter, we briefly present the general features of these models and the peculiarities of our approach. Over the recent years, transfer-learning models have gain significant popularity in extracting text sentiments. These models are based on the BERT model (Devlin et al., 2018), which has the power to generate context-aware word and document embeddings because it has memory of past inputs (Rayl and Sinha, 2022). In doing this, BERT relies on Transformers architecture (Vaswani et al., 2017). Also, compared to other word-embedding models, i.e., Word2vec or Embeddings from Language Model, BERT uses a bidirectional Transformer encoder to process the entire sequence of the words on both the left and right context simultaneously.

The input framework of the model is represented by token embeddings, segment embeddings, which are separate parts of a sentence, and position embeddings. The output is a sequence of vectors, where each vector corresponds to an input token with the same index (Devlin et al., 2018). The BERT model is pretrained in two unsupervised tasks: Masked Language Modeling and Next Sentence Prediction. The former randomly masks a fraction of 15% of all tokens in each sentence and then predicts the masked words. The latter predicts if two sentences follow each other. The pre-training for the BERT model is performed on a corpus with more than 3,300,000 million words – BookCorpus and English Wikipedia – with over 110 million parameters. Due to these particularities, BERT has outperformed the state-of-the-art machine learning models (Devlin et al., 2018).

The BERT model is appropriate for extracting sentiment predictions for standard text corpus, like the one on which it was pre-trained. For domain-specific text, the model must be further fine-tuned on a similar corpus in order to improve classification performance (Howard and Ruder, 2018). Therefore, we propose a novel fine-tuned BERT by re-training it on a dataset of central bank sentiment-labeled sentences, i.e., BERT-CBSI. For implementing BERT-CBSI and for computing central bank sentiment indicators, we will follow the next steps.

First, we randomly select a sample of sentences from the corpus of central banks minutes. Afterwards, we manually label each sentence. Specifically, we annotate each sentence by considering three sentiment measures, i.e., hawkishness, dovishness, and neutral. Generally, the hawkish sentiment is related to the expansionary phase of the economic cycle, being associated with the probability of a more aggressive monetary policy. Therefore, a sentence is labeled as hawkish if it refers, for example, to rising consumption, higher economic growth, rising investments, increasing resource utilization, higher inflation rate, low unemployment, high employment, policy tightening or balance sheet normalization. Conversely, the dovish sentiment is linked to the negative stage of the economic cycle, being correlated with an expansionary monetary policy. Consequently, a sentence is labeled as dovish if it refers, for example, to subdued consumption, decelerating economic growth, lower investments, falling resource

[2] The Flesch–Kincaid score measure the readability of a document in terms of years of education necessary to understand it.

Table 1
Descriptive statistics for central banks minutes.

	Start date	End date	No. of documents	Average words per document	Average sentences per document	Average words per sentence	Flesch–Kincaid readability score
Czechia	Jan.'07	Feb.'22	127	1301	48	25	14
Hungary[a]	Jan.'07	Jan.'22	180	4573	168	25	14
Poland	Apr.'07	Jan.'22	166	1956	60	30	17
Romania's press releases	Feb.'07	Aug.'22	77	651	16	38	21
Romania's minutes[b]	Sep.'16	Feb.'22	41	2767	57	45	23

Note: For Czechia, Hungary, and Poland the corpus includes the central bank minutes. In Romania, the central bank started to publish the minutes in September 2016. Until then, we used the press releases on the monetary policy decisions.

[a]In Hungary, the central bank minutes were structured, within the same document, into two main sections until 2013. The former details the evolution of the domestic economy and the external macroeconomic environment, while the latter discusses the views and the decision of the board on inflation, economic evolution, and forecasts. Starting with 2014, the central bank has published these two sections in separate documents. For a more comprehensive view, in our analysis, we will consider both documents.

[b]In Romania, the central bank publishes the minutes of the board meetings starting from September 2016. Until then, a press release was used to disseminate the board decisions.

Table 2
Number of manually labeled hawkish, dovish, and neutral sentences for each central bank.

	Hawkish sentences	Dovish sentences	Neutral sentences	Total sentences
Czechia	196	182	229	607
Hungary	198	205	197	600
Poland	133	149	108	390
Romania	160	115	126	401
Total	**687**	**651**	**660**	**1,998**

utilization, lower inflation rate, higher unemployment, lower employment, policy easing or unconventional monetary policy tools. Finally, the neutral measure refers to sentences that signals a balanced economic and inflation outlook, in line with forecasts, and no future policy change. Our approach is quite similar with the three broad categories used by Gertler and Horvath (2018) to classify central bank communication items and in line with the hawkish and dovish measures of Apel et al. (2021). The final sample comprises labeled sentences on which all there authors have agreed on their hawkish, dovish, and neutral stance and consists of 1,998 labeled sentences. The number of manually annotated sentences for each of the four central banks is listed in Table 2. The sample represents approximately 4.0% of the total corpus sentences. The complete data are presented in the .csv file in Harvard Database.[3]

Second, by employing the Python architecture for Transformers, we rely on the BERT-base model, which uses 12 transformers block, with a hidden size of 768, and a number of self-attention heads of 12, and has around 110 million trainable parameters. We fine-tune the BERT model by re-training it on our dataset of 1,998 labeled sentences on central bank monetary policy stance[4] and we obtain a deep learning domain-specific model, ready for sentiment predictions, the BERT-CBSI model.

Third, we calculate the tone of each minute. Specifically, each document is divided in sentences. Afterwards, the BERT model classifier predicts the probabilities (logits) of each sentence of being hawkish, dovish or neutral and computes the final sentiment score for each sentence as follows:

$$Sentence_{sentiment} = Logit_{Hawkish} - Logit_{Dovish} \quad (1)$$

Finally, the sentiment for each document is calculated by averaging sentence sentiment for all the sentiment in the document:

$$Minute_{sentiment} = \frac{\sum_{i=1}^{n} Sentence\_sentiment_i}{n} \quad (2)$$

All the computations are run by using the Python programming language.[5]

4.2. Predictive power of the BERT-CBSI

Once we have estimated the central banks sentiment index, we test whether the minutes tone signals the future interest rate decisions. Similar with Apel and Blix-Grimaldi (2014), we use an ordered probit model in which the dependent variable is a discrete measure of policy rate decision. Comparable specifications were also applied by Apel et al. (2021), Picault and Renault (2017), and Lee et al. (2019). Specifically, we first estimate the following basic equation:

$$PRD_{t+1} = \beta_1 PRD_t + \beta_2 SI_t + \varepsilon_t \quad (3)$$

where PRD_{t+1} is the future policy rate decision at $t+1$, which can take three values: -1 for easing the interest rate, 0 for no change, and $+1$ for tightening the interest rate; SI_t is the BERT sentiment index for the central bank minutes at t; and ε_{t+1} is the stochastic error term. Alternatively, for checking the results robustness and acquiring a more comprehensive analysis, we replace SI_t with a hawkish and dovish measure.

The main drawback of the above specifications is the exclusion of the available macroeconomic indicators, which could modify the results of the model. Specifically, the prediction power of the central banks tone could be lower if we consider other available indicators into the model. Therefore, we adopt a standard Taylor-type rule, by including the inflation gap and output gap into the equations[6]:

$$PRD_{t+1} = \beta_1 PRD_t + \beta_2 SI_t + \beta_3 \left(\pi_t - \pi^*\right) + \beta_4 \left(y_t - y^*\right)_t + \beta_5 ER_t + \varepsilon_t \quad (4)$$

3 https://doi.org/10.7910/DVN/40JFEK

4 The training corpus in divided in two samples: 90% is used for training and 10% for validation. For fine-tuning the BERT model, we use TensorFlow and PyTorch deep learning frameworks, with the following parameters: the batch size is equal to six; the number of training epochs is three; the learning rate and the epsilon parameter of the Adam optimizer are set to 2e−5 and, respectively, 1e−8.

5 To ensure the replicability of the results, the data and the methodology set-up are available at https://sites.google.com/view/bert-cbsi/.

6 Most of the Taylor-type rule specifications account for inflation expectations. Due to the lack of data on expected inflation for all countries, for the period 2007–2022, we did not include inflation expectations in our Taylor rule equation.

where $(\pi_t - \pi^*)$ is the inflation gap, computed as difference between the inflation rate and inflation target,[7] $(y_t - y^*)_t$ represents the output gap defined as difference between industrial production and potential output,[8] ER_t is the exchange rate. To ensure stationarity, we use in our estimations the difference for inflation gap, output gap, and exchange rate.[9]

4.3. Robustness

For checking the robustness of BERT-CBSI, we follow three approaches. First, we check whether the domain-specific FinBERT[10] performs better in extracting the sentiment from central bank communication compared to the BERT model. For doing this, we fine-tune the FinBERT model by re-training it on our corpus of manually labeled sentences, used to classify the sentiments of central bank minutes (hereinafter, we refer to this model as FinBERT-CBSI) and we extract the sentiment of each document, similarly with the steps used for the BERT-CBSI model.

Second, we check the robustness of our labeled sentences by using an automated method for labeling sentences. Specifically, we use the keywords, the hawkish modifiers, and the dovish modifiers proposed by Apel et al. (2021). For the neutral stance we use the same keywords and neutral modifiers, i.e., in line with forecasts, balanced, as expected, wait-and-see, moderate, temporary, negligible, unchanged, prudent, maintain or within the target. We use a 10-word window to find sentences that include keywords associated with specific hawkish, dovish, and neutral modifiers. We search the instances when the keyword and the modifier are adjacent, accounting also the instances when the keyword and the modifier are separated by other words in the sentence. Finally, we extract the sentences that include keywords and hawkish, dovish, and neutral modifiers. Even if this method is easy to implement, it has several drawbacks. First, numerous sentences include several modifiers. Therefore, they are automatically included in at least two categories. Second, the automatically extracted corpus includes a high number of extracted sentences, which can be time-consuming when it is used to fine-tune the BERT model. In order to limit these drawbacks, we briefly revise the corpus by clarifying the label for some sentences and by limiting the similar sentences. The final sample consists of 9,154 automated labeled sentences, which is used to fine-tune the BERT model.

Third, one could argue that using a fine-tuned BERT model to extract central bank sentiment on a corpus of which a fraction was used to fine-tune the BERT model could lead to biased results. Therefore, we also fine-tune the BERT model by re-training it on the corpus of labeled sentences of the Czech National Bank (CNB), Hungarian National Bank (HNB), National Bank of Poland (NBP), and National Bank of Romania (NBR). Consequently, we obtain four fine-tuned BERT-CBSI models. Afterwards, each of these models is used, alternatively, to extract the central bank sentiment for each of the four central banks. We obtain 16 sentiment indices, which are compared with the BERT-CBSI, obtained from the BERT model fine-tuned on all central banks annotated sentences.

5. Empirical results

5.1. Time-varying paths of BERT-CBSI and comparison with other indicators

This section discusses the BERT-CBSI sentiment indices, together with the trajectory of other measures of the economic environment. Despite that the ups and downs are more pronounced for some countries (e.g., Poland and Romania), the sentiment indices naturally follow similar paths (see Fig. 1). Also, the BERT-CBSI mimics the general economic cycles. The tone of the minutes is mainly dovish, as negative values are predominant. This result is intuitive bearing in mind that, generally, between 2005 and 2021, the inflation has been stable and fluctuated around and below the lower bounds of the central banks targets. Additionally, the global financial crisis, the European sovereign debt crisis, and the Covid-19 pandemic have negatively influenced the economic output and triggered deep recessions in the region. As the plots for all countries reveal, the crises deteriorated the BERT-CBSI sentiment indices, which reached the lowest scores. For most of the countries, we identify a hawkish period between 2017 and 2019, when the BERT-CBSI becomes positive. However, the pandemic crisis triggers a strong downward trend. The movement is short-lived, as the index increases gradually around the tail end of the sample period due to overriding fears of inflation.

Fig. 1 also co-plots our central bank sentiment index with another measure of economic uncertainty, namely the country-level index of financial stress.[11] The two time-series mirror each-other, higher values of the country-level index of financial stress corresponding to lower values of central bank sentiment index and vice versa. The upwards of the country-level index of financial stress are associated with systemic financial stress periods, while the downwards of the BERT-CBSI are associated with dovish periods. Therefore, the two measures perform similarly in capturing the uncertainty. The result is revealing considering that the two indices are constructed very differently (for more details on methodology of the country-level index of financial stress, see Duprey et al., 2017).

Fig. 2 compares the BERT-CBSI with an indicator that describes the monetary policy decisions. This variable takes a value of 0 when there is no change in monetary policy stance, +1 for a hawkish monetary policy decision, which corresponds to an increase in the policy rate, and −1 for a dovish monetary policy decision, which corresponds to a policy rate cut. The plots reveal that the two series are positively correlated, a dovish sentiment index being associated with policy rate cuts, and vice versa. Such results lead us to conclude that the BERT-CBSI captures accurately both the uncertainty and changes in monetary policy stance.

5.2. BERT-CBSI and monetary policy forward guidance

This section investigates whether the BERT-CBSI sentiment indices have any predictive power for the future monetary decisions. For checking the robustness of these results, we use alternative sentiment measures of the central bank minutes tone. All tables report the marginal effects obtained from an ordered

7 The inflation target of each central banks was extracted from its website. For Czechia, the inflation target was 3% for the period between 2007 to 2009, and 2% afterwards. For Hungary, the inflation target was 3%. For Poland, the inflation target was 2.5%. For Romania, the inflation target was 4% for 2007, 3.8% for 2008, 3.5% for the period between 2009–2010, and 3% afterwards.

8 The potential output was proxied by using the trend of a Hodrick–Prescott filter, with a smoothing parameter set to 129 600, as recommended by Ravn and Uhlig (2002).

9 The Augmented Dickey–Fuller test show that inflation gap, output gap, and exchange rate have a unit root, while sentiment indicators are stationary.

10 For the financial domain corpus, Araci (2019) developed FinBERT, which is built on the BERT architecture. Besides the BERT pre-training on BookCorpus and English Wikipedia corpus, further re-pre-training was done on a specific financial corpus. For re-pre-training the BERT model, Araci (2019) used Thomson Reuters financial corpus, which consists of approximately 1,800,370 news articles published over the period 2008 to 2010.

11 The country-level index of financial stress measures the movements of the financial cycle and includes six financial stress indices that capture three financial market segments: equity markets, bond markets, and foreign exchange markets.

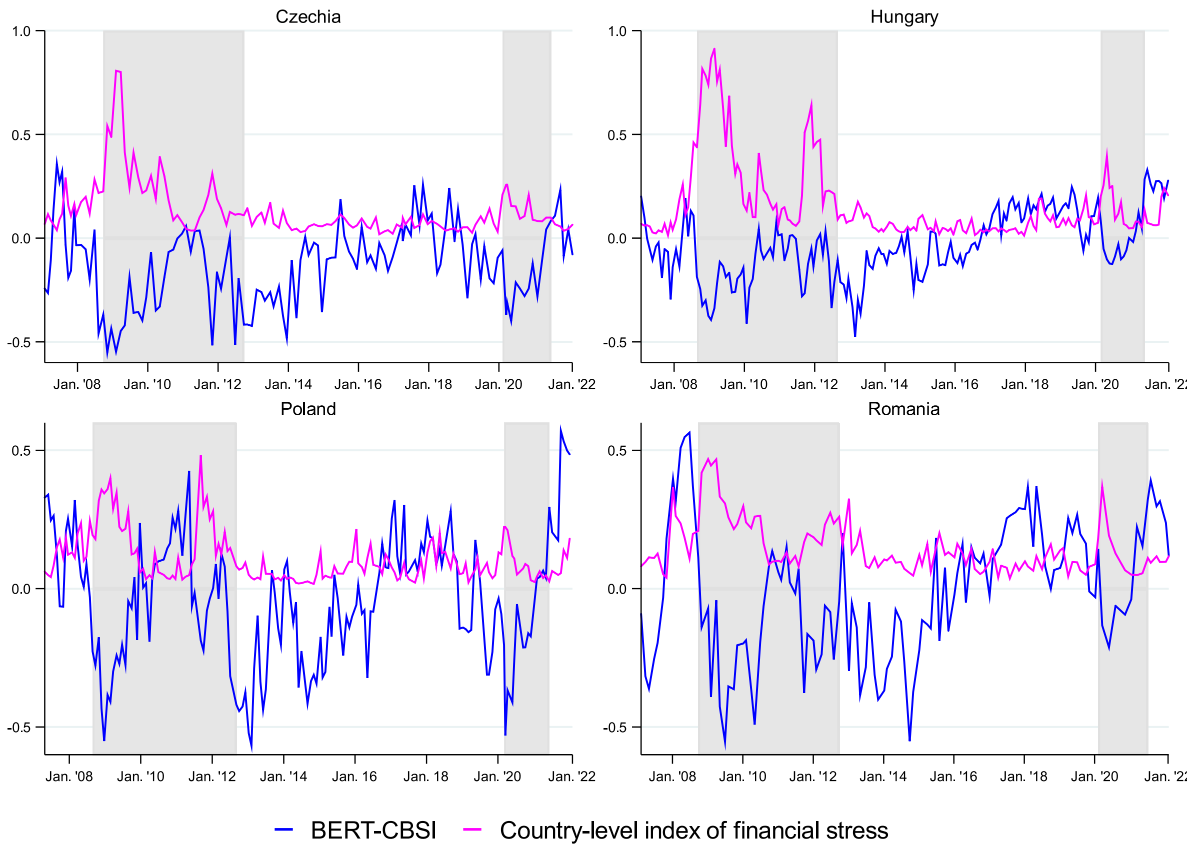

Fig. 1. BERT-CBSI vs. country-level index of financial stress.

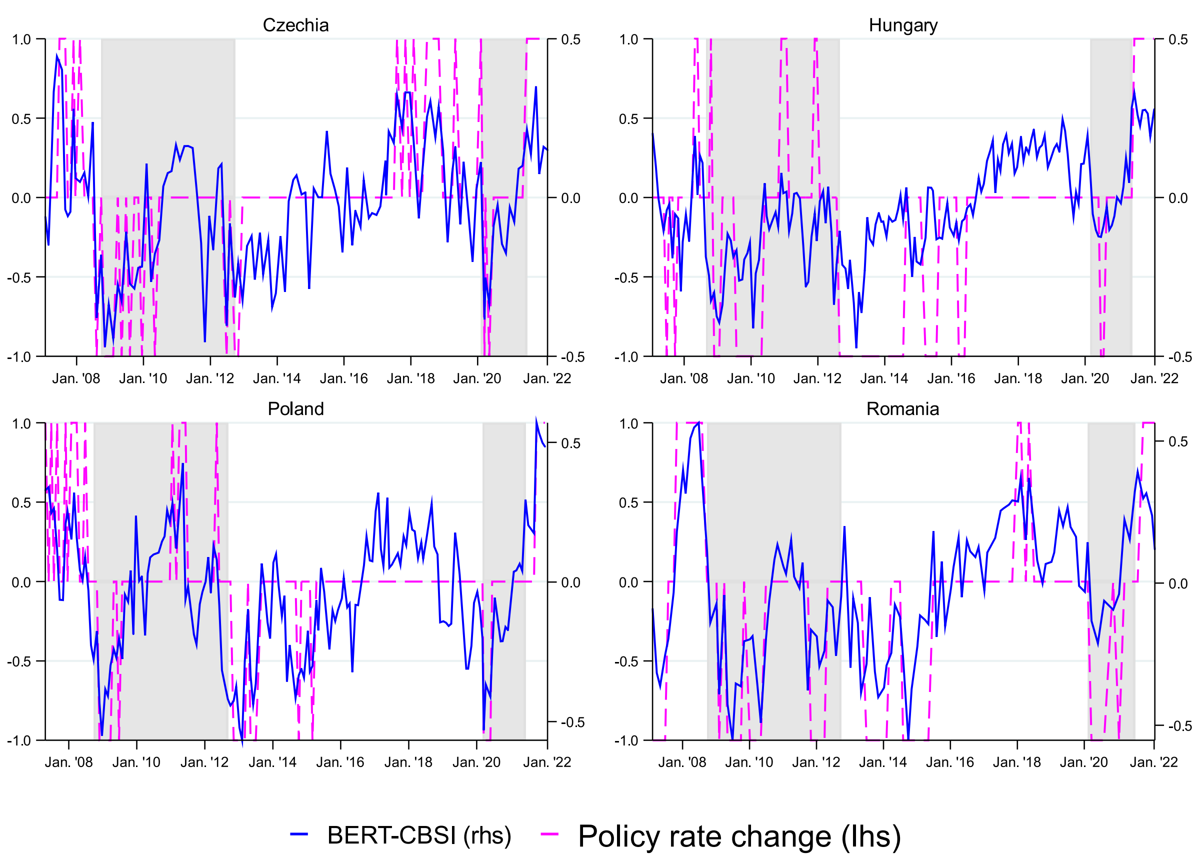

Fig. 2. BERT-CBSI vs. changes in policy rate.

Table 3
Ordered probit estimations for policy rate changes in Czechia.

	(1)	(2)	(3)	(4)	(5)	(6)
PRD_t	0.204	0.227	0.407	0.189	0.245	0.363
	(0.301)	(0.293)	(0.289)	(0.313)	(0.305)	(0.303)
$Sent_t^{BERT-CBSI}$	3.063***			3.261***		
	(0.822)			(0.853)		
$Hawkish_t^{BERT-CBSI}$		6.261***			6.424***	
		(1.587)			(1.636)	
$Dovish_t^{BERT-CBSI}$			−4.170***			−4.619***
			(1.410)			(1.462)
$\Delta(\pi_t - \pi^*)$				0.644*	0.562*	0.705**
				(0.358)	(0.342)	(0.358)
$\Delta(y_t - y^*)$				0.037	0.034	0.036
				(0.041)	(0.040)	(0.040)
ΔER_t				0.589*	0.558*	0.553*
				(0.334)	(0.334)	(0.329)
$\hat{\gamma}_{-1}$	−1.570***	0.109	−6.781***	−1.599***	0.135	−7.383***
	(0.185)	(0.388)	(1.864)	(0.191)	(0.399)	(1.936)
$\hat{\gamma}_1$	1.177***	2.862***	−4.127**	1.234***	2.954***	−4.641**
	(0.163)	(0.476)	(1.779)	(0.171)	(0.494)	(1.836)
N	126	126	126	125	125	125
$Pseudo-R^2$	0.170	0.181	0.139	0.202	0.209	0.172

Notes: Specifications 1–3 are estimated by using Eq. (3), while specifications 4–6 are computed by using Eq. (4). ***, **, * correspond to the 1%, 5%, and 10% levels of significance, respectively. Standard errors in parentheses.

probit model with the decision of the succeeding monetary board meeting as endogenous variable.

Table 3 reports the predictive power of the CNB sentiment index in anticipating future monetary policy decisions. The first three columns show the findings for the baseline model (see Eq. (3)). Surprisingly, the contemporaneous monetary interest rate is not a significant predictor of future policy decisions. This result might be a consequence of a wait-and-see strategy of the central bank, as the board prefers the status quo in order to evaluate the effects. The coefficient of the sentiment index shows that it has a pronounced predictive power for future policy rate decisions. A higher net index is positively associated with higher policy rates and negatively linked with lower interest rates. We go further into our analysis and replace the net sentiment index with a hawkish and dovish measure (see columns 2 and 3). The findings reveal that the hawkish tone is negatively (positively) associated with a future decision of easing (tightening) policy rates. Conversely, a more dovish tone of the central bank's minutes is positively (negatively) related to lower (higher) interest rates.

We deepen our analysis and include the available macroeconomic information in our specifications (columns 4, 5, and 6) to test whether the minutes tone keeps their predictive power when other relevant information is taken into account. Adding these variables to the models raises the value of Pseudo-R^2. The results show that contemporaneous policy rates do not convey significant information for future policy rate decisions. However, the sentiment index continues to explain future policy decisions, as a higher BERT-CBSI is positively correlated with tightened interest rates (column 4). Also, the cut-off points around which the dependent variable is ordered, i.e., easing (−1), unchanged (0), and tightened are significant for BERT-CBSI in both specifications (1 and 4). Therefore, we conclude that net sentiment index is a good forward guidance proxy for CNB future policy decisions. When hawkish and dovish measures are included separately into specifications (columns 5 and 6), we find that both are significant and relevant for future policy rates decisions. Specifications 4–6 show that the findings for inflation gap are in line with the theory, emphasizing the focus of CNB on inflation. Also, we notice that the depreciation (appreciation) of the local currency is associated with a higher (lower) interest rate. Horvath (2008) documented a similar link between policy rate and exchange rate for Czechia, while Geršl and Holub (2006) emphasized the role played by CNB's foreign exchange rate interventions within the inflation targeting framework.

Table 4 summarizes the findings for the predictive power of the HNB minutes tone. Columns 1, 2, and 3 report the results for the basic model. We find that both contemporaneous policy interest rate and BERT-CBSI are good predictors for future monetary policy decisions. Specifically, a decrease of the interest rate is more likely to be followed by a decline in the interest rate than by an increase. Conversely, a rise in the interest rate is more likely to determine an increase of future rates. The sentiment index reveals that higher hawkishness is negatively (positively) associated with future easing (tightening). Also, we notice that the sentiment index has a more pronounced predictive power compared to contemporaneous policy rate, as the coefficient is larger. Columns 2 and 3 report similar results for the predictive power of contemporaneous interest rate on future decisions. Moreover, a more hawkish (dovish) tone of the minutes is more likely to be followed by a tight (ease) monetary policy.

After controlling for the available macroeconomic variables (columns 4, 5, and 6), the predictive power of the contemporaneous policy rate and sentiment measures remains significant and in line with expectations. Also, the cut-off points around which the dependent variable is ordered are significant for BERT-CBSI in both specifications (1 and 4). Including the inflation gap, output gap, and exchange rate into the models marginally raises the value of Pseudo-R^2. A higher sentiment score is negatively (positively) associated with a future decision to lower (increase) the interest rate (Model 4). Furthermore, the findings revealed by specifications 5 and 6 show that a more dovishness tone for the central bank's minutes signals a future ease of the policy rate, while an increase in the hawkishness tone is linked to higher interest rates. For specifications 4–6, the results reveal no significant coefficients for inflation gap, output gap, and exchange rate.

Table 5 displays the predictive power of the NBP sentiment index. Columns 1, 2, and 3 reports the results for the basic model, while columns 4, 5, and 6 show the results for the Taylor rule model. The findings for the basic model show that policy rate cuts (hikes) are likely to be followed by similar policy commitments. The coefficient for the BERT-CBSI reveals that a decrease (increase) of the sentiment index is associated with future lower (higher) policy rates. The coefficients for the hawkish and dovish

Table 4
Ordered probit estimations for policy rate changes in Hungary.

	(1)	(2)	(3)	(4)	(5)	(6)
PRD_t	2.016***	2.073***	1.975***	1.979***	2.045***	1.928***
	(0.228)	(0.225)	(0.230)	(0.231)	(0.227)	(0.234)
$Sent_t^{BERT-CBSI}$	3.268***			3.500***		
	(0.863)			(0.934)		
$Hawkish_t^{BERT-CBSI}$		5.220***			5.404***	
		(1.590)			(1.692)	
$Dovish_t^{BERT-CBSI}$			−7.260***			−7.959***
			(1.784)			(1.963)
$\Delta(\pi_t - \pi^*)$				0.119	0.141	0.124
				(0.240)	(0.237)	(0.242)
$\Delta(y_t - y^*)$				−0.031	−0.029	−0.033
				(0.032)	(0.031)	(0.032)
ΔER_t				0.017	0.011	0.023
				(0.021)	(0.021)	(0.022)
$\hat{\gamma}_{-1}$	−1.563***	0.143	−11.16***	−1.570***	0.198	−12.10***
	(0.182)	(0.470)	(2.441)	(0.188)	(0.504)	(2.678)
$\hat{\gamma}_1$	1.746***	3.353***	−7.759***	1.771***	3.425***	−8.638***
	(0.190)	(0.559)	(2.296)	(0.193)	(0.590)	(2.519)
N	179	179	179	178	178	178
$Pseudo - R^2$	0.496	0.482	0.507	0.502	0.486	0.514

Notes: Specifications 1–3 are estimated by using Eq. (3), while specifications 4–6 are computed by using Eq. (4). ***, **, * correspond to the 1%, 5%, and 10% levels of significance, respectively. Standard errors in parentheses.

Table 5
Ordered probit estimations for policy rate changes in Poland.

	(1)	(2)	(3)	(4)	(5)	(6)
PRD_t	0.511*	0.534*	0.525*	0.287	0.453	0.369
	(0.326)	(0.322)	(0.315)	(0.366)	(0.356)	(0.361)
$Sent_t^{BERT-CBSI}$	5.868***			5.869***		
	(1.098)			(1.222)		
$Hawkish_t^{BERT-CBSI}$		9.615***			8.940***	
		(1.902)			(2.040)	
$Dovish_t^{BERT-CBSI}$			−11.65***			−12.44***
			(2.221)			(2.591)
$\Delta(\pi_t - \pi^*)$				1.191***	1.166***	1.236***
				(0.424)	(0.395)	(0.441)
$\Delta(y_t - y^*)$				0.110*	0.105*	0.123*
				(0.065)	(0.058)	(0.068)
ΔER_t				−2.731	−2.456	−3.170
				(1.842)	(1.714)	(1.965)
$\hat{\gamma}_{-1}$	−2.442***	0.977**	−18.02***	−2.678***	0.615	−19.43***
	(0.318)	(0.484)	(3.241)	(0.376)	(0.521)	(3.812)
$\hat{\gamma}_1$	1.970***	4.880***	−13.31***	2.209***	4.862***	−14.04***
	(0.234)	(0.722)	(2.768)	(0.281)	(0.782)	(3.211)
N	165	165	165	164	164	164
$Pseudo - R^2$	0.440	0.393	0.452	0.526	0.475	0.551

Notes: Specifications 1–3 are estimated by using Eq. (3), while specifications 4–6 are computed by using Eq. (4). ***, **, * correspond to the 1%, 5%, and 10% levels of significance, respectively. Standard errors in parentheses.

indices strengthen the robustness of the findings, as a higher hawkish (dovish) index are associated with future policy hikes (cuts).

The Taylor rule models confirm the predictive power of the sentiment indices. Including inflation gap, output gap, and exchange rate into the specifications raises the value for Pseudo-R^2. The results reveal that the predictive power for the contemporaneous policy rate is no longer significant for future policy rate decisions. However, the coefficients for our interest variables – the sentiment index, the hawkish index, and the dovish index – are significant and reveal that a more positive (negative) tone of the minutes anticipates a policy rate hike (cut). For Poland, we see that both inflation gap and output gap are positive and significant, implying that the policy rate tightening is associated with higher inflation and output gaps, while the coefficient of the exchange rate is not significant. Also, similarly to Czechia, we notice that Poland's central bank places a greater emphasis on inflation gap, compared to the output gap growth rate.

Table 6 presents the estimates for the predictive power of the NBR sentiment index. The results in column 1 suggest that the current policy rate and the sentiment index help to predict future policy rate hikes and cuts. Specifically, an increase (decrease) in the policy rate is more likely to be followed by similar policy commitments. A higher BERT-CBSI is positively associated with a tighter policy rate. Also, a more hawkish (dovish) tone in the minutes of the central bank anticipates a higher (lower) policy rate. The cut-off points around which the dependent variable is ordered are significant for BERT-CBSI in both specifications (1 and 4).

Table 6
Ordered probit estimations for policy rate changes in Romania.

	(1)	(2)	(3)	(4)	(5)	(6)
PRD_t	1.271***	1.522***	1.382***	1.317***	1.563***	1.447***
	(0.324)	(0.288)	(0.331)	(0.360)	(0.321)	(0.362)
$Sent_t^{BERT-CBSI}$	2.925***			2.819***		
	(0.909)			(0.979)		
$Hawkish_t^{BERT-CBSI}$		4.172***			4.044***	
		(1.402)			(1.548)	
$Dovish_t^{BERT-CBSI}$			−4.540***			−4.391**
			(1.726)			(1.803)
$\Delta\,(\pi_t - \pi^*)$				0.690***	0.697***	0.674***
				(0.263)	(0.264)	(0.256)
$\Delta\,(y_t - y^*)$				0.0664	0.0669	0.0723
				(0.0451)	(0.0439)	(0.0457)
ΔER_t				0.378	0.714	−0.153
				(2.461)	(2.460)	(2.454)
$\hat{\gamma}_{-1}$	−1.324***	−0.187	−7.073***	−1.417***	−0.311	−6.997***
	(0.201)	(0.394)	(2.226)	(0.221)	(0.441)	(2.328)
$\hat{\gamma}_1$	1.782***	2.891***	−4.121*	2.048***	3.135***	−3.679*
	(0.248)	(0.547)	(2.129)	(0.305)	(0.637)	(2.222)
N	117	117	117	116	116	116
$Pseudo-R^2$	0.432	0.423	0.412	0.490	0.482	0.476

Notes: Specifications 1–3 are estimated by using Eq. (3), while specifications 4–6 are computed by using Eq. (4). ***, **, * correspond to the 1%, 5%, and 10% levels of significance, respectively. Standard errors in parentheses.

The Taylor rule specifications raise the value of Pseudo-R^2. The findings for the contemporaneous policy rate and the sentiment indices are similar with those from the base model. The coefficients for inflation gap are in line with the theory, emphasizing the importance of inflation rate changes within the central bank's monetary policy strategy. Also, the findings show that output gap and exchange rate changes are not significant in setting the policy rate. Considering that the central banks included in our sample apply an inflation targeting strategy, the greater importance on inflation gap is in line with theory.

Overall, the sentiment of central banks minutes prove to be a useful tool for predicting the forthcoming monetary policy decisions. Our results are consistent with previous findings, which documented that central banks communications contain relevant information for the expected path of the monetary policy stance in addition to the classic Taylor rule (e.g., Picault and Renault, 2017; Lee et al., 2019; Apel et al., 2021).

5.3. Robustness

5.3.1. Different training strategies

We test the robustness of our BERT-CBSI by using several approaches. We fine-tune the domain-specific FinBERT by re-training it on our corpus of manually labeled sentences on monetary policy stances. Furthermore, we fine-tune the BERT model by re-training it on a dataset of 9,154 automated labeled sentences. Fig. 3 compares BERT-CBSI trained on our manual labeled sentences, FinBERT-CBSI trained on our manual labeled sentences, and BERT-CBSI trained on automated labeled sentences. The time-varying paths of the three indices are quite similar. Also, Table A.1, listed in Appendix, compares the predictive power of FinBERT-CBSI (columns 3 and 4) and BERT-CBSI trained on automated labeled sentences (columns 5 and 6) with BERT-CBSI trained our manual labeled sentences (columns 1 and 2). The results reveal that the three indices have similar predictive power in anticipating future policy changes and that there no major difference between the coefficients for the three indices.

In order to test which of these three models performs better for the NLP analysis of monetary policy narratives, we also compare them by having in view several specific NLP evaluation metrics, i.e., loss, accuracy, and macro average F1 scores (Table 7).

Table 7
Evaluation metrics for different training strategies.

Model	Training dataset	Loss	Accuracy	F1 score
BERT-CBSI	Manually labeled sentences	0.23	0.89	0.88
FinBERT-CBSI	Manually labeled sentences	0.29	0.88	0.88
BERT-CBSI	Automated labeled sentences	0.32	0.81	0.80

The evaluation scores show that the highest accuracy is obtained for the BERT model, trained on our corpus of manually labeled sentences. The accuracy score reveals that after pre-training the model manages to label correctly 89% of the validation sample. However, the difference compared with the FinBERT model is minor, emphasizing the strong performance of the BERT model. Peng et al. (2021) also showed that continual pre-training of the original BERT model seems to be the most effective option, while domain-specific pre-training is less effective for financial tasks. The BERT model trained on the dataset of automated labeled sentences reveal the lowest accuracy score.

5.3.2. Training on a different corpus of annotated sentences

As we have mentioned in the Methodology Section, it is important to see if the deep learning index remains robust when the model is fine-tuned on a different corpus of annotated sentences. We individually fine-tuned the model on the individual sentence corpus of CNB, HNB, NPB, and NBR. Compared to the overall sample, the individual corpus of each of the four banks is much smaller. When we fine-tune the BERT model on the individual small dataset corpus of the four central banks, we notice lower accuracy scores and a certain degree of instability in the fine-tuning. These findings are similar with those of Zhang et al. (2020) and Peng et al. (2021), who highlighted fluctuations in the results when BERT is fine-tuned on small datasets. Therefore, for this robustness check, we prefer to rely on the FinBERT model, which performs better on small datasets. After fine-tuning the model on the individual corpus of the central banks, we obtain four FinBERT-CBSI models. Afterwards, each of these models is used to obtain a sentiment index for each of the four central banks, resulting 16 sentiment indices. For each of these indices, we run a simple correlation with the FinBERT-CBSI, obtained from the

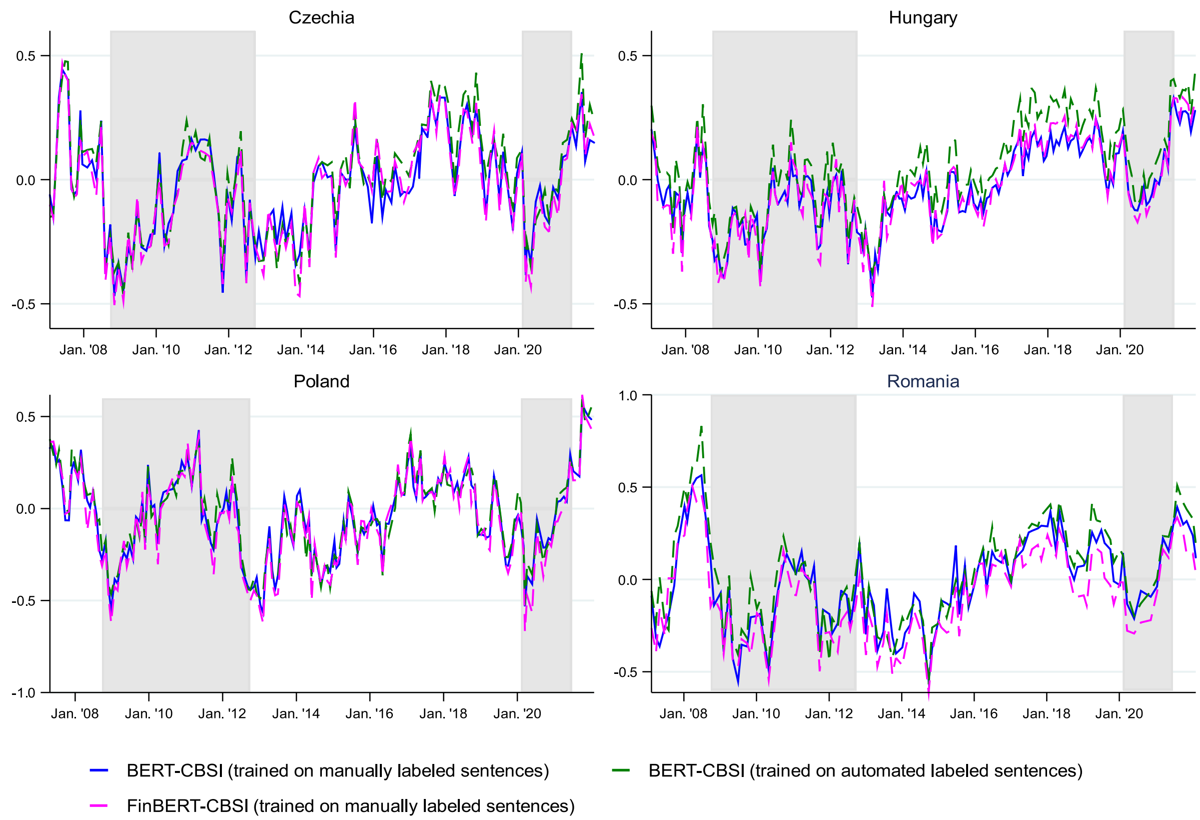

Fig. 3. BERT-CBSI robustness to different training strategies.

Table 8
The correlation between FinBERT-CBSI sentiment indices obtained by fine-tuning FinBERT on different corpus of annotated sentences.

	FinBERT-CBSI fine-tuned with CNB sentences to predict the sentiment index for …	FinBERT-CBSI fine-tuned with HNB sentences to predict the sentiment index for …	FinBERT-CBSI fine-tuned with NBP sentences to predict the sentiment index for …	FinBERT-CBSI fine-tuned with NBR sentences to predict the sentiment index for …
	Czechia			
FinBERT-CBSI Czechia	0.961***	0.944***	0.924***	0.917***
	Hungary			
FinBERT-CBSI Hungary	0.910***	0.987***	0.982***	0.959***
	Poland			
FinBERT-CBSI Poland	0.965***	0.971***	0.977***	0.962***
	Romania			
FinBERT-CBSI Romania	0.907***	0.924***	0.907***	0.916***

Notes: Columns 2–5 reveal the correlation coefficient between, on the one hand, the FinBERT-CBSI obtained after fine-tuning the FinBERT model on the corpus of all 1,998 annotated sentences and, on the other hand, the FinBERT-CBSI obtained after individually fine-tuning the FinBERT model on the corpus of Czechia (607 labeled sentences), Hungary (600 labeled sentences), Poland (390 labeled sentences), and Romania (401 labeled sentences) ***, **, * correspond to the 1%, 5%, and 10% levels of significance, respectively.

FinBERT model trained on all central banks annotated sentences (see Table 8).

The findings are impressive. All the correlation coefficients are significant and higher than 0.90. For example, extracting a sentiment index for Czechia with FinBERT-CBSI model trained on annotated sentences from the HNB minutes will lead to similar results to the FinBERT-CBSI obtained for Czechia after training the FinBERT-CBSI model on a corpus of 1,998 sentences from all central banks minutes, as the correlation coefficient is 0.944. Furthermore, fine-tuning FinBERT on a smaller corpus of annotated sentences does not change the results. Specifically, fine-tuning FinBERT on the corpus of 390 annotated sentences of the NBP will lead to similar results with the fine-tuned FinBERT on the corpus of 1,998 sentences (see column 3). To conclude, our methodological framework is robust and can be further used to compute a deep learning sentiment index for other central banks. Hereinafter, for the sake of the brevity, we will only refer to the BERT-CBSI computed after fine-tuning the BERT model on the corpus of 1,998 annotated sentences.[12]

5.3.3. Comparison with other text-based sentiment indices

One may wonder how does our BERT-CBSI perform compared to other central bank lexicon-based sentiment indicators. To clarify this research question, we compute other three sentiment measures,[13] i.e., $Sent^{GT}$, $Sent^{LM}$, and $Sent^{FSS}$ and we compare them with BERT-CBSI.

For building $Sent^{GT}$, we follow Gonzalez and Tadle (2021), who used Latent Dirichlet Allocation to create specific hawkish and dovish keywords for each central banks that adopted inflation targeting as their monetary policy strategy. Having in view that three of our four countries, i.e., Hungary, Poland, and Romania, are covered in the paper of Gonzalez and Tadle (2021), we use their list of hawkish keywords, dovish keywords, positive modifiers, and negative modifiers. For Czechia, we follow a similar procedure and we apply Latent Dirichlet Allocation to build a specific list of hawkish and dovish keywords. Tables A.2 and A.3 reveal the list of hawkish and dovish keywords for each central banks, and the positive and negative modifiers. Afterwards, to build sentiment indices, we use a 10-word window to find hawkish and dovish keywords associated with specific positive and negative modifiers. We count the instances when the keyword and the modifier are adjacent, accounting also the instances when the keyword and the modifier are separated by other words in the sentence. The second measure, $Sent^{LM}$, is computed based on the general financial dictionary proposed by Loughran and McDonald (2011). Finally, $Sent^{FSS}$ is based on the specific central banks financial stability sentiment dictionary developed by Correa et al. (2021). Compared to $Sent^{LM}$ and $Sent^{FSS}$, $Sent^{GT}$ has two main advantages. First, the list of modifiers and keywords is created considering each central bank particularities. Second, it relies on multiple words search taking into consideration the context of the sentence, by using a 10-words window.

Fig. A.2, included in Appendix, compares the time-varying pattern of BERT-CBSI to that of $Sent^{GT}$, $Sent^{LM}$, and $Sent^{FSS}$. The plots unanimously reveal that our BERT-CBSI is less noisy, especially when we compare it with $Sent^{LM}$ and $Sent^{FSS}$. Also, BERT-CBSI captures better the stance of the monetary policy, considering that in some cases the alternative indices have only negative values, e.g., $Sent^{LM}$ for Czechia, Hungary, and Poland. Finally, in some cases the paths for the BERT-CBSI mirrors with other indicators, e.g., BERT-CBSI and $Sent^{FSS}$ for Czechia and Romania. Which of these patterns better predict future policy changes will be revealed in the following sub-section.

Table A.1 presents the predictive power of BERT-CBSI (columns 1 and 2), compared to alternative sentiment indices (columns 7–10). Unanimously, for all the countries, the findings reveal that the BERT-CBSI index performed better compared to the other text-based sentiment indicators, both in terms of statistical significance of the coefficients and Pseudo-R^2. For all countries, columns 7 and 8 show that the estimates obtained with $Sent^{GT}$ are similar with those of BERT-CBSI in terms of statistical significance. However, the coefficients for BERT-CBSI are larger and the Pseudo-R^2 is higher (see columns 1 and 2). The predictive power of $Sent^{LM}$ (columns 9 and 10) and $Sent^{FSS}$ (columns 11 and 12) fails to predict future policy changes. Except for $Sent^{LM}$ in case of Czechia, the coefficients of these indicators are either not statistically significant or not in line with expectations. Furthermore, the Pseudo-R^2 is low. Considering these findings, we can firmly conclude that BERT-CBSI outperforms other existing text-based indicators, confirming our endeavor to construct a new tailored central bank sentiment index.

6. Conclusions

Central banks minutes represent a valuable source of information both for policymakers and market participants, providing a guidance about the future monetary policy. In this paper, we proposed a new a NLP technique to extract the tone of central banks narratives and to generate a novel deep learning sentiment index: BERT-CBSI.

We contribute to the literature by fine-tuning a BERT model by re-training it on a manually sentiment labeled sentences on monetary policy stances. By creating this dataset of field-specific sentences, we capture better the predictive power of the information encompassed in central banks communications. As most of the existent studies are focused on central banks from developed economies, we turned our attention to a group of central banks from Central and Eastern Europe that were neglected so far in the sentiment analysis literature. However, our framework, which is more tailored on the idiosyncrasies of central bank language, enables future possible research in developing more accurate sentiment indicators for Fed, ECB, or other central banks. The extension of BERT-CBSI could better capture the effects central bank communication on both high- and low-frequency indicators.

Ensuing, we assessed the new central bank sentiment indices by testing whether they can anticipate future monetary policy decisions. Our evidence reveals that BERT-CBSI is a significant predictor of central banks' forward guidance measures. Additionally, when the hawkish and dovish scores are considered, we find that both are relevant for future policy rates decisions. More specifically, a hawkish sentiment is negatively (positively) associated with a future decision of easing (tightening) the monetary policy, while a more dovish sentiment is positively (negatively) related to lower (higher) policy rates. Our index is robust to different fine-tuning strategies and surpass other lexicon-based indicators. Moreover, the BERT-CBSI sentiment score is a good measure of the financial uncertainty, leading to similar signals as the country-level index of financial stress. The tools and the findings provided by this paper can help financial market participants to form an opinion about central bank initiatives faster and with less effort than by reading and analyzing all the communication materials. Furthermore, central banks can use these sentiment indices to adjust their monetary policy communications in order to deliver the proper message to the public about their future decisions.

Our analysis has some caveats. The monetary policy sentences could be labeled by having in mind a broader range of scales, given that some positive (negative) sentiments are more positive (negative) than others. Also, the predictive power of the sentiment indicators can be improved by dividing them into several categories, for example, economic outlook, inflation outlook, and monetary policy outlook. Finally, the dataset of manually labeled sentences could be enlarged with sentences from other central banks in order to improve its applicability. We leave these caveats for future research.

12 The other indices are uploaded on the dedicated website for our central bank deep learning sentiment index (https://sites.google.com/view/bert-cbsi/).

13 For these indicators, the sentiment index is upon the following formula: $Sentiment_{index} = \frac{hawk_t - dove_t}{hawk_t + dove_t}$, where $hawk_t$ and $dove_t$ are the total number of hawkish and dovish occurrences in a given set of minutes.

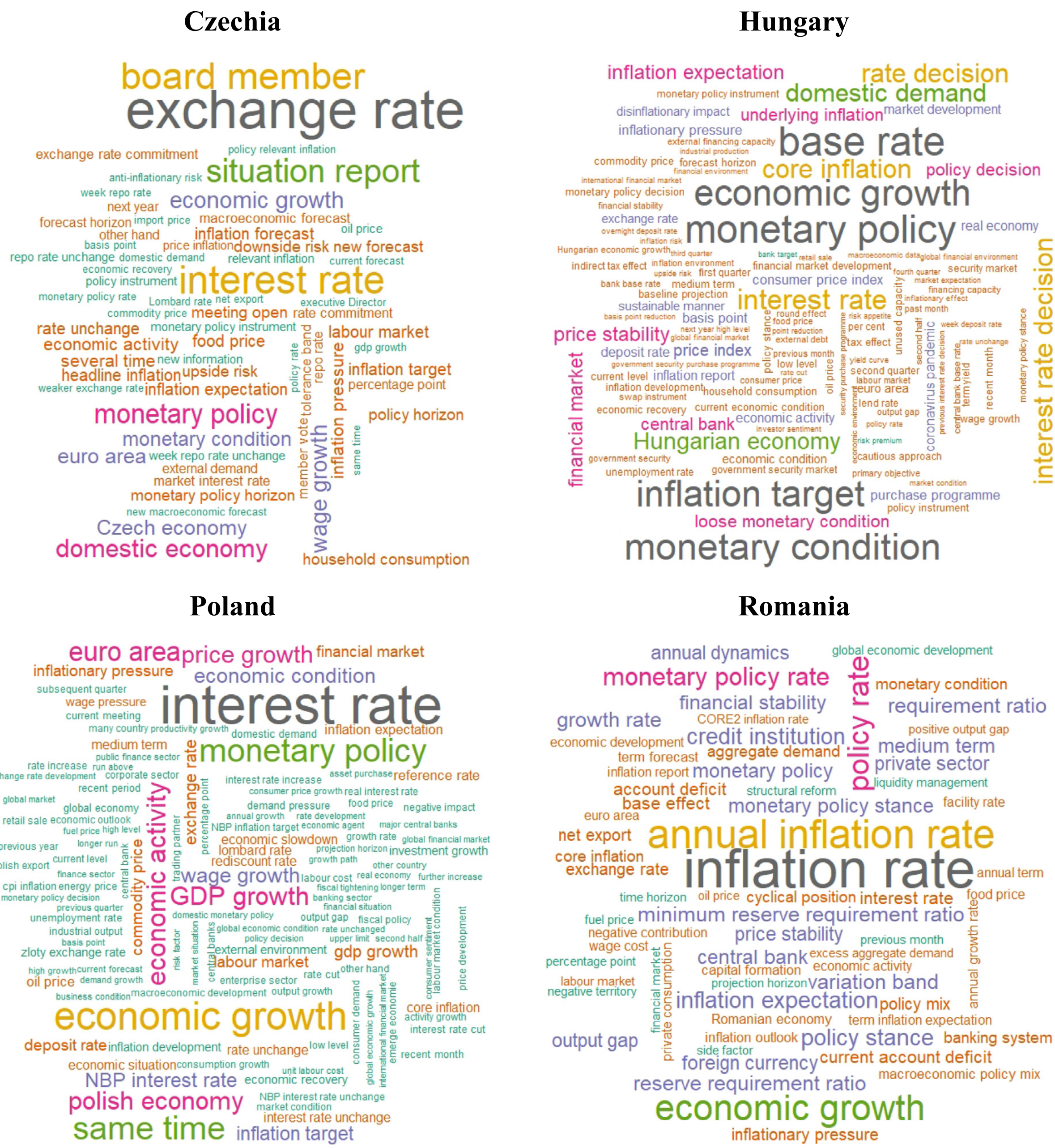

Fig. A.1. Main topics covered by central banks minutes. Note: The word-clouds include n-gram higher than one, used more than 50 times. The word-clouds are pictured by using Google's PageRank algorithm to extract keywords, included in UDPipe package in R (Straka and Straková, 2017).

CRediT authorship contribution statement

Mihai Niţoi: Conceptualization, Investigation, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review & editing, Project administration. **Maria-Miruna Pochea:** Conceptualization, Investigation, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft, Visualization, Writing – review & editing, Funding acquisition. **Ştefan-Constantin Radu:** Software, Data curation, Formal analysis, Writing – original draft.

Acknowledgment

We are grateful to Stefan Palan (the editor) and the two anonymous referees for their comments that greatly helped improve our manuscript.

Appendix

See Figs. A.1, A.2 and Tables A.1–A.3.

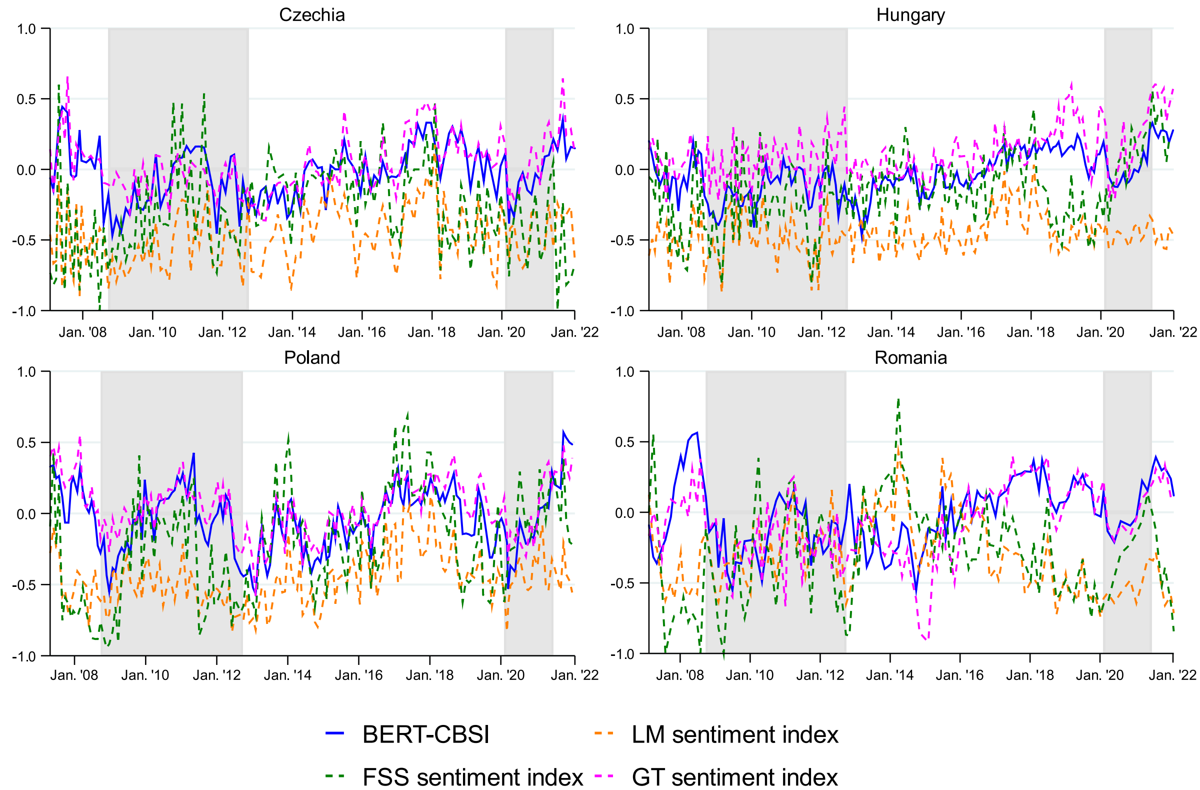

Fig. A.2. BERT-CBSI vs. other sentiment indicators.

Table A.1
Predictive power of different central bank sentiment indicators.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Czechia												
$Sent_t^{BERT-CBSI}$	3.063***	3.261***										
	(0.822)	(0.853)										
$Sent_t^{FinBERT-CBSI}$			2.654***	2.923***								
			(0.758)	(0.797)								
$Sent_t^{BERT-CBSI(als)}$					3.223***	3.482***						
					(0.792)	(0.828)						
$Sent_t^{GT}$							2.530***	2.801***				
							(0.738)	(0.761)				
$Sent_t^{LM}$									1.158**	1.296**		
									(0.536)	(0.545)		
$Sent_t^{FSS}$											0.125	0.253
											(0.341)	(0.355)
N	126	125	126	125	126	125	126	125	126	125	126	125
$Pseudo-R^2$	0.170	0.202	0.160	0.194	0.187	0.222	0.157	0.194	0.115	0.146	0.091	0.118
Hungary												
$Sent_t^{BERT-CBSI}$	3.268***	3.500***										
	(0.863)	(0.934)										
$Sent_t^{FinBERT-CBSI}$			2.708***	2.808***								
			(0.776)	(0.826)								
$Sent_t^{BERT-CBSI(als)}$					2.872***	2.895***						
					(0.756)	(0.784)						
$Sent_t^{GT}$							3.566***	3.591***				
							(0.990)	(1.030)				
$Sent_t^{LM}$									0.448	0.396		
									(0.755)	(0.844)		
$Sent_t^{FSS}$											0.390	0.388
											(0.410)	(0.462)
N	179	178	179	178	179	178	179	178	179	178	179	178
$Pseudo-R^2$	0.496	0.502	0.488	0.492	0.496	0.499	0.493	0.496	0.445	0.451	0.447	0.453

(continued on next page)

Table A.1 (*continued*).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Poland												
$Sent_t^{BERT-CBSI}$	5.868***	5.869***										
	(1.098)	(1.222)										
$Sent_t^{FinBERT-CBSI}$			5.176***	5.178***								
			(0.980)	(1.097)								
$Sent_t^{BERT-CBSI(als)}$					6.054***	5.927***						
					(1.120)	(1.221)						
$Sent_t^{GT}$							3.668***	3.596***				
							(0.841)	(0.900)				
$Sent_t^{LM}$									0.649	0.354		
									(0.588)	(0.642)		
$Sent_t^{FSS}$											0.668**	0.474
											(0.307)	(0.332)
N	165	164	165	164	165	164	165	164	165	164	165	164
$Pseudo-R^2$	0.440	0.526	0.431	0.518	0.450	0.529	0.343	0.451	0.240	0.362	0.257	0.371
Romania												
$Sent_t^{BERT-CBSI}$	2.925***	2.819***										
	(0.909)	(0.979)										
$Sent_t^{FinBERT-CBSI}$			3.654***	3.306***								
			(0.980)	(1.047)								
$Sent_t^{BERT-CBSI(als)}$					3.423***	3.467***						
					(0.937)	(1.012)						
$Sent_t^{GT}$							1.311**	1.425**				
							(0.607)	(0.660)				
$Sent_t^{LM}$									−0.888*	−0.757		
									(0.500)	(0.547)		
$Sent_t^{FSS}$											−0.272	−0.246
											(0.382)	(0.412)
N	117	116	117	116	117	116	117	116	117	116	117	116
$Pseudo-R^2$	0.432	0.490	0.456	0.501	0.453	0.514	0.402	0.471	0.393	0.456	0.381	0.448

Notes: ***, **, * correspond to the 1%, 5%, and 10% levels of significance, respectively. Standard errors in parentheses.

Table A.2
The list of hawkish and dovish keywords used to construct GT sentiment index.
Source: For Czechia, we apply Latent Dirichlet Allocation to build a specific keywords of hawkish and dovish keywords. For Hungary, Poland, and Romania, we use the lists from Gonzalez and Tadle (2021) who applied Latent Dirichlet Allocation to build a specific list of hawkish and dovish keywords for each central bank.

Country	Hawkish keywords	Dovish keywords
Czechia	Consumer price, Czech economy, discount rate, domestic demand, domestic economy, economic activity, economic growth, economy, employment, exchange rate, external demand, GDP growth, household consumption, inflation, inflation expectations, inflation forecast, inflation pressure, interest rate, labor market, Lombard rate, net export, outlook, output gap, potential output, wage growth,	Risk, slowdown, uncertainty, unemployment
Hungary	Activity, assessment, condition, consumer price, core inflation, cost, cost shock, council's assessment, demand, development, domestic, domestic demand, economic agent, economic growth, economy, euro area, financial market, growth, household consumption, Hungarian economy, inflation, inflation expectation, inflation target, inflationary pressure, labor market, market, output, price, price stability, private sector, wage	Disinflationary impact, reduction, risk, slowdown, unused capacity
Poland	Activity, demand, deposit, economic condition, economy, employment, exchange rate, growth, growth rate, household, inflation, inflation expectation, interest, loan, price, price growth, production, wage	Deposit rate
Romania	Banking system, consumer price, credit, credit institution, current account, development, domestic, economic, economic growth, financial stability, foreign currency, foreign exchange, global economic, growth, inflation expectation, inflation report, lending, liquidity, loan, price, stability	Deficit, disinflation, risk, uncertainty

Table A.3
The list of positive and negative modifiers used to construct GT sentiment index.
Source: Gonzalez and Tadle (2021).

Positive modifiers
Above, accelerate, accelerated, accelerates, accelerating, accommodate, accommodated, accommodates, accommodating, added, augment, augmented, augmenting, augments, benign, best, better, biggest, boost, boosted, boosting, boosts, brighter, buoy, buoyant, buoyed, buoying, buoys, calm, calmed, calming, calms, climb, climbed, climbing, climbs, depreciate, depreciated, depreciates, depreciating, dynamic, elevate, elevated, elevates, elevating, encouraging, escalate, escalated, escalates, escalating, exceed, exceeded, exceeding, exceeds, expand, expanded, expanding, expands, expansionary, expansive, fast, faster, fastest, favourable, favourable, firmer, good, great, greater, greatest, grew, grow, growing, grown, grows, healthier, high, higher, highest, improve, improved, improves, improving, impulse, impulsed, impulses, impulsing, increase, increased, increases, increasing, inflationary, large, larger, largest, lift, lifted, lifting, lifts, loose, loosen, loosened, loosening, loosens, looser, maximum, mitigate, mitigated, mitigates, mitigating, more, mount, mounted, mounting, mounts, optimistic, outperform, outperformed, outperforming, outperforms, peak, peaked, peaking, peaks, pick, picked, picking, picks, positive, raise, raised, raises, raising, ramp, ramped, ramping, ramps, rapid, recover, recovered, recovering, recovers, reinforce, reinforced, reinforces, reinforcing, restore, restored, restores, restoring, rise, risen, rises, rising, rose, satisfactory, skyrocket, skyrocketed, skyrocketing, skyrockets, spike, spiked, spikes, spiking, spur, spurred, spurring, spurs, stabilise, stabilised, stabilises, stabilising, stabilize, stabilized, stabilizes, stabilizing, stable, steady, stimulate, stimulated, stimulates, stimulating, stimulative, stimulatory, strengthen, strengthened, strengthening, strengthens, strong, stronger, strongest, successful, surge, surged, surges, surging, swifter, upper, upside, upswing, upswinging, upswings, upswung, uptrend, upturn, upturned, upturning, upturns, upward, vigorous, widen, widened, widening, widens, wider

(continued on next page)

Table A.3 (*continued*).

Negative modifiers
Adverse, aggravate, aggravated, aggravates, aggravating, appreciate, appreciated, appreciates, appreciating, appreciatory, bad, bottom, bottomed, bottoming, bottoms, concern, concerned, concerning, concerns, conservative, constrain, constrained, constraining, constrains, contract, contracted, contracting, contractionary, contracts, cut, cuts, cutting, dampen, dampened, dampening, dampens, decelerate, decelerated, decelerates, decelerating, decline, declined, declines, declining, decrease, decreased, decreases, decreasing, deepen, deepened, deepening, deepens, deflationary, descend, descended, descending, descends, destabilizing, deteriorate, deteriorated, deteriorates, deteriorating, difficult, diminish, diminished, diminishes, diminishing, disappointing, disinflationary, dovish, down, downside, downsize, downsized, downsizes, downsizing, downward, downwards, drop, dropped, dropping, drops, erode, eroded, erodes, eroding, fade, faded, fades, fading, fail, failed, failing, fails, fall, fallen, falling, falls, fell, fewer, flatten, flattened, flattening, flattens, fluctuate, fluctuated, fluctuates, fluctuating, fragile, harm, harmed, harming, harms, inconsistent, jeopardise, jeopardised, jeopardises, jeopardising, jeopardize, jeopardized, jeopardizes, jeopardizing, lacklustre, least, less, low, lower, lowered, lowering, lowers, lowest, mild, minimal, minimum, minor, moderate, moderated, moderates, moderating, modest, negative, pessimistic, poor, recessionary, reduce, reduced, reduces, reducing, restrictive, riskier, risky, sank, shorten, shortened, shortening, shortens, shrink, shrinking, shrinks, shrunk, shrunken, sink, sinking, slow, slowed, slower, slowest, slowing, slows, sluggish, small, smaller, smallest, soften, softened, softening, softens, speculate, speculated, speculates, speculating, stress, stressed, stresses, stressing, stringent, subdued, subprime, sunk, suppress, suppressed, suppresses, suppressing, threaten, threatened, threatening, threatens, tighten, tightened, tightening, tightens, tighter, tougher, turbulent, uncertain, unclear, undermine, unfavourable, unfavourable, unstable, volatile, vulnerable, wane, waned, wanes, waning, weak, weaken, weakened, weakening, weakens, weaker, weakest, worse, worsen, worsened, worsening, worsens, worst

References

Apel, M., Blix-Grimaldi, M., 2014. How informative are central bank minutes? Rev. Econ. 65, 53–76.

Apel, M., Blix-Grimaldi, M., Hull, I., 2021. How much information do monetary policy committees disclose? Evidence from the FOMC's minutes and transcripts. J. Money Credit Bank. 54, 1459–1490.

Araci, D., 2019. FinBERT: Financial sentiment analysis with pre-trained language models. http://dx.doi.org/10.48550/arXiv.1908.10063, arXiv preprint.

Armelius, H., Bertsch, C., Hull, I., Zhang, X., 2020. Spread the word: International spillovers from central bank communication. J. Int. Money Finance 103, 102116.

Bennani, H., 2020. Central bank communication in the media and investor sentiment. J. Econ. Behav. Organ. 176, 431–444.

Berger, H., De Haan, J., Sturm, J.E., 2011. Does money matter in the ECB strategy? New evidence based on ECB communication. Int. J. Finance Econ. 16, 16–31.

Born, B., Ehrmann, M., Fratzscher, M., 2014. Central bank communication on financial stability. Econ. J. 124, 701–734.

Boschen, J.F., Mills, L.O., 1995. The relation between narrative and money market indicators of monetary policy. Econ. Inq. 33, 24–44.

Burkhard, L., Fischer, A.M., 2009. Communicating policy options at the zero bound. J. Int. Money Finance 28, 742–754.

Cieslak, A., Schrimpf, A., 2019. Non-monetary news in central bank communication. J. Int. Econ. 118, 293–315.

Clements, M.P., Reade, J.J., 2020. Forecasting and forecast narratives: The Bank of England inflation reports. Int. J. Forecast. 36, 1488–1500.

Conrad, C., Lamla, M.J., 2010. The high-frequency response of the EUR-USD exchange rate to ECB communication. J. Money Credit Bank. 42, 1391–1417.

Correa, R., Garud, K., Londono, J.M., Mislang, N., 2021. Sentiment in central banks' financial stability reports. Rev. Finance 25, 85–120.

Daudert, T., 2021. Exploiting textual and relationship information for fine-grained financial sentiment analysis. Knowl.-Based Syst. 230, 107389.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. http://dx.doi.org/10.48550/arXiv.1810.04805, arXiv preprint.

Dewachter, H., Erdemlioglu, D., Gnabo, J.Y., Lecourt, C., 2014. The intra-day impact of communication on euro-dollar volatility and jumps. J. Int. Money Finance 43, 131–154.

Duprey, T., Klaus, B., Peltonen, T., 2017. Dating systemic financial stress episodes in the EU countries. J. Financ. Stab. 32, 30–56.

Ehrmann, M., Fratzscher, M., 2007. Communication by central bank committee members: Different strategies, same effectiveness? J. Money Credit Bank. 39, 509–541.

Gardner, B., Scotti, C., Vega, C., 2022. Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements. J. Econom. 231, 387–409.

Gerlach, S., 2007. Interest rate setting by the ECB, 1999–2006: Words and deeds. Int. J. Central Bank. 3, 1–46.

Geršl, A., Holub, T., 2006. Foreign exchange interventions under inflation targeting: the Czech experience. Contemp. Econ. Policy 24, 475–491.

Gertler, P., Horvath, R., 2018. Central bank communication and financial markets: New high-frequency evidence. J. Financ. Stab. 36, 336–345.

Gonzalez, M., Tadle, R.C., 2021. Monetary Policy Press Releases: An International Comparison. Central Bank of Chile Working Papers No. 912, Central Bank of Chile.

Hansen, S., McMahon, M., 2016. Shocking language: Understanding the macroeconomic effects of central bank communication. J. Int. Econ. 99, S114–S133.

Heinemann, F., Ullrich, K., 2007. Does it pay to watch central bankers' lips? The information content of ECB wording. Swiss J. Econ. Stat. 143, 155–185.

Horvath, R., 2008. Asymmetric monetary policy in the Czech Republic? Czech J. Econ. Finance 58, 470–481.

Howard, J., Ruder, S., 2018. Universal language model fine-tuning for text classification. http://dx.doi.org/10.48550/arXiv.1801.06146, arXiv preprint arXiv.

Jansen, D.J., De Haan, J., 2007. The Importance of being Vigilant: Has ECB Communication Influenced Euro Area Inflation Expectations? CESifo Working Papers Series No. 2134, CESifo.

Lee, Y., Kim, S., Park, K.Y., 2019. Deciphering monetary policy board minutes with text mining: The case of South Korea. Korean Econ. Rev. 35, 471–511.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. J. Finance 66, 35–65.

Lucca, D.O., Trebbi, F., 2009. Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements. NBER Working Paper No. 15367, National Bureau of Economic Research.

Parle, C., 2022. The financial market impact of ECB monetary policy press conferences—A text based approach. Eur. J. Political Econ. 102230.

Peng, B., Chersoni, E., Hsu, Y.Y., Huang, C.R., 2021. Is domain adaptation worth your investment? Comparing BERT and FinBERT on financial tasks. In: Proceedings of the Third Workshop on Economics and Natural Language Processing. pp. 37–44.

Picault, M., Pinter, J., Renault, T., 2022. Media sentiment on monetary policy: Determinants and relevance for inflation expectations. J. Int. Money Finance 124, 102626.

Picault, M., Renault, T., 2017. Words are not all created equal: A new measure of ECB communication. J. Int. Money Finance 79, 136–156.

Poole, W., 1971. Rules-of-thumb for guiding monetary policy. In: Open Market Policies and Operating Procedures-Staff Studies. Board of Governors of the Federal Reserve System, pp. 135–189.

Potts, G.T., Luckett, D.G., 1978. Policy objectives of the Federal Reserve system. Q. J. Econ. 52, 5–534.

Ravn, M.O., Uhlig, H., 2002. On adjusting the Hodrick–Prescott filter for the frequency of observations. Rev. Econ. Stat. 84, 371–376.

Rayl, N., Sinha, N.R., 2022. Integrating Prediction and Attribution to Classify News. Finance and Economics Discussion Series No. 2022-042, Board of Governors of the Federal Reserve System.

Rosa, C., 2013. Market efficiency broadcasted live: ECB code words and euro exchange rates. J. Macroecon. 38, 167–178.

Rosa, C., Verga, G., 2007. On the consistency and effectiveness of central bank communication: Evidence from the ECB. Eur. J. Political Econ. 23, 146–175.

Sadique, S., In, F., Veeraraghavan, M., Wachtel, P., 2013. Soft information and economic activity: Evidence from the Beige Book. J. Macroecon. 37, 81–92.

Schmeling, M., Wagner, C., 2019. Does Central Bank Tone Move Asset Prices? CEPR Discussion Paper No. DP13490, Centre for Economic Policy Research.

Shapiro, A.H., Sudhof, M., Wilson, D.J., 2022. Measuring news sentiment. J. Econometrics 228, 221–243.

Shapiro, A.H., Wilson, D., 2022. Taking the fed at its word: a new approach to estimating central bank objectives using text analysis. Rev. Econ. Stud. 89, 2768–2805.

Straka, M., Straková, J., 2017. Tokenizing, pos tagging, lemmatizing and parsing UD 2.0 with UDPipe. In: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. Association for Computational Linguistics, pp. 88–99.

Tripathy, A., Agrawal, A., Rath, S.K., 2016. Classification of sentiment reviews using n-gram machine learning approach. Expert Syst. Appl. 57, 117–126.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 6000–6010.

Zhang, T., Wu, F., Katiyar, A., Weinberger, K.Q., Artzi, Y., 2020. Revisiting few-sample BERT fine-tuning. arXiv preprint doi: 10.48550/arXiv.2006.05987.