CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Center for Economic Research and Graduate Education

Master's Thesis

2025 **Bernhard Brunner**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Center for Economic Research and Graduate Education

Predicting Policy Rate Changes from Central Bank Minutes using Machine Learning: Evidence from the Czech Republic (1998-2024)

Master's Thesis

Author of the thesis: Bernhard Brunner

Study program: Master in Economic Research

Supervisor: Sebastian Ottinger, Ph.D.

Year of the defense: 2025

Declaration

1. I hereby declare that I have compiled this thesis using the listed literature and resources only.
2. I hereby declare that my thesis has not been used to gain any other academic title.
3. I fully agree to my work being used for study and scientific purposes.
4. During the preparation of this thesis, the author used the AI tools chatgpt.com and grammarly.com in order to edit and rephrase parts of this thesis. After using this tool/service, the author reviewed and edited the content as necessary and takes full responsibility for the content of the publication

In Prague on 4.1.2025 Bernhard Brunner

References

BRUNNER, Bernhard. *Predicting Policy Rate Changes from Central Bank Minutes using Machine Learning: Evidence from the Czech Republic (1998-2024).* Praha, 2025. Master's thesis. Charles University, Center for Economic Research and Graduate Education – Economics Institute, A joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences. Supervisor Sebastian Ottinger, Ph.D.

Length of the Thesis: 92,150

Abstract

This thesis investigates how the sentiment of central bank minutes predicts upcoming policy rate changes in the Czech Republic from 1998 to 2024. I fine-tuned the RoBERTa language model on central bank communications to create a sentiment index of the Czech National Bank (CNB) Board meeting minutes. I then estimated an ordered probit model by regressing the upcoming policy rate change on this sentiment index while controlling for the current policy rate change and macroeconomic indicators. My results show that the sentiment of CNB minutes is a sizeable and statistically significant predictor of upcoming policy rate changes. Robustness checks using two alternative sentiment measures – a BERT sentiment index and RoBERTa sentiment dummies – and re-estimation of the model for subperiods further support my findings.

Abstrakt

Tato diplomová práce zkoumá, jak sentiment zápisů z jednání rady České národní banky (ČNB) předpovídá budoucí změny měnověpolitických sazeb v České republice v letech 1998 až 2024. Dotrénoval jsem jazykový model RoBERTa na komunikaci centrální banky, abych vytvořil index sentimentu zápisů z jednání rady ČNB. Následně jsem odhadl ordered probit model regresí budoucí změny měnověpolitické sazby na tomto indexu sentimentu, přičemž jsem kontroloval současnou změnu sazby a makroekonomické ukazatele. Výsledky ukazují, že sentiment zápisů z jednání rady ČNB je významným a statisticky významným prediktorem budoucích změn měnověpolitických sazeb. Robustnost výsledků je ověřena pomocí dvou alternativních měr sentimentu – indexu sentimentu BERT a sentimentových dummy proměnných RoBERTa – a přeodhadnutím modelu pro dílčí období, což dále podporuje mé závěry.

Keywords

Czech National Bank, Policy Rates, Sentiment Analysis, Transfer Learning, RoBERTa, BERT

Klíčová slova

Česká národní banka, měnověpolitické sazby, analýza sentimentu, transfer learning, RoBERTa, BERT

Title

Predicting Policy Rate Changes from Central Bank Minutes using Machine Learning: Evidence from the Czech Republic (1998-2024)

Název práce

Predikce změn měnověpolitických sazeb ze záznamů centrální banky pomocí strojového učení: Evidence z České republiky (1998–2024)

Acknowledgment

I want to express my gratitude to my supervisor, Sebastian Ottinger, Ph.D., who allowed me to pursue this topic for my Master's thesis. He supported me with his expertise at every step and motivated me to keep pushing toward the finish line. I also want to express my gratitude to RNDr. Milan Straka, Ph.D. His machine-learning courses inspired me, and he was available for consultation whenever I had a coding question related to my thesis. I would like to extend a special thank you to Deborah Nováková, M.A., for dedicating her time to editing my entire thesis. Additionally, I want to thank Doc. Nikolas Mittag, Ph.D., and Byeongju Jeong, Ph.D., for consultations. Lastly, I want to thank my girlfriend Tereza for her patience and emotional support throughout the process of writing this thesis.

Table of Contents

Introduction

Predicting when central banks will change policy rates has become increasingly important due to their broad impact on financial markets and the real economy. Unfortunately, central banks often avoid communicating explicitly about the future path of policy rates, and when they do communicate, their statements tend to remain vague and inconclusive. However, with advances in deep learning and natural language processing, it is now possible to extract information from text with greater sophistication, potentially allowing researchers to infer the future path of policy rates from current central bank communication.

A growing literature in economics assesses the predictive power of central bank communications, with the majority of papers focusing on the sentiments expressed in regular central bank publications, such as meeting minutes. The definition of sentiment depends on the research question and the type of information researchers aim to extract from texts. In the context of central bank communication, sentiment often reflects an interest rate tilt, indicating whether meeting minutes imply a future monetary policy tightening or easing. Central bank sentiment is then typically used as an explanatory variable in regression analysis, in which a shifted outcome variable of interest is regressed on central bank sentiment and control variables.

This thesis studies how the sentiments in central bank minutes predict upcoming policy rate changes in the Czech Republic from 1998 to 2024. To quantify the sentiments of Czech National Bank (CNB) minutes, I create a CNB Sentiment Index (CNBSI) by conducting sentiment analysis using a RoBERTa language model (Liu et al., 2019) that I fine-tune on central bank communications. I then estimate an ordered probit model to assess how the CNBSI influences the probability of an upcoming policy rate change being a rate cut, no change, or a rate hike while controlling for macroeconomic indicators and the current policy rate change. To validate my results, I re-estimate my model using two alternative measures of sentiment. Additionally, I examine the heterogeneity of my model over time by re-estimating it for three distinct periods.

My research is closely related to Nitoi, Pochea, and Radu (2023) who fine-tuned a BERT language model on annotated sentences issued by four European central banks, including the Czech National Bank. They predict upcoming policy rate changes by estimating an ordered probit model. My thesis improves on their prior analysis along several dimensions: First, I fine-tune a more advanced language model that addresses the issue of underfitting associated with

BERT. Second, I extend the period of analysis by 11 years, including data from 1998 to 2006 and 2022 to 2024, which allows me to obtain more precise and robust estimates of the predictive power of CNB sentiment. Third, I directly compare Nitoi, Pochea, and Radu's (2023) BERT sentiment index with my RoBERTa-based CNBSI for classifying central bank communications and predicting upcoming policy rate changes to understand the potential advantages of utilizing more advanced language models. Lastly, I assess how the predictive power of the CNBSI changes over time and offer insights into the heterogeneity of CNB sentiment across different periods.

The remainder of this thesis is structured as follows: the next section provides a background on central bank communications, reviews relevant literature, and offers an overview into the workings of language models. I then present the Czech National Bank and its institutional setting. Subsequently, I present the data and methodology, and explain the creation of the CNBSI and my empirical strategy. Next, I present my empirical results, including the estimation of the baseline model, extensions, robustness checks, and heterogeneity over time. Finally, I conclude by summarizing my findings.

1. Background

1.1 Central Bank Communications

Central banks are important institutions in many developed economies due to their role in conducting monetary policy by setting interest rates. They set interest rates in accordance with a predefined mandate, typically focused on price stability and, in the case of the Federal Reserve (Fed) (Federal Reserve, 2021), also on maintaining low unemployment rates. Setting interest rates is not the only tool available to central banks for conducting monetary policy. During periods when interest rates reach the lower bound, they may turn to unconventional tools, such as quantitative easing or forward guidance, to achieve their policy objectives. The following paragraphs offer an overview of central bank communications and explain their importance, characteristics, and the methods through which they occur. The section primarily draws on Blinder et al. (2008) who present a comprehensive analysis of modern central bank communication. In addition, I include up-to-date information on the communications practices of the Federal Reserve and European Central Bank, sourced from their official websites.

1.1.1 Why is Central Bank Communication Important?

Blinder et al. (2008) define central bank communications as the public release of information by the central bank about its policy objectives and strategies, economic forecasts, and forecasts of future policy rates. The authors state that, by communicating with the public, central banks increase available information that is necessary for forming expectations and predicting the path of future policy rates and, thereby, reduce uncertainty about their actions. Furthermore, they argue that central bank communications matter if the economy or monetary policy rules change over time and the public is able to adapt to those changes. They also argue that central bank communications matter when the public has non-rational expectations or if information asymmetries exist between the central bank and the public. Clearly, these conditions are often fulfilled in the real world.

According to Blinder et al. (2008), central banks' ability to affect the economy by communicating with the public depends on the degree to which they can influence the public's expectations about the path of future policy rates. The authors state that this is due to the term structure of interest rates, where the shape of the yield curve is predominantly determined by the public's expectations about future policy rates. Consequently, central banks can affect the economy not only through the direct effects of the current policy rate on aggregated demand but also through their ability to shape short-term and long-term interest rate expectations via their communications and the associated indirect effects on aggregated demand (Blinder et al., 2008).

1.1.2 What Do Central Banks Communicate About?

Blinder et al. (2008) note that many central banks communicate with the public by releasing various types of information; however, what and how this information is communicated differs across central banks. According to Blinder et al. (2008), central banks can release information about their objectives and strategies, policy decisions, economic forecasts, and the future path of policy rates.

The authors emphasize that independent central banks should be given well-defined mandates about their objectives and strategies, such as inflation and output targets, which can anchor the public's expectations and mitigate shocks. However, they note that many central banks have not explicitly defined such targets, which are often inferred by the public and learned over time, given the central bank's communications and actions. According to Blinder et al. (2008), examples of central banks with well-defined objectives include the Federal Reserve and the European Central Bank. They state that the Federal Reserve's mandate is to maintain low inflation and low unemployment, though it does not have an explicit inflation target. In contrast, the European Central Bank has an explicitly defined inflation target that it sets for itself (Blinder et al., 2008).

Blinder et al. (2008) point out that central banks also announce monetary policy decisions, which constitute another form of communication. Nowadays, according to the authors, many central banks release their policy decisions on the same day as they are made, eliminating the public's need to infer what the decisions will be from the central bank's open-market operations. While many central banks announce policy decisions immediately, they differ in the extent to which they release additional information accompanying their decisions (Blinder et al., 2008). They also note that additional central bank communications can include publishing meeting minutes and/or voting records or holding press conferences. Currently, both the Federal Reserve and the European Central Bank are releasing the minutes of their monetary policy meetings; the Federal Reserve releases theirs three weeks after announcing policy decisions (Federal Reserve, 2024), and the European Central Bank after four weeks (European

Central Bank, 2024a). Further, the European Central Bank holds press conferences on the days it announces its decisions (European Central Bank, 2024b).

In addition to publishing information about their objectives and policy decisions, central banks can also share information about their current economic outlooks, including forecasts of inflation rates and economic output (Blinder et al., 2008). The authors point out that central banks differ in terms of the types of information they provide about their economic outlooks. Blinder et al. (2008) also note that many inflation-targeting central banks regularly publish inflation rate forecasts. The European Central Bank publishes its macroeconomic projections four times a year (European Central Bank, 2024c). These include forecasts of the inflation rate and economic growth in the Eurozone. The Federal Reserve publishes economic projections made by its policymakers four times per year (Federal Reserve, 2022). These include forecasts of economic growth, unemployment, and the inflation rate.

Lastly, central banks can publish their outlook on the path of future policy rates in an attempt to provide some degree of forward guidance (Blinder et al., 2008). As noted by Blinder et al. (2008), such communication can range from vague and implicit signals to publishing a quantitative forecast. However, the authors note that neither the Federal Reserve nor the European Central Bank have provided such a quantitative forecast to date. They stress that communicating future policy rates has the potential to cause confusion, as the public may interpret such communication as a commitment by the central bank. However, according to Blinder et al. (2008), the path of future policy rates is not static and can always change in response to a dynamic economy.

1.1.3 How Do Central Banks Communicate?

Blinder et al. (2008) point out the importance of the content of central bank communications and the ways their communications take place. The authors note that central banks can communicate through committees or their members.

Blinder et al. (2008) explain that central banks can communicate through committees on the days monetary policy decisions are taken, via legally required reporting, such as congressional testimonies, and through regular publications, such as annual reports. The most direct way for a committee to communicate is on the day monetary policy decisions are taken (Blinder et al., 2008). For example, the authors note that the Federal Reserve releases a brief statement on its decision day, announcing a policy decision and providing a short explanation of the rationale behind it. In contrast, the European Central Bank does not issue a press release but holds a televised press conference where journalists can ask follow-up questions about the policy decision (Blinder et al., 2008). The authors point out that the European Central Bank's press conferences tend to convey less accurate information than the Federal Reserve's press releases. However, they note that televised press conferences allow central banks to directly reach out to the public by making information easily accessible, unlike press releases.

According to Blinder et al. (2008), communications between individual members can reveal a committee's inner workings and have the advantage of greater flexibility than do communications through committees. Furthermore, they note that speeches and interviews by committee members allow the central bank to communicate quickly with the public.

1.2 Related Literature

To provide an overview of the literature related to my research question, I present the most relevant economic studies investigating the predictive power of central bank sentiments on policy rates and other economic outcome variables. I begin by presenting the methods of sentiment analysis implemented in these studies, which can be divided into dictionary and deep learning methods. In addition, I include key studies that implement innovative methods of sentiment analysis in economics that are unrelated to central bank communications. I then outline the empirical strategies of the studies that investigate the predictive power of central bank sentiments.

1.2.1 Sentiment Analysis of Central Bank Communication

1.2.1.1 Dictionary Methods

Dictionary methods are widely used for sentiment analysis in economics because they rely only on pre-defined word lists and are relatively easy to implement. These methods deterministically analyze the sentiments of texts by counting the number of occurrences of specific words and phrases in a sentence that belong to a pre-defined sentiment class, such as positive, negative, or neutral. The overall sentiment of a document is then derived from the share of each sentiment class within the text. Prior to the recent rise in the popularity of machine-learning techniques, most economic studies used dictionary methods for sentiment analysis.

A prominent example of the use of a dictionary method for analyzing the sentiment of central bank communication is Apel and Grimaldi (2014). The authors created a sentiment index that measures the net hawkishness of the Swedish Riksbank's monetary policy committee meeting minutes and estimate how this sentiment index predicts the direction of upcoming policy rate changes. To analyze the sentiments of the minutes, they created a context-specific word list that contained hawkish and dovish adjective-noun combinations and counted their occurrences by processing bigrams from the left to the right. They then calculated the sentiment score for a given minute using the difference between the share of hawkish and dovish bigrams.

A disadvantage of dictionary methods is their dependence on pre-defined word lists that describe the sentiment of specified words or phrases. The accuracy with which they are able to determine the sentiment of a document depends on how closely word lists are designed to fit the language and subject of the documents they are used to perform sentiment analysis on.

Loughran and McDonald (2011) address the limitations of commonly used word lists in dictionary methods, such as the Harvard Psychological Dictionary, for sentiment analysis of financial texts. The authors point out that the limitations mainly arise because the meanings of domain-specific words used in finance differ from the meanings of the words in dictionaries used for sentiment analysis. They propose two solutions: constructing word lists that include domain-specific words and their corresponding sentiments and weighting terms based on their frequency. Loughran and McDonald (2011) find that both solutions improve the accuracy of text classification with dictionary methods.

Over the past decade, dictionary methods for sentiment analysis in economics have become increasingly sophisticated. Authors have begun to expand word lists, filter text, and increase the number of n-grams they consider when classifying text.

Picault and Renault (2017) are an example of a study that employs a more advanced dictionary method for central bank sentiment analysis. They propose a novel way to design word lists for dictionary methods by extending Loughran and McDonald's (2011) and Apel and Grimaldi's (2014) approach to sentiment analysis. The authors create their word list by first manually classifying the introduction of each European Central Bank press statement into two categories (monetary policy and economic outlook) and three sentiments (hawkish/positive, neutral, and dovish/negative). Then, they compute the probability that each n-gram (ranging from one to ten words) that appears at least twice in their text corpus belongs to a specific category and sentiment. To classify the sentiment and category of each ECB press statement, the authors aggregate the probabilities of the n-grams and select the sentiment and category that is most probable. To improve the accuracy of their predictions, they only consider n-grams with a probability of at least 50%. When overlapping n-grams appear, only the highest-order ones are kept. Picault and Renault (2017) use their word lists to create two ECB sentiment indices: one that tracks the sentiments of monetary policy communications and another that tracks the sentiments of economic outlook communications. The authors use both indices to predict upcoming ECB policy rate changes.

In addition to expanding word lists, filtering text, and increasing the number of n-grams, economists have begun to consider the *context* of words when classifying them. A basic example of a context-sensitive dictionary method for sentiment analysis is Tadle (2022), who adjusts the sentiment of words based on surrounding negations.

Tadle (2022) implements a dictionary method to analyze the sentiment of US Federal Reserve press releases and meeting minutes from December 2004 to December 2015. He then uses his method to create a continuous sentiment index to measure the Federal Reserve's policy rate tilt (whether it tends toward tightening or easing future monetary policy). The sentiment of a document is determined by classifying its sentences using a combination of keywords and polarizing terms. Hawkish keywords, such as growth and expansion, indicate improving economic conditions and rising inflationary pressure, while dovish keywords, such as recession and downturn, indicate worsening conditions and deflationary tendencies. A sentence is hawkish if it includes a hawkish keyword with more positive than negative terms or a dovish keyword with more negative terms. Vice-versa, a sentence is dovish if it contains a dovish keyword with more positive than negative terms or a hawkish keyword with more negative terms. The overall sentiment of a document is calculated by averaging the sentiments of all sentences containing hawkish or dovish keywords. Tadle (2022) then estimates the relationship between the sentiment index and financial instruments, such as Federal Reserve funds futures and exchange rates against the U.S. dollar.

While dictionary methods are useful and relatively easy to implement, they have significant limitations when performing sentiment analysis. They rely on pre-defined word lists and are unable to consider more complex context-dependent meanings of words, which manifests in the inability to understand language in a more complex way. This inability highlights the need for more advanced methods that not only capture the meanings of individual words but also their relationships with other words. Consequently, an increasing number of economic studies have started to perform sentiment analysis using deep learning to leverage the ability of neural networks to capture non-linear relationships. These methods allow sentiment analysis not only by considering the meanings of individual words but also their context and position within a sentence. In this way, sentiment can be modeled not by relying on pre-defined word lists but by interpreting sentences as a whole.

1.2.1.2 Deep Learning Methods

Deep learning methods for sentiment analysis come in various forms due to the wide variety of model architectures. Broadly, deep learning can be defined as a form of machine learning that relies on neural networks as building blocks and combines them to form complex architectures that are able to learn patterns in high-dimensional data, such as text. Due to the potentially large amount of data required to train deep learning architectures, the majority of economic studies conducting central bank sentiment analysis rely on transfer learning of pre-trained language models. Transfer learning is a form of deep learning in which a pre-trained model, such as a language model, which performs well on a general task like understanding text, is fine-tuned on an additional dataset, such as annotated central bank communications. This approach extends the language model's ability to understand text to the central bank communications domain and allows economists to use its text understanding capabilities for sentiment analysis.

A notable example of utilizing transfer learning for natural language processing tasks in finance is Araci (2019). The author develops FinBERT, a language model pre-trained on a financial text corpus and fine-tuned for sentiment analysis. FinBERT is built on the BERT architecture (Devlin et al., 2019), and demonstrates superior performance in sentiment analysis of financial text compared to other state-of-the-art machine-learning methods, even when only parts of the model are fine-tuned on a subset of the data. Araci (2019) finds that FinBERT, when trained on roughly 3,000 observations, performs 15% more accurately than other models. However, the author finds inconclusive evidence that pre-training FinBERT on a financial corpus significantly improves its performance on downstream tasks because only fine-tuning BERT's last two layers achieves comparable results in classifying the sentiment of financial text. FinBERT is fine-tuned with the Financial PhraseBank dataset (Malo et al., 2014), which consists of 4,845 English sentences sampled from financial news and labeled by 16 annotators with finance or business domain knowledge. The labels are positive, neutral, and negative and signal future developments of stock prices.

In recent years, FinBERT has become the go-to model for most economic studies conducting central bank sentiment analysis. This is because FinBERT can be directly used for classifying sentiment without fine-tuning, making it an easy option to implement for deep-learning-based sentiment analysis. However, additional fine-tuning is possible and has been implemented by a couple of authors.

Gössi et al. (2023) propose a novel way of fine-tuning FinBERT to improve its accuracy when classifying the sentiment of complex financial sentences. The authors claim that fine-tuning FinBERT with a strategy called sentiment focus increases FinBERT's overall accuracy by 5% and its accuracy in classifying complex financial sentences by 17.4%. Gössi et al.'s (2023) sentiment focus strategy consists of introducing one pre-processing step before fine-tuning: simplifying complex sentences by keeping the parts most relevant to the sentiment of the sentences. In practice, this is done by removing the parts followed by disjunctive or contrastive conjunctions. The authors state that they improve FinBERT's performance by isolating complex sentences from their dataset and then simplifying the sentences through sentiment focus. They then use FinBERT to reclassify the complex sentences, combine them with the simpler ones, and fine-tune FinBERT on this larger dataset. They show that their method makes their FinBERT version perform better on unseen data than the original FinBERT. However, their method and results should be interpreted with caution.

Another study that utilizes FinBERT is Hilscher, Nabors, and Raviv (2024), who develop an index to track the sentiment of Federal Reserve and European Central Bank communication over time. They analyze various central bank publications, such as speeches, press statements, and economic outlooks, but focus primarily on Fed minutes and ECB press conferences due to their stronger correlation with other central bank measures. They classify the sentiment of each document by classifying the sentiment of its individual sentences and then computing their average. Their sentence preprocessing follows Gössi et al.'s (2023) methodology. Furthermore, the authors apply the Hodrick-Prescott (HP) filter (Hodrick & Prescott, 1997) to smooth their sentiment index. By doing so, they aim to better align their sentiment index to changes in economic conditions and monetary policy. The authors then use their sentiment indices to investigate the relationships between Fed and ECB communications and their sentiment indices' relationship with other economic variables, such as policy rates.

In addition to the economic studies that utilize FinBERT, there are a few that fine-tune a language model from scratch for central bank sentiment analysis. Fine-tuning a language model from scratch requires annotated datasets specific to the central bank domain, which are scarce, but which offer the advantage of being able to choose state-of-the-art pre-trained language models as a backbone.

An example of a study that fine-tunes a language model from scratch is Nitoi, Pochea, and Radu (2023), who fine-tuned BERT for their sentiment analysis of central bank minutes to create sentiment indices of four European central banks. The authors fine-tune BERT with 1,998 manually annotated sentences sampled from the minutes of the Czech, Polish, Hungarian, and Romanian central banks. The authors use their own domain knowledge to annotate their dataset with three labels: hawkish, neutral, and dovish. Nitoi, Pochea, and Radu (2023) show that their fine-tuned version of BERT outperforms dictionary methods and FinBERT when classifying the sentiment of central bank minutes. However, they also observe that FinBERT achieves better results than their fine-tuned BERT version when it has been trained on a smaller dataset. The authors use their sentiment indices to predict the direction of future monetary policy, and they have published their dataset of annotated sentences.

1.2.2 Predicting Future Policy Rates from Sentiment

Creating a central bank sentiment index is typically aimed at investigating its predictive power regarding specified economic outcome variables. In most cases, authors focus on predicting the direction of upcoming policy rate changes. This is typically achieved by regressing the upcoming policy rate change on the current sentiment score while controlling for the current policy rate change. The following paragraphs show how the economic studies presented in Section 1.2.1 incorporate their sentiment indices into their empirical strategy.

Apel and Grimaldi (2014) analyze how the sentiments of Riksbank minutes predict the direction of future monetary policy. They assess their sentiment index's predictive power by estimating an ordered probit model in which they regress the upcoming policy rate change on the current sentiment score while controlling for the current policy rate change. The authors categorize their outcome variable (the upcoming policy rate change) based on its magnitude, with cutoff points of -25, 0, and 25 basis points. Their sentiment index reflects the minutes' net hawkishness over time, which indicates how much they signal a monetary policy tightening.

Apel and Grimaldi (2014) find that their sentiment index significantly predicts upcoming policy rate changes, with more hawkish (dovish) minutes being associated with rate hikes (cuts). Furthermore, the authors find that the overall sentiment of minutes, regardless of their degree of hawkishness or dovishness, predicts the upcoming policy rate change. They show this by regressing the upcoming policy rate change on sentiment dummy variables, indicating whether a minute is hawkish or dovish. To ensure the robustness of their results, the authors expand their dictionary to include less common two-word combinations and computed a new sentiment index. Then, they re-estimate their model using the new sentiment index and confirm their original findings.

Apel and Grimaldi (2014) further extend their analysis by adding GDP growth and inflation to their model. The authors then re-estimate their previous findings and robustness checks. They find that the estimates remain significant even when controlling for macroeconomic indicators. However, when they substitute their sentiment index with sentiment dummy variables, they find that the dovish sentiment dummy becomes insignificant.

Picault and Renault (2017) utilize their monetary policy and economic outlook sentiment indices to analyze how they predict current and future policy rate changes. The authors assess the predictive power of their sentiment indices by estimating an ordered probit model in which they regress an upcoming policy rate change m periods ahead on a current sentiment score, while controlling for the current policy rate change. They also extend their analysis by adding the inflation gap, the output gap, inflation expectations, and GDP growth expectations to their model. The authors further estimate models for two ordinal outcome variables: the first difference of the ECB's Main Refinancing Operation (MRR) interest rate and a second ordinal outcome variable that includes non-standard monetary policy after the financial crisis, such as quantitative easing. The first outcome variable has six possible outcomes, ranging from -75 to +25 basis points in steps of 25 basis points. The second outcome variable has four possible outcomes, ranging from -2 to 1 in steps of 1.

Picault and Renault (2017) assess the predictive power of their sentiment indices in two settings: the predictive power of prior sentiment regarding current policy rate changes, and the predictive power of current sentiment regarding future policy rate changes. For current policy rate changes, they find that the economic outlook sentiment index significantly predicts both outcome variables at the 1% level, while the monetary policy sentiment index is less significant for the first outcome variable (MRR) at the 5% level. The authors conclude that a hawkish (dovish) sentiment in previous ECB communications is associated with current rate hikes (cuts). However, a robustness check using Loughran and McDonald’s (2011) dictionary method indicates that Picault and Renault’s (2017) sentiment index does not have greater predictive power for current policy rate changes. For future policy rate changes, they find that the economic outlook sentiment index significantly predicts both outcome variables at the 1% level, while the monetary policy sentiment index has no predictive power. The authors estimate their ordered probit model for policy rate changes one and two periods ahead and show that the positive (negative) sentiment of current ECB communication predicts rate hikes (cuts). Furthermore, they show that their method produces a sentiment index with greater predictive power regarding future policy rate changes compared to other methods.

Nitoi, Pochea, and Radu (2023) estimate an ordered probit model, similar to the approaches of Apel and Grimaldi (2014) and Picault and Renault (2017), to assess how their sentiment index predicts upcoming policy rate changes. They regress an upcoming policy rate change on the current sentiment score while controlling for the current policy rate change. The authors' ordinal outcome variable takes the values -1 for a policy rate decrease, 0 for no change, and +1 for a policy rate increase. Similar to Apel and Grimaldi (2014), their sentiment index reflects the net hawkishness of the minutes over time, which indicates the degree to which they signal a monetary policy tightening.

Nitoi, Pochea, and Radu (2023) show that their sentiment index significantly predicts upcoming policy rate changes. A hawkish (dovish) sentiment is associated with an upcoming rate hike (cut). Notably, their results show that the current policy rate change does not predict upcoming policy rate changes. As a robustness check, the authors substitute their sentiment index with two variables that capture hawkish and dovish sentiment separately. Both variables significantly predict upcoming policy rate changes. Furthermore, the authors' sentiment index remains significant after they add macroeconomic indicators (the inflation gap, the output gap, and the exchange rate) to their model.

Hilscher, Nabors, and Raviv (2024) examine the relationship between the sentiment of Fed minutes and ECB press statements, as well as their relationship with macroeconomic indicators, such as future policy and Taylor-rule rates. The authors obtain their findings by conducting a local projection analysis. They show a strong statistically significant relationship between Fed and ECB sentiment indices, each influencing the other. Specifically, ECB sentiment leads Fed sentiment by one period, but negatively impacts it three to four quarters later. They also find that sentiment is positively associated with an immediate increase in Taylor-rule rates. The authors state that this is due to sentiment signaling the central bank's decision to change policy rates as a response to changing inflation expectations or economic growth. They also observe a similar relationship between sentiment and policy rates. Moreover, the strength of both effects increases over time, as sentiment increasingly predicts Taylor-rule and policy rates further into the future. The authors conclude that their findings confirm that central bank sentiments help to predict future policy rate changes as they carry forward-looking information.

1.3 Technical Background

To follow the overview of the most relevant economic literature, I continue by introducing key advances in deep learning and natural language processing related to my research question. I begin by outlining the transformer architecture, which is foundational to many state-of-the-art deep-learning architectures, and has significantly improved performance in natural language processing tasks. I then present the transformer-based language models BERT and RoBERTa, which allow fine-tuning of their pre-trained language representations for downstream tasks, such as sentiment analysis. These language models are of great importance to the analysis of central bank communication because they can understand the linguistic nuances and context of text, which enables improved sentiment analysis compared to the dictionary methods presented in section 1.2.1.1.

1.3.1 Terminology

The following paragraph provides a high-level explanation of key terms and definitions used in the subsequent sections, to familiarize readers with deep learning and natural language processing terminology.

A **token** is a word, such as *bank*, or a sub-word, such as *bank-* in *banking*. A **sequence** is a sentence and consists of one or more tokens. A sequence can be the input to or output from a neural network. A **feed-forward network** is a neural network in which information flows from one direction to the other. Feed-forward networks consist of an input layer, one or more hidden layers, and an output layer. A typical example of a feed-forward network is the multilayer perceptron. **Residual connections** allow information to skip layers within a neural network and are used to stabilize the training process and improve generalization. **Layer normalization** standardizes the values within each layer and has similar benefits as do residual connections. **Token embeddings** map tokens to a latent space by representing them as vectors. Tokens with similar meanings are mapped close to one another. It is sufficient to understand **key, query,**

and value matrices as being necessary to compute self-attention. **Self-attention** relates all tokens of an input sequence to one another. A **linear layer** performs a linear transformation. An **activation function** performs a nonlinear transformation. The **ReLU activation function** is used between layers and transforms all negative values to 0. The **softmax** activation function transforms values into a categorical distribution. For further reading, please refer to Goodfellow et al. (2016) and Zhang et al. (2023).

1.3.2 The Transformer Architecture

The transformer architecture introduced in the paper 'Attention is All You Need' by Vaswani et al. (2017) laid the foundation for many state-of-the-art deep-learning architectures and is widely used in natural language processing and other areas today. It was designed to overcome two major shortcomings of recurrent neural networks (RNNs): their sequential processing of input tokens and declining performance when they encounter longer input sequences. Unlike RNNs, the transformer can process all tokens of an input sequence simultaneously, allowing each word in a sentence to relate to all other words in that sentence.

The original transformer architecture relies on an encoder-decoder structure. The encoder stack, which is just a sequence of multiple encoders, typically consists of 6 identical layers, each containing a multi-head self-attention mechanism and a feed-forward network, with residual connections and layer normalization. The decoder stack, which is just a sequence of multiple decoders, is similar to the encoder stack. However, in addition to the multi-head self-attention mechanism and a feed-forward network, the decoder stack also includes a masked multi-head self-attention mechanism.

The encoder stack's inputs are the sum of the input sequence's token embeddings and their corresponding positional encodings. The token embeddings map tokens to a latent space by representing them as vectors of size d_{model}. The positional encodings represent the position of the tokens within the input sequence and are computed once and then reused across input sequences.

The combined input representations (token embeddings plus positional encodings) are then passed into the encoder stack, from which the key, query, and value matrices are created by multiplying them with weight matrices. The key, query, and value matrices are then projected into h attention heads, each with a dimension of d_{model}/h. Afterward, the multi-head self-attention mechanism computes each attention head's self-attention, allowing the transformer to relate tokens within the input sequence to each other. Self-attention is calculated by applying the softmax function to the dot product of the query and key matrices (attention scores), divided by the square root of the attention head size, and then multiplied by the value matrix.

Next, the outputs of the attention heads are concatenated and passed through a linear layer. The output of the self-attention mechanism is then passed through a residual connection followed by a layer normalization. The residual connection adds the hidden representations to the original input representations. The layer normalization then normalizes the resulting representations for each token along the embedding axis. To complete the encoder stack, the hidden representations are passed through a feed-forward network consisting of two linear layers and a ReLU activation function. The output of this network is then passed through another residual connection and layer normalization.

The decoder stack's inputs are the sum of the output sequence's token embeddings and positional encodings. The output sequence corresponds to the target sequence the transformer is trained on, shifted one position to the right by a start token. The combined representations (token embeddings plus positional encodings) are fed into the decoder stack, where the masked multi-head self-attention mechanism computes self-attention for each attention head. Due to the transformer's autoregressive nature of predicting the next output token given the tokens already predicted, the masked multi-head self-attention mechanism restricts self-attention to valid token pairs by masking invalid token pairs, ensuring that each token can only attend to previous tokens in the output sequence.

The masked self-attention (used as queries) is then fed into the decoder's regular multi-head self-attention mechanism, along with the encoder's output (used as keys and values). This allows the transformer to relate tokens in the output sequence to those in the input sequence. Finally, the output of the decoder's regular multi-head self-attention mechanism is passed through a feed-forward network. As in the encoder, after each decoder sublayer, the output passes through a residual connection, followed by a layer normalization. The decoder's output is then passed through a linear layer that maps the token embeddings to the transformer's vocabulary (collection of all tokens). A softmax activation function then produces a probability distribution over the entire vocabulary, predicting the next output token by choosing the token with the highest probability. This process continues until the transformer predicts an end token, signaling the end of the output sequence.

1.3.3 The BERT Language Model

BERT, or Bidirectional Encoder Representation from Transformers, is a language model introduced by Devlin et al. (2019). It is designed to pre-train context-based language representations from unlabeled text by considering input sequences' left and right contexts. This is achieved using the transformer architecture's self-attention mechanism, initially introduced by Vaswani et al. (2017). By adding one linear layer to the model, BERT allows fine-tuning of its pre-trained language representation for downstream tasks, such as sentiment analysis. By doing so, BERT achieves state-of-the-art results in many natural language processing tasks, and outperforms task-specific architectures.

BERT consists of multiple encoders stacked on top of each other. Each encoder is identical to the original transformer encoder, which includes a multi-head self-attention mechanism and a feed-forward network, with a residual connection and layer normalization between each sublayer. There are two versions of BERT: BERT-base and BERT-large. BERT-base consists of 12 encoders, has an embedding size of 768, 12 attention heads, and 110 million parameters. BERT-large consists of 24 encoders, has an embedding size of 1,024, 16 attention heads, and 340 million parameters. In addition to its architecture, BERT uses WordPiece embeddings (Wu et al., 2016) and a vocabulary of size 30,000 in its input sequences. A classification token [CLS] is added to the beginning of every input sequence, whose final hidden state serves as an aggregated input for downstream classification tasks. Furthermore, BERT's input representation can handle both single sentences and pairs of sentences within one input token sequence, using a separating token and learned segment embeddings to indicate which sentence each token belongs to. The final input representation for each token is given by summing the respective token embeddings, segment embeddings, and positional encodings.

BERT is pre-trained using two unsupervised tasks: masked language modeling and next-sentence prediction (NSP). During masked language modeling, 15% of the tokens of an input sequence are randomly masked, and the goal is for BERT to predict the correct token for the masked position, given the unmasked tokens. The final hidden representations of the masked tokens are passed through a linear layer with a softmax activation function, which produces a probability distribution over the entire vocabulary, enabling the prediction of the masked tokens by choosing the tokens with the highest probability. To mitigate the absence of masked tokens while fine-tuning BERT, 80% of the tokens chosen for masking are replaced with the mask token [MASK], 10% are replaced with random tokens, and the remaining 10% are left unchanged. In addition to predicting words given their context, BERT is trained to understand relationships between sentences. In NSP, BERT is trained on sentence pairs to predict whether the following sentence is consecutive. The sentences are consecutive in 50% of the training examples, while in the other 50%, they are not. A respective label indicates whether the following sentence is consecutive. During pre-training, the final model achieves between 97% and 98% accuracy in predicting whether the following sentence is or is not consecutive. BERT is pre-trained on BooksCorpus (Zhu et al., 2015) and English-language Wikipedia, which contain around 800 million and 2.5 billion words, respectively. The pre-training is performed with a batch size of 256 and a maximum sequence length of 512, spanning one million training steps (approximately 40 epochs). Devlin et al. (2019) use the Adam optimizer (Kingma & Ba, 2014) with a peak learning rate of 1e-4, β_1 of 0.9, β_2 of 0.999, L2 weight decay of 0.01, and linear decay of the learning rate with 10,000 warmup steps. BERT utilizes GeLU (Hendrycks & Gimpel, 2016) instead of ReLU activation functions.

Fine-tuning BERT for downstream tasks, such as sentiment analysis, only requires adding one linear layer to the pre-trained model. For classification tasks, this additional linear layer maps the last hidden state of the classification token, located at the start of each input sequence, to a vector representing the number of classes. This vector contains the logits, which are the unnormalized predictions of the model. Predictions are made by selecting the class with the highest value using the arg max of the logits. Devlin et al. (2019) suggest using the same hyperparameters for fine-tuning as those used in pre-training. The batch size for fine-tuning should be 16 or 32, with learning rates of 5e-5, 3e-5, or 2e-5, and the model should be trained for 2 to 4 epochs.

1.3.4 The RoBERTa Language Model

Although BERT (Devlin et al., 2019) outperforms many other models on natural language processing tasks, Liu et al. (2019) demonstrate that it was significantly undertrained during pre-training and did not reach its full potential. To address this, the authors introduce RoBERTa (Robustly Optimized BERT approach), and show that BERT's performance can be enhanced by modifying its pre-training strategy. RoBERTa consistently outperforms BERT on the development set, with an average improvement of 4.88 percentage points across eight out of nine tasks on the General Language Evaluation benchmark (GLUE) (Wang et al., 2019), 4.25 percentage points on the Stanford Question Answering Dataset (SQuAD) 1.1 (Rajpurkar et al., 2016), 7.55 percentage points on SquAD 2.0 (Rajpurkar et al., 2018), and 10.77 percentage points on the ReAding Comprehension Dataset from Examinations (RACE) (Lai et al., 2017). RoBERTa consistently outperforms BERT thanks to its improved pre-training strategy: it has an extended pre-training period, uses larger batch sizes and more data, removes the NSP, and applies dynamic masking and different text encodings.

While BERT is pre-trained for 1 million steps with a batch size of 256, RoBERTa is pre-trained for 500,000 steps with a batch size of 8,000. Thus, RoBERTa processes approximately 16 times more data than BERT during pre-training. Additionally, RoBERTa is pre-trained on a dataset 10 times larger than BERT's, with 160 GB of data instead of BERT's 16 GB. Like BERT, RoBERTa is pre-trained on BookCorpus and English Wikipedia, but it is also trained on three additional datasets: CC-News (Liu et al., 2019), OpenWebText (Gokaslan & Cohen, 2019), and Stories (Trinh & Le, 2018). CC-News is based on the CommonCrawl News dataset (Nagel, 2016) and consists of 63 million English news articles downloaded between September 2016 and February 2019. OpenWebText consists of roughly 8 million English documents sourced from URLs shared on Reddit. The Stories dataset consists of roughly 1 million story-like texts from the CommonCrawl dataset.

RoBERTa is pre-trained on full sentences from one or more documents, with a maximum sequence length of 512 tokens. Unlike BERT, RoBERTa does not use the NSP during pre-training, because Liu et al. (2019) show that removing it improves performance on downstream tasks after fine-tuning. Additionally, RoBERTa modifies BERT's masked language modeling task. While both models use masked language modeling, RoBERTa uses dynamic masking. In contrast to BERT, in which input sequences are masked only once with a static mask before pre-training, RoBERTa applies 10 masks to each input sequence. Therefore, each mask is processed four times during 40 pre-training epochs. Finally, RoBERTa encodes input sequences differently from BERT by using Byte-Pair Encoding (BPE) (Sennrich et al., 2016). BPE tokenizes words by splitting them into sub-words based on a statistical analysis of the training data. To avoid potential issues with Unicode characters, which can make up a significant portion of the vocabulary, RoBERTa uses byte-level BPE (Radford et al., 2019). This approach uses bytes instead of Unicode characters as sub-word units, allowing RoBERTa to handle text input without encountering unknown tokens. Furthermore, RoBERTa is pre-trained using the Adam optimizer with a peak learning rate of 4e-4, ε of 1e-6, β_1 of 0.9, β_2 of 0.998, L2 weight decay of 0.01, and linear decay of the learning rate with 30,000 warmup steps.

Fine-tuning RoBERTa for downstream tasks is identical to fine-tuning BERT, as both models share the same architecture, and only adding one linear layer to the pre-trained model is required. However, fine-tuning RoBERTa differs from fine-tuning BERT in its hyperparameter settings. RoBERTa uses early stopping while training for up to 10 epochs instead of training for a smaller fixed number of epochs. This allows the model to stop the training when performance on the development set declines within a specified tolerance. Additionally, RoBERTa uses a smaller learning rate warm-up ratio than BERT, with 6% of training examples used instead of 10%. The batch sizes for fine-tuning RoBERTa can be 16, 32, or 48, and weight decay can be 0.01 or 0.1. RoBERTa always uses a linear learning rate decay. Similar to BERT, there are two versions of RoBERTa: RoBERTa-base and RoBERTa-large. Both models share the same architecture as BERT-base and BERT-large.

2. The Institutional Setting of the Czech National Bank

The Czech National Bank is the Czech Republic's central bank and was created as an institution in January 1993 during the country's transition from communism to a free market economy. It is responsible for the Czech Republic's monetary policy and is independent of political influence, as regulated in the CNB Act (No. 6/1993). All the information in Section 2 of this thesis is sourced from the official website of the Czech National Bank.

2.1. The Mandate

The Czech National Bank's mandate (Czech National Bank, 2024a) includes maintaining price stability, ensuring financial stability and proper functioning of the financial system in the Czech Republic, issuing banknotes and coins, regulating the circulation of money, overseeing clearing between banks, and monitoring the organizations and institutions active in the Czech financial market. The CNB maintains price stability by using monetary policy to keep inflation low and stable. To ensure financial stability and proper functioning of the financial system, the Czech National Bank applies a macroprudential policy, which it considers essential for maintaining price stability. Additionally, by combining monetary and macroprudential policies and managing foreign exchange reserves, the CNB aims to preserve confidence in the Czech koruna and the stability of the macroeconomy.

2.2 CNB Monetary Policy

The Czech National Bank conducts monetary policy independently to fulfill its constitutional task of maintaining price stability, which it defines as low and stable price growth (Czech National Bank, 2024b). The CNB uses inflation targeting to keep inflation near its 2% annual target and sets interest rates as its main monetary policy tool to maintain price stability. The interest rates set by the CNB's Bank Board influence the market rates in the Czech Republic and thereby steer the economy and inflation in the desired direction. Since January of 2008, the Bank Board meets eight times per year. Prior to that, the Bank Board met on a monthly basis. The Bank Board can also meet in exceptional cases, such as a crisis, to decide on monetary policy outside its regular meetings. During Bank Board meetings, its members discuss the uncertainties and risks associated with the current macroeconomic forecast and decide whether to change interest rates. The CNB's monetary policy is based on the current macroeconomic forecast created by its Monetary Department.

2.3 Communication with the Czech Public

The CNB Bank Board meets eight times per year to decide on interest rate setting. Its decision is announced at 2:30 p.m. on the same day as the meeting (Czech National Bank, 2024c). At 3:45 p.m., the Bank Board holds a press conference to release a statement detailing the voting records and the rationale behind the decision. The CNB also releases the minutes of the Bank Board meeting on the same day as the meeting. The Bank Board meeting minutes provide detailed information on the discussions, including who contributed to specific topics and any differing opinions expressed within the committee. The Czech National Bank publishes forecasts for key macroeconomic indicators, including headline inflation, monetary policy-relevant inflation, annual GDP growth, short-term market interest rates, and the koruna-to-euro exchange rate (Czech National Bank, 2024d). These forecasts are released quarterly in the CNB’s monetary policy report. The report also provides outlooks for foreign economies and explains the Bank Board's decisions in light of the current domestic and international economic environment.

In addition to the CNB Bank Board's official communications as a committee, individual members frequently give interviews to the press to share their views on the current monetary policy situation (Czech National Bank, 2024e). These interviews, often televised, address a variety of topics and potential questions the public may have about the current economic environment and the CNB’s future monetary policy.

3. Data

I employ three types of data to investigate how the sentiment of CNB minutes predicts the direction of upcoming policy rate changes in the Czech Republic. I use annotated sentences to fine-tune the language model that classifies the sentiment of CNB minutes, CNB Bank Board meeting minutes as the central bank communication for sentiment analysis, and economic data to estimate how the sentiment of CNB minutes predicts upcoming policy rate changes while controlling for current policy rate changes and macroeconomic indicators.

3.1 Data for Fine-Tuning the Language Model

Fine-tuning a language model for sentiment analysis is a supervised learning task that requires labeled data. However, labeled data for sentiment analysis in the central bank domain is scarce. Therefore, I fine-tune my language model on sentences from central bank minutes annotated by Nitoi, Pochea, and Radu (2023), who were among the first to publish their data. The authors annotate 1,998 randomly selected sentences from 591 central bank minutes from the Czech Republic, Poland, Hungary, and Romania between January 2007 and January 2022. Their dataset consists of 687 sentences labeled hawkish and 651 and 660 sentences labeled dovish and neutral, respectively. They note that sentences designated as hawkish indicate a contractionary monetary policy stance associated with an overheating economy. Vice-versa, sentences labeled dovish indicate an expansionary monetary policy stance associated with a sluggish economy, and sentences labeled neutral indicate a stable monetary policy stance associated with a balanced economy.

Sentence	**Annotation**	**Date**
“The domestic economy was continuing to decline from the peak of the business cycle.”	dovish	6.11.2008, 5.2.2009
“The Board then discussed economic growth.”	neutral	1.11.2012
“Workforce shortages could therefore increase pressures on inflation.”	hawkish	29.3.2007

Table 1: Annotated Sentences from CNB Minutes (Nitoi, Pochea, and Radu, 2023)

Table 1 presents three example sentences from the annotated dataset by Nitoi, Pochea, and Radu (2023). The sentences' minutes can be sourced from the official CNB website (Czech National Bank, 2024c).

To ensure the robustness of my results, I divide the 1,998 annotated sentences into training, development, and testing sets. I use the training set to fine-tune the language model to perform sentiment analysis, the development set to tune the model's hyperparameters, and the testing set to assess the model's performance on previously unseen data. The training set contains 1,618 labeled sentences, and the development and testing sets contain 180 and 200 sentences each. Like Nitoi, Pochea, and Radu (2023), I use 90:10 splits due to the limited size of the dataset and my objective of maximizing the accuracy of the fine-tuned language model. Therefore, I divide the annotated sentences into 90% training and 10% testing data and then divide the training data again into 90% training and 10% development data, resulting in 81% training, 9% development, and 10% testing data. Further, I use stratified sampling to ensure an equal distribution inside the training, development, and testing data of labels and central banks which the annotated sentences are sampled from.

3.2 CNB Bank Board Meeting Minutes

I use 265 CNB Bank Board meeting minutes from January 1998 to September 2024 to investigate how the sentiment of central bank minutes predicts policy rate changes in the Czech Republic. My corpus for sentiment analysis is based on Nitoi, Pochea, and Radu's (2023) published corpus of 127 CNB minutes from January 2007 to February 2022. I extend their corpus by manually collecting and adding 138 CNB minutes, from January 1998 to December 2006 and March 2022 to September 2024 (Czech National Bank, 2024c). This enables me to analyze the sentiment of CNB minutes across the past 26 years, extending Nitoi, Pochea, and Radu's (2023) analysis period by 11 years.

In my corpus, the average CNB minute has a word count per document of 1,007 words, an average sentence per document of 47, and an average number of words per sentence of 21. A typical minute starts by assessing the Czech Republic's then-current monetary and economic situation, and ends with a vote and decision about the current policy rate. Usually, the board discusses the current economic situation and how this economic situation is contributing to the decision the CNB will take. It also analyzes the position of the Czech Republic internationally and domestically.

3.3. Economic Data

My empirical analysis uses data on CNB interest rates, inflation, real GDP, and the koruna-to-euro exchange rate from January 1998 to September 2024.

For interest rates, I use the CNB's two-week repo rate, the central bank's main policy rate, which is the interest rate at which it lends to commercial banks. The data is sourced from the official CNB website (Czech National Bank, 2024f). To capture changes in the interest rate over time, I use the first difference of the two-week repo rate.

Inflation is measured using the harmonized indices of consumer prices (HICP). According to the Czech Statistical Office (Czech Statistical Office, 2024), the HICP is a consumer price index that allows the comparison of consumer price level changes across EU member states. The main difference between the HICP and the Czech Republic's national CPI is in the composition of the consumer basket. The HICP includes revenue from non-residents spendings in the Czech Republic, but does not include housing prices. Conversely, the national CPI does not include revenue from non-residents, but includes housing prices. HICP data is sourced from the official ECB website (European Central Bank, 2024d).

Real GDP is measured using the B1GM gross domestic product. This quarterly, seasonally adjusted time series is based on constant prices. I decompose the series into its trend and cyclical components using the HP filter, while using the trend in my empirical analysis. The data is sourced from the official CNB website (Czech National Bank, 2024g).

The koruna-to-euro exchange rate is also obtained from the official CNB website (Czech National Bank, 2024h). Because my analysis covers the period from 1998 to 2024 and the euro was introduced in January 1999, I reconstruct the koruna-to-euro exchange rate for 1998. I achieved this by using the koruna-to-deutsche mark exchange rate for 1998 (Czech National Bank, 2024h) and dividing it by 1.95583, which is the official deutsche mark-to-euro exchange rate.

4. Methodology

4.1 A Czech National Bank Sentiment Index

4.1.1 Transfer Learning with Language Models

I employ transfer learning to develop a Czech National Bank Sentiment Index based on the Bank Board's meeting minutes. Transfer learning enables fine-tuning of pre-trained models, such as language models, to utilize their pre-existing capabilities for domain-specific downstream tasks. Unlike most deep-learning techniques, which require large amounts of data and computational capabilities, transfer learning requires less data and training to generalize well on unseen data. This is particularly true for natural language processing, where transfer learning with pre-trained language models achieves state-of-the-art performance in many downstream tasks, and dominates the leaderboards for sentiment analysis. Transfer learning, therefore, constitutes a cost-effective alternative for training a model from scratch, and enables state-of-the-art performance in natural language processing. Further, pre-trained language models are widely available on platforms like Hugging Face, which can be easily downloaded and fine-tuned using Python.

4.1.2 Fine-Tuning RoBERTa for Central Bank Sentiment Analysis

To develop a Czech National Bank sentiment index based on the Bank Board's meeting minutes, I first downloaded the pre-trained RoBERTa language model using Hugging Face's Transformers library (Wolf et al., 2020). Then, I fine-tuned the RoBERTa backbone model with data from Nitoi, Pochea, and Radu (2023) using the Pytorch library (Paszke et al., 2019). Their dataset consists of 1,998 sentences and their corresponding labels (hawkish, neutral, and dovish), randomly sampled from the minutes issued by four European central banks. Then, I wrote three Python scripts[1] to implement the fine-tuning of RoBERTa and classify the CNB minutes. The first script prepares the data for training, development, and testing. The second script defines the model architecture, processes the data, and fine-tunes RoBERTa. The third script processes the CNB minutes and predicts their sentiment using the fine-tuned model.

I selected RoBERTa-large as the backbone model for developing the Czech National Bank sentiment index because of its superior performance. RoBERTa-large outperforms RoBERTa-base, due to its approximately three times larger number of parameters and therefore, its greater

[1] https://github.com/bebrunn/Predicting-Policy-Rate-Changes-from-Central-Bank-Minutes

ability to learn language representations. RoBERTa-base and RoBERTa-large outperform both BERT versions thanks to their optimized training regimen (Liu et al., 2019).

Table 2 compares the performance of the RoBERTa and BERT language models fine-tuned on the dataset by Nitoi, Pochea, and Radu (2023) for classifying the sentiment of unseen central bank communications. The accuracy column represents the proportion of correctly classified sentences compared to all sentences in the development set.

Language model	**Accuracy in %**
RoBERTa-large[2]	92.22
RoBERTa-base[3]	91.67
BERT-large[4]	91.11
BERT-base[5]	91.11

Table 2: Fine-tuned Language Models in Comparison

4.1.3 Model Implementation and Training

I fine-tuned RoBERTa-large by adding a fully connected linear layer to the backbone model, as suggested by Devlin et al. (2019). This layer projects the output of the last hidden layer of the backbone model to the number of classes, which, in this case, is three for classifying CNB minutes into hawkish, neutral, and dovish classes. I pass the representations of the classification token [CLS] at the beginning of each sequence through the linear layer instead of using the representations of the entire sequence. Devlin et al. (2019) included the [CLS] token as an aggregated sequence representation, meaning that it captures information from the entire input sequence, and it is specifically intended for classification tasks.

I follow Liu et al. (2019) for fine-tuning RoBERTa, using the Adam optimizer with a weight decay of 0.01 and label smoothing for 10% of the labels to prevent overfitting and improve generalization. Further, I use a linear learning rate schedule with a warm-up for 6% of the total input sequences to improve convergence while minimizing loss, and trained the model for up to 10 epochs with a learning rate of 3e-05. In addition, I used early stopping to obtain the

[2] RoBERTa-large was trained for 10 epochs with a batch size of 32 and a learning rate of 3e-05.
[3] RoBERTa-base was trained for 6 epochs with a batch size of 32 and a learning rate of 3e-05.
[4] BERT-large was trained for 4 epochs with a batch size of 16 and a learning rate of 3e-05.
[5] BERT-base was trained for 2 epochs with a batch size of 16 and a learning rate of 5e-05.

weights from the best-performing epoch by stopping the training if the development loss failed to decrease for two consecutive epochs.

4.2 The Sentiment Analysis of CNB Minutes

4.2.1 Preprocessing Text Inputs for Inference

I pre-processed every input sequence with Hugging Face's pre-trained RoBERTa AutoTokenizer, which I downloaded with the RoBERTa model from the Transformers library. Wolf et al. (2020) note that the AutoTokenizer splits each input string into sub-word token strings and maps each to an input ID. It also extends the model's vocabulary by adding new tokens; for RoBERTa, this is done by Byte-Pair Encoding. Additionally, the AutoTokenizer handles special tokens, such as mask [MASK] and start-of-sentence [SOS] tokens, and RoBERTa-specific tokens, such as the separator [SEP] and classification [CLS] tokens. Because RoBERTa requires input sequences in a batch to have the same length, the AutoTokenizer provides padding by adding a padding token [PAD] to the end of each sequence that is shorter than the longest sequence in the batch.

I construct the Czech National Bank Sentiment Index by analyzing the sentiment of each CNB minute from January 1998 to September 2024 using my fine-tuned RoBERTa language model. Before employing the model for inference, I preprocess the minutes by splitting them into sentences using the Spacy natural language processing library (Honnibal et al., 2020). Splitting the minutes into sentences is necessary because RoBERTa was fine-tuned on Nitoi, Radu, and Pochea's (2023) sentence-level annotated dataset. As a result, the model achieves optimal performance when predicting sentiment for input sequences that are similar in length to the input sequences used during training. I then tokenize each sentence with the RoBERTa AutoTokenizer before predicting its sentiment.

4.2.2 Sentence-Level and Minute-Level Sentiment

For each CNB minute, I predict the sentiment of its sentences. For each sentence, RoBERTa returns the raw logits over the number of classes, in this case, three. I then utilize the raw logits to determine the class of a sentence (hawkish, dovish, or neutral) and obtain its net hawkishness. A sentence is assigned to the class with the maximum predicted logit value, and the net hawkishness is computed by taking the difference between the hawkish and dovish logits (Nitoi, Pochea, and Radu, 2023). This log-odds ratio measures how much more likely a sentence is to be hawkish than dovish, and allows me to obtain a continuous CNB sentiment index. To compute the overall sentiment of a minute, I aggregate the sentiment of its sentences. To obtain the categorical sentiment of a minute, I automatically count the predicted classes of its sentences and choose the class with the most occurrences. Further, I obtain the net hawkishness of a minute as the average net hawkishness of its sentences. Figure 1 shows the Czech National Bank Sentiment Index (CNBSI) plotted against a variable indicating rate hikes, rate cuts, and no changes in the policy rate from January 1998 to September 2024. Here, the CNBSI is normalized to fall within the range of [-1, 1].

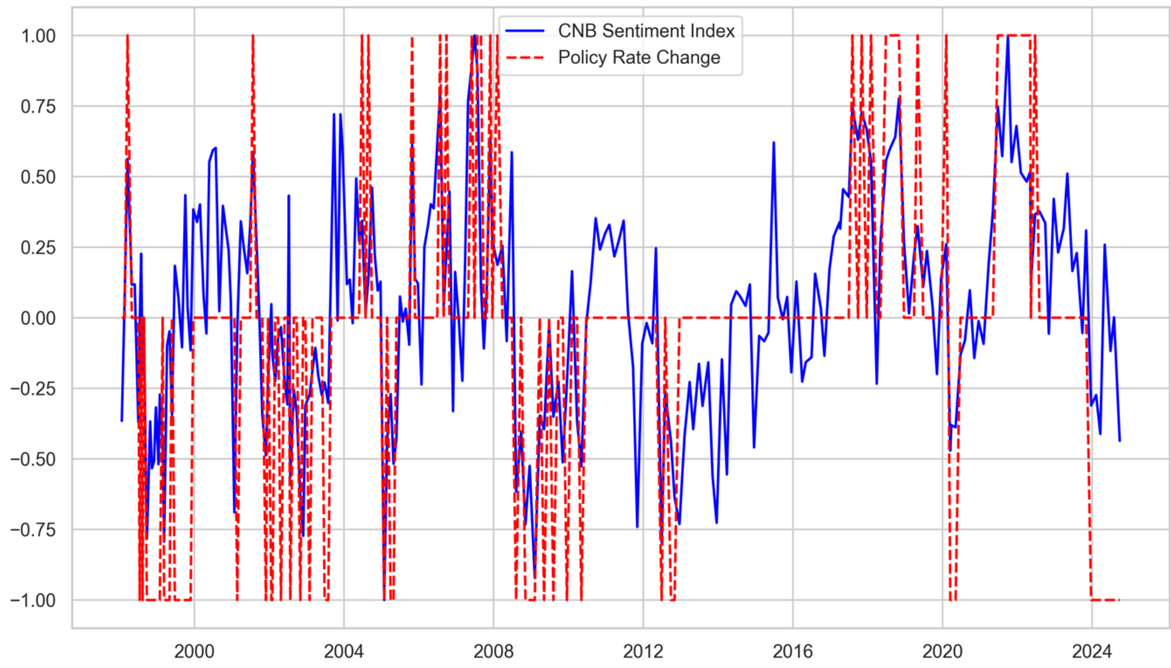

Figure 1: The CNBSI plotted against the Policy Rate Change

4.3 Estimating the Predictive Power of the CNBSI

To evaluate how the Czech National Bank Sentiment Index predicts upcoming policy rate changes, I begin my empirical analysis by estimating two ordered probit models. Ordered probit models estimate how an ordered outcome variable relates to a set of explanatory variables by treating the outcome as a latent variable (Cameron & Trivedi, 2005). Furthermore, Cameron and Trivedi (2005) define the probability of observing a specific outcome as the probability that the latent variable falls between the two cutoff points corresponding to that outcome. In an ordered probit model, the error term of the latent variable is assumed to follow a normal distribution, and the probability of different outcomes is computed using the cumulative normal distribution as a link function.

The estimated coefficients of an ordered probit model are difficult to interpret because they do not represent the average change in the outcome variable associated with a one-unit or one-standard-deviation change in the explanatory variables (Apel & Grimaldi, 2014). Instead, one must compute the marginal effects, which measure how the probability of an ordinal outcome changes when the explanatory variables change. In my analysis, I compute the conditional marginal effects at the mean value of each explanatory variable, which allows me to obtain estimates of how the dependent variables change if the explanatory variables change by one unit or one standard deviation from their mean.

4.3.1 The Baseline Model

To estimate the extent to which the Czech National Bank Sentiment Index predicts upcoming policy rate changes, I follow the literature and begin my empirical analysis by estimating the baseline model (Apel & Grimaldi (2014), Picault and Renault (2017), Nitoi, Pochea, and Radu (2023))

$$\Delta r_{t+1} = \beta_1 \Delta r_t + \beta_2 CNBSI_t + \varepsilon_t. \tag{1}$$

In the baseline model (1), I regress the upcoming policy rate change Δr_{t+1} on the sentiment score of the current minute, captured by $CNBSI_t$, while controlling for the current policy rate change Δr_t. The error term is represented by ε_t.

4.3.2 Extending the Baseline Model

I extend the baseline model (1) by including additional macroeconomic controls to account for economic factors that may influence upcoming policy rate changes. Similar to Apel and Grimaldi's (2014) and Nitoi, Pochea, and Radu's (2023) approach, I expand the baseline model to the Taylor-rule (Taylor, 1993) like regression equation

$$\Delta r_{t+1} = \beta_1 \Delta r_t + \beta_2 CNBSI_t + \beta_3 \Delta(\pi_t - \pi^*) + \beta_4 \Delta y_t + \beta_5 CZK_t + \varepsilon_t. \tag{5}$$

The extension of the baseline model (2) includes the inflation gap change $\Delta(\pi_t - \pi^*)$, Δy_t, the change in the percentage deviation of real GDP from its trend, and the koruna-to-euro exchange rate CZK_t. π^* denotes the CNB's inflation target of 2% annually.

The rationale behind including the inflation gap change is due to the CNB's mandate of price stability, which it defines as keeping the inflation rate close to its target of 2%. If the CNB credibly commits to its target, it is expected to raise the interest rates when inflation exceeds the target and to lower them when inflation falls below the target.

The inclusion of the change in the percentage deviation of real GDP from its trend follows Taylor (1993), who shows that the Federal Reserve's monetary policy decisions can be modeled with the inflation gap and the percentage deviation of real GDP from its trend. For my analysis, I assume that the CNB implicitly considers the state of the economy when it makes interest rate decisions. Therefore, I include the change in the percentage deviation of real GDP from its trend in my regression equation.

Including the koruna-to-euro exchange rate is based on the CNB's exchange rate commitment of 27 koruna to the euro, which was in place from November 2013 until April 2017. Moreover, shocks to the koruna exchange rate likely impact the Czech Republic's real economy and inflation rate through various channels, such as net exports. This highlights the potential importance of the koruna-to-euro exchange rate in the CNB's decision making process and justifies its inclusion in the model.

5. Empirical Results

The empirical results of my thesis are comprised of three parts: the results of estimating the baseline model and its extension as presented in Section 4.3, two robustness checks which re-estimate the baseline model and its extension with alternative measures of CNB sentiment, and an analysis of heterogeneity over time for three distinct periods.

5.1 Results of Estimating the Baseline Model and its Extensions

Table 3 presents the results of estimating the baseline model (1) and its extensions (2) through (5). I obtained the results by regressing the ordinal outcome variable (the CNB's upcoming policy rate change) on the current sentiment score (measured by the CNBSI), while including the current policy rate change and three macroeconomic indicators as control variables. The macroeconomic indicators are: the change in the inflation gap, the change in the percentage deviation of real GDP from its trend, and the koruna-to-euro exchange rate. I constructed the ordinal outcome variable by grouping the CNB's interest rate decisions into three ordinal categories based on the direction of their change: a *rate hike*, which represents a policy tightening; a *rate cut*, which represents a policy easing; and *no change*, indicating that the interest rate remains unchanged. The outcome variable is ordered in terms of rate hikes.

To assess the specific impact of each macroeconomic indicator on the regression results, I include them individually and then jointly into the baseline model (1). This ensures model comparability and allows me to analyze how each macroeconomic indicator predicts upcoming policy rate changes. Including the control variables step-by-step allows me to observe how the coefficient on the CNBSI changes in each estimated model.

The results show that the coefficients of the current policy rate change and the CNBSI are statistically significant at the 1% level and positive for all models except (5), where the coefficient on the current policy rate change is statistically significant at the 5% level. Furthermore, the change in the inflation gap and the koruna-to-euro exchange rate are statistically significant at the 5% level – positive for the inflation gap change and negative for the koruna-to-euro exchange rate. The coefficients for the change in the percentage deviation of real GDP from its trend are insignificant.

The cutoff points for models (1) through (3) are statistically significant: the lower cutoff points at the 5% level and the upper cutoff points at the 1% level. In models (4) and (5), the lower cutoff points are statistically significant at the 1% level, while the upper cutoff points are insignificant. The upper cutoff points become insignificant after including the koruna-to-euro exchange rate as a control variable, which indicates that the boundary between the *no change* and *rate hike* categories may be ill-defined.

Table 3: Ordered Probit Estimates of the Baseline Model and its Extensions

	(1)	(2)	(3)	(4)	(5)
Δr_t	0.696^{***}	0.597^{***}	0.694^{***}	0.585^{***}	0.485^{**}
	(0.192)	(0.203)	(0.192)	(0.199)	(0.212)
$CNBSI_t$	0.417^{***}	0.415^{***}	0.406^{***}	0.440^{***}	0.425^{***}
	(0.106)	(0.109)	(0.107)	(0.108)	(0.111)
$\Delta(\pi_t - \pi^*)$		0.266^{**}			0.252^{**}
		(0.112)			(0.111)
Δy_t			0.0533		0.0524
			(0.0488)		(0.0486)
CZK_t				-0.0520^{**}	-0.0517^{**}
				(0.0211)	(0.0210)
$0\|1$	-0.400^{**}	-0.519^{**}	-0.405^{**}	-2.022^{***}	-2.133^{***}
	(0.193)	(0.208)	(0.194)	(0.712)	(0.726)
$1\|2$	2.135^{***}	2.061^{***}	2.138^{***}	0.549	0.483
	(0.231)	(0.241)	(0.232)	(0.673)	(0.679)
N	264	264	264	264	264
$pseudo-R^2$	0.174	0.187	0.176	0.190	0.203

Standard errors in parentheses. $0|1$ denotes the cutoff point from *rate cut* to *no change* and $1|2$ from *no change* to *rate hike*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The positive coefficient for the current policy rate change, the CNBSI, which indicates the net hawkishness of the current CNB minute, and the change in the inflation gap indicate that an increase in either variable increases the probability of the upcoming policy rate change falling into a higher category. If the CNB raises interest rates, releases more hawkish minutes, or the change in the inflation gap increases, the probability of a rate hike during the next Bank Board meeting rises. On the contrary, the negative coefficient for the koruna-to-euro exchange rate indicates that an increase in the exchange rate decreases the probability of the upcoming policy rate change falling into a higher category. If the koruna appreciates, the probability of a rate hike at the next Bank Board meeting decreases.

To deepen my analysis, I compute the conditional marginal effects at the mean value of each explanatory variable to assess how a one-unit or one-standard-deviation increase affects the probability of the upcoming policy rate change falling into each ordinal category. A 100-basis point increase of the current policy rate lowers the probability of an upcoming rate cut by 10.9 percentage points, raises the probability of no change by 5 percentage points, and raises the probability of an upcoming rate hike by 5.9 percentage points. A one-standard-deviation increase in the CNBSI lowers the probability of an upcoming rate cut by 9.6 percentage points, increases the probability of no change by 4.4 percentage points, and increases the probability of an upcoming rate hike by 5.1 percentage points. A 100-basis point increase in the inflation gap lowers the probability of an upcoming rate cut by 5.6 percentage points, increases the probability of no change by 2.6 percentage points, and increases the probability of an upcoming rate hike by 3 percentage points. Finally, a one-koruna increase in the koruna-to-euro exchange rate raises the probability of an upcoming rate cut by 1.2 percentage points, lowers the probability of no change by 0.5 percentage points, and lowers the probability of an upcoming rate hike by 0.6 percentage points.

My findings show that the current sentiments of the CNB Bank Board meetings minutes, measured by the Czech National Bank Sentiment Index (CNBSI), are a significant predictor of upcoming policy rate changes, alongside other variables. A one-standard-deviation increase in the hawkishness of a current minute decreases the probability of an upcoming rate cut by 9.6 percentage points, increases the probability of no change by 4.4 percentage points, and increases the probability of an upcoming rate hike by 5.1 percentage points. These findings suggest that the CNBSI provides valuable insights for predicting upcoming policy rate changes and support the potential to forecast such changes based on central bank minutes in the Czech Republic.

5.2 Running Robustness Check with Alternative Sentiment Measures

To ensure the robustness of my results, I re-estimated models (1) through (5) with two alternative measures of CNB sentiment: a BERT net hawkishness index and RoBERTa sentiment dummy variables. Similarly to the CNBSI, the BERT net-hawkishness index measures how much more likely a minute is to be hawkish than dovish, and is derived from the averaged sentence-level differences between hawkish and dovish logits. The raw logits come from the BERT-large model, fine-tuned on the same dataset as the RoBERTa-large used to obtain the CNBSI. The RoBERTa sentiment dummy variables indicate the overall sentiment of a minute, categorized as hawkish, dovish, or neutral. It is derived for a minute by selecting the most frequently predicted class among its sentences. The raw logits come from the same fine-tuned RoBERTa-large model used to obtain the CNBSI. To avoid multicollinearity, I omit the neutral sentiment dummy variable.

5.2.1 The BERT Net Hawkishness Index

Table 4 presents the results of estimating the baseline model (1) and its extensions (2) through (5) using the BERT sentiment index as an alternative sentiment measure. The results in Table 4 closely resemble those in Table 3.

The coefficients of the current policy rate changes are positive and statistically significant at the 1% level for (1) through (4), and at the 5% level for (5). The changes in the inflation gap and the koruna-to-euro exchange rate are statistically significant at the 5% level – positive for the inflation gap changes and negative for the exchange rate. The coefficients of the changes in the percentage deviation of real GDP from its trend are insignificant. All lower cutoff points are statistically significant at the 1% level (2, 4, 5) or at the 5% level (1, 3). The upper cutoff points are statistically significant at the 1% level for (1) through (3).

The coefficients of the BERT sentiment index are all positive and statistically significant at the 1% level. These findings indicate that an increase in the BERT sentiment index increases the probability of the upcoming policy rate change falling into a higher category. This is supported by the estimated marginal effects, which show that a one-standard-deviation increase in hawkishness of minutes, measured by the BERT sentiment index, lowers the probability of an upcoming rate cut by 10 percentage points, increases the probability of no change by 4.7 percentage points, and increases the probability of an upcoming rate hike by 5.4 percentage points.

Re-estimating equations (1) through (5) using the BERT sentiment index as an alternative sentiment measure supports the findings from Section 5.1 that the sentiment of CNB minutes is a significant predictor of upcoming policy rate changes. A one-standard-deviation increase in the net hawkishness of CNB minutes, as measured by the CNBSI and the BERT sentiment index, similarly affects the probabilities of each ordinal outcome. This indicates that both sentiment indices are equally effective predictors of upcoming policy rate changes, although RoBERTa demonstrates slightly better performance in classifying the sentiment of central bank communications. My findings suggest that using a more advanced language model, like RoBERTa, compared to BERT does not significantly improve the predictive power of sentiment indices. To achieve that, utilizing a model that greatly outperforms both RoBERTa and BERT may be necessary.

Table 4: Ordered Probit Estimates of the Baseline Model and its Extensions with BERT Sentiment Index

	(1)	(2)	(3)	(4)	(5)
Δr_t	0.658^{***}	0.550^{***}	0.656^{***}	0.558^{***}	0.450^{**}
	(0.198)	(0.210)	(0.198)	(0.203)	(0.218)
$CNBSI_t^{BERT}$	0.442^{***}	0.445^{***}	0.430^{***}	0.457^{***}	0.448^{***}
	(0.109)	(0.111)	(0.109)	(0.110)	(0.113)
$\Delta(\pi_t - \pi^*)$		0.278^{**}			0.265^{**}
		(0.113)			(0.112)
Δy_t			0.0514		0.0501
			(0.0492)		(0.0490)
CZK_t				-0.0497^{**}	-0.0495^{**}
				(0.0207)	(0.0207)
$0\|1$	-0.440^{**}	-0.570^{***}	-0.444^{**}	-1.986^{***}	-2.109^{***}
	(0.199)	(0.215)	(0.199)	(0.696)	(0.714)
$1\|2$	2.105^{***}	2.026^{***}	2.108^{***}	0.594	0.522
	(0.234)	(0.244)	(0.234)	(0.661)	(0.670)
N	264	264	264	264	264
$pseudo-R^2$	0.177	0.191	0.179	0.191	0.206

Standard errors in parentheses. $0|1$ denotes the cutoff point from *rate cut* to *no change* and $1|2$ from *no change* to *rate hike*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.2.2 RoBERTa Sentiment Dummies

Table 5 presents the results of estimating the baseline model (1) and its extensions (2) through (5) using the RoBERTa sentiment dummy variables as an alternative sentiment measure. The results in Table 5 closely resemble those in Table 3.

Table 5: Ordered Probit Estimates of the Baseline Model and its Extensions with RoBERTa Sentiment Dummies

	(1)	(2)	(3)	(4)	(5)
Δr_t	0.885***	0.769***	0.871***	0.782***	0.657***
	(0.172)	(0.181)	(0.174)	(0.178)	(0.191)
I_{dovish}	-0.175	-0.151	-0.165	-0.277	-0.240
	(0.214)	(0.214)	(0.215)	(0.214)	(0.216)
$I_{hawkish}$	0.677***	0.713***	0.665***	0.653***	0.675***
	(0.240)	(0.249)	(0.240)	(0.241)	(0.249)
$\Delta(\pi_t - \pi^*)$		0.293***			0.276**
		(0.112)			(0.111)
Δy_t			0.0785		0.0777
			(0.0514)		(0.0509)
CZK_t				-0.0475**	-0.0469**
				(0.0208)	(0.0207)
$0\|1$	-0.105	-0.231	-0.124	-1.611**	-1.732**
	(0.176)	(0.188)	(0.178)	(0.682)	(0.698)
$1\|2$	2.363***	2.289***	2.361***	0.886	0.825
	(0.227)	(0.235)	(0.229)	(0.651)	(0.658)
N	264	264	264	264	264
$pseudo-R^2$	0.158	0.173	0.161	0.171	0.188

Standard errors in parentheses. 0|1 denotes the cutoff point from *rate cut* to *no change* and 1|2 from *no change* to *rate hike*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The coefficients of the current policy rate changes are positive and statistically significant at the 1% level for all models. The coefficients of the changes in the inflation gap are positive and statistically significant at the 1% level for (2) and at the 5% level for (5). Further, the coefficients of the koruna-to-euro exchange rate are all negative and statistically significant at the 5% level. The coefficients of the changes in the percentage deviations of real GDP from its trend are insignificant. In (1) through (5) only one cutoff point is statistically significant. In (4) and (5), the lower cutoff point is statistically significant at the 5% level, while the upper cutoff point is insignificant. In all other models, the upper cutoff point is statistically significant at the 1% level, while the lower cutoff point is insignificant.

Notably, the coefficients of the hawkish sentiment dummy are all positive and statistically significant at the 1% level, whereas the coefficients of the dovish sentiment dummy are all negative but insignificant. These results indicate that only hawkish minutes increase the probability of the upcoming policy rate change falling into a higher category, while dovish minutes have no effect. This is supported by the estimated marginal effects, which show that a hawkish minute lowers the probability of a policy rate decrease by 15.5 percentage points, increases the probability of no change by 6.9 percentage points, and increases the probability of a policy rate increase by 8.5 percentage points. The estimated marginal effects of the current policy rate change and the hawkish sentiment dummy closely resemble each other, indicating that both variables similarly predict upcoming policy rate changes.

Re-estimating (1) through (5) using the RoBERTa sentiment dummies shows that hawkish minutes are a predictor of upcoming policy rate changes, while dovish minutes have no predictive power. A hawkish minute decreases the probability of an upcoming rate cut by 15.5 percentage points, increases the probability of no change by 6.9 percentage points, and increases the probability of an upcoming rate hike by 8.5 percentage points. These findings suggest that primarily hawkish minutes contain signals about upcoming policy rate changes.

5.3 Checking for Heterogeneity Over Time

To assess whether the predictive power of the Czech National Bank Sentiment Index varies over time, I re-estimate models (1) and (5) for three different periods: 1998–2006, 2007–2016, and 2017–2024. I choose these periods because they include varying economic conditions that the Czech economy experienced. The 1998–2006 period includes the dot-com bubble, the 2007–2016 period includes the Global Financial Crisis, the Great Recession, and the European Debt Crisis, and 2017–2024 period includes the COVID-19 pandemic, the Russo-Ukrainian war, and post-COVID-19 inflation. Table 6 presents the ordered probit estimates for models (1) and (5) across the three periods.

Table 6: Ordered Probit Estimates for the Baseline Model and its Extension Over Time

	(1) 1998-2006	(5) 1998-2006	(1) 2007-2016	(5) 2007-2016	(1) 2017-2024	(5) 2017-2024
Δr_t	0.530^{**}	0.193	0.0301	-0.525	0.765	0.751
	(0.239)	(0.259)	(0.409)	(0.520)	(0.485)	(0.493)
$CNBSI_t$	0.375^{**}	0.410^{***}	0.538^{**}	0.497^{**}	0.646^{**}	0.699^{**}
	(0.146)	(0.152)	(0.241)	(0.229)	(0.286)	(0.297)
$\Delta(\pi_t - \pi^*)$		0.456^{*}		0.551		0.0933
		(0.244)		(0.387)		(0.137)
Δy_t		0.0752		0.573^{**}		-0.00629
		(0.269)		(0.231)		(0.0288)
CZK_t		-0.118^{***}		0.697^{***}		0.229^{*}
		(0.0458)		(0.190)		(0.130)
$0\|1$	-0.311	-4.568^{***}	-1.377^{***}	15.89^{***}	-0.199	5.608^{*}
	(0.236)	(1.618)	(0.527)	(4.850)	(0.413)	(3.378)
$1\|2$	2.145^{***}	-1.930	1.709^{***}	20.26^{***}	2.051^{***}	7.926^{**}
	(0.305)	(1.524)	(0.471)	(5.259)	(0.537)	(3.384)
N	116	116	84	84	64	64
$pseudo - R^2$	0.116	0.181	0.129	0.344	0.264	0.281

Standard errors in parentheses. $0|1$ denotes the cutoff point from *rate cut* to *no change* and $1|2$ from *no change* to *rate hike*. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In contrast to the estimates in Table 3, the coefficients for the current policy rate changes are statistically significant only for (1) in 1998–2006, and remain insignificant across all other models and periods. The coefficients for the CNBSI are statistically significant at the 5% level for all models and periods, and at the 1% level for (5) in 1998–2006. The coefficients for the inflation gap change are statistically significant only in 1998–2006 at the 10% level, while the koruna-to-euro exchange rate is statistically significant at the 1% level in 1998-2006 and 2007-2016, and at the 10% level in 2017-2024. Notably, the coefficient for the change in the percentage deviations of real GDP from the trend is statistically significant at the 5% level in 2007–2016, in contrast to its estimates in Table 3. In 1998-2006, the upper cutoff point for (1) and the lower cutoff point for (5) are statistically significant at the 1% level. In 2007–2016, the cutoff points are statistically significant at the 1% level. In 2017–2024, both cutoff points are statistically significant only for (5).

As in Section 5.1, I compute the conditional marginal effects at the mean value of each explanatory variable to assess how a one-unit or one-standard-deviation increase affects the probability of the upcoming policy rate change falling into each ordinal category. For the 1998-2006 period, a one-standard-deviation increase in the hawkishness of a minute increases the probability of an upcoming rate hike by 2.7 percentage points. In 2007–2016, all estimated marginal effects except for the koruna-to-euro exchange rate are insignificant. Moreover, the koruna-to-euro exchange rate is only statistically significant for the ordinal category *rate cut*. A one-koruna increase in the exchange rate reduces the probability of the upcoming policy rate change being a rate cut by 5.4 percentage points. This differs from the findings in 5.1, where an appreciation of the current koruna-to-euro exchange rate increased the probability of the upcoming policy rate change being a rate cut. The marginal effects for 2017–2024 are similarly insignificant as those in 2007–2016. Only the marginal effects for the *rate hike* category of the CNBSI are statistically significant. A one-standard-deviation increase in the CNBSI raises the probability of the upcoming policy rate change being rate hike by 19.2 percentage points.

There may be several reasons the marginal effects are mostly insignificant, despite the statistically significant ordered probit estimates in Table 6. Possible reasons for this could be that the sample size for (5) in 2007-2016 is approximately one-third of the sample size in Table 3, and one-quarter in 2017-2024. Because I compute the conditional marginal effects at the mean, it may be that the marginal effects are significant at values other than the mean.

My findings in Section 5.3 show that the predictive power of the Czech National Bank Sentiment Index (CNBSI) varies over time. While the CNBSI, alongside other variables, predicts upcoming policy rate changes across all three periods, its marginal effects lack statistical significance, and limit precise estimates of its predictiveness to the 1998-2006 and 2017–2024 period. During these periods, a one-standard-deviation increase in the hawkishness of minutes raises the probability of the upcoming policy rate change being a rate hike by 2.7 and 19.2 percentage points, respectively. Compared to the findings in Section 5.1, the CNBSI is a significantly stronger predictor of policy rate increases in 2017–2024 than in the full time span of 1998–2024. These findings suggest that CNB sentiment provides valuable information for predicting upcoming policy rate changes over time, with its predictive power being particularly strong in the 2017–2024 period.

Conclusion

This thesis investigates how the sentiments expressed in central bank meeting minutes predict upcoming policy rate changes in the Czech Republic from 1998 to 2024. I create a Czech National Bank Sentiment Index to quantify the net hawkishness of CNB Bank Board meeting minutes by conducting sentiment analysis using fine-tuned RoBERTa and BERT language models. I then estimate an ordered probit model to assess how the CNBSI influences the probability of an upcoming policy rate change being a rate cut, no change, or a rate hike, while controlling for macroeconomic indicators and the current policy rate change.

My results show that CNB sentiment, captured by the CNBSI, significantly predicts upcoming policy rate changes. For the entire period from 1998 to 2024, a one-standard-deviation increase in the hawkishness of minutes decreases the probability of an upcoming rate cut by 9.6 percentage points, increases the probability of no change by 4.4 percentage points, and increases the probability of an upcoming rate hike by 5.1 percentage points.

A robustness check with two alternative sentiment measures, a BERT sentiment index and RoBERTa sentiment dummies, confirm my results, and show slightly larger effects. A one-standard-deviation increase in the hawkishness of minutes, captured by the BERT sentiment index, increases the probability of a rate hike by 5.4 percentage points, while a generally hawkish meeting minute increases the probability of a rate hike by 8.5 percentage points. Furthermore, my results suggest that the predictive power of CNB minutes is mainly driven by hawkish sentiments, due to statistically insignificant dovish sentiment dummies. I also find no clear evidence that fine-tuning a more advanced language model than BERT significantly increases the predictive power of sentiment indices.

Analyzing the heterogeneity over time further supports my results, as the CNBSI predicted upcoming policy rate changes for all three periods: 1998-2006, 2007-2016, and 2017-2024. However, I only observe statistically significant marginal effects at means for 1998-2006 and 2017-2024. In 1998-2006, a one-standard-deviation increase in the hawkishness of minutes increases the probability of an upcoming rate hike by 2.7 percentage points. In contrast, in 2017-2024, a one-standard-deviation increase in the hawkishness of minutes increases the probability of an upcoming rate hike by 19.2 percentage points, suggesting that the predictive power of CNB sentiment has become significantly stronger in recent periods.

My thesis has several important implications. First, it shows that the sentiments in CNB minutes do predict upcoming policy rate changes in the Czech Republic. As a result, CNB minutes allow market participants to adjust their expectations for the future path of policy rates based on the sentiments expressed within them. Second, the thesis demonstrates that central bank communications contain valuable information that can be extracted using machine learning, making unconventional data sources accessible and highlighting their potential in economic research.

Future research could extend my analysis by assessing the predictive power of CNB sentiment not only for an upcoming policy rate change, but also for changes beyond that. Analyzing multiple periods ahead would show how far current CNB sentiments predict the future path of policy rates. While my thesis uses aggregated sentence-level sentiments to create a sentiment score for each minute, future research may consider the relationships between sentences within a document. This has the potential to significantly improve the accuracy of sentiment analysis to better capture the underlying sentiments of CNB minutes.

List of References

Apel, M., & Grimaldi, M. (2014). How Informative Are Central Bank Minutes? *Review of Economics*, 53-76.

Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1908.10063

Blinder, A. S., Ehrmann, M., Fratzscher, M., De Haan, J., & Jansen, D.-J. (2008, May 1). Central Bank Communication and Monetary Policy. *ECB Working Paper No. 898*.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications.* Cambridge University Press.

Czech National Bank. (2024a). *The mandate of the Czech National Bank*. Retrieved December 18, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/about_cnb/mission/

Czech National Bank. (2024b). *Monetary policy*. Retrieved December 19, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/monetary-policy/

Czech National Bank. (2024c). *Bank Board decisions*. Retrieved December 20, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/monetary-policy/bank-board-decisions/

Czech National Bank. (2024d). *Forecast*. Retrieved December 20, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/monetary-policy/forecast/

Czech National Bank. (2024e). *Interviews, articles*. Retrieved December 20, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/public/media-service/interviews-articles/

Czech National Bank. (2024f). *Monetary policy instruments*. Retrieved December 21, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/monetary-policy/instruments/

Czech National Bank. (2024g). *GDP Expenditures: Quarterly, Constant Prices, B1GM Gross domestic product, Nominal values in CZK mill., Seasonally adjusted*. Retrieved December 12, 2024, from Arad Data Series System Web site: https://www.cnb.cz/arad/#/en/search/b1gm

Czech National Bank. (2024h). *Selected exchange rates*. Retrieved December 21, 2024, from Czech National Bank Web site: https://www.cnb.cz/en/financial-markets/foreign-exchange-market/central-bank-exchange-rate-fixing/central-bank-exchange-rate-fixing/selected_form.html

Czech Statistical Office. (2024). *Consumer Price Indices - Detailed Information*. Retrieved December 23, 2024, from Czech Statistical Office Web site: https://csu.gov.cz/produkty-archiv/1-b50028db67

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1810.04805

European Central Bank. (2024a). *Monetary policy accounts*. Retrieved December 19, 2024, from European Central Bank Web site: https://www.ecb.europa.eu/press/accounts/html/index.en.html

European Central Bank. (2024b). *Monetary policy statements*. Retrieved December 19, 2024, from European Central Bank Web site: https://www.ecb.europa.eu/press/press_conference/monetary-policy-statement/html/index.en.html

European Central Bank. (2024c). *Macroeconomic projections*. Retrieved December 19, 2024, from European Central Bank Web site: https://www.ecb.europa.eu/press/projections/html/index.en.html

European Central Bank. (2024d). *HICP - Overall index, Czech Republic, Monthly*. Retrieved December 10, 2024, from European Central Bank Web site: https://data.ecb.europa.eu/data/datasets/ICP/ICP.M.CZ.N.000000.4.ANR

Federal Reserve. (2021, July 29). *Monetary Policy Principles and Practice*. Retrieved December 1, 2024, from Federal Reserve Board Web site: https://www.federalreserve.gov/monetarypolicy/monetary-policy-what-are-its-goals-how-does-it-work.htm

Federal Reserve. (2022, January 28). *Federal Open Market Committee: Transcripts and other historical materials*. Retrieved December 20, 2024, from Federal Reserve Board Web site: https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm

Federal Reserve. (2024, December 18). *Federal Open Market Committee: Meeting calendars, statements, and minutes (2019-2026)*. Retrieved December 19, 2024, from Federal Reserve Board Web site: https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm

Gokaslan , A., & Cohen, V. (2019). OpenWebText Corpus. http://Skylion007.github.io/OpenWebTextCorpus.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* MIT Press.

Gössi, S., Chen, Z., Kim, W., Bermeitinger, B., & Handschuh, S. (2023). FinBERT-FOMC: Fine-Tuned FinBERT Model with Sentiment Focus Method for Enhancing Sentiment Analysis of FOMC Minutes. *ICAIF '23: Proceedings of the Fourth ACM International Conference on AI in Finance* (pp. 357 - 364). New York, NY, USA: Association for Computing Machinery.

Hendrycks, D., & Gimpel, K. (2016). Gaussian Error Linear Units (GELUs). *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1606.08415

Hilscher, J., Nabors, K., & Raviv, A. (2024, September 13). Information in Central Bank Sentiment: An Analysis of Fed and ECB Communication. doi:http://dx.doi.org/10.2139/ssrn.4797935

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 1-16. doi:https://doi.org/10.2307/2953682

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. doi:10.5281/zenodo.1212303

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1412.6980

Lai, G., Xie, Q., Liu, H., Yang, Y., & Hovy, E. (2017). RACE: Large-scale ReAding Comprehension Dataset From Examinations. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 785–794). Copenhagen, Denmark: Association for Computational Linguistics.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., . . . Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1907.11692

Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance, 66*(1), 35-66.

Malo, P., Sinha, A., Korhonen, P. J., Wallenius, J., & Takala, P. (2014). Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts. *Journal of the American Society for Information Science and Technology, 65*(4), 782-796. doi: https://doi.org/10.1002/asi.23062

Nagel, S. (2016). CC-news. http://web.archive.org/save/http://commoncrawl.org/2016/10/news-dataset-available.

Nitoi, M., Pochea, M.-M., & Radu, Ş.-C. (2023). Unveiling the sentiment behind central bank narratives: A novel deep. *Journal of Behavioral and Experimental Finance*, 38.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . Chintala, S. (2019). Pytorch: An Imperative Style, High-Performance Deep Learning Library. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, (pp. 8026 - 8037). Vancouver, Canada.

Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. *Journal of International Money and Finance, 79*(C), 136-156.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. Technical Report, OpenAI.

Rajpurkar, P., Jia, R., & Liang, P. (2018). Know What You Don't Know: Unanswerable Questions for SQuAD. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 784–789). Melbourne, Australia: Association for Computational Linguistics.

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 2383–2392). Austin, Texas: Association for Computational Linguistics.

Sennrich, R., Haddow, B., & Birch, A. (2015). Neural Machine Translation of Rare Words with Subword Units. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1508.07909

Tadle, R. C. (2022). FOMC minutes sentiments and their impact on financial markets. *Journal of Economics and Business, 118*(C).

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy, 39*, 195-214.

Trinh, T. H., & Le, Q. V. (2018). A Simple Method for Commonsense Reasoning. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1806.02847

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention Is All You Need. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1706.03762

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1804.07461

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., . . . Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Online: Association for Computational Linguistics.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., . . . Wu, Y. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv preprint*. doi:https://doi.org/10.48550/arXiv.1609.08144

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into Deep Learning* (1st ed.). Cambridge University Press.

Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books. *2015 IEEE International Conference on Computer Vision (ICCV)*, (pp. 19-27). Santiago, Chile.

Attachments

The STATA do file, and the dataset used for the empirical section of this thesis are available in the GitHub repository (https://github.com/bebrunn/Predicting-Policy-Rate-Changes-from-Central-Bank-Minutes) under the filenames: *master_brunner.do* and *master_brunner_data.dta*.

Additionally, I have provided the estimated marginal effects for Section 5 as an attachment to the thesis. For Section 5.1, the marginal effects file is named *marginal_effects_probit.txt*. The file for Section 5.2.1 is called *marginal_effects_bert.txt*, while the file for Section 5.2.2 is named *marginal_effects_dummy.txt*. For Section 5.3, the marginal effects files are organized by period: *marginal_effects_1998_2006.txt* for 1998-2006, *marginal_effects_2007_2016.txt* for 2007-2016, and *marginal_effects_2017_2024.txt* for 2017-2024.