SVERIGES RIKSBANK
WORKING PAPER SERIES

417

Central Bank Mandates and Monetary Policy Stances: through the Lens of Federal Reserve Speeches*

Christoph Bertsch, Isaiah Hull, Robin L. Lumsdaine, and Xin Zhang

October 2022 (Updated September 2024)

Central Bank Mandates and Monetary Policy Stances: through the Lens of Federal Reserve Speeches*

Christoph Bertsch[a], Isaiah Hull[b], Robin L. Lumsdaine[†c], and Xin Zhang[a]

[a]Sveriges Riksbank, Brunkebergstorg 11, SE-103 37 Stockholm, Sweden
[b]BI Norwegian Business School, NO-0442 Oslo, Norway; CogniFrame, 255 Duncan Mill Rd, North York ON M3B 3H9, Canada
[c]Kogod School of Business, American University, 4400 Massachusetts Avenue, NW Washington, DC 20016, United States; Erasmus University Rotterdam; National Bureau of Economic Research (NBER); Tinbergen Institute; Center for Financial Stability

Sveriges Riksbank Working Paper Series
No. 417
September 2024

Abstract

The Federal Reserve System has an institutional mandate to pursue price stability and maximum sustainable employment; however, it remains unclear whether it can also pursue secondary objectives. The academic literature has largely argued that it should not. We characterize the Fed's interpretation of its mandate using state-of-the-art methods from natural language processing, including a collection of large language models (LLMs) that we modify for enhanced performance on central bank texts. We apply these methods and models to a comprehensive corpus of Fed speeches delivered between 1960 and 2022. We find that the Fed perceives financial stability to be the most important policy concern that is not directly enumerated in its mandate, especially in times when the debt-to-GDP ratio is high, but does not generally treat it as a separate policy objective. In its policy discourse, it has frequently discussed the use of monetary policy to achieve financial stability, which we demonstrate generates movements in asset prices, even after rigorously controlling for macroeconomic and financial variables.

Keywords: Large Language Models, Machine Learning, Central Bank Communication, Financial Stability.

JEL classification: C55, E42, E5, E61, G28.

---

*We are grateful for comments from the Editor, Associate Editor, two anonymous referees, Francesco Bianchi, James Chapman, Ricardo Correa, Michael Ehrmann, Juri Marcucci, Tho Pham, seminar participants at American University, George Washington University, Linköping University, Sveriges Riksbank, Universidad Carlos III, AI Sweden: NLP Seminar Series, the Applied Machine Learning, Economics and Data Science (AMLEDS) seminar, the Transdisciplinary Econometric & Data Science (TEDS) seminar, and conference participants at the 2022 Society for Financial Econometrics (SoFiE) Annual Meeting, the 2022 CARMA Conference, the 2022 Open Data Science Conference, the 2022 Data Science Global Summit, the 2023 Royal Economic Society Annual Meeting, the 2023 International Association for Applied Econometrics (IAAE) Annual Meeting, the 2024 Africa Meeting of the Econometric Society, and the 5th Conference on "Nontraditional Data, Machine Learning, and Natural Language Processing in Macroeconomics". Nicole Brynjolffson provided excellent assistance in collecting and organizing some of the speeches used in this paper as part of the Center for Financial Stability's Summer Internship Program. This work was conducted while Lumsdaine was on partial secondment to the Office of Financial Research, US Department of the Treasury, under an Intergovernmental Personnel Agreement (IPA). We thank Edvin Ahlander and Joakim Jigling for excellent research assistance. The opinions expressed in this article are the sole responsibility of the authors and should not be interpreted as reflecting the views of Sveriges Riksbank, the US Department of the Treasury, or any of the other institutions they represent. The authors have no relevant conflicts of interest to declare. All remaining errors are our own.

†Corresponding author. Email: Robin.Lumsdaine@american.edu. Tel: (202) 885-1964.

1 Introduction

A central bank's mandate is a legal directive from the government that specifies its responsibilities as an institution. Nearly all central banks have a mandate that contains an explicit reference to price stability. The next most common objective is financial stability, followed by various requirements to target economic growth, employment, or welfare (BIS, 2009). The standard interpretation of the Federal Reserve's "dual mandate," codified via a change to the Federal Reserve Act in 1977, is that it requires the pursuit of both price stability and maximum (sustainable) employment.[1]

While there is broad agreement about which objectives are specified in the Fed's mandate, disagreement remains over how it should be executed and what, if anything, it prohibits. In this paper, we focus on an important source of disagreement: namely, can and should the Fed pursue objectives that are not explicitly enumerated in its dual mandate? Setting the legal question aside, the academic literature has largely concluded that the Fed should not, at least not with respect to financial stability (Vollmer, 2022; Schularick et al., 2021).

In contrast to academics, Federal Reserve officials appear less convinced that pursuing secondary objectives, such as financial stability, conflicts with the dual mandate. Eric Rosengren, former President of the Federal Reserve Bank of Boston, for example, has indicated on several occasions that financial stability does and should influence monetary policy.[2] Other members of the FOMC have argued for the pursuit of financial stability more explicitly during their tenure. Lael Brainard, in a 2014 speech, claimed that monetary policy should constitute a "second line of defense" against financial instability and suggested that the Fed would be faster to use monetary policy to pursue financial stability than some other central banks because "... [the Fed's] regulatory perimeter is narrower, the capital markets are more important, and the macroprudential toolkit is not as extensive."[3]

Although Brainard argues that dual mandate and financial stability concerns typically coincide, she also concedes that the two may conflict under certain circumstances. A clear example of this was the failures of Silicon Valley Bank and Signature Bank, during a period of high inflation in the Spring of 2023. In response to these adverse developments in the financial sector, Austan Goolsbee, President of the Federal Reserve Bank of Chicago, acknowledged that "[in] moments like this, of financial stress, the right monetary approach calls for prudence and patience–for assessing the potential impact of financial stress on the real economy."[4]

While anecdotal evidence suggests that Federal Reserve officials may pursue goals beyond the dual mandate, it is unclear how widespread this practice is and whether it affects monetary policy or influences financial markets. We contribute to this debate by applying large language models (LLMs) and state-of-the-art textual analysis methods to analyze a newly collected textual database (corpus) of Federal Reserve speeches, which is the most comprehensive one available to our knowledge. We

[1] See, for example, https://www.chicagofed.org/research/dual-mandate/dual-mandate.

[2] In a 2016 speech during his tenure on the FOMC, Rosengren claimed that: "...financial stability concerns could be a consideration in how long policymakers wait before resuming the gradual removal of monetary accommodation." The speech was given at the Shanghai Advanced Institute of Finance, Beijing, China, August 31, 2016.

[3] Lael Brainard in a speech to the Hutchins Center on Fiscal and Monetary Policy, The Brookings Institution, Washington, D.C., December 3, 2014.

[4] Austan D. Goolsbee in a speech to the Economic Club of Chicago Forum Luncheon, Chicago, IL, April 11, 2023.

characterize the Fed's interpretation of its mandate over much of its institutional history, and examine the implications for monetary policy and the impact on financial markets.

The application of LLMs allows us to identify subtle features of central bank discussions, such as the Fed's expression of interest in policy objectives beyond the dual mandate. Focusing on speeches delivered between 1960 and 2022, the long time series dimension of the corpus allows us to measure changes in the Fed's discussion of its mandate over time and in response to events, such as financial crises and recessions.[5] And the variation across districts and across Federal Reserve officials enables us to examine heterogeneity and control for speaker and district-specific idiosyncracies. This contrasts, for example, with dictionary-based methods, which count the frequency with which certain words appear in a document and are often used in economics to measure latent features of text.[6] However, we are interested in measuring concepts relevant to policymakers that are embodied in sequences of text, such as sentences or paragraphs, rather than word frequencies. Each distinct concept would require its own dictionary, which would rely on us having a strong prior about how to measure such concepts through the frequency of word use and without examining words in the context of their usage. The approach we take makes use of the flexibility and precision of large language models, which can achieve state-of-the-art performance on a variety of language tasks.

There are several important differences between the LLMs used in this paper and generative pre-trained transformer models that have garnered substantial attention of late.[7] First, the LLMs we employ have around 65 million parameters, rather than tens or hundreds of billions of parameters. This allows us to perform natural language tasks on a large central bank corpus without incurring substantial time, computational, or monetary costs. Second, the LLMs used in this paper are discriminative, rather than generative, which means that they are trained to perform specific tasks, such as text classification, rather than to generate new text data (e.g., act as a chatbot, follow instructions, etc.). Because we are interested in performing discriminative tasks, the approach we take increases transparency, consistency, and computational efficiency in the production of text features. And third, the LLMs we use are open source, allowing us to determine how the models were pre-trained and fine-tuned, which is not typically possible when using proprietary models. This also makes it possible to extend the pre-training and perform fine-tuning without restriction.

Employing these LLMs, we first attempt to distinguish between content in the Federal Reserve speech corpus that discusses the dual mandate and content that discusses other topics. For those paragraphs that are classified as being unrelated to the dual mandate, we apply an LLM to identify the excerpt in each paragraph that corresponds to the speaker's most significant concern.[8] Parsing

[5]The corpus extends back to the founding of the Federal Reserve System in 1913; however, we focus on speeches from 1960 onward, as speech frequency is much higher after this date.

[6]Dictionary-based methods are best employed when the researcher has a strong prior about the words used to describe a topic and when the text data are not inherently informative (Gentzkow et al., 2019). Good examples of dictionary-based methods in economics include the measurement of Economic Policy Uncertainty (Baker et al., 2016) and sentiment in central bank texts (Loughran and McDonald, 2011; Apel and Blix Grimaldi, 2014; Correa et al., 2021; Apel et al., 2022).

[7]For instance, the GPT-3.5 and GPT-4 models that underpin OpenAI's ChatGPT.

[8]Note that we do not use the term "dual mandate" to identify references to the Fed's mandate, since it was not codified into law until the late 1970s. The term is described in The Federal Reserve Reform Act of 1977 and the Full Employment and Balanced Growth Act of 1978. Additionally, the term itself was not commonly used by the Fed until

the output, we find that the most frequently discussed non-dual mandate concerns relate to financial stability and the financial sector. While the Fed also discusses other topics, these do not appear to account for a substantial share of the speech content.

Sorting paragraphs into "dual mandate" and "non-dual mandate" groups provides us with a coarse division of the speech corpus. We also construct a more granular partition that places each paragraph into a class based on the economic or financial topic it discusses. Because we are interested in the Fed's interpretation of its mandate, including the admissibility of secondary objectives, we focus on paragraphs that have high financial stability classification scores. We then employ a sentence transformer model to determine whether the speaker advocates for the use of monetary policy to achieve financial stability. We repeat this exercise to determine whether the speaker advocates for banking regulation to achieve financial stability. Around the time of the Great Recession, we observe a growing endorsement for the use of both methods to achieve financial stability; however, the support for the use of monetary policy begins to rise earlier – even prior to the Great Recession – and remains elevated until the end of our sample period. In contrast, support for the use of banking regulation to achieve financial stability appears to be much more episodic and transitory.

We next attempt to determine what is associated with the Fed's support for the use of bank regulation or monetary policy to achieve financial stability, making use of the two aforementioned text features: advocacy for monetary policy as a means to achieve financial stability and advocacy for banking regulation towards the same objective. In both cases, we find statistically significant associations between the discussion of the components of the dual mandate (inflation and employment) and advocacy for the use of either monetary policy or banking regulation to achieve financial stability.

In related exercises, we examine the explanatory power of the variables used in the regression exercises by calculating Shapley values. We find that high debt levels and concerns about bank capital appear to be associated with advocacy for the use of banking regulation (but not monetary policy) to achieve financial stability; whereas dual mandate-related discussion and concerns about financial crises tend to be accompanied by calls for both the use of monetary policy and banking regulation. These results suggest that the Fed does not always see monetary policy as the correct instrument for achieving financial stability. Rather, it holds a more nuanced view that is closer to what Lael Brainard articulated in her 2014 speech.

We also show that the appearance of such content in Fed speeches is associated with policy decisions and significant movements in asset prices across the board.[9] In particular, a one standard deviation increase in the discussion of financial crises is associated with a 9-16 basis point reduction in the rate of return on bonds, risky and safe assets, even after the inclusion of macroeconomic and financial controls. Additionally, discussion of financial stability is associated with a reduction in asset returns, suggesting that it operates through expectations about monetary conditions.

Finally, we extend our main findings by examining the impact of communication during the long Zero Lower Bound (ZLB) episode that followed the Great Recession, which was speculated to have

the mid-1990s.

[9]This contributes to a growing literature on the subject (Clarida et al., 2000; Bianchi, 2012; Bianchi et al., 2022) and recent work on the risk-taking channel of monetary policy (Jiménez et al., 2014; Dell'Ariccia et al., 2017).

amplified the importance of communication in policymakers' toolkits (see Hansen and McMahon, 2016). Making use of the text features described earlier, we estimate an augmented Taylor rule (Taylor, 1993, 1999) to determine whether they have explanatory power during this period.

2 Data

We use three different types of data. The first is a collection of speeches given by Federal Reserve Bank presidents and members of the Board of Governors, which we use to measure features of Federal Reserve officials' communication, including discussion of financial stability and its institutional mandate. The second is a corpus of journal articles and working papers related to central banking, which we use to refine the natural language processing (NLP) models, allowing us to extract higher quality features from the Fed speeches. The third is a set of macroeconomic and financial variables, which are used as either controls or the dependent variable in different regression exercises. For the sake of consistency and to ensure coverage over the long sample period, we take most of these control variables from the Macrohistory Database, introduced by Jordà et al. (2016).

2.1 Federal Reserve Speeches

Our primary source of text data is a novel collection of speeches given by presidents of Federal Reserve Banks and members of the Board of Governors of the Federal Reserve System. It includes 7,351 speeches and 152 speakers and spans the period between 1914 and 2022.[10] While the coverage of speeches given is not complete over the sample period, it is, as far as we are aware, the most comprehensive collection assembled.

Speech frequency increases over the sample period, but declines near the end of the sample period. This is a consequence of several factors. First, the Fed increased the frequency of its public communication over the sample. And second, works that were produced more recently are more likely to have been digitized and incorporated into an online collection.[11] There are fewer than 20 speeches in the corpus for most years between 1914 and the early 1960s. However, after 1960, the number of speeches given and available annually tends to rise over time. The number also jumps again in the late 1990s.[12] We also see a decline near the end of the sample – and especially in the final two years – primarily due to incompleteness in coverage. We will focus primarily on the period between 1960 and 2022. This period includes 6,639 speeches delivered by 112 speakers.

[10]We build on the sample by van Dieijen and Lumsdaine (2019), containing speeches of members of the Board of Governors from 1997-2016, which we extend over a longer time period and augment with speeches of the Federal Reserve Bank presidents.

[11]See Figure A1 of the Appendix for a plot of the speech counts by year and Table A1 for speech counts by institution.

[12]In the late 1990s under Chairman Greenspan, there was a gradual shift to more transparency, which is not only reflected in a large increase in the frequency of speeches by presidents of Federal Reserve Banks and members of the Board of Governors, but also in other communication, such as the introduction in 1994 of FOMC statements in conjunction with all rate changes and later on in 2000 for all FOMC meetings.

2.2 Journal Articles and Working Papers

The second text corpus consists of abstracts and metadata from journal articles and working papers that were drawn from the Semantic Scholar Open Corpus (S2ORC), introduced in Lo et al. (2020). As of July 2020, this dataset contained metadata for 136M papers, including titles, abstracts, and citations. Within this corpus, we focus on the 2.3M journal articles and working papers that were identified by S2ORC as being from the field of economics.

We next identify the subset of articles that discuss topics related to macroeconomics, monetary economics, and financial markets. We then filter those articles using two additional criteria. First, the article must have an abstract available, because abstracts will be used in the training process. And second, it must be published in a journal or working paper series that has at least 500 entries in the S2ORC database. The second criterion is intended to filter out articles based on journal quality. The final sample comprises 328,370 articles.

From this final sample, we construct two datasets. The first consists of the full article abstract texts. This is used to extend the pre-training of the LLMs we use in this paper, which are already pre-trained on large text corpora, including the full English language Wikipedia.[13] The purpose of extending the pre-training is to improve the model's comprehension of text related to macroeconomics, monetary economics, and financial markets.

To construct the second dataset, we start by selecting abstracts that reference central banks, which reduces the sample to 29,781 articles. We then divide each abstract into sentences and identify two types of sentence pairs: 1) sentences in the same abstract; and 2) sentences in different abstracts.[14] We randomly select an equal number of both types of pairs, yielding a total sample size of 194,227. We use the aforementioned text data and the approach introduced in Reimers and Gurevych (2019) to fine-tune an LLM that is based on the BERT model (Vaswani et al., 2017). The purpose of the extended fine-tuning is to improve the model's performance in comparing the similarity of two text sequences in a central banking context.

Table A2 in the Appendix provides article counts for the journals and working paper series that appear most frequently in the sentence pair corpus. The most common working paper repository is the Social Science Research Network (SSRN), which accounts for almost 14% of the 29,781 articles. The most common journal is the *Journal of Banking and Finance*. In total, there are 283 journals and working paper series included in the corpus.

2.3 Macroeconomic and Financial Data

In addition to the text data, we also use macroeconomic and financial data in regression exercises. These include bond returns, CPI inflation, financial crisis classifications, the debt-to-GDP ratio, return on equity, house prices, the loan-to-deposit ratio, returns to risky assets, returns to safe assets, the interest rate, loan origination volume, and the output gap. With the exception of the output gap, all

[13] See Section 3.2.1 for a description of the models and pre-training process.

[14] Using randomly selected sentences from the abstract increases the probability that the two sentences will be closely related. In contrast, using the text of the full paper risks the possibility of a weak connection, making it difficult to gain traction on the learning task.

variables are taken from the Macrohistory Database (Jordà et al., 2016). The output gap is measured as the percentage difference between actual and potential (real) GDP. Actual GDP is taken from the Bureau of Economic Analysis (BEA) and potential GDP is measured by the Congressional Budget Office (CBO). All variables are measured at an annual frequency and span the period between 1960 and 2020. The variables used are listed in Table A3 in the Appendix.

3 Methods

Our objective is to examine the Fed's interpretation of its own mandate. We do this by extracting text features from Federal Reserve speeches and examining their variation across time, district, and speaker. Many of the more informative features involve sequences of words, such as sentences and paragraphs, rather than individual words. As such, we use LLMs that are capable of modeling sequences, building on recently introduced variants of transformer models (Vaswani et al., 2017). This section provides a brief overview of the NLP tasks we perform, the LLMs we use, and the types of features we extract. In all cases, we select pre-trained and fine-tuned LLMs that were among the top open source performers for their respective linguistic task at the time of this paper's writing. We then attempt to improve their performance further in our domain of interest: central bank communication.[15] We also discuss the pre-training and fine-tuning process for these models, but relegate the technical details about the NLP models to Section A.3.2 of the Appendix.

3.1 The NLP Task

Many natural language processing tasks can only be performed using a sequence-to-sequence (S2S) model, which maps one sequence of words to another. Others do not require an S2S model, but achieve better performance when one is employed. Machine translation tasks, for instance, yield low quality results when performed at the word level. An entire sentence typically needs to be processed and interpreted before a suitable translation can be generated. This is, in part, because words have different meanings in different contexts and the grammatical ordering of words (e.g., subject-verb-object) differs across languages. See Section A.3.1 in the Appendix for an overview of S2S modeling.

Another example of an inherently sequential NLP task is extractive question answering. This involves finding a subsequence of text that contains an answer to a question. An NLP model would receive a "context" sequence (the original text) and a query as inputs. It would then yield a subsequence of the context as an output. Both the inputs and output are sequences. State-of-the-art S2S models, such as the LLMs we use in this paper, are naturally suited to such problems.

For our purposes, S2S modeling will mostly involve the transformation of a sequence of words in a paragraph into a sequence of "contextualized" words. The words themselves will be represented using dense vectors called embeddings.[16] The contextualized embeddings will encode information about the meaning of the word in the context in which it was used. That is, vectors that are close together in

[15] Related work by Araci (2019) and Liu et al. (2020) trains BERT-type models for financial texts and fine-tunes them for sentiment classification.

[16] See Gentzkow et al. (2019) for an overview of embeddings.

the embedding space contain similar semantic content. For example, the word “run” will be encoded differently in the following two sequences:

Sequence 1: “If depositor concerns are not addressed, there could be a bank *run*.”
Sequence 2: “If stock prices increase again tomorrow, it will be the longest bull *run* in history.”

3.2 Large Language Models

We make use of a collection of LLMs called transformer models (Vaswani et al., 2017; Devlin et al., 2019; Liu et al., 2019) and modify them for improved performance on central bank texts.[17] These models process text sequences, such as sentences and paragraphs, and are pre-trained on general language tasks, such as masked language learning, which involves predicting omitted words in a sentence. Transformer models introduced several innovations that have proven useful for natural language processing (see Section A.3.2 of the Appendix for details). They also make use of extensive pre-training. The purpose of pre-training is to create a baseline or “foundation” model that understands a given language using large amounts of text. This pre-training protocol enables the automatic generation of extensive training sets from large text corpora, such as the entirety of Wikipedia, that are unrelated to the NLP task of interest. This model can then be fine-tuned on a small amount of text that is closely related to the language task of interest (e.g., instruction following, text summarization, or text classification), often with proficiency equal to, or even surpassing, human performance.[18]

Transformer models output sequences of contextualized embeddings, which can be used to perform a wide variety of textual analysis tasks, including zero shot classification (classification without training), textual similarity measurement, machine translation, contextual embedding generation, and extractive text summarization. Transformers also can be extended and trained to perform supervised learning tasks, such as sentiment classification. Furthermore, such extensions can be fine-tuned at a low computational cost to yield further improved performance in a specific domain, such as central banking communication.

We use transformer models to extract features from Fed speeches, focusing on elements related to the Fed’s mandate discussions. Using transformer models that were fine-tuned for different language tasks enables us to measure subtle features of a speech, such as the speaker’s primary concern in a given paragraph or the extent to which a statement expresses approval of monetary policy as a means of achieving financial stability. It also allows us to identify whether certain content – such as concern about bank liquidity – is present in a paragraph.

Because we use a variety of pre-trained and fine-tuned models in this paper – and also extend the pre-training and fine-tuning process – we discuss both topics in the remainder of this subsection.

[17] The details about the models used and the pre-training and fine-tuning processes are given in this section and Appendix A.3.3.

[18] Generative versions of LLMs, like the GPT-3.5 and GPT-4 models, can write text that is indistinguishable from human-produced work and are even capable of passing rigorous licensing exams in medicine and law (OpenAI, 2023).

3.2.1 Pre-training

The NLP literature has shown that pre-training a language model can yield substantial performance benefits on downstream tasks (Dai and Le, 2015; Peters et al., 2018; Radford et al., 2018; Howard and Ruder, 2018). Indeed, some of the most substantial gains in the development of language models have come from changes in the pre-training procedure and data, rather than changes in model architecture. The BERT and RoBERTa models described below achieved state-of-the-art performance on NLP benchmarks using the same transformer architecture introduced in Vaswani et al. (2017).

BERT. The success of supervised transfer learning – that is, training on a corpus in one domain and performing prediction in a different domain – was one of the motivations for the construction of the BERT model, introduced in Devlin et al. (2019).[19] It proposed several modifications to the training process used in Vaswani et al. (2017). In the context of BERT, the pre-training process concentrated on two tasks: 1) masked language modeling (MLM); and 2) next sentence prediction (NSP). Both tasks automatically generate labels, allowing for pre-training on arbitrarily large text datasets that have not been labeled by a human.

The MLM task involves masking randomly-selected words in a sequence and then training the model to predict them. Another benefit of pre-training with MLM is that it allows for bidirectionality in the interpretation of sequences. Rather than training BERT to predict the next word in a sequence, conditional on the preceding words, it is instead trained to predict a missing word in a sequence, conditional on all words before and after it. This results in a pre-trained model that has a substantially expanded capacity to understand language.

The other training task, next sentence prediction (NSP), also allows for the automatic generation of labels. This task presents the model with a sequence of two sentences drawn from the corpus. The model must determine whether the second sentence follows the first or whether it is drawn from a different place in the document.

Because neither the MLM nor NSP tasks require human-labeled data, Devlin et al. (2019) were able to train BERT on two very large text corpora: the 800M word Toronto Book Corpus from Zhu et al. (2015) and a 2500M-word corpus constructed from English language Wikipedia articles.[20] This increase in the size of the training corpus allowed for a corresponding increases in model size.[21]

RoBERTa. Liu et al. (2019) introduced the RoBERTa model, a "robustly optimized" version of BERT, arguing that the gains from BERT are primarily attributable to the modification of the pre-training process, which allows for an increase in model size by several orders of magnitude.[22]

[19] See Conneau et al. (2017) and McCann et al. (2017) for examples of transfer learning. For an early example of transfer learning in economics NLP applications, see Apel et al. (2022).

[20] Note that the training sets can be generated programmatically and without the need for a human to perform labelling.

[21] Making use of the extended training corpus, Devlin et al. (2019) pre-trained two versions of BERT: 1) $\texttt{BERT}_{\texttt{BASE}}$, which has 12 transformer blocks, a hidden dimension size of 768, 12 attention heads, and 110M parameters; and 2) $\texttt{BERT}_{\texttt{LARGE}}$, which has 24 transformer blocks, a hidden dimension size of 1024, 16 attention heads, and 340M parameters.

[22] They also argue that this is true of most landmark language models, such as ELMo (Peters et al., 2018), GPT (Radford et al., 2018), XLM (Conneau and Lample, 2019), and XLNet (Yang et al., 2019).

The RoBERTa model attempts to provide further improvements over the pre-training process in BERT. Specifically, it employs larger datasets, removes the NSP task, trains on larger sequence lengths, employs a dynamic masking pattern, and uses a new dataset (CC-News).

Liu et al. (2019) find that the resulting modifications to the pre-training process boost the "robustly optimized" version of BERT to match or outperform the models that were introduced after BERT. We use the RoBERTa model for many of the NLP-related tasks in this paper.

3.2.2 Fine-tuning

Once a deep learning model, such as a transformer model, has been pre-trained, we can fine-tune it to perform a task in our domain of interest. Such models are made up of a sequence of layers, which transform the inputs from the previous layer. Fine-tuning is typically performed by *freezing* all of the layers, except the output layer and a few of the layers that directly precede it. Note that freezing a layer prevents its parameters from being updated. This means that in $\texttt{BERT}_{\texttt{BASE}}$, for instance, fewer than 100K of the 340M parameters need to be trained. The rest of the model can then serve as a state-of-the-art text feature extractor.

Devlin et al. (2019) show that the pre-trained BERT model can be fine-tuned to achieve state-of-the-art performance on downstream tasks, such as question answering and language inference. This entails training a model to take the contextualized embedding output from BERT to use as an input to a supervised learning task. This can be done with a few hours of GPU training and does not require modifications to the model's architecture that are specific to the task. The authors used this approach to achieve state-of-the-art performance on 11 NLP benchmarks. See Devlin et al. (2019) for information about the benchmarks and model performance.

The fine-tuned versions of BERT introduced in Devlin et al. (2019) remain near the state-of-the-art on the General Language Understanding Evaluation (GLUE) (Wang et al., 2019) and Stanford Question Answering Dataset (SQuAD) (Rajpurkar et al., 2018) benchmarks, which are used to assess the performance of NLP models on specific language tasks. The models that have since surpassed them also use transformer model architectures. Fine-tuned BERT models have also demonstrated better performance on domain-specific NLP tasks than models that were trained exclusively using text from that domain, which is one of the reasons why we use them as base or "foundation" models.

In some cases, we extend the pre-training further. For example, when using a type of LLM called a "sentence transformer" (Devlin et al., 2019; Reimers and Gurevych, 2019), we use the Semantic Scholar Open Corpus (Lo et al., 2020) to extend the model's pre-training and to enhance its understanding of language used in the context of central banking.[23] Once pre-trained and fine-tuned, we then use our models to perform supervised learning on a domain-specific task and with domain-specific data, in our case the corpus of Federal Reserve texts.

[23] A sentence transformer maps a text sequence, such as a paragraph, to an embedding, which is a dense vector representation of the text. In our case, the model yields 768-dimensional vectors that provide a numerical representation of the underlying semantic content of the text. Vectors that are close together in the embedding space, as measured by cosine similarity, contain similar semantic content.

3.3 Text Feature Generation

In this subsection, we discuss three different types of text tasks that our models use for feature extraction.

3.3.1 Zero Shot Classification (ZSC)

For some exercises in this paper, we perform text classification. This task often involves training a supervised learning model with labelled text categories. However, in the absence of an appropriate dataset or labels, it is not possible to perform supervised classification. An alternative approach, introduced in Pushp and Srivastava (2017), trains a general model that can perform classification on arbitrarily chosen categories without first labelling the data.[24]

To perform zero shot classification, we make use of a model that was trained on the Multi-genre Natural Language Inference (MNLI) dataset (Williams et al., 2018).[25] The model performs zero shot classification by embedding candidate labels and sequences of interest in the same space. This can then be used as a basis for assessing the closeness of the embeddings and, thereby, producing a classification. This will yield the probability that the sequence and candidate label match, even without training the model to perform classification on that specific set of labels.

In Appendix A.4, we provide an example of ZSC with the aforementioned model. As we will see in later exercises, transformer models perform well on simple ZSC tasks for our corpus of Fed speeches.

3.3.2 Extractive Question Answering (EQA)

The original BERT model (Devlin et al., 2019) was trained to perform EQA on the Stanford Question Answering Dataset (Rajpurkar et al., 2016), which consists of 100,000 human-labelled questions, answers, and context passages. The context is a passage taken from a Wikpedia article. The model must correctly predict the subsegment of the text that contains the correct answer to the question. BERT exceeded human-level performance on both benchmarks for that task.

We use extractive question answering with pre-trained BERT models in several exercises in this paper.[26] For some exercises, we attempt to determine the speaker's most significant concern in a passage. Two specific examples demonstrating this purpose are given in Appendix A.4. Once the most significant concerns have been extracted from a speech with EQA, they can then be converted to a set of numerical features, such as contextualized word embeddings via BERT or a sentence embedding via sentence-BERT (SBERT), that can then be further analyzed via econometric methods.

[24] See Yogatama et al. (2017), Zhang et al. (2019), and Yin et al. (2019) for alternative approaches to ZSC.

[25] The model we use is `distilbert-base-uncased-mnli`. The training set consists of 433k sentence pairs that were annotated with information about textual entailment. More specifically, each text example is associated with a hypothesis about that text. The model classifies the text-hypothesis pair using one of the following three labels: entailment, neutral, or contradiction. The model was trained at a learning rate of 2e-5 for 5 epochs. See Williams et al. (2018) for more information.

[26] We use the `distilbert-base-uncased-distilled-squad` model, which was fine-tuned on `SQuAD` v1.1. It has 67 million parameters and was trained using the `BERT` model as a teacher. See Rajpurkar et al. (2016) for additional details of the `SQuAD` benchmark.

3.3.3 Semantic Textual Similarity (STS)

For several exercises, we will need to measure the similarity between pairs of passages from speeches using a measure called semantic textual similarity. The BERT and RoBERTa models, which we use throughout the paper, achieve state-of-the-art performance on the STS task; however, they require each pair of passages to be input into the model simultaneously to produce an STS score and hence are computationally intensive. For 10,000 sentences, this would require around 50M STS pair computations. To compute STS efficiently, we make use of SBERT models which dramatically reduces the computational cost of comparing a pair of passages. We also extend the pre-training process and perform fine-tuning to improve performance on our corpus. For a detailed description of this process and the use of cosine similarity to measure STS scores, see Section A.3.3 of the Appendix. Cosine similarity is often used in natural language processing to quantify the similarity between two embedding vectors and returns a number between -1 and $+1$, where two identical sentences score $+1$, two unrelated sentences score around 0, and two sentences with opposite meanings score -1.

3.4 LLM Training Summary

In Figure A.2 in the Appendix, we visually summarize the training process for the LLMs used in this paper. We start with the pre-trained BERT model (Devlin et al., 2019). In exercises that involve zero shot classification, we use a model that was fine-tuned to perform ZSC using the multi-genre language inference (MNLI) dataset (Williams et al., 2018). For exercises that involve extractive question answering, we use a model that was fine-tuned using the `SQuAD` v1.1 dataset (Rajpurkar et al., 2016). And in exercises that make use of cosine similarities produced by sentence transformers, we extend the pre-training process and fine-tune the model on the semantic textual similarity task using a custom dataset based on the S2ORC corpus (Lo et al., 2020). A list of the pre-trained base models used is given in Table A4 of the Appendix. Additionally, Appendix A.3.3 describes the training procedures for the sentence transformer models. See Table A5 in the Appendix for a performance comparison between our version of the model with extended pre-training and fine-tuning, the base model, and another state-of-the-art sentence transformer.

4 Interpretation of Text Features

The previous section discussed three methods for extracting text features using transformer models: zero shot classification, extractive question answering, and semantic textual similarity measurement. This section describes a selection of the features we extracted from Federal Reserve speeches using those methods. Our objective is to examine whether these features adequately represent the concept we intended to measure in the text. We describe how some of the more informative features evolved over time and across district.

Most of the features discussed in this section are constructed from paragraph-length sequences. For relatively simple features, we use zero shot classification with the RoBERTa model to determine whether a paragraph discusses a certain concept, such as financial stability. For more subtle concepts, we extend the pre-training and fine-tuning of the model to improve its capacity to understand central bank texts. We then measure cosine similarity between paragraphs of speech text and the statement we want to evaluate. Finally, we standardize the raw scores by subtracting the sample mean and dividing by the sample standard deviation.

Using S2S models – and, in particular, the RoBERTa model – has at least three advantages. First, in contrast to dictionary-based methods, RoBERTa's classifications are based on the entire paragraph, taking long-run dependencies, negations, and modifiers into account. Second, RoBERTa automatically identifies related terms and, thus, does not rely on the ex-ante identification of all relevant terms. And third, unlike average word embeddings, RoBERTa accounts for the context in which words are used.

For an example of passages identified using zero shot classification (ZSC) and semantic textual similarity (STS), see Table A8 in the Appendix. The table is constructed by drawing a random speech passage that is assigned the financial crisis label using ZSC. We next use the sentence transformer model to compare that statement to the other statements in the corpus using cosine similarity to construct STS scores. The table consists of the original randomly drawn statement, identified using ZSC, followed by 10 similar statements, identified using cosine similarity.

4.1 Time Variation in Text Features

Because we are interested in measuring changes in the Federal Reserve System's interpretation of its mandate over time, we will start by examining the time variation in the text features. This subsection discusses what features we measure, how they perform, and what insights they provide into the Federal Reserve's interpretation of its mandate.

4.1.1 Dual Mandate Content

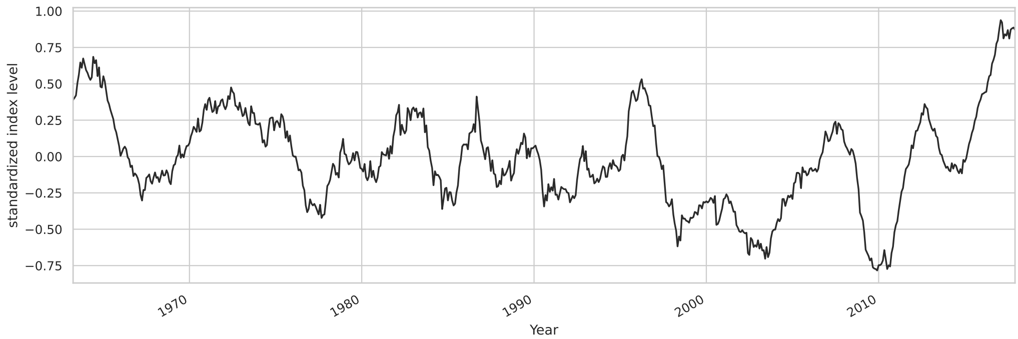
We first partition the text into content that is related to the Fed's "dual mandate" and content that is not. Because the sample starts prior to the introduction of the dual mandate, we measure whether paragraphs discuss topics related to inflation, employment, or output growth, rather than trying to identify references to the dual mandate itself. The evolution of the series is shown in Figure 1a. Table A9 of the Appendix gives examples of passages classified to be about "dual mandate" content.[27]

The figure indicates that Fed officials' discussion of dual mandate-related concerns preceded the codification of the mandate in the Federal Reserve Act as well as the use of the specific phrase. In addition, there is both short- and long-term variation in the dual mandate content of speeches. We examine the content of paragraphs that have low dual mandate content and find that they are largely related to the financial sector, financial crises, and banking regulation. To demonstrate this informally, we use extractive question answering with the RoBERTa model to identify the speaker's concern in each paragraph that has a low dual mandate content score. We then construct word clouds for the pre-Great Moderation period (1960-1983) and the Great Moderation period (1984-2022), shown in Figure 2. In both cases, the discussion is dominated by topics related to finance and banking regulation.

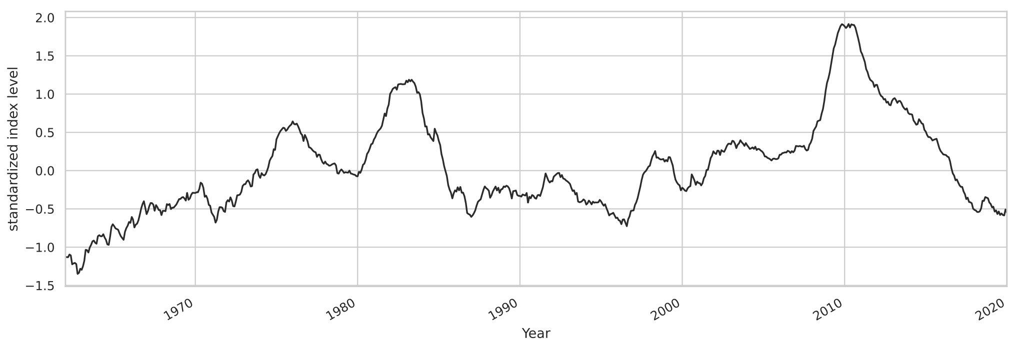
[27] The table provides examples of statements that had the highest cosine similarity scores with the statement "inflation, employment and output growth," which we use to identify dual mandate-related content.

Figure 1: Text Features

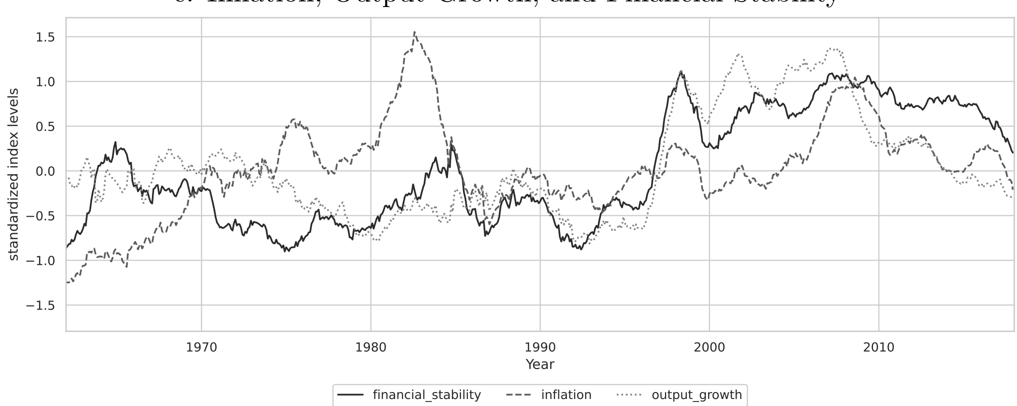
a. Dual Mandate Content



b. Financial Crisis



c. Inflation, Output Growth, and Financial Stability

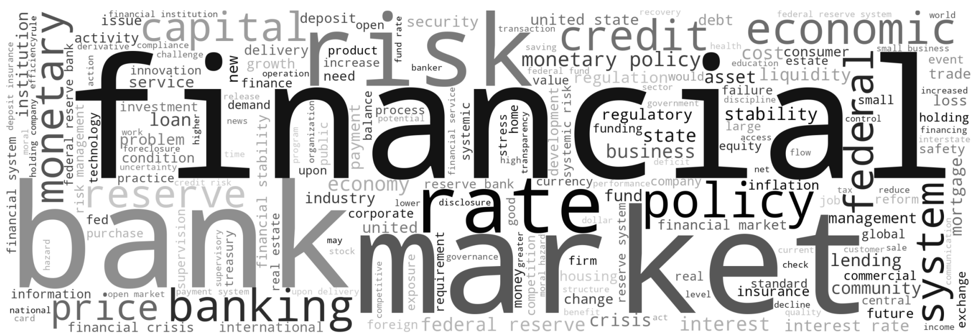


Notes: The figures show 2-year rolling means of text feature indices. Figure 1a shows the rolling mean of the standardized cosine similarity score from BERT for “inflation, employment, and output growth” and each paragraph in the text. Figure 1b shows the standardized classification score from BERT for the term “financial crisis.” This is computed by classifying each paragraph as describing “financial crisis” or not. We then compute the average score for each month, standardize it, and plot the rolling mean. Finally, Figure 1c shows the rolling mean of the standardized classification score for the terms “financial stability,” “output growth,” and “inflation.”

Figure 2: Non-Dual Mandate Content Word Cloud

a. 1960-1983

b.1984-2022



Notes: The figures above show word clouds of terms and groups of terms that were identified as the speaker's main concern in paragraphs with low dual mandate content scores. Such paragraphs are identified using extractive question answering.

4.1.2 Financial Stability Content

Figure 1b shows the 24-month rolling mean of the financial crisis feature. By comparing it with the analogously-constructed financial stability feature shown in Figure 1c, we are able to differentiate between discussions about the Fed's role in achieving financial stability and deliberations about specific financial crises.

As expected, the highest levels of the financial stability classification score coincide with the Great Recession. It is preceded by a spike around 1998, coinciding with the Asian and Russian financial crises and the collapse of hedge fund Long-Term Capital Management (LTCM), which resulted in a bailout of LTCM by a group of private banks that was orchestrated by the Federal Reserve Bank of New York. We can see that financial stability content is declining in the late 1960s and 1970s, followed by another spike around the Latin American Debt Crisis in 1982, when the nine major U.S. banks were heavily exposed to Latin American debt, amounting to 176.5 percent of their capital (Sachs, 1987). Thereafter, the financial stability index declined during the first half of the Great Moderation.

In contrast to the financial stability discussion shown in Figure 1c, there is a general upward trend in financial crisis content prior to the start of the Great Moderation in the mid-1980s. There is also a lull between 2000 and 2007. Interestingly, the Asian and Russian financial crises, as well as the collapse of LTCM do not generate a spike in the financial crisis index. This may be explained by the effectiveness of the "Greenspan put" in combination with the much lower exposure of the major U.S. banks to the 1998 external crisis events than the 1982 external crisis events. Also interesting is the peak of the financial crisis index in 2010, the time of the European sovereign debt crisis, and the way in which the index declines rapidly after that peak. In contrast, the financial stability index remains high. Taken together, these contrasting patterns illustrate that while discussion of the global financial crisis itself quickly subsided, discussions of financial stability in its aftermath continued.

4.1.3 Mandate Description

During the period we examine, the Federal Reserve tended to talk about either the components of its dual mandate or financial stability. In Figure 1c, we also show the series for inflation and output growth to examine how the Fed's focus shifted from dual mandate content to financial stability over time and across cycles. Consistent with movements in inflation, textual content about inflation in speeches increases during the period in the late 1970s and early 1980s, an era that has been classified as discretionary with large deviations from the Taylor rule (see, e.g., Nikolsko-Rzhevskyy et al., 2014).

Another notable feature of the figure is that financial stability and output growth content diverge prior to the Great Moderation, but then positively comove thereafter. This suggests that the Great Moderation could have been a turning point with respect to Fed communication. For this reason, we will include sample splits in our regression exercises in Section 5, in order to evaluate the pre-Great Moderation and Great Moderation periods separately. It is also noteworthy that the inflation and output growth features appear to comove in the early 1970s and again during the Volcker years.

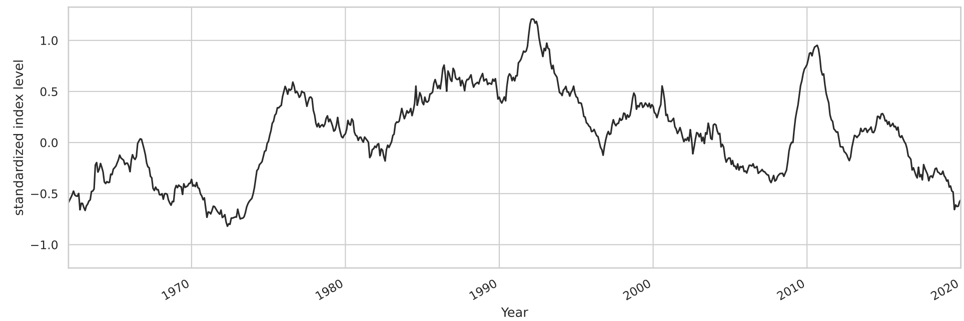
4.1.4 Financial Stability Position

In addition to measuring broad text features related to financial stability, we also construct two features that provide a direct description of the Fed's evolving perception of its mandate. The first feature, shown in the top panel of Figure 3, measures the semantic textual similarity (STS) between each paragraph in our speech corpus and the statement "banking regulation should be used to achieve financial stability," as described in Section 3.3.3. Peaks for this series are typically closely related to crisis events and some resemble those seen in the financial crisis text feature in Figure 1b, such as the peaks associated with the aftermath of the Global Financial Crisis and the European Sovereign Debt Crisis. Yet there are some important differences: the peak in the late 1980s, for instance, seems to follow a similar peak in Figure 1b, while the peak around 1992, the year the Basel I capital regulations were implemented and the year after a significant tightening of banking regulation under the Federal Deposit Insurance Corporation Improvement Act of 1991, is only apparent in the STS score capturing banking regulation advocacy.

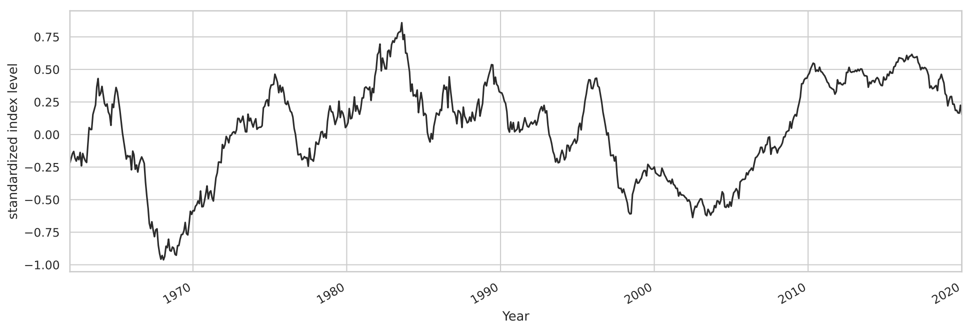
In contrast, the bottom panel of Figure 3 measures the STS score for each paragraph and the statement "monetary policy should be used to achieve financial stability," providing us with a measure of how inclined the Fed is to advocate for the use of use monetary policy as a means to achieve financial

Figure 3: Financial Stability Advocacy

a. STS Score: Banking Regulation and Financial Stability



b. STS Score: Monetary Policy and Financial Stability



Notes: Both panels plot semantic textual similarity, as measured by standardized cosine similarity. Cosine similarity is often used in natural language processing to quantify the similarity between two embedding vectors and returns a number between −1 and +1, where two identical sentences score +1, two unrelated sentences score around 0, and two sentences with opposite meanings score −1. The top panel shows the semantic textual similarity (STS) score computed between the statement "banking regulation should be used to achieve financial stability" and the contextualized embeddings for paragraphs with high classification scores for "financial stability." STS scores are repeated for the same exercise in the bottom panel, using the statement "monetary policy should be used to achieve financial stability" instead of "banking regulation..."

stability. Examples of passages with a high semantic textual similarity score for both features are given in Tables A10 and A11 of the Appendix. We can see a clear difference between the evolution of the two series in Figure 3, suggesting that the Fed takes a nuanced approach to achieving financial stability, at times favoring the use of monetary policy and at others the use of banking regulation. The more jagged nature of the first plot suggests a more episodic role for banking regulation to achieve financial stability, rather than being a sustained tool. We can see, for example, that the Fed's discussion of banking regulation in relation to financial stability peaked sharply around the time of the financial crisis, but declined thereafter; whereas the discussion of monetary policy as a means of achieving financial stability has remained elevated since the crisis.

Looking further at the semantic textual similarity depicted in the bottom panel of Figure 3, it is also apparent that support for the idea that monetary policy should be used to achieve financial stability varies according to the level of accommodation in the Fed's stance, and in particular, the cosine similarity appears to increase as the Fed tightens and decreases (or, in the last two decades remains low or constant) during easing periods. This pattern is consistent with an endorsement of a forward looking "leaning against the wind" view, rather than a "financial instability is caused by monetary tightening" view. We will explore this aspect in our regressions in Section 5.

4.1.5 Concern Type and Level

Not all content contained in Fed speeches is equally informative. In order to identify statements and terms that reveal internal concerns, we make use of extractive question answering, as described in Section 3.3.2. Specifically, for each paragraph, we query the model with "What is the most significant concern in the passage?" This yields a concern, which is extracted from the text, along with a score that indicates the model's uncertainty. A low score, for instance, might indicate that no specific concern was stated or that the model was unable to classify the stated concern accurately.

4.1.6 Temporal Focus

In addition to distinguishing between the content of financial stability concerns, it is also useful to identify the tense or the temporal focus of a statement. Differences in verb tense could indicate whether a paragraph is discussing a past crisis, an unfolding crisis, or the prospect of a future crisis and may indicate the extent to which Federal Reserve officials rely on historical precedent, are reactive, or are proactive in making their decisions. We use zero shot learning to classify each tense separately, allowing for the possibility that no clear tense is established in a given statement.

The standardized indices for past, present, and future focus are plotted in the top panel of Figure 4. The bottom panel shows the difference between future and past focus, which has been rising since the start of the Great Moderation. The increased use of future tense coincides with the Fed's emphasis on increased transparency and forward guidance. Prior to the Great Moderation, there is evidence that an increased use of future tense coincides with tightening episodes, whereas the shift towards a focus on the past coincided with the easing that began in April 1989. Table A12 in the Appendix provides example speech passages that are classified as containing a dominant tense.

4.2 Institutional Variation in Text Features

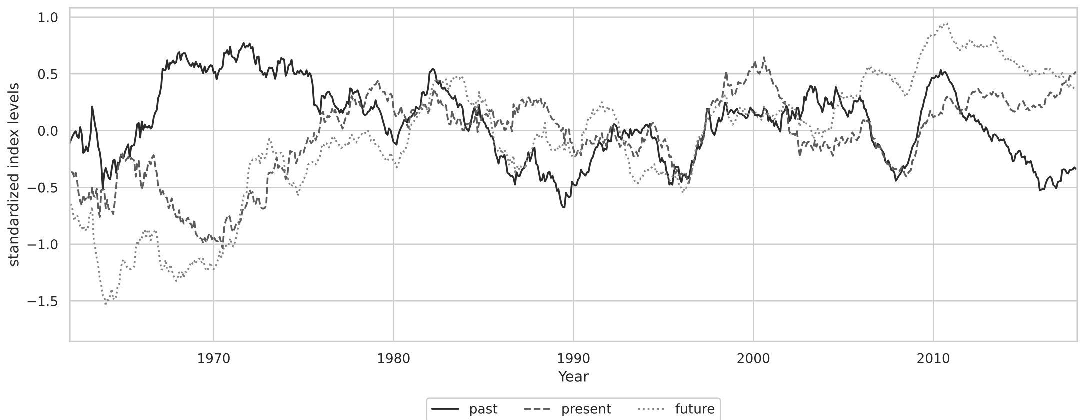
Another source of variation in text features arises from differences in the institutional arrangements and concerns of different Federal Reserve Banks and the Board of Governors. For instance, the Cleveland and New York Federal Reserve Banks may have different concerns, since there are differences in the compositions of their respective regional economies and the bank holding companies they supervise. Furthermore, given the local appointment processes, some district banks may favor presidents with academic profiles, while others may favor those with private sector experience. In this subsection, we will attempt to document the variation we were able to measure across districts and the Board of Governors.[28]

Financial Stability Focus. First, we find that the Federal Reserve Banks of New York and Richmond have the highest average financial stability scores, consistent with the fact that most of the

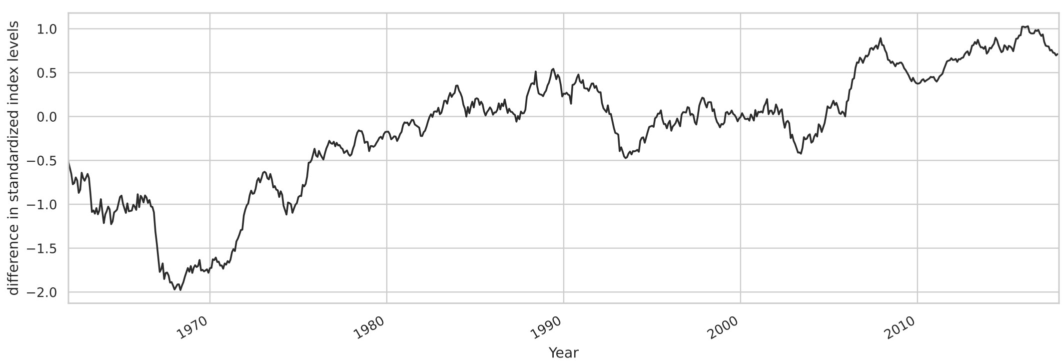
[28]In the interest of space, we summarize our findings for some of the individual speakers; details of these results are available on request.

Figure 4: Text Features

a. Text Features: Past Focus, Present Focus, Future Focus



b. Text Features: Difference Between Future and Past Focus



Notes: The figure in the top panel shows the standardized classification score for the speech tense for a focus on the "past", "present" and "future." This is computed by classifying each paragraph and then computing the average score for each month, standardizing it, and then plotting the 2-year rolling mean. The figure in the bottom panel shows the 2-year rolling mean of the difference between the "future" and "past" focus scores.

systemically important financial institutions (SIFIs) are in those two districts. In addition, many of the speakers with high scores (e.g., Bernanke, Dudley, Evans, Geithner, George, Kocherlakota, Kohn, Kroszner, Lacker, Minehan, Plosser, Stern, Tarullo, and Warsh) were officials in the period surrounding the Global Financial Crisis of 2007-2009 (GFC), a time when the overall financial stability content of speeches was near its highest levels. Many of the other speakers with high scores were members of the Board of Governors; in general, reserve bank presidents' speeches have less financial stability content, although the three highest index scores are associated with reserve bank presidents.[29] The most negative scores are also associated with reserve bank presidents, many of whom served near the start of the Great Moderation (e.g., Eastburn, Kimbrel, Parry, Roos, and Willes). Overall, the results align well with our intuition.

[29]In addition to Minehan and Lacker, Loretta Mester, President of the Cleveland Fed, has a very high score; the Cleveland Fed co-hosts an annual financial stability conference.

Future Focus. Second, as shown in Table A7 in the Appendix, we find that the Federal Reserve Banks of New York and Richmond also have the highest scores on future focus. This is consistent with the previous finding, given the strong connection between future focus and financial stability documented in the cosine similarity results. As noted previously, these districts supervise the largest, most systemically important financial institutions in the US. The NY Fed also has the highest scores for past and present focus; taken together, these results indicate that tense usage is clearer and more distinct in speeches by its presidents than in speeches by presidents of other districts.

Looking across the individual speakers' scores, there is a strong positive correlation between the future focus and financial stability scores (0.60), suggesting that discussions of financial stability involve the use of more forward-looking language.[30] As a result, many of the speakers with high future focus and financial stability scores are the same ones, encompassing many of those that were on the FOMC during the GFC.

Academic Inclination. There is a long-standing debate in the literature over the extent to which academic discussions influence central bank deliberations and policy-making. Much of the work in this discussion concludes that academic work positively influenced central bank decisions. Some, such as Mankiw (2006), diverge from the majority position and argue that policy-making is largely uninformed by the academic literature. Others, such as Howitt (2012) argue that central bankers often face crises that have not been adequately studied by the academic literature and, thus, typically lead the literature.

On an informational level, Orphanides (2001, 2003) documents the salience of misinformation at the time when the monetary policy decisions are made, while Meltzer (2009) highlights the role of misconceptions about economic theory that led the Fed to chase the Phillips curve in an attempt to lower unemployment during the 1970s.

With respect to financial stability, the academic literature has largely argued that macroprudential policy – not monetary policy – should be used to achieve financial stability (Vollmer, 2022). As such, the Fed's interpretation of its mandate and its belief about what may be achievable with monetary policy may also depend on its openness to dialogue with academics. To capture this, we use zero shot learning to identify whether a paragraph refers to an "academic debate or the academic literature."

Just as the NY and Richmond districts have the highest future focus and financial stability scores, so also do they have a high academic focus score. In the case of Richmond, this likely reflects the academic backgrounds of the two main presidents in our corpus (Broaddus and Lacker), while for the NY Fed, the presidents' backgrounds are more mixed, but still reflect a high reliance on academic debate and literature, consistent with its large research division of PhD economists. The Federal Reserve Board also has a high academic focus score, unsurprisingly given the more than 200 staff economists that support its work.

Communication Clusters. Finally, we examine clustering in the financial stability content of Federal Reserve Bank speeches. We cluster on the content of speeches, which we represent using

[30] In contrast, the correlation between past or present focus and financial stability is 0.07 and 0.21, respectively.

contextualized sequence embeddings produced using the RoBERTa model. For each paragraph of each speech, this yields a 768-dimensional vector, where each dimension corresponds to a text feature. We then compute the average vector for each district-month.

To prepare the data for visualization, we first perform dimensionality reduction using the t-distributed stochastic neighbor embedding (t-SNE) introduced in van der Maaten and Hinton (2008). The t-SNE approach is a nonlinear dimensionality reduction technique that preserves both local and global structure and is intended to produce representations that can be visualized in 2 or 3 dimensions. It is an effective tool for differentiating between and visualizing clusters; however, the distances across clusters are not readily interpretable.

We first apply t-SNE to the full sample, which spans the period between 1960 and 2022, and visualize it in Figure 5. The markers in the visualization represent the different Federal Reserve Banks, as well as the Board of Governors (FRB). The Federal Reserve Bank of New York (denoted NY in the Figure) appears to have a nearly self-contained cluster (at the far left of the figure), suggesting that the content of their financial stability discussions is closer to their own past discussions than it is to contemporaneous communication at other institutions of the Federal Reserve System.

In contrast to the tight and distinct cluster for NY, the Federal Reserve Bank of St. Louis's (STL) communication (the light blue squares, many near the bottom of the figure) forms a long cluster that overlaps with the communication of many other Federal Reserve banks. Other regional banks have tight clusters (e.g., ATL, DAL), similar to the NY one, but are considerably closer to the clusters of other districts, indicating that their communication is dominated more by contemporaneous views than by past discussions.

5 Econometric Results

We next move to the empirical tests. Sections 5.1 and 5.2 analyze the determinants of financial stability concerns. Section 5.3 explores the association between the text features and financial ratios. And Section 5.4 conducts extended Taylor rule regressions to examine whether our text features explain time variation in the historical conduct of monetary policy.

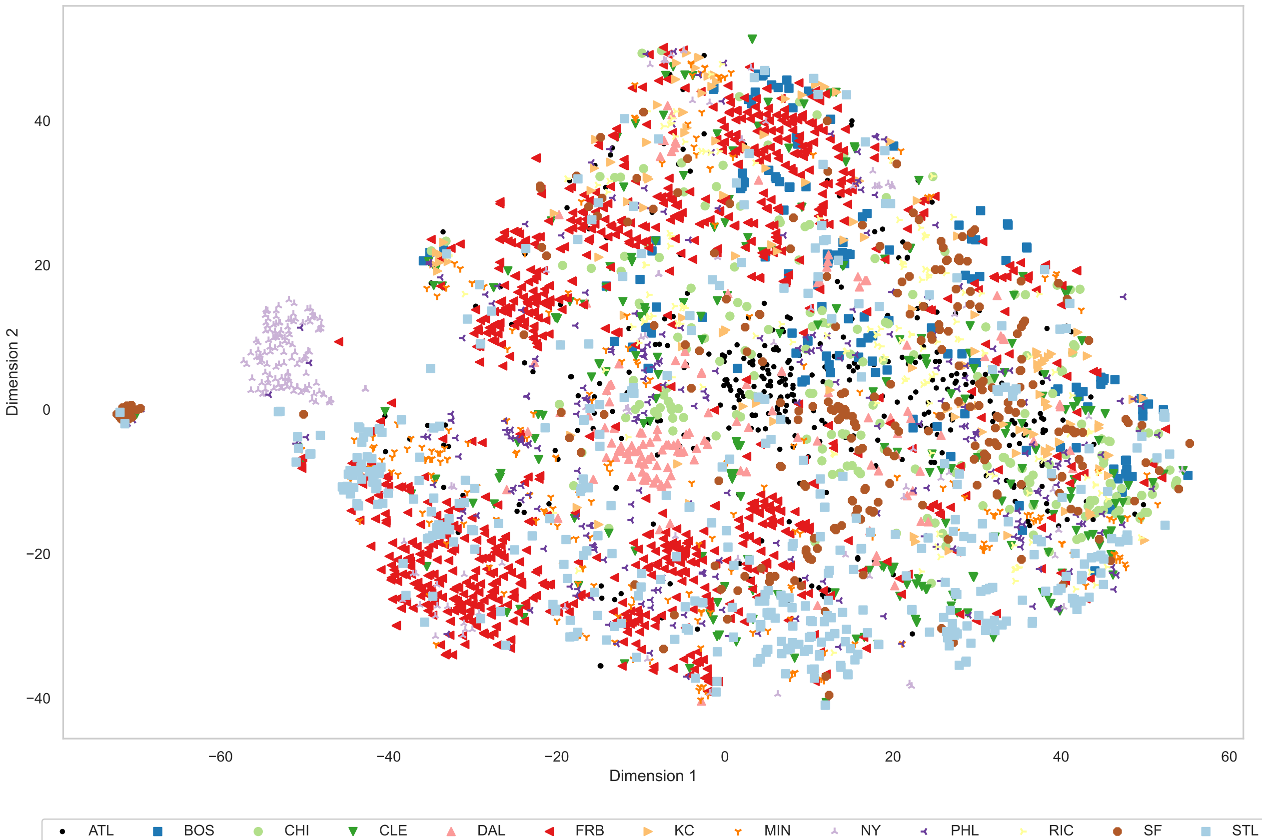
5.1 Empirical Specification

The primary specification is given in Equation (1):

$$y_{jt} = \beta_0 + \beta_1 \tau^{fc}_{jt} + \beta_2 \tau^{\pi}_{jt} + \beta_3 \tau^{e}_{jt} + \beta_4 \tau^{bc}_{jt} + \beta_5 \tau^{bl}_{jt} + \beta_6 \nu^{m}_{t-1} + \beta_7 \nu^{f}_{t-1} + \zeta_k + \gamma_t + e_{jt} \tag{1}$$

In the first set of regressions, the dependent variable, y_{jt}, is a measure of the financial stability content of a particular paragraph. The explanatory variables include various text features indicated by the variable τ. Such features are computed at the paragraph level and, thus, have both paragraph (j) and time (t) variation. The superscripts on τ indicate what text control it encompasses: financial crisis (fc), inflation (π), employment (e), bank capital (bc), and bank liquidity (bl). Summary statistics for the text variables used in the regression exercises below are provided in Tables A6 and A7.

Figure 5: t-SNE: Sequence Embeddings for Paragraphs Related to Financial Stability



Notes: The figure visualizes the output of the t-stochastic nearest neighbors (t-SNE) algorithm applied to sequence embeddings for paragraphs that discuss financial stability. We compute embeddings for all documents in the corpus over the period between 1960 and 2022. The sequence embeddings are produced for each paragraph using a RoBERTa sentence transformer model with extended pre-training on abstracts from the S2ORC corpus, as well as fine-tuning on sentence pairs from the S2ORC corpus. Each dot corresponds to a 2-dimensional representation of the average sequence embedding for a given month and institution.

Additionally, institution fixed effects and time fixed effects are indicated as ζ_k and γ_t, respectively.

We additionally include macroeconomic and financial controls, indicated by ν; these only have time variation (t). The macro variables (ν^m) include inflation, the output gap, house prices, and the debt-to-gdp ratio. Financial variables (ν^f) include the short term interest rate, a financial crisis indicator, the loan-to-deposit ratio, and the natural log of the total loan volume to the non-financial sector. We refer the interested reader to Table A3 in the Appendix for details about the macro and financial data definitions and sources. Because speeches are delivered throughout a quarter and macro variables are typically measured at the end of the quarter, we use lagged values for the macro variables.[31]

All regressions use Newey-West standard errors with 4 lags. Additionally, all regressions include institutional fixed effects, macroeconomic controls, and financial controls. The results are also robust to replacing macroeconomic and financial controls with year-month fixed effects.[32]

[31]The regression results are not sensitive to using contemporaneous values of macro variables, rather than lagged values.

[32]An earlier version of the paper contained complete sets of these robustness results; these are available from the

5.2 Regression results

We first attempt to measure the associations between financial stability discussion and advocacy, and three sets of variables: text features, macroeconomic controls, and financial controls; the empirical results are presented in Table 1. For each group of three columns, we present full sample results in the first column, followed by results for the two subsamples, 1960-83 (pre-Great Moderation) and 1984-2020 (Great Moderation). In columns (1)-(3), the dependent variable is the text feature "financial stability," which measures the extent and intensity of financial stability discussion in a speech passage.

In columns (4)-(6) of Table 1, the dependent variable measures advocacy for the use of monetary policy to achieve financial stability. As noted above, this variable is constructed from the cosine similarity (STS score) between embeddings generated for the statement "monetary policy should be used to achieve financial stability" and the content of a speech passage. In columns (7)-(9), we use an analogous measurement of advocacy for the use of banking regulation to achieve financial stability. This is measured by computing cosine similarity between each speech passage and the statement "banking regulation should be used to achieve financial stability."

Because we are interested in examining the way in which the Federal Reserve System discusses its mandate, several of our regression exercises will have a dependent variable and explanatory variables that are both text features. To enhance interpretability, we standardize all text variables included in the regression exercises. This allows for an interpretation of the size of an association in terms of the variation in the variables around their respective means.

In addition to the regression results, we plot the Shapley values (Sundararajan et al., 2020; Buckmann and Joseph, 2022) for the five variables with the most explanatory power in each regression in Figures 6a-6c. Shapley values are a game-theoretic notion for measuring the importance of each individual variable. The plots show the impact of each of these variables in a 1% random subsample of our dataset. A larger magnitude *SHAP* value indicates a greater impact on the predicted values for a given observation.[33]

Financial Crises and Financial Stability. Our first finding, given in column (1) of Table 1, indicates that a one standard deviation increase in the discussion of financial crises is associated with a 0.0521 standard deviation increase in the discussion of financial stability. This suggests that a large increase in the discussion of financial crises within a speech paragraph is typically accompanied by a small, but strongly statistically significant increase in the discussion of financial stability *within the same paragraph.*

The result also holds in the subsample prior to the Great Moderation (column (2)) and in the subsamble including the Great Moderation (column (3)). Moreover, the result holds even in the presence of macroeconomic and financial controls. Additionally, it is robust to the use of year-month fixed effects instead of macroeconomic and financial controls, as well as to whether we use contemporaneous or lagged controls.[34] Thus, irrespective of the state of the economy and financial

authors on request.

[33]We are grateful to an anonymous referee for the suggestion to use Shapley values.

[34]To simplify the presentation, we condense the main regression results into a single table; however, both the results

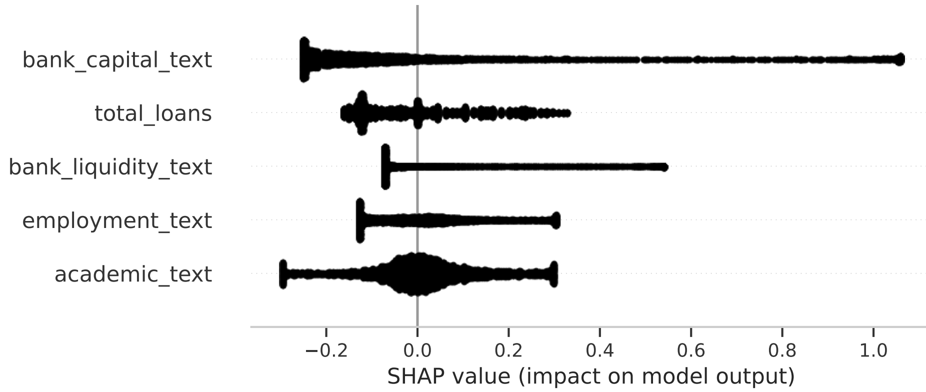
Table 1: Financial Stability and Policy Advocacy

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent Variable	**FS**	**FS**	**FS**	**MP-FS**	**MP-FS**	**MP-FS**	**BR-FS**	**BR-FS**	**BR-FS**
inflation_{jt} [text]	0.0285***	-0.0388***	0.0536***	0.2072***	0.2407***	0.1943***	-0.3103***	-0.2185***	-0.3353***
	(0.0024)	(0.0043)	(0.0029)	(0.0021)	(0.0043)	(0.0023)	(0.0020)	(0.0036)	(0.0024)
employment_{jt} [text]	0.1177***	0.1467***	0.1071***	-0.1821***	-0.1628***	-0.1895***	-0.1888***	-0.1651***	-0.1962***
	(0.0018)	(0.0035)	(0.0021)	(0.0022)	(0.0044)	(0.0026)	(0.0020)	(0.0035)	(0.0023)
financial crisis_{jt} [text]	0.0521***	0.0276***	0.0622***	0.1405***	0.1441***	0.1369***	0.2307***	0.1395***	0.2569***
	(0.0024)	(0.0045)	(0.0028)	(0.0025)	(0.0051)	(0.0029)	(0.0025)	(0.0044)	(0.0029)
bank liquidity_{jt} [text]	0.1494***	0.1665***	0.1419***	-0.1736***	-0.1826***	-0.1686***	-0.1728***	-0.1484***	-0.1752***
	(0.0035)	(0.0067)	(0.0040)	(0.0040)	(0.0081)	(0.0045)	(0.0047)	(0.0087)	(0.0055)
bank capital_{jt} [text]	0.3147***	0.3591***	0.2995***	0.0676***	0.0579***	0.0717***	0.3997***	0.3107***	0.4219***
	(0.0037)	(0.0073)	(0.0043)	(0.0038)	(0.0078)	(0.0043)	(0.0046)	(0.0086)	(0.0054)
past focus_{jt} [text]	0.0019	0.0116***	-0.0013	-0.0916***	-0.1088***	-0.0865***	-0.0855***	-0.0880***	-0.0859***
	(0.0014)	(0.0026)	(0.0017)	(0.0019)	(0.0038)	(0.0022)	(0.0018)	(0.0033)	(0.0022)
present focus_{jt} [text]	0.0669***	0.0493***	0.0749***	0.0189***	0.0396***	0.0103***	0.0954***	0.0890***	0.0976***
	(0.0015)	(0.0026)	(0.0018)	(0.0020)	(0.0037)	(0.0024)	(0.0019)	(0.0032)	(0.0023)
future focus_{jt} [text]	0.0786***	0.0663***	0.0816***	0.0694***	0.0516***	0.0756***	-0.0360***	-0.0323***	-0.0356***
	(0.0018)	(0.0033)	(0.0021)	(0.0022)	(0.0042)	(0.0025)	(0.0020)	(0.0036)	(0.0024)
academic focus_{jt} [text]	0.1180***	0.1255***	0.1149***	-0.0344***	-0.0676***	-0.0231***	-0.0020	0.0006	-0.0023
	(0.0017)	(0.0035)	(0.0020)	(0.0022)	(0.0047)	(0.0025)	(0.0022)	(0.0040)	(0.0025)
debt-to-gdp ratio_{t-1}	0.1908***	-0.3003	0.1519***	-0.1229***	1.8765***	-0.4149***	0.0174	2.2750***	0.0858
	(0.0241)	(0.2139)	(0.0342)	(0.0475)	(0.4168)	(0.0677)	(0.0441)	(0.3136)	(0.0650)
loan-to-deposit ratio_{t-1}	0.0026***	0.0068**	0.0040***	-0.0096***	-0.0240***	-0.0073***	0.0063***	-0.0031	0.0022*
	(0.0005)	(0.0027)	(0.0007)	(0.0009)	(0.0056)	(0.0013)	(0.0008)	(0.0042)	(0.0013)
Sample Period	1960-2020	1960-1983	1984-2020	1960-2020	1960-1983	1984-2020	1960-2020	1960-1983	1984-2020
R-squared Adj.	0.5125	0.5115	0.5143	0.0937	0.1048	0.0967	0.1830	0.1373	0.2018
N	363,154	95,855	267,299	363,154	95,855	267,299	363,154	95,855	267,299

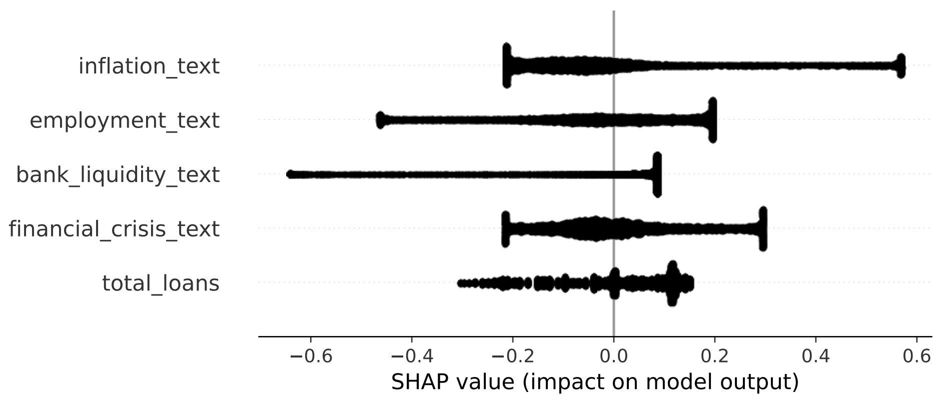
Notes: Three different dependent variables are used: `FS`, `MP-FS`, and `BR-FS`. Columns (1)-(3) use `FS`, the financial stability text feature. Columns (4)-(6) use `MP-FS`, the cosine similarity between the embedding for a given speech passage and the embedding that corresponds to the sequence "monetary policy should be used to achieve financial stability". And columns (7)-(9) use `BR-FS`, the cosine similarity between the embedding for a given speech passage and the embedding that corresponds to the sequence "banking regulation should be used to achieve financial stability." All controls that include [*text*] indicate that they are text features measured using zero shot classification. For instance "inflation [text]" is the classification score for whether a sequence describes inflation. The features "past focus [text]," "present focus [text]," and "future focus [text]" classify the tense of a sequence to determine whether the speaker was discussing the past, present, or future. "Academic focus" classifies whether a speaker was discussing an academic debate or the academic literature. All regressions include macro controls, financial controls, the log of house prices, and institutional fixed effects. Macro controls include the output gap and consumer price inflation. Financial controls include the log of total loans to the non-financial sector, the short term interest rate, an indicator for whether a financial crisis occurred in a given year, the debt-to-gdp ratio, and the loan-to-deposit ratio. The financial controls and house prices are taken from the *Macrohistory Database*, introduced by Jordà et al. (2016). To save space, we exclusively report estimates for the two controls of interest: the debt-to-gdp ratio and the loan-to-deposit ratio. The remaining estimates are available on request. Note that j indexes paragraph and t indexes time. Standard errors are Newey-West (NW), but clustering at the district level does not change the results. * $p < .1$, ** $p < .05$, *** $p < .01$. The results are not sensitive to using contemporaneous macroeconomic and financial controls or replacing the controls with month x year fixed effects.

Figure 6: Shapley Values: Variables with the most Explanatory Power

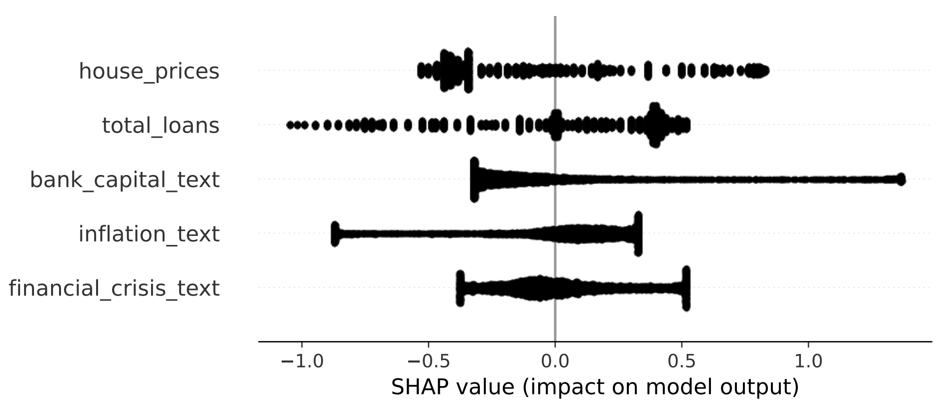
(a) Financial Stability **(FS)** Regression



(b) Advocacy for Monetary Policy to Achieve Financial Stability **(MP-FS)** Regression



(c) Advocacy for Banking Regulation to Achieve Financial Stability **(BR-FS)** Regression



Notes: The figure above shows a beeswarm plot of a 1% random sample of Shapley values from a regression of the cosine similarity score that measures advocacy for the use of monetary policy to achieve financial stability on text features, macroeconomic variables, and financial variables. A beeswarm plot visualizes the impact of a given variable on the dependent variable. The impact may be positive or negative, depending on the sign of the regression coefficient and the value of the variable. Larger absolute values indicate a larger absolute impact on the dependent variable. Points are visualized to be non-overlapping, so a greater width corresponds to a greater density. We display the five most important variables for explaining variation in the dependent variable.

system, an uptick in discussion of financial crises appears to be strongly and statistically significantly associated with discussion of financial stability.

Discussion of financial crises does not, however, explain the highest share of the variation in financial stability discussion among all variables included in our regression. As shown in Figure 6a, which visualizes the Shapley values for a 1% random subsample of our dataset, discussion of bank capital accounts for the highest share of variation in financial stability discussion. Additionally, high indebtedness, as measured by total lending, is one of the most useful variables for explaining discussion of financial stability in speeches.

Dual Mandate and Financial Stability. Another important question about financial stability is whether it is typically discussed in the context of the Fed's dual mandate or whether it is treated as a separate concern or objective. In the first column of Table 1, the coefficients on the inflation and employment text features are positive and strongly statistically significant, indicating that financial stability discussion is positively associated with dual mandate related concerns over the full sample. Furthermore, we can see that the association with financial stability is stronger for employment, where the effect size is more than four times as large as that for inflation.

Initially, this result may seem at odds with literature suggesting that aggressive inflation-targeting is conducive to financial stability (Bernanke and Gertler, 2001). However, when the sample is split into two sub-periods (columns (2) and (3)), the relationship between inflation and the discussion of financial stability appears to be negative prior to the Great Moderation, when there were episodes of high inflation and output growth was less stable. In addition, in this first subsample (1960-1983), the employment effect is about 20% stronger than over the full sample. In contrast, during the Great Moderation period, the relationship between inflation and financial stability becomes strongly positive. In this sub-period inflation targeting also became more prominent at the Federal Reserve, suggesting that, in line with the literature, stabilizing prices and the financial sector are not treated as separate concerns or as being in conflict. In addition, during the Great Moderation, the effect size for employment declines while that for inflation increases.

Academic Focus. Next, in column (1) of Table 1, we find that a focus on academic debates and the academic literature is strongly associated with increased discussion of financial stability. In all financial stability specifications (columns (1)-(3)), a one standard deviation increase in academic focus is associated with an approximately 0.12 standard deviation increase in financial stability content, an effect magnitude equivalent to that of the employment component. This suggests that, among the topics that Federal Reserve officials discuss, financial stability appears to have an above average concentration on the academic discussion, possibly reflecting the need for a conceptual framework. We can also see from Figure 6a that academic focus is the fifth most important feature for explaining variation in the discussion of financial stability in Federal Reserve System speeches.

In addition, the significant negative coefficient on the academic focus variable in Columns (4)-(6) in Table 1 indicates that Fed speeches that reference the academic literature tend to oppose the use

with fixed effects and the results with contemporaneous rather than lagged controls are available on request.

of monetary policy to achieve financial stability (Vollmer, 2022). Columns (7)-(9) in Table 1 suggest little association between references to the academic literature and advocacy for the use of banking regulation to achieve financial stability.

Comparing the various subsample regressions suggests that, although there is little change in the association between academic focus and financial stability overall, the use of academic literature to voice opposition to the use of monetary policy to achieve financial stability appears to have declined during the Great Moderation (relative to the period before the Great Moderation).

Figures 6b and 6c present plots of the Shapley values for the regressions in columns (4) and (7) of Table 1, respectively. Figure 6b suggests that the most useful variables for explaining variation in the advocacy for the use of monetary policy to achieve financial stability are other text variables, including the discussion of inflation, employment, bank liquidity, financial crises, and total loans. As shown in Figure 6c, economic variables, such as total loans and house price growth, explain a high share of variation in advocacy for the use of banking regulation to achieve financial stability. The text features with the most explanatory power capture the discussion of bank capital, inflation, and financial crises.

Speech Tense. Column (1) of Table 1 suggests that statements about financial stability tend to be framed in terms of the present and future, perhaps suggesting that they are in response to ongoing or anticipated events. This tendency increased during the Great Moderation (compare columns (2) and (3)). Advocacy for the use of monetary policy to achieve financial stability also typically hinges on the use of future hypotheticals, as indicated by column (4) of Table 1; this advocacy increased during the Great Moderation (compare columns (5) and (6)). The present is also positively associated with advocacy for the use of monetary policy to achieve financial stability while the past tends to be used to argue against it. In contrast, advocacy for the use of bank regulation is typically focused on events in the present, as indicated by column (7) of Table 1, while the past and future tenses are used to argue against it. This result is evident in both subsamples as well (see columns (8) and (9)).

Financial Variables and Discussion of Financial Stability. Turning to the financial variables in column (1) of Table 1, we find that an increase in the debt-to-GDP ratio or an increase in the loan-to-deposit ratio, which tends to indicate a deterioration in the liquidity position of banks, appears to be associated with a considerable increase in discussion of financial stability; however, this increased discussion does not necessarily lead to advocacy for a particular approach to policy. Columns (5) and (8) of Table 1 show that a higher debt-to-GDP ratio was strongly associated with increased advocacy of monetary policy and banking regulation as a means of achieving financial stability prior to the Great Moderation. Consistent with the view in Meltzer (2009) that the Fed's policy was driven by the forces of political pressure, the association between the advocacy for the use of monetary policy and the public debt-to-GDP ratio is more prominent in Fed communication during the pre-Volcker period when the Fed was subservient to fiscal dominance. In the second subsample, however, a higher debt-to-GDP ratio is actually associated with less advocacy for the use of monetary policy to achieve financial stability.

The results for the loan-to-deposit ratio, however, are more mixed. An increase in total lending is associated with less advocacy for the use of monetary policy and more advocacy for the use of banking regulation to achieve financial stability over the full sample, but in the first subsample the effect is stronger in advocating against the use of monetary policy while there is no significant evidence that total lending influences the advocacy for the use of banking supervision to achieve financial stability. In the second subsample, the effect sizes are smaller in magnitude than over the full sample – approximately one-forth less for advocacy of monetary policy and only one-third as strong for the use of banking supervision. The reduced advocacy for the use of banking regulation to achieve financial stability during this period is consistent with the overall deregulation environment that began during the Reagan presidency and continued into the Greenspan era.

5.3 Financial Markets

We next demonstrate that the concerns articulated in Federal Reserve speeches matter for financial ratios and returns. We do this by estimating a modified version of Equation (1), where the dependent variable is a measure of annual asset returns and the text feature financial stability is introduced as an additional explanatory variable (τ_{jt}^{fs}). Table 2 reports the association between the variables of interest and returns on equity, bonds, risky assets, and safe assets. For sources and return definitions, see Table A3 in the Appendix.

Conceptually, we expect the content of Fed speeches to matter for asset returns, potentially in a differential way through the risk-taking channel of monetary policy. The impact of Fed communication is likely to occur through market participants' risk perceptions and risk attitudes; and through expectations about monetary conditions, which affect the riskiness of bank lending, valuations, and risk measures (Jiménez et al., 2014; Dell'Ariccia et al., 2017). Our evidence offers support for a broad effect of the discussion in speeches by Federal Reserve officials on all asset returns, after including a battery of macro and financial controls and institutional fixed effects.

We find a positive association between discussion of inflation and the returns to equity, risky assets and safe assets. Financial crisis discussion has a positive impact on equity returns (see columns (1) and (2)), which is consistent with the communication of monetary easing associated with a financial crisis, while it has a negative impact on bond and safe asset returns, as well as the more loosely defined risky asset category. Also financial stability discussion has a significant negative association with bond, risky, and safe asset returns, but not with equity returns. When interacted with future focus (column (2)), financial stability discussion is also negatively associated with equity returns at a 10% level of significance, which is consistent with signalling of monetary and/or regulatory tightening. As discussed in Section 5.2, the focus of this type of discussion is often academic, which may contribute to the negative association of academic focus with equity returns in columns (1) and (2).

Notably, the text features have significant effects on asset returns even after controlling for macro and financial variables, which arguably capture some of the discussion in Fed speeches. Table 2 also reports the estimated coefficients for the lagged debt-to-GDP and loan-to-deposit ratios, which show a strong positive association between leverage and asset returns.

One reason why the concerns expressed in Federal Reserve speeches may be important to financial

Table 2: Federal Reserve Speech Impact on Asset Returns, 1960-2020 (N=363,154)

	(1)	(2)	(3)	(4)	(5)
Return type	Equity	Equity	Bond	Risky	Safe
inflation$_{jt}$ [text]	0.0010**	0.0010**	0.0000	0.0007***	0.0003**
	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0002)
employment$_{jt}$ [text]	-0.0008**	-0.0008**	0.0000	0.0007***	0.0003**
	(0.0003)	(0.0003)	(0.0003)	(0.0001)	(0.0001)
financial stability$_{jt}$ [text]	-0.0002	-0.0002	-0.0010***	-0.0003**	-0.0004***
	(0.0004)	(0.0004)	(0.0003)	(0.0001)	(0.0001)
financial crisis$_{jt}$ [text]	0.0013***	0.0013***	-0.0009***	-0.0016***	-0.0010***
	(0.0004)	(0.0004)	(0.0003)	(0.0002)	(0.0002)
bank liquidity$_{jt}$ [text]	-0.0009	-0.0008	-0.0015***	-0.0008***	-0.0011***
	(0.0006)	(0.0006)	(0.0004)	(0.0002)	(0.0002)
bank capital$_{jt}$ [text]	-0.0005	-0.0005	0.0020***	-0.0003	0.0011***
	(0.0006)	(0.0006)	(0.0004)	(0.0002)	(0.0002)
past focus$_{jt}$ [text]	-0.0004	-0.0003	-0.0004**	-0.0001	-0.0003**
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0001)
present focus$_{jt}$ [text]	0.0012***	0.0014***	0.0000	0.0006***	0.0001
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0001)
future focus$_{jt}$ [text]	-0.0005	-0.0006*	0.0004*	0.0001	0.0003**
	(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0001)
past focus$_{jt}$ $*$ financial stability$_{jt}$ [text]		0.0004			
		(0.0003)			
present focus$_{jt}$ $*$ financial stability$_{jt}$ [text]		0.0004			
		(0.0004)			
future focus$_{jt}$ $*$ financial stability$_{jt}$ [text]		-0.0006*			
		(0.0003)			
academic focus$_{jt}$ [text]	-0.0014***	-0.0013***	0.0005**	0.0001	0.0003**
	(0.0004)	(0.0004)	(0.0002)	(0.0001)	(0.0001)
debt-to-gdp ratio$_{t-1}$	0.2875***	0.2875***	0.2171***	0.1143***	0.0113***
	(0.0094)	(0.0094)	(0.0061)	(0.0039)	(0.0030)
loan-to-deposit ratio$_{t-1}$	0.0029***	0.0029***	0.0039***	0.0022***	0.0003***
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
Adj. R-squared	0.2189	0.2189	0.2743	0.3376	0.3227

Notes: The dependent variable in each regression is a measure of annual asset returns. The returns are taken from Jordà et al. (2016) and include total equity, total bond, risky assets, and safe assets, as specified in the "return type" row of the table. All controls labeled [*text*] indicate that they are text features measured using zero shot classification. For instance inflation [text] is the classification score for whether a sequence describes inflation. The features "past focus," "present focus," and "future focus" classify the tense of a sequence to determine whether the speaker was discussing the past, present, or future. "Academic focus" classifies whether a speaker was discussing an academic debate or the academic literature. All regressions include macro controls, financial controls, the log of house prices, and institution fixed effects. Macro controls include the output gap and consumer price inflation. Financial controls include the log of total loans to the non-financial sector, the short term interest rate, an indicator for whether a financial crisis occurred in a given year, the debt-to-gdp ratio, and the loan-to-deposit ratio. The financial controls, house prices, debt-to-gdp ratio, and loan-to-deposit ratio are taken from the *Macrohistory Database*, introduced by Jordà et al. (2016). Note that j indexes paragraph and t indexes time. Standard errors are Newey-West (NW). * $p < .1$, ** $p < .05$, *** $p < .01$. Lagging the macroeconomic and financial controls or using clustered standard errors rather than Newey West standard errors does not substantially change the results.

markets is that they are informative about the future conduct of monetary policy. Next, we examine how our text features relate to the historical conduct of monetary policy.

5.4 Taylor Rule Regressions

Building on our earlier results, we explore time variation in the historical conduct of monetary policy and whether Fed officials' views as measured through speech features are reflected in their policy decisions. We do this by conducting regressions that extend the Taylor rule (Taylor, 1993, 1999) to include macroeconomic time series data and semantic variables.

Existing work has documented low frequency time variation in the Fed's monetary policy rule (Clarida et al., 2000; Orphanides, 2003; Boivin, 2006; Hamilton et al., 2011; Bianchi et al., 2022; Bauer et al., 2024). Our exercises aim to uncover key drivers of policy rate adjustments by examining the Federal Reserve's speeches. Specifically, we investigate whether non-dual mandate related concerns, as articulated in speeches of Federal Reserve officials, contribute to changes in the policy rule.

5.4.1 Specification of the Monetary Policy Rule

In a recent paper, Carvalho et al. (2021) argue in favor of using ordinary least squares (OLS) to estimate the Taylor rule and document that the endogeneity bias in OLS estimates is small. Using the Taylor rule specification, notation, and results in Table 2 of Carvalho et al. as a starting point, we estimate versions of an extended Taylor rule, which includes text features from Federal Reserve speeches:[35]

$$r_t = \theta_0 + \theta_{1,1} r_{t-1} + \theta_{1,2} r_{t-2} + \theta_2 E_t[\pi_{t+1}] + \theta_3 E_t[x_{t+1}] + \theta_{4,i} \tau_t^i + e_t, \tag{2}$$

This specification uses the same interest rate rule as in Clarida et al. (2000), which allows for interest rate smoothing. The dependent variable r_t is the interest rate. For the subsample prior to 2008, we use the federal funds rate for the interest rate; and for the period from 2008 to 2020, when monetary policy was constrained by the zero lower bound and/or inflation was below target, we use the end-of-quarter shadow rate from Wu and Xia (2016).

As explanatory variables, we follow Carvalho et al. (2021) and use lagged values of the interest rate, as well as lagged values of the real-time core CPI inflation and for the output gap as proxies for expectations, $E_t[\pi_{t+1}]$ and $E_t[x_{t+1}]$.[36] Text feature variables are denoted by τ_t^i, where the superscripts indicate the specific features used: $i = nd$ refers to non-dual mandate, $i = fs$ financial stability, and $i = mf$ refers to advocacy for the use of monetary policy to achieve financial stability.

[35] See Peek et al. (2016) for a study of the ternary mandate using an extended Taylor rule and dictionary-based methods to measure financial stability concerns from FOMC meeting transcripts. In a related paper Istrefi et al. (2021) show that a more negative tone in Federal Reserve speeches on financial stability topics is associated with a more accommodative policy stance.

[36] In addition to real-time data, Carvalho et al. (2021) also use one-quarter ahead Greenbook forecasts for the output gap and inflation, which introduces data limitations. The data are obtained from the U.S. Congressional Budget Office's (CBO) History and Projections for Key Economic Variables. Core CPI inflation includes all items less food and energy, and the output gap is constructed using the CBO's estimate of potential real GDP.

The choice of including a policy inertia term can be justified by both empirical evidence and theoretical studies (Coibion and Gorodnichenko, 2012). Importantly, the inclusion of interest rate smoothing (via lags of the interest rate) increases the explanatory power due to an excessive volatility in interest rates predicted by the standard Taylor rule.

Following Carvalho et al. (2021), we define $\rho \equiv \theta_{1,1} + \theta_{1,2}$, $\beta \equiv \theta_2/(1-\rho)$, $\gamma \equiv \theta_3/(1-\rho)$ and $\pi^* \equiv (\theta_0 - (1-\rho)rr^*)/((1-\rho)(1-\beta))$, where ρ is the interest rate smoothing coefficient, β is the coefficient on the inflation gap, γ is the coefficient on the output gap, rr^* is the equilibrium real interest rate and π^* is the inflation target.

5.4.2 Extended Taylor Rule Regressions

Column (1) of Table 3 shows the standard Taylor rule specification used by Carvalho et al. (2021) for the sample period 1987Q3–2007Q4 under the Fed chairmanships of Alan Greenspan and Ben Bernanke. We report the estimates for the coefficients β on the inflation gap and γ on the output gap, as well as ρ for the interest rate smoothing and π^* for the inflation target. The coefficient for the inflation gap is highly significant and greater than one for the full sample, indicating that the "Taylor principle" of raising the nominal interest rate more than one-for-one is satisfied, which is an important condition for the existence of a stable inflation rate in macro models. The coefficient for the output gap is also highly significant and close to one. In columns (2) and (3) we extend this Taylor rule regression by including the text features in the explanatory variable τ_t^i.

Column (2) reports the estimated coefficient for the non-dual mandate text feature variable τ_t^{nd}, which is defined analogously as $\delta^{nd} \equiv \theta^{nd}/(1-\rho)$, and column (3) reports the estimated coefficient for the financial stability text feature variable τ_t^{fs}. The coefficient estimate for the non-dual mandate text feature is negative and significant, meaning that non-dual mandate discussion is negatively associated with the federal funds rate – that is, it enters the expanded policy rule in an accommodative fashion. In contrast, the coefficient estimate for the financial stability text feature is insignificant. This changes in the subsequent ZLB/low inflation period.

Next, we consider the ZLB/low inflation subsample (2008Q1-2020Q4) in columns (4)-(8). Due to the zero lower bound episode, there is little variation in the dependent variable in the extended sample. As such, columns (4)-(8) of Table 3 report results when we replace the federal funds rate with the Wu and Xia (2016) shadow federal funds rate starting in 2008Q1, which also captures the effects of quantitative easing. Again, we first report the results of a standard Taylor rule regression in column (4), which we expand in column (5) by including the non-dual mandate text feature.[37] We find that the coefficient estimate for the non-dual mandate text feature is again negative and significant. Notably, the accommodative role of non-dual mandate related discussions appears to be more prominent with a more negative coefficient estimate, when compared to the earlier sample period in column (2), potentially indicating that non-dual mandate related discussions by Federal Reserve officials are more likely associated with changes in the conduct of monetary policy during the low inflation subsample, when the policy discussion was shaped by the experience of the Global

[37] Notably, the inflation gap coefficient becomes insignificant in the ZLB/low inflation subsample. We also run (unreported) regressions using the federal funds rate instead of the shadow rate, with similar results.

Table 3: Taylor Rule Regressions

	(1) FFR	(2) FFR	(3) FFR	(4) Shadow FFR	(5) Shadow FFR	(6) Shadow FFR	(7) Shadow FFR	(8) Shadow FFR
β	1.19***	1.20***	1.46***	-0.60	-0.89	-0.28	-0.51	-0.96
	(0.32)	(0.30)	(0.40)	(0.81)	(0.82)	(0.46)	(0.56)	(0.83)
γ	0.91***	0.92***	0.88***	0.44	0.21	0.33*	0.37**	0.52*
	(0.21)	(0.19)	(0.19)	(0.37)	(0.26)	(0.19)	(0.18)	(0.30)
ρ	0.82***	0.81***	0.81***	0.87***	0.87***	0.82***	0.82***	0.86***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.07)	(0.06)	(0.06)
π^*	0.81	0.05	2.37	2.28*	1.61*	2.79***	2.54***	2.27***
	(5.09)	(4.29)	(2.84)	(1.24)	(0.90)	(0.85)	(0.78)	(0.87)
non-dual mandate$_t$ [text]		-0.12**			-0.21***			
		(0.05)			(0.07)			
financial stability$_t$ [text]			0.07			-0.21	-0.29**	
			(0.07)			(0.14)	(0.14)	
monetary financial$_t$ [text]							0.18	
							(0.12)	
financial stability$_{t-1}$ [text]								-0.21
								(0.14)
monetary financial$_{t-1}$ [text]								0.28***
								(0.10)
Sample	1987Q3-2007Q4	1987Q3-2007Q4	1987Q3-2007Q4	2008Q1-2020Q4	2008Q1-2020Q4	2008Q1-2020Q4	2008Q1-2020Q4	2008Q1-2020Q4
Adj. R-squared	0.96	0.96	0.96	0.91	0.92	0.92	0.92	0.92
N	82	82	82	52	52	52	52	52

Notes: We estimate versions of the specification in equation (2). The dependent variable is the end-of-quarter federal funds rate (FFR) and, where indicated, the Wu and Xia (2016) shadow federal funds rate (Shadow FFR) from 2008Q1 onwards. Controls include the lagged values of the interest rate, as well as lagged values of core CPI inflation and for the output gap. All controls labeled [text] indicate that they are text features measured using zero shot classification or cosine similarity. The text feature "non-dual mandate [text]" is the classification score for whether a sequence is non-dual mandate related, the text feature "financial stability [text]" is the classification score for whether a sequence is financial stability related, and the text feature "monetary financial [text]" is the classification score for whether a sequence supports the view that monetary policy should be used to achieve financial stability. Each text feature is aggregated up to a quarterly frequency by computing the mean feature score over all paragraphs in the dataset within a given quarter. The properties of the series constructed are not highly sensitive to changes in the aggregation procedure. Robust standard errors are reported in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Financial Crisis and when the prevailing low level of inflation did not command a more narrow focus on the dual mandate.

To explore the role played by financial stability discussions, we next estimate equation (2), including the variables τ_t^{fs} and τ_t^{mf}; results are reported in columns (6)-(7). Interestingly, financial stability related discussion appears to play a prominent role. We find that, like non-dual mandate talk, financial stability discussion also enters the extended policy rule in an accommodative fashion (as seen by its negative coefficient), significantly so after controlling for advocacy for the use of monetary policy to achieve financial stability objectives (column (7)). Conversely, the text feature that measures whether monetary policy should be used to achieve financial stability is associated with less accommodative policy decisions, but the coefficient is only significant when it is included with a lag in column (8), which is consistent with a lingering "lean against the wind" rationale.

Taken together, these results on financial stability related considerations contrast with the result for the earlier pre-Global Financial Crisis sample period in column (3), where the financial stability text feature is insignificant.[38] Combined with the lower importance of the non-dual mandate related discussions in the earlier sample period in column (2), these findings point to the Federal Reserve's reliance on communication as an additional policy driver during the ZLB/low inflation subsample.

6 Conclusion

This paper evaluates important dimensions of Federal Reserve policy – namely the extent of financial stability considerations, the advocacy for using either monetary policy or banking regulation in pursuit of these objectives, and the implications for policy decisions and market participants – using state-of-the-art methods from natural language processing. In doing so, we attempt to determine whether these dimensions predict movements in policy decisions and financial markets. We approach this by constructing variables that capture the Fed's latent position about its own mandate. More specifically, we assemble the largest corpus of Fed speeches and apply a collection of large language models to extract a variety of paragraph-level text features. We also perform validation exercises and demonstrate that the models perform well on central bank texts.

We next partition the speech content into paragraphs that discuss the dual mandate and paragraphs that do not. We find that paragraphs that do not discuss the dual mandate usually discuss financial stability. This raises the question of whether Federal Reserve officials perceive financial stability to be a third component of its mandate. We find that they do not, even though they sometimes advocate for the use of monetary policy to achieve financial stability. Rather, they typically discuss financial stability as a means of achieving the objectives of the dual mandate or in the context of banking regulation.

We then examine whether the Fed's interpretation of its mandate predicts changes in monetary policy. Estimating an augmented Taylor rule with text features that measure the Fed's latent positions about its mandate, we find that a rise in speech content that is unrelated to the dual mandate is typically associated with more accommodative monetary policy. If we focus specifically on financial stability content, which accounts for much of the speech material that is unrelated to the dual mandate, this association increases in magnitude. In contrast, a rise in advocacy for monetary policy as a means of achieving financial stability is typically associated with tightening, which is consistent with an endorsement of a forward looking "leaning against the wind" view, rather than a "financial instability is caused by monetary tightening" view. We also show that the Fed's discussion of financial stability affects asset prices. In particular, we find a negative association with asset returns, even after controlling for both macroeconomic and financial variables, and discussion about financial crises.

Taken together, these results demonstrate how machine learning can be used in the context of policy evaluation. In the aftermath of the ZLB era, where central banks continue to contemplate the use of communication as an effective policy alternative, we demonstrate that natural language processing tools can be employed to shed light on the Federal Reserve's interpretation of its mandate, above and beyond more traditional approaches. Understanding these dynamics will help policymakers

[38] This also holds in (unreported) extended Taylor rule regressions for the earlier sample that include our measure of advocacy for the use of monetary policy to achieve financial stability or lagged text features.

to make better-informed decisions, formulate policies that are in line with the central bank's objectives, and communicate them effectively, thereby enhancing the consistency and transparency of central bank decision-making processes.

References

Apel, Mikael and Marianna Blix Grimaldi, "How Informative are Central Bank Minutes?," *Review of Economics*, 2014, *65*, 53–76.

—, —, and Isaiah Hull, "How Much Information Do Monetary Policy Committees Disclose? Evidence from the FOMC's Minutes and Transcripts," *Journal of Money, Credit and Banking*, 2022, *54* (5), 1459–1490.

Araci, Dogu, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," *arXiv*, 2019, *https://arxiv.org/abs/1908.10063.*

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in Yoshua Bengio and Yann LeCun, eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, "Measuring Economic Policy Uncertainty," *Quarterly Journal of Economics*, 2016, *131*, 1593–1636.

Bauer, Michael D., Carolin E. Pflueger, and Adi Sunderam, "Perceptions about Monetary Policy," *The Quarterly Journal of Economics (forthcoming)*, 2024.

Bernanke, Ben S. and Mark Gertler, "Should Central Banks Respond to Movements in Asset Prices?," *American Economic Review*, 2001, *91* (2), 253–257.

Bianchi, Francesco, "Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics," *The Review of Economic Studies*, 10 2012, *80* (2), 463–490.

—, Martin Lettau, and Sydney C Ludvigson, "Monetary Policy and Asset Valuation," *The Journal of Finance*, 2022, *77* (2), 967–1017.

BIS, *Issues in the Governance of Central Banks*, Basel, Switzerland: Bank for International Settlements, 2009.

Boivin, Jean, "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data," *Journal of Money, Credit and Banking*, 2006, *28* (5), 1149–1173.

Bowman, Samuel R., Gabor Angeli, Christopher Potts, and Christopher D. Manning, "A Large Annotated Corpus for Learning Natural Language Inference," in "Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics Lisbon, Portugal September 2015, pp. 632–642.

Buckmann, Marcus and Andreas Joseph, "An Interpretable Machine Learning Workflow with An Application to Economic Forecasting," *SSRN Electronic Journal*, 2022.

Carvalho, Carlos, Fernanda Nechio, and Tiago Tristão, "Taylor Rule Estimation by OLS," *Journal of Monetary Economics*, 2021, *124*, 140–154.

Cer, Daniel et al., "Universal Sentence Encoder for English," in "Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations" Association for Computational Linguistics Brussels, Belgium November 2018, pp. 169–174.

Cho, Kyunghyun et al., "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in "Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)" Association for Computational Linguistics Doha, Qatar October 2014, pp. 1724–1734.

Clarida, Richard, Jordi Gali, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, 2000, *115* (1), 147–180.

Coibion, Olivier and Yuriy Gorodnichenko, "Why Are Target Interest Rate Changes so Persistent?," *American Economic Journal: Macroeconomics*, 2012, *4* (4), 126–162.

Conneau, Alexis and Guillaume Lample, "Cross-lingual Language Model Pretraining," in Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, eds., *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019*, Vancouver, BC, Canada 2019, pp. 7057–7067.

_, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes, "Supervised Learning of Universal Sentence Representations from Natural Language Inference Data," in "Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics Copenhagen, Denmark September 2017, pp. 670–680.

Correa, R., K. Garud, J.M. Londono, and N. Mislang, "Sentiment in Central Banks' Financial Stability Reports," *Review of Finance*, 2021, *75* (1), 85–120.

Dai, Andrew M. and Quoc V. Le, "Semi-supervised Sequence Learning," in Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, eds., *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, 2015, pp. 3079–3087.

Dell'Ariccia, Giovanni, Luc Laeven, and Gustavo A. Suarez, "Bank Leverage and Monetary Policy's Risk-Taking Channel: Evidence from the United States: Bank Leverage and Monetary Policy's Risk-Taking Channel," *The Journal of Finance*, April 2017, *72* (2), 613–654.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Jill Burstein, Christy Doran, and Thamar Solorio, eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics 2019, pp. 4171–4186.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, September 2019, *57* (3), 535–574.

Hamilton, James D., Seth Pruitt, and Scott C. Borger, "Estimating the Market-Perceived Monetary Policy Rule," *American Economic Journal: Macroeconomics*, 2011, *3* (3), 1–28.

Hansen, Stephen and Michael McMahon, "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication," *Journal of International Economics*, March 2016, *99*, S114–S133.

Hill, Felix, Kyunghyun Cho, and Anna Korhonen, "Learning Distributed Representations of Sentences from Unlabelled Data," in "Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies" Association for Computational Linguistics San Diego, California June 2016, pp. 1367–1377.

Hochreiter, Sepp and Jürgen Schmidhuber, "Long Short-Term Memory," *Neural Computation*, November 1997, *9* (8), 1735–1780.

Howard, Jeremy and Sebastian Ruder, "Universal Language Model Fine-tuning for Text Classification," in "Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)" Association for Computational Linguistics Melbourne, Australia July 2018, pp. 328–339.

Howitt, Peter, "What Have Central Bankers Learned from Modern Macroeconomic Theory?," *Journal of Macroeconomics*, 2012, *34* (1), 11–22.

Hull, Isaiah, *Machine Learning for Economics and Finance in TensorFlow 2: Deep Learning Models for Research and Industry*, Berkeley, CA, USA: Apress, 2021.

Istrefi, Klodiana, Florens Odendahl, and Giulia Sestri, "FED Communication on Financial Stability Concerns and Monetary Policy Decisions: Revelations from Speeches," *Banco de España Documentos de Trabajo. N.º 2110*, 2021.

Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina, "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say About the Effects of Monetary Policy on Credit Risk-Taking?," *Econometrica*, 2014, *82* (2), 463–505.

Jordà, Òscar, Moritz Schularick, and Alan M. Taylor, *Macrofinancial History and the New Business Cycle Facts*, Chicago, IL, USA: University of Chicago Press, May 2016.

Kiros, Ryan, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler, "Skip-Thought Vectors," in "Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2" NIPS'15 MIT Press Cambridge, MA, USA 2015, pp. 3294–3302.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *CoRR*, 2019, *abs/1907.11692.*

Liu, Zhuang, Degen Huang, Kaiyu Huang, Zhuang Li, and Jun Zhao, "FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining," in Christian Bessiere, ed., *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, International Joint Conferences on Artificial Intelligence Organization 7 2020, pp. 4513–4519. Special Track on AI in FinTech.

Lo, Kyle, Lucy Lu Wang, Mark Neumann, Rodney Kinney, and Daniel Weld, "S2ORC: The Semantic Scholar Open Research Corpus," in "Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics" Association for Computational Linguistics Online July 2020, pp. 4969–4983.

Loughran, T. and B. McDonald, "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10Ks," *Journal of Finance*, 2011, *LXVI*, 35–36.

Luong, Thang, Hieu Pham, and Christopher D. Manning, "Effective Approaches to Attention-based Neural Machine Translation," in "Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics Lisbon, Portugal September 2015, pp. 1412–1421.

Mankiw, N. Gregory, "The Macroeconomist as Scientist and Engineer," *Journal of Economic Perspectives*, December 2006, *20* (4), 29–46.

McCann, Bryan, James Bradbury, Caiming Xiong, and Richard Socher, "Learned in Translation: Contextualized Word Vectors," in I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., *Advances in Neural Information Processing Systems*, Vol. 30 Curran Associates, Inc. 2017.

Meltzer, Allan H., *A History of the Federal Reserve*, Vol. Volume 1, 1951-1969 and Volume 2, 1970-1986, Chicago, IL, USA: University of Chicago Press, 2009.

Nikolsko-Rzhevskyy, Alex, David H. Papell, and Ruxandra Prodan, "Deviations from Rules-Based Policy and their Effects," *Journal of Economic Dynamics and Control*, December 2014, *49*, 4–17.

OpenAI, "GPT-4 Technical Report," *arXiv*, 2023, *https://arxiv.org/abs/2303.08774.*

Orphanides, Athanasios, "Monetary Policy Rules Based on Real-Time Data," *American Economic Review*, September 2001, *91* (4), 964–985.

—, "Historical Monetary Policy Analysis and the Taylor Rule," *Journal of Monetary Economics*, 2003, *50* (5), 983–1022.

Peek, Joe, Eric S. Rosengren, and Geoffrey M. B. Tootell, "Does Fed Policy Reveal a Ternary Mandate?," Working Papers 16-11, Federal Reserve Bank of Boston September 2016.

Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer, "Deep Contextualized Word Representations," in Marilyn A. Walker, Heng Ji, and Amanda Stent, eds., *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, Association for Computational Linguistics 2018, pp. 2227–2237.

Pushp, Pushpankar Kumar and Muktabh Mayank Srivastava, "Train Once, Test Anywhere: Zero-Shot Learning for Text Classification," *CoRR*, 2017, *abs/1712.05972.*

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever, "Improving Language Understanding with Unsupervised Learning," *Technical Report, Open AI*, 2018.

Rajpurkar, Pranav, Jian Zhang, Konstantin Lopyrev, and Percy Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," in "Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics Austin, Texas November 2016, pp. 2383–2392.

—, Robin Jia, and Percy Liang, "Know What You Don't Know: Unanswerable Questions for SQuAD," in Iryna Gurevych and Yusuke Miyao, eds., *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers*, Association for Computational Linguistics 2018, pp. 784–789.

Reimers, Nils and Iryna Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, eds., *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, Association for Computational Linguistics 2019, pp. 3980–3990.

Sachs, Jeffrey D., "International Policy Coordination: The Case of the Developing Country Debt Crisis," *Working Paper No. 2287*, June 1987.

Schroff, Florian, Dmitry Kalenichenko, and James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *CoRR*, 2015, *abs/1503.03832.*

Schularick, Moritz, Lucas ter Steege, and Felix Ward, "Leaning against the Wind and Crisis Risk," *American Economic Review: Insights*, 2021, *3* (2), 199–214.

Sundararajan, Mukund, Kedar Dhamdhere, and Ashish Agarwal, "The Shapley Taylor Interaction Index," in Hal Daumé III and Aarti Singh, eds., *Proceedings of the 37th International Conference on Machine Learning*, Vol. 119 of Proceedings of Machine Learning Research PMLR 13–18 Jul 2020, pp. 9259–9268.

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le, "Sequence to Sequence Learning with Neural Networks," in "Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2" NIPS'14 MIT Press Cambridge, MA, USA 2014, pp. 3104–3112.

Taylor, John B., "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 1993, *39*, 195–214.

__ , "A Historical Analysis of Monetary Policy Rules," in "Monetary Policy Rules" NBER Chapters, National Bureau of Economic Research, Inc, May 1999, pp. 319–348.

van der Maaten, Laurens and Geoffrey Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, 2008, *9* (86), 2579–2605.

van Dieijen, Myrthe and Robin L. Lumsdaine, "What Say They About Their Mandate? A Textual Assessment of Federal Reserve Speeches," 2019. Mimeo, Available at SSRN: www.ssrn.com/abstract_id=3455356.

Vaswani, Ashish et al., "Attention is All you Need," in Isabelle Guyon et al., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 2017, pp. 5998–6008.

Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *J. Mach. Learn. Res.*, December 2010, *11*, 3371–3408.

Vollmer, Uwe, "Monetary Policy or Macroprudential Policies: What Can Tame the Cycles?," *Journal of Economic Surveys*, December 2022, *36*, 1510–1538.

Wang, Alex et al., "GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding," in "7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019" OpenReview.net 2019.

Wang, Kexin, Nils Reimers, and Iryna Gurevych, "TSDAE: Using Transformer-based Sequential Denoising Auto-Encoder for Unsupervised Sentence Embedding Learning," *CoRR*, 2021, *abs/2104.06979.*

Williams, Adina, Nikita Nangia, and Samuel Bowman, "A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference," in "Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)" Association for Computational Linguistics New Orleans, Louisiana June 2018, pp. 1112–1122.

Wu, Jing Cynthia and Fan Dora Xia, "Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound," *Journal of Money, Credit, and Banking*, 2016, *48* (2-3), 253–291.

Wu, Yonghui et al., "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation," *CoRR*, 2016, *abs/1609.08144.*

Yang, Zhilin, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," in H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds., *Advances in Neural Information Processing Systems*, Vol. 32 Curran Associates, Inc. 2019.

Yin, Wenpeng, Jamaal Hay, and Dan Roth, "Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach," in Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, eds., *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, Association for Computational Linguistics 2019, pp. 3912–3921.

Yogatama, Dani, Chris Dyer, Wang Ling, and Phil Blunsom, "Generative and Discriminative Text Classification with Recurrent Neural Networks," *CoRR*, 2017, *abs/1703.01898.*

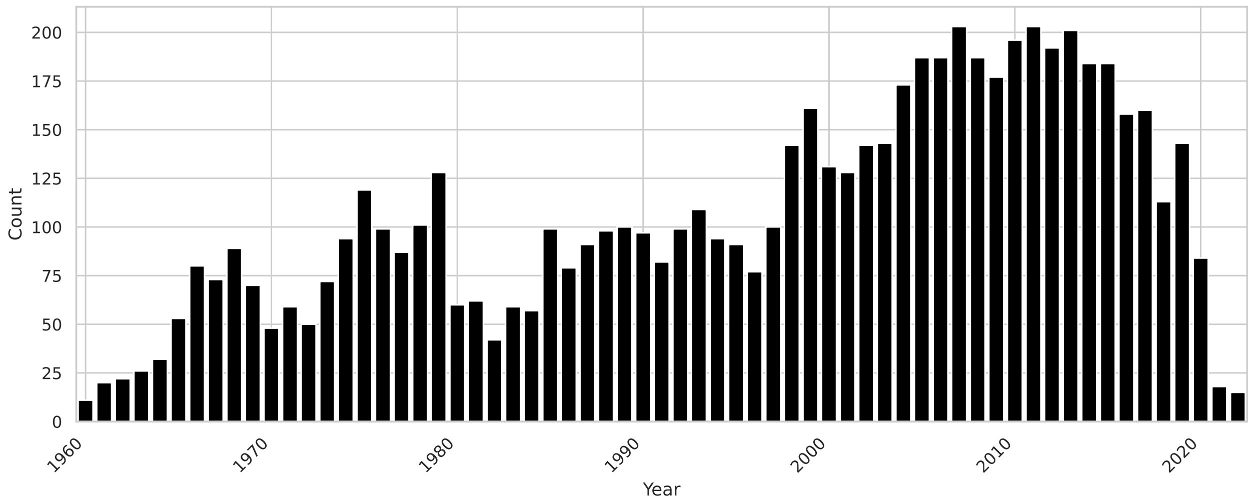
Zhang, Jingqing, Piyawat Lertvittayakumjorn, and Yike Guo, "Integrating Semantic Knowledge to Tackle Zero-shot Text Classification," in Jill Burstein, Christy Doran, and Thamar Solorio, eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Association for Computational Linguistics 2019, pp. 1031–1040.

Zhu, Yukun et al., "Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books," in "2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015" IEEE Computer Society 2015, pp. 19–27.

A Appendix

A.1 Additional Figures and Tables Describing the Data

Figure A1: Federal Reserve Speech Count



Notes: The figure above plots the annual count of Federal Reserve speeches in the corpus. While the corpus spans the period between 1914 and 2022, both coverage and speech frequency increase considerably in the 1960s and again in the 1990s. Additionally, the completeness of coverage declines near the end of the sample, especially in the last two years.

Table A1: Speech Count by Institution

Institution	**Count**
FRB	2676
ATL	605
STL	472
SF	416
CHI	370
PHL	367
MIN	343
CLE	329
NY	305
DAL	219
RIC	200
BOS	183
KC	154

Notes: This table provides the speech count by institution. Each entity in the table is either a Federal Reserve Bank or the Federal Reserve Board (FRB).

Table A2: Most Common Journals in Sentence Pair Corpus

Journal	**Article Count**
SSRN Electronic Journal	4,164
Journal of Banking and Finance	1,388
Journal of Money, Credit and Banking	618
IMF Working Papers	590
Journal of Finance	585
National Bureau of Economic Research	427
Applied Economics	369
Econometric Reviews	357
Economic Modelling	356
Journal of International Money and Finance	348
Other journals	20,579
Total number of articles	29,781

Notes: This table provides article counts for the 10 most common journals and working paper series in the corpus we construct to train our NLP models. In total, the final corpus includes 283 journals and 29,781 articles. It is a subset of the 2.3M economics articles in the S2ORC corpus (Lo et al., 2020), selected according to the criteria described in Section 2.2.

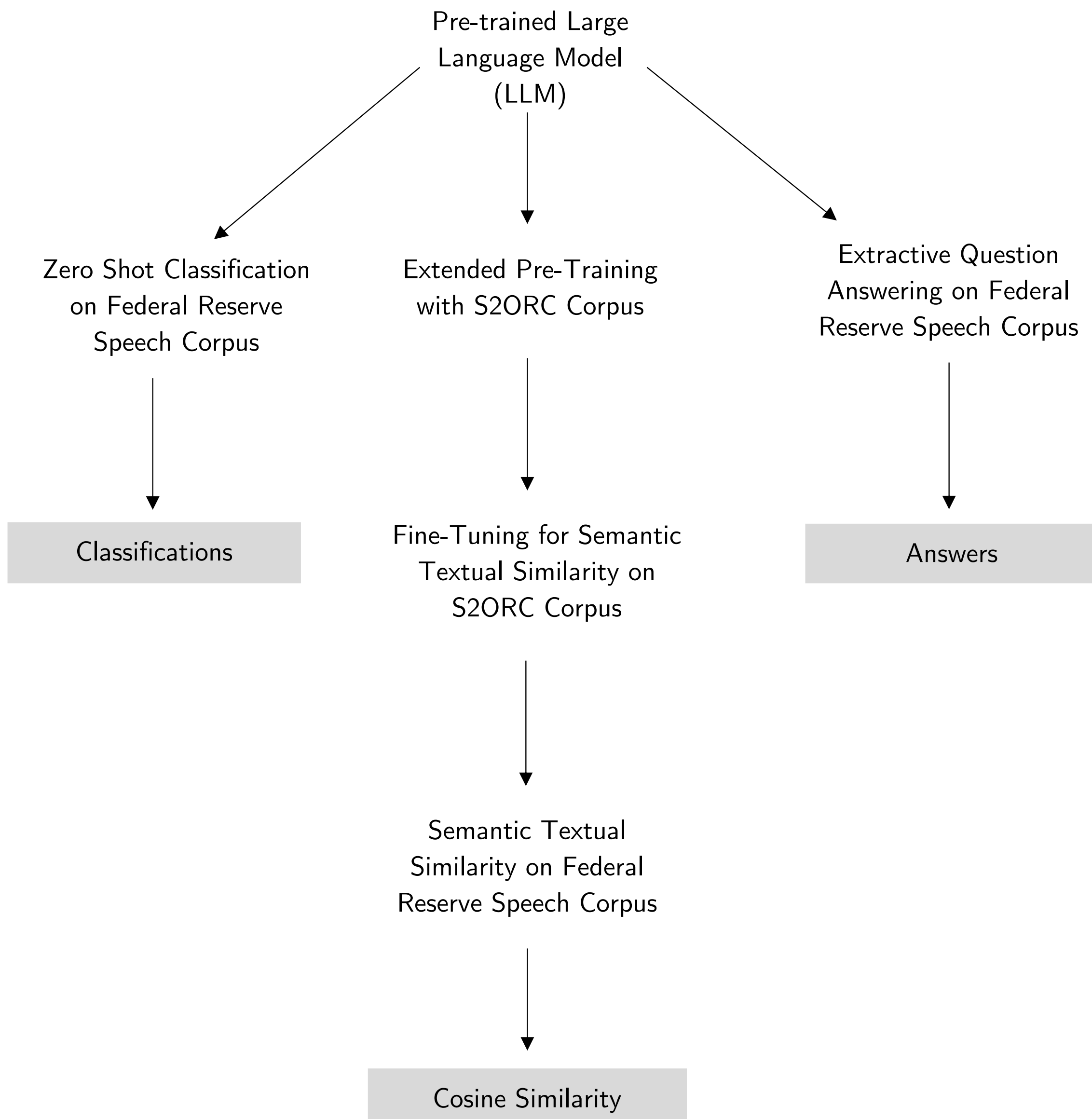
Table A3: Macroeconomic and Financial Series

Variable	**Description**	**Source**
bond return	total return on government bonds	Jordà et al.
cpi inflation	consumer price index (1990 = 100)	Jordà et al.
core cpi inflation	includes all items less food and energy (1982-84=100)	CBO
crisis	dummy for financial crisis	Jordà et al.
debt-to-gdp ratio	public debt-to-GDP ratio	Jordà et al.
equity return	total return to equity	Jordà et al.
house prices	nominal house prices	Jordà et al.
interest rate	short term interest rate	Jordà et al.
ltd ratio	loan-to-deposit ratio	Jordà et al.
output gap	percentage difference between actual and potential (real) GDP	CBO, BEA, FRED
risky return	total return on risky assets	Jordà et al.
safe return	total return on safe assets	Jordà et al.
total loans	total loans originated to non-financial sector	Jordà et al.

Notes: The output gap variable is taken from the St. Louis Fed's FRED database. It is computed as the percentage difference between actual and potential (real) GDP. The underlying measures of actual and potential GDP are computed by the Bureau of Economic Analysis (BEA) and Congressional Budget Office (CBO), respectively. The remaining macroeconomic and financial variables are taken from the Macrohistory Database (Jordà et al., 2016).

A.2 Overview of the Training Process and Models

Figure A2: LLM Training Summary



Notes : This figure provides an overview of the training process for the LLM models used in this paper. All models start with the BERT or RoBERTa model, pre-trained on the Toronto BookCorpus (800 million words) and the English language Wikipedia (2500 million words). In exercises that involve zero shot classification or extractive question answering, the models can be directly applied to the Fed speech corpus. In exercises that make use of cosine similarities produced by sentence transformers, we extend the pre-training process and fine-tune the model on the semantic textual similarity task using the S2ORC corpus before applying the models to the Fed speech corpus. In some cases, we combine multiple approaches and models. For example, in some exercises, we use zero shot classification to identify paragraphs that contain content about financial stability. We then compute the cosine similarity between those paragraphs and other natural language statements (e.g., monetary policy).

Table A4: Pre-Trained Model List

Model	Task
`distilbert-base-uncased-distilled-squad`	Extractive Question Answering
`distilbert-base-uncased-mnli`	Zero Shot Classification
`stsb-distilroberta-base-v2`	Sentence Embedding Generation
`t5-xxl`	Sentence Embedding Generation

Notes: The table above lists the pre-trained base models used in this paper, along with the linguistic tasks for which they were employed. All models are available via HuggingFace.

A.3 Additional Details on the NLP Methods Used

A.3.1 Sequence-to-Sequence Modeling

Sequence-to-sequence (S2S) modeling poses challenges that are not present in other NLP tasks. Among other problems, dense neural networks require fixed-length inputs and outputs and, thus, cannot be used for S2S tasks, such as machine translation, which requires variation in input and output length. This is true even for a given translated sentence, which may be best expressed using a different number of words in two languages. Consequently, models that are suitable for text classification purposes, such as dense neural networks, are not usable for S2S tasks.

Early breakthroughs in S2S modeling centered around the use of a variant of recurrent neural networks (RNN) called a long-short term memory (LSTM) model (Hochreiter and Schmidhuber, 1997).[39] LSTMs explicitly treat input data, such as words in a sentence, as a sequence, rather than as a set of features.[40] This allows for a more parsimonious parameterization than a dense neural network would permit. LSTMs are also able to handle variable length input sequences, which provides an advantage over dense neural networks for S2S modeling.

The initial innovation in S2S modeling with LSTMs involved the use of an encoder-decoder architecture (Sutskever et al., 2014; Cho et al., 2014). The encoder maps a sequence of symbols to a latent vector, which can be viewed as a compressed representation of the input text. The decoder then maps the latent vector to a sequence of symbol predictions. Since the output sequence must also be permitted to have a variable length in many applications, the model is trained to output an end-of-sentence (`[EOS]`) token, which terminates the sequence of predictions.

LSTM-based models with an encoder-decoder architecture provided the initial means of performing high-quality machine translation and also generated many spillover benefits for related NLP tasks. However, the introduction of the attention mechanism (Bahdanau et al., 2015) and later transformer models (Vaswani et al., 2017) fundamentally changed how sequences are modeled in S2S contexts and in NLP more broadly. This provided a foundation for the set of NLP tools we use in this paper to measure the content of central bank communications.

[39] See Apel et al. (2022) for an application of LSTM models to natural language processing tasks in economics.

[40] LSTMs also allow for long-term dependence between words in a sequence and correct a version of the vanishing gradient problem. See Hull (2021) for an overview of LSTMs in the context of economics.

A.3.2 The Transformer Model

In this subsection, we provide a detailed overview of transformer models and their advantages over earlier S2S modeling techniques.

Attention in Transformer Models. Transformer models apply three distinct forms of scaled dot product attention: 1) encoder-decoder attention; 2) encoder self-attention; and 3) decoder self-attention. Encoder-decoder attention uses the query vectors, $\mathcal{Q}$, from the previous decoder layer and key and value vectors, $\mathcal{K}$ and $\mathcal{V}$, from the current encoder layer.

Self-attention, in contrast, is applied to words in the same sequence and in the same encoder or decoder layer. We make a distinction between encoder and decoder self-attention because decoder self-attention only uses the sequence of words up to and including the word being predicted; whereas encoder self-attention uses the entire sequence. Figure A3 illustrates the self-attention mechanism from the RoBERTa model applied to word sequences in the Federal Reserve speeches in our corpus.[41]

The Attention Mechanism. Bahdanau et al. (2015) argued that the latent vector in LSTM-based encoder-decoder models created a bottleneck that made it difficult to improve model architecture and training. As a solution to this problem, they proposed using the attention mechanism, which eliminated the need to encode the entire input sequence in a single latent vector. For a given symbol, such as a word, the attention mechanism determines which symbols are related to it without explicitly considering the temporal ordering of the sequence. This allows for symbols that are not close together to be closely related. Luong et al. (2015) demonstrated how this could be used effectively on machine translation tasks. Figure A3 illustrates the application of the attention mechanism to sequences taken from two Federal Reserve speeches.

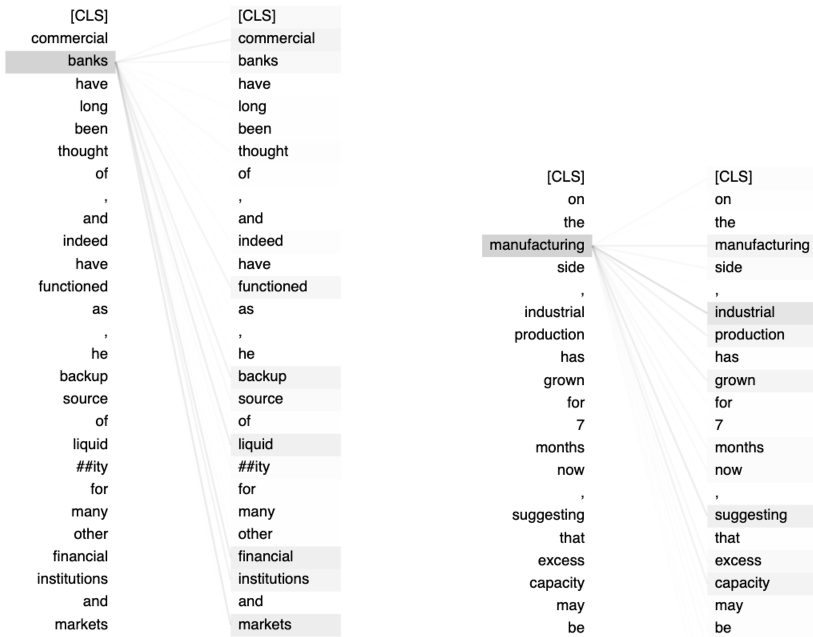
In practice, attention proved to be an invaluable construct for NLP. We will briefly describe the attention mechanism below in an LSTM encoder-decoder context, focusing on the scaled dot product variant, which was later used in transformer models. For concreteness, consider the passage given in the quote below, taken from a speech given by Gary Stern, then-President of the Minneapolis Federal Reserve Bank, in January 2009. We will first convert the sequence of words to a sequence of embedding vectors, as shown in Equation (3).

> *"Commercial banks have long been thought of, and indeed have functioned as, the backup source of liquidity for many other financial institutions and markets. Banks continue to play this role, but it has become more challenging today to do so because some lenders find themselves capital-constrained as a result of recent losses and/or sizable, unanticipated additions to their balance sheets of formerly off-balance-sheet instruments."*

In Equation (3) below, M is the embedding dimension and T is the sequence length. The mapping between words and embeddings is constructed outside of the model, typically through a separate

[41] The RoBERTa model is a "robustly optimized" version of the BERT model, which is described in detail below.

Figure A3: Attention Mechanism



Notes: The left panel illustrates the attention mechanism applied to the word "banks" in a sequence from a speech given by Gary Stern in January 2009. The words "liquid(ity)," "financial institutions," and "markets" are not close to "banks" in the sequence, but the attention mechanism determines that they are highly relevant for contextualizing banks. The right panel illustrates self-attention applied to the word "manufacturing" in a sequence taken from a speech given in September 2002 by Cathy Minehan, then-President of the Federal Reserve Bank of Boston. The attention mechanism identifies the importance of "industrial," "production," and "capacity" for contextualizing "manufacturing" in the sequence. Note that the `[CLS]` "word" indicates the start of a sequence and `##` indicates that a word is outside of the corpus.

training process.[42] In a minimal LSTM model with a single cell, the sequence of embeddings is processed in order with each step yielding a hidden state, which is then combined with the next input embedding in the sequence. Equation (4) provides the sequence of hidden states, which have the same dimension as the embedding vector in this architecture.

$$\left\{e_1, e_2, ..., e_t, ..., e_T\right\} = \left\{\begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1M} \end{bmatrix} \begin{bmatrix} e_{21} \\ e_{22} \\ \vdots \\ e_{2M} \end{bmatrix} \cdots \begin{bmatrix} e_{t1} \\ e_{t2} \\ \vdots \\ e_{tM} \end{bmatrix} \cdots \begin{bmatrix} e_{T1} \\ e_{T2} \\ \vdots \\ e_{TM} \end{bmatrix}\right\} \tag{3}$$

$$\left\{h_1, h_2, ..., h_t, ..., h_T\right\} = \left\{\begin{bmatrix} h_{11} \\ h_{12} \\ \vdots \\ h_{1M} \end{bmatrix} \begin{bmatrix} h_{21} \\ h_{22} \\ \vdots \\ h_{2M} \end{bmatrix} \cdots \begin{bmatrix} h_{t1} \\ h_{t2} \\ \vdots \\ h_{tM} \end{bmatrix} \cdots \begin{bmatrix} h_{T1} \\ h_{T2} \\ \vdots \\ h_{TM} \end{bmatrix}\right\} \tag{4}$$

[42] For some models, such as BERT – which we use in most NLP exercises – we will use embeddings based on sub-word units, such as the WordPiece embeddings (Wu et al., 2016). This allows for the use of full words, individual characters, and multi-character strings. The word "growing," for example, can be split into the word "grow" and the multi-character string "ing."

The relationship between the contemporaneous hidden state, h_t, the previous hidden state, h_{t-1}, and the input embedding, e_t, is given by Equation (5). Note that $\mathcal{W}^E$ and $\mathcal{W}^H$ are shape-preserving linear transformations of e_t and h_{t-1} and $\mathcal{G}$ is an elementwise nonlinear activation function.

$$h_t = \mathcal{G}\left(\mathcal{W}^E e_t + \mathcal{W}^H h_{t-1}\right) \tag{5}$$

In a standard encoder-decoder architecture, the terminal state, which we denote h_T, is a fixed-length vector that encodes a summary of the entire sequence. Attention modifies the standard LSTM construction by retaining the hidden states and scoring them. It then applies the softmax function (defined below) to the hidden states, and then multiplies the softmaxed states by the original (un-transformed) states. The procedure introduces three additional sets of trainable weights: $\mathcal{W}^Q$, $\mathcal{W}^K$, and $\mathcal{W}^V$.

For each hidden state h_t, there are associated query (q_t), key (k_t), and value (v_t) vectors, which are computed as $\mathcal{W}^i h_t$, where $i \in \{Q, K, V\}$. Stacking those row vectors into matrices $\mathcal{Q}$, $\mathcal{K}$, and $\mathcal{V}$, we can define scaled dot product attention for h_t in Equation (6), where $D = dim(k_t)$.

$$\mathcal{Z} = \sigma\left(\frac{\mathcal{Q}\mathcal{K}^T}{\sqrt{D}}\right)\mathcal{V} \tag{6}$$

Note that σ is a rowwise softmax function, defined in Equation (7), where X_t is row t in $\mathcal{Q}\mathcal{K}^T/\sqrt{D}$.

$$\sigma(X_{t\hat{d}}) = exp(X_{t\hat{d}}) / \sum_{d \in D} exp(X_{td}) \tag{7}$$

The elementwise product of the row vectors $\mathcal{Q}_t$ and $\mathcal{K}_s$ measures the extent to which they are related to or *attend to* each other. Dividing by $\sqrt{D}$ improves computational performance, but is not strictly necessary. Applying the rowwise softmax function σ amplifies the strength of strong associations and also normalizes the sum of the attention weights to be equal to one.

The weights are then multiplied by $\mathcal{V}$, which is a matrix of linear transformations of the hidden vectors. Each row of the vector $\mathcal{Z}_t$ is a weighted sum of the embedding vectors, where each weight depends on the extent to which a given hidden vector attends to another. Since each hidden vector is most closely associated with a specific embedding in the sequence (i.e., h_t is closest to e_t), attention provides us with a contextualized embedding for each input word. That is, rather than using a fixed embedding, we incorporate the context of other words that are most closely related to it in a sentence.

To be consistent with the subsection that follows on transformer models, we have provided a description of how to compute attention for all hidden vectors. However, for most LSTM-based encoder-decoder models, we will exclusively use $\mathcal{Z}_T$, which is the row vector associated with the final hidden state, h_T. $\mathcal{Z}_T$ is sometimes called the *context* vector and is concatenated with h_T and passed to the decoder. Similar to the earlier variant of S2S models, such as Bahdanau et al. (2015) and Luong et al. (2015), the decoder takes the output of the encoder as an input and then sequentially predicts symbols until it terminates with an (`[END]`) token.

Multi-headed Attention. One innovation of transformer models is to make use of multi-headed attention, which is enabled by the removal of sequential elements from the model. This amounts to an h-way partition of the query, key, and value matrices, such that $\mathcal{W}^j = \{\mathcal{W}_1^j, ..., \mathcal{W}_h^j\}$ for $j \in \{Q, K, V\}$, where $\mathcal{W}_i^j \in \mathbb{R}^{DxD/h}$. For the original transformer model, $M = 512$, $D = 64$, which yields a model with $h = 8$ attention heads. The model then attends to each subspace separately and in parallel, yielding a computational time that is similar to that of a single-headed model.

Positional Encodings. Rather than using sequential elements like recurrence or convolution, transformers modify input embeddings by encoding positional information. For each input embedding in a sequence, positional encodings are generated using Equation (8).

$$p(t,m) \quad = \quad \begin{cases} sin\left(\frac{t}{10000^{2m/M}}\right) & \text{if } m \text{ even} \\ cos\left(\frac{t}{10000^{2m/M}}\right) & \text{if } m \text{ odd} \end{cases} \tag{8}$$

The positional encodings are then added to the input embeddings to create positional embeddings, which contain both information about word features and position within the embedding and sequence. The positional embedding is given in Equation (9) and is the input to the model.

$$\left\{\tilde{e}_1, ..., \tilde{e}_T\right\} \quad = \quad \left\{ \begin{bmatrix} e_{11} + p(1,1) \\ e_{12} + p(1,2) \\ \vdots \\ e_{1M} + p(1,M) \end{bmatrix} \quad ... \quad \begin{bmatrix} e_{T1} + p(T,1) \\ e_{T2} + p(T,2) \\ \vdots \\ e_{TM} + p(T,M) \end{bmatrix} \right\} \tag{9}$$

Time Complexity. One advantage of the transformer models is that they reduce the computational complexity of certain components of the training process relative to LSTM-based S2S models. In particular, the attention mechanism has a time complexity of $\mathcal{O}(T^2D)$; whereas recurrent operations, such as LSTM cells, have a time complexity of $\mathcal{O}(TD^2)$, where T is the length of the sequence and D is dimension of the key, query, and value vectors. This suggests that transformer models will tend to have a training time advantage when embeddings are large.

Another important advantage of transformer models is that they require $\mathcal{O}(1)$ sequential operations; whereas LSTM-based S2S models require $\mathcal{O}(T)$. This implies a substantial training time advantage for transformer models, since they can parallelize operations that must be performed in sequence for LSTM-based S2S models.

A.3.3 Sentence BERT Models

In this Appendix section, we discuss additional details related to sentence BERT models used in this paper. The models produce sentence embeddings, which can be used to the measure semantic textual similarity (STS) between a pair of passages. We start by defining STS and then discuss the pre-training and fine-tuning process for the sentence BERT models used in this paper.

Semantic Textual Similarity. An alternative to using BERT directly to compute STS is to construct sentence embeddings, which we can compute individually for each passage, and then measure the cosine similarity between pairs of embeddings. This provides a more computationally efficient means of performing this comparison. See Equation (10) for the construction of cosine similarity and note that S is a sentence embedding.

$$sim(S_i, S_j) = \frac{S_i \cdot S_j}{\|S_i\| \|S_j\|} \tag{10}$$

Sentence embeddings have been explored in Kiros et al. (2015), Conneau et al. (2017), and Cer et al. (2018). We will make use of the approach in Reimers and Gurevych (2019), which modifies the pre-trained BERT and RoBERTa models to produce contextualized sentence embeddings, rather than contextualized word embeddings. This approach uses Siamese and triplet networks (Schroff et al., 2015) to train the model, which have objective functions that are comparable to cosine similarity. It is trained using the SNLI (Bowman et al., 2015) and NLI (Williams et al., 2018) datasets.

SBERT Pre-training. The sentence BERT (SBERT) models employed in the paper were pre-trained first by Devlin et al. (2019) and then Reimers and Gurevych (2019). The specific variant of the model we used was `stsb-distilroberta-base-v2`, which has 67 million parameters, and was trained to generate a cosine embedding loss that matches the values generated by the `RoBERTa` model. The `RoBERTa` model was pre-trained on eight 16GB V100 GPUs for 90 hours. See Reimers and Gurevych (2019) for the training information for sentence transformer models. See model repositories, such as `https://huggingface.co/`, for the pre-trained versions of models.

We refine the pre-training using an unsupervised learning process called a Transformer-based Sequential Denoising Auto-Encoder (TSDAE). Specifically, we use the training process in Wang et al. (2021), along with the data described in Section 2, which is compiled from the S2ORC corpus (Lo et al., 2020). The TSDAE approach to training sentence embeddings was based on earlier work by Vincent et al. (2010) and Hill et al. (2016). The training task entails injecting noise into the input embeddings and then training the model to recover the denoised embeddings. Much like the masked language modeling (MLM) and next sentence prediction (NSP) tasks for BERT, TSDAE does not require labels, making it an attractive choice for refining the pre-trained model on domain-specific text. It also achieves state-of-the-art performance, which approaches that of models trained with supervised methods on domain-specific texts.

We extend the pre-training of the model for one epoch at a learning rate of 1e-5 and with a batch size of 128 on an A100 GPU. We checkpoint every 25 steps and retain the best version of the model.

SBERT Fine-tuning. In addition to pre-training the SBERT model, we also fine-tune it on the semantic textual similarity task using 194,227 pairs of sentences drawn from paper abstracts in the S2ORC corpus (Lo et al., 2020). See Section 2 for an overview of the construction of the dataset. We fine-tune the model for one epoch at a learning rate of 1e-5 on an A100 GPU with a cosine similarity loss. We use a batch size of 64 and checkpoint every 50 steps, and retain the best version of the model.

Table A5 shows validation set performance for the fine-tuning task for three sentence transformer models. The first is `stsb-distilroberta-v2`, which is the base model that we perform extended pre-training and fine-tuning on for this paper to produce `distil-roberta-v2-central-bank`. We also consider performance on another state of the art sentence transformer, `t5-xxl`. The validation task involves producing cosine similarity scores for a subset of the S2ORC corpus that is not in the training set for the fine-tuning task. The spearman correlation and point-biserial correlation scores are computed between the cosine similarity scores and a dummy variable for whether the sentences are in the same abstract. The version of the model we construct in this paper, `distil-roberta-v2-central-bank`, achieves substantially improved performance on the validation task, as measured by higher correlations, than the other two models considered.

Table A5: Validation: Performance on Fine-Tuning Task

Model	**Spearman Corr.**	**PBS Corr.**
`stsb-distilroberta-v2`	0.5337	0.5310
`t5-xxl`	0.5680	0.5613
`distil-roberta-v2-central-bank`	0.7534	0.7471

Notes: The table above shows validation set performance for three sentence transformer models. The first is `stsb-distilroberta-v2`, which is the base model that we perform extended pre-training and fine-tuning on for this paper to produce `distil-roberta-v2-central-bank`. We also consider performance on another state of the art sentence transformer, `t5-xxl`. The validation task involves producing cosine similarity scores for a subset of the S2ORC corpus that is not in the training set for the fine-tuning task. The spearman correlation and point-biserial (PBS) correlation scores are computed between the cosine similarity scores and a dummy variable for whether the sentences are in the same abstract.

A.4 Examples Applying NLP Methods to the Federal Reserve Corpus

For concreteness, we demonstrate how the BERT model would use the passage given in the quote below in the training process for the masked language modeling (MLM) and next sentence prediction (NSP) tasks. The text is taken from a speech given by Gary Stern, then-President of the Minneapolis Federal Reserve Bank, in January 2009.

> *"Commercial banks have long been thought of, and indeed have functioned as, the backup source of liquidity for many other financial institutions and markets. Banks continue to play this role, but it has become more challenging today to do so because some lenders find themselves capital-constrained as a result of recent losses and/or sizable, unanticipated additions to their balance sheets of formerly off-balance-sheet instruments."*

Recall that the MLM task involves masking randomly-selected words in a sequence and then training the model to predict them. For example, in the quote above, MLM might generate the following sequence and labels.

Sequence: "Commercial $\texttt{[MASK]}_1$ have long been thought of, and indeed have functioned as, the backup source of $\texttt{[MASK]}_2$ for many other financial institutions and markets."
Labels: $\texttt{[MASK]}_1$ = banks, $\texttt{[MASK]}_2$ = liquidity.

The other training task, next sentence prediction (NSP), presents the model with a sequence of two sentences drawn from the corpus. The model must determine whether the second sentence follows the first or whether it is drawn from a different place in the document. Again, returning to the speech from Gary Stern, consider the following three sentences.

> **Sentence A:** "Commercial banks have long been thought of, and indeed have functioned as, the backup source of liquidity for many other financial institutions and markets."
> **Sentence B:** "Banks continue to play this role, but it has become more challenging today to do so because some lenders find themselves capital-constrained as a result of recent losses and/or sizable, unanticipated additions to their balance sheets of formerly off-balance-sheet instruments."
> **Sentence C:** "On the positive side, term funding is more readily available than at the height of the crisis, and risk premia have diminished through much of the financial sector."

In the speech, Sentence B follows Sentence A; whereas Sentence C is selected from a random location in the text. The next sentence prediction (NSP) task might yield sequence (A,B), which could be passed to BERT with the training label `IsNextSentence`. Alternatively, it could yield the sequence (A,C), which could be passed to BERT with the training label `IsNotNextSentence`.

We next provide an example of zero shot classification with the aforementioned model. We attempt to classify whether the statement, which is again taken from Gary Stern's January 2009 speech, is about one of the following four categories of topics: financial stability, output, inflation, or the labor market. The scores can be interpreted as a probability distribution over the categories.

> **Sequence:** "Banks continue to play this role but it has become more challenging today to do so because some lenders find themselves capital constrained as a result of recent losses and or sizable unanticipated additions to their balance sheets of formerly off balance sheet instruments."
> **Candidate Labels:** ['financial stability', 'output', 'inflation', 'labor market']
> **Scores:** [0.718, 0.203, 0.048, 0.031]

The model is able to identify "financial stability" as the most probable label, even though it was not specifically trained on a central bank communication corpus.

Finally, we use extractive question answering with pre-trained BERT models in several exercises in this paper. For example, in some exercises we attempt to determine the speaker's most significant concern in a passage, as shown in the next two examples. Below we specify the query, context, and model output for passages in a speech given by then-President of the Federal Reserve Bank of St. Louis, Darryl Francis, in February 1972.

> **Query 1:** What is the most significant concern in the passage?
> **Context 1:** "The suspension of the convertibility of the dollar into gold and the imposition of a 10 percent import surcharge last summer ran the risk of mass foreign retaliation in the form of destructive trade barriers."
> **Output 1:** mass foreign retaliation

> **Query 2:** What is the most significant concern in the passage?
> **Context 2:** "Another significant aspect of the President's new policies announced August 15 are the measures taken to reverse the deteriorating US balance of payments."
> **Output 2:** deteriorating US balance of payments

A.5 Additional Information on the Text Features

Table A6: Text Feature Descriptive Statistics (N=363,154)

	Financial Stability	**Inflation**	**Employment**	**Financial Crisis**	**Bank Liquidity**	**Bank Capital**	**Past Focus**	**Present Focus**	**Future Focus**	**Academic Focus**	**Cos. Sim. MP & FS**	**Cos. Sim. BR & FS**
Min	−1.239	−0.980	−1.226	−1.638	−0.450	−0.702	−2.936	−2.249	−1.606	−2.582	−3.414	−2.642
25%	−0.815	−0.654	−0.820	−0.677	−0.449	−0.601	−0.407	−0.465	−0.766	−0.495	−0.730	−0.715
50%	−0.077	−0.264	−0.103	−0.168	−0.440	−0.334	0.419	0.512	−0.207	−0.050	0.040	−0.145
75%	0.471	0.152	0.478	0.545	−0.166	0.069	0.772	0.756	0.797	0.485	0.743	0.577
Max	2.516	2.684	2.538	2.064	3.590	3.450	0.794	0.786	1.669	2.386	3.008	4.297

Notes: The table above provides the descriptive statistics for the text features used in regression exercises. The feature values are computed at the paragraph level. We then standardize the features. The descriptive statistics reported are for the standardized features included in the regressions. Each feature has a mean of approximately 0 and a standard deviation of approximately 1. We report the minimum, maximum, and quartiles.

Table A7: Mean Text Feature Values by Federal Reserve District

District	**Past Focus**	**Present Focus**	**Future Focus**	**Financial Stability**	**Academic Focus**
ATL	−0.019	−0.015	0.072	−0.042	−0.049
BOS	0.048	−0.024	−0.023	−0.048	−0.029
CHI	−0.005	0.026	0.016	−0.055	−0.053
CLE	−0.058	0.022	−0.019	0.015	−0.052
DAL	0.073	−0.086	−0.016	−0.101	0.004
FRB	0.003	0.016	−0.019	0.024	0.021
KC	−0.004	0.041	0.097	0.093	0.017
MIN	−0.080	−0.102	−0.049	−0.036	−0.035
NY	0.085	0.168	0.208	0.162	0.202
PHL	−0.060	0.000	0.024	−0.030	−0.099
RIC	−0.003	−0.073	0.171	0.249	0.106
SF	0.071	−0.034	−0.012	−0.117	−0.020
STL	−0.041	−0.117	−0.131	−0.028	−0.036

Notes: The table above provides the mean values of selected text features for each Federal Reserve district bank over the entire sample period. All text features are measured at the paragraph level. The first column lists the Federal Reserve district. The next three columns provide measures of tense. Note that tense usage is not mutually exclusive and not all passages have a clear focus on a single tense. As such, it is possible for a given district bank to have positive or negative scores for all three tenses. The remaining two columns provide mean values for features that 1) indicate whether a paragraph is about financial stability; and 2) indicate whether a paragraph references academic work or an academic discussion.

Table A8: Text Feature Examples: Financial Crisis

Date	Speaker	Passage
December 4, 2009	Charles Plosser	Unfortunately, rather than limiting moral hazard and the too-big-to-fail problem, we have made them worse during the crisis. In trying to stabilize the financial system, we have led creditors of large financial institutions to expect that the government will protect them from losses, which in turn means they need not monitor risk-taking by these firms.
March 4, 2013	Jerome Powell	As I said earlier, reforms to end too big to fail must wage the fight on two fronts. First, we need enhanced regulation to make large financial institution failures much less likely. Second, we need a credible mechanism to manage the failure of even the largest firms, without causing or amplifying a systemic crisis.
March 29, 2012	Jeffrey Lacker	We should take a very rigorous approach to the Dodd-Frank provisions requiring credible resolution plans for large financial firms. To improve the credibility of a commitment to greater market discipline, we should further restrict the means available to use public funds to rescue private creditors.
June 16, 2010	Charles Plosser	During the World Financial Crisis, several governments bailed out ailing financial firms through fiscal transfers and other mechanisms because they feared that these firms were too large or too systemic to fail without catastrophic costs. Many of our recommendations are intended to create a robust financial system in which any troubled financial company ...
March 18, 2010	Daniel Tarullo	Entrenching too-big-to-fail status obviously risks imposing significant costs on the taxpayer. It undermines market discipline, competitive equality among financial institutions of different sizes, and normal regulatory and supervisory expectations.
September 15, 2011	Daniel Tarullo	In performing these kinds of analyses, we will draw on the extensive work on systemic risk we have already done in connection with our development of capital requirements, the designation of systemically important firms by the Financial Stability Oversight Council, and other matters.
March 31, 2009	Charles Plosser	We also need more systematic policies for handling financial firms whose financial condition is deteriorating. One lesson learned from the savings and loan crisis was that insolvent firms permitted to remain open make poor decisions. The regulatory forbearance that did not close insolvent institutions in a timely manner contributed to the crisis.
March 30, 2010	Charles Evans	However, even with such a structure, it would be hubris on the part of policymakers to assume that we would be able to prevent financial stress at all financial institutions. Therefore, we also need to contain the disruptive spillovers that result from the failure of systemically important institutions without resorting to bailouts or ad hoc rescues.
April 26, 2001	Roger Ferguson	We recommend that the risks to individual firms and to the financial system could be reduced by stepped-up efforts to understand the implications of working out a large and complex financial institution. Because no institution 1s too big to fail, I believe that regulators should develop a clearer understanding of, for example, the administration of ...
April 5, 1988	Lee Hoskins	For whatever reason, forbearance in closing insolvent institutions, relaxed regulatory tests of performance, and debt guarantees to uninsured creditors of banks and bank holding companies have worsened an already difficult situation. Despite six years of a remarkably robust economic expansion, the incidence of troubled institutions has not diminished.
June 30, 2009	Thomas Hoenig	Although we have a legal framework for dealing with failing institutions, we have learned that, in a crisis, the "systemic spillover" that can emerge from failures of our largest institutions and the threat to the broad economy require additional consideration. The most recent examples of this have led to the suspension of normal bankruptcy and bank resolution ...

Notes: The table above is constructed using the following procedure. First, we draw a random speech passage that is assigned the financial crisis label using zero shot classification (ZSC). This is the first statement shown in the table. Next, we use the sentence transformer model to compare that statement to the other statements in the corpus by computing cosine similarity scores. The remaining statements in the table are those that had high cosine similarity scores with the original statement.

Table A9: Text Feature Examples: Inflation, Employment, and Output Growth

Date	Speaker	Passage
April 11, 2012	Janet Yellen	To illustrate this puzzle, figure 5 plots changes in the unemployment rate against real GDP growth–a simple portrayal of the relationship known as Okun's law. It is evident from the figure that 2011 is something of an outlier, with the drop in the unemployment rate last year much larger than would seem consistent with real GDP.
March 2, 1983	Paul Volcker	The slower increases in nominal wages have been fully consistent with higher real wages for the average worker precisely because the inflation rate has been declining. Continuation of that benign interaction among lower inflation, lower nominal wages, and higher real wages – combined with recovery in profits – must be a central part of a non-inflationary.
November 17, 2005	William Poole	Putting all these indicators together, core inflation and inflation expectations have been contained, but underlying determinants of inflation suggest caution. Depending on what measure is used, wage change has been about steady or has risen. The profit share of GDP has risen, suggesting that firms have increased pricing power. Fortunately, productivity growth remains robust.
January 12, 2018	Eric Rosengren	Figure 6 provides another reason to expect lower real interest rates, which is the reduction in the growth in the civilian labor force. With slower population growth, fewer immigrants, and an aging population, the growth in the labor force is expected to be much lower for quite some time than it has been in previous decades. Again, this implies lower real interest ...
February 25, 2008	Frederic Mishkin	Fluctuations in inflation and economic activity induced by variation over time in sources of economic inefficiency, such as changes in the markups in goods and labor markets or inefficiencies in labor market search, could also drive a wedge between the goals of stabilizing inflation and economic activity (Blanchard and Gali, 2006; Gali, Gertler, and Lopez-Salido, 2007). For example, ...
September 28, 2007	Janet Yellen	... leads to an increase in unemployment. Of course, if productivity growth is high, as it has been on average since the mid-1990s, then downward nominal wage rigidity becomes a less important issue." Behavioral considerations thus point to the possibility of a long-run tradeoff between inflation and unemployment at very low inflation rates.
January 16, 2002	Laurence Meyer	The connection between the output and unemployment gaps is often expressed in an equation referred to as Okun's Law: y = kUGAP, with k typically estimated at about 0.5. Empirically, versions of Okun's Law that allow for lags in the relationship between output and unemployment gaps seem to fit better.
June 10, 2008	Eric Rosengren	The first empirical question involves the role that domestic excess capacity (for example, unemployed workers) plays in reducing inflation. Does sizable excess capacity in The economy significantly reduce future inflation? This question is at the heart of discussions of the Phillips Curve and its relevance.
November 4, 2013	Eric Rosengren	The relationship between GDP growth and the unemployment rate can be analyzed using a modified Okun's Law, which is an approximation of how much unemployment falls when the economy grows faster than its potential. While this relationship is only an approximation of what might happen with a given GDP growth rate, it does provide some context for how long it would take to get to my estimate of full ...
February 19, 2014	James Bullard	... cycle effects ("cyclical") or it might just be noise about the fundamental hump-shaped trend. Relatively minor cyclical effects on labor force participation would likely be too small to have major macroeconomic implications given everything else going on in a macroeconomic model. Consequently, the thinking would go, maybe we do not need to worry too much about the ...
May 22, 1992	Lawrence Lindsay	By contrast, the 1981-1990 cycle saw a deceleration in inflation from 10.3 percent to 5.4 percent, with an average rate of less than 5 percent. Certainly these two periods should provide a test of the hypothesis that inflationary policies are good for opportunity and income distribution.

Notes: The table provides examples of statements that had the highest cosine similarity scores with the statement "inflation, employment and output growth," which we use to identify dual mandate-related content.

Table A10: Text Feature Examples: Banking Regulation and Financial Stability

Date	Speaker	Passage
May 18, 2001	Alan Greenspan	It seems clear that, building on bank practice, we are in the process of developing an improvement in both lending and supervisory policies that will not only foster better risk management but could also reduce the pro-cyclical pattern of easing and tightening of bank lending, and accordingly increase bank shareholder values. It is not an easy road, but it seems that we are well along it.
April 25, 2002	William McDonough	... on measures of market risk and a second based on the treatment of corporate loans. In either case, we intend for both methods to result in comparable capital requirements, which we will seek to confirm through our next impact study. Responsibility would lie with the national supervisor to determine the approach most suitable for its banks.
February 26, 1998	Alan Greenspan	... the setting of soundness standards should achieve a complex balance – remembering that the goals of prudential regulation should be weighed against the need to permit banks to perform their essential risk-taking activities. Thus, capital standards should be structured to reflect the lines of business and degree of risk-taking in which the individual bank chooses to engage.
November 11, 2019	Eric Rosengren	In summary, I am not sure that recent developments and proposals in bank regulation properly reflect the risks we are likely to face in a low interest rate environment that challenges bank profitability and provides less by way of monetary policy buffers. Specifically, capital buffers should be rising now so that there is more room for them to decline if the economy ...
April 17, 1974	Robert Holland	A third issue under the safety and soundness heading is that of capital adequacy for banks (and bank holding companies). The limits that are placed by the regulatory agencies on capital leveraging by banks are necessitated by the special requirements for capital adequacy under which banks must operate in order adequately to safeguard the public interest.
November 30, 1995	Michael Moskow	An excellent example is the work we're doing on bank capital standards for market risk. Now, most banks will not be affected by the new capital requirements for trading-account risk. Probably only the largest 30 to 35 banks will be affected. However, the regulatory approach should be of interest to all banks, and I hope it can be extended to other areas.
October 28, 1994	Robert Forrestal	Aside from purely domestic considerations, safe and sound banking practices are a prerequisite for full participation in the global financial system. That is why the Federal Reserve has strongly supported the original Basle Accord and, more recently, the work being done in Basle on capital standards for all banks.
June 19, 2000	William McDonough	The primary tool of capital regulation currently is the set of minimum ratios that were devised in 1988 by the Basel Committee. The 1988 Capital Accord has truly become a global standard for banks worldwide. Its adoption by over 100 countries has helped to strengthen the safety and soundness of the international banking system and has contributed to the achievement of competitive equality.
May 13, 1986	Paul Volcker	Taken as a whole, the banking system has responded constructively and resiliently to these pressures. There is, indeed, highly encouraging evidence that the system as a whole is now gaining strength. Specifically, for most banks, capital ratios have improved, earnings have increased and nonperforming ...
October 30, 1997	Susan Phillips	In conclusion, the history of banking and of bank supervision shows a long and rather close relationship between the health of the banking system and the economy, a connection that reflects the role of banks in the credit intermediation process. We can expect that relationship to continue and for bank earnings and asset quality to fluctuate as economic conditions change.
September 18, 1998	Laurence Meyer	Why do we supervise and regulate banks? The reasons are very straightforward. Initially, when banks were the dominant financial institutions, the major purpose was to reduce "systemic risk" and the impact on the economy of bank failure. Then as the safety net was created to reduce systemic risk, the resultant moral hazard created an even more important reason to supervise and regulate banks.

Notes: The table provides examples of statements that had the highest cosine similarity scores with the statement "banking regulation should be used to achieve financial stability."

Table A11: Text Feature Examples: Monetary Policy and Financial Stability

Date	Speaker	Passage
October 18, 1984	Martha Seger	... help assure the soundness of the financial system, these problems can best be addressed over time by prudent management decisions on the part of the institutions themselves. Their task would be facilitated by a sustained period of growth in the economy and stability in financial markets. In my view, such a period of stability can not be ensured without a continuation of monetary policies designed to prevent a resurgence of inflation.
May 18, 2017	Loretta Mester	Committee has indicated for some time is likely to be appropriate. This upward policy path will help prolong the expansion, not curtail it. It will help avoid a build-up of risks to macroeconomic stability that could arise if the economy is allowed to overheat. It will help avoid a build-up of risks to financial ...
May 26, 2010	Ben Bernanke	In undertaking financial reforms, it is important that we maintain and protect the aspects of central banking that proved to be strengths during the crisis and that will remain essential to the future stability and prosperity of the global economy. Chief among these aspects has been the ability of central banks to make monetary policy decisions based on what is good for the economy in the longer run ...
March 4, 2005	Michael Moskow	As many of you know, the Federal Reserve's overall mission is to foster financial conditions that allow for maximum sustainable economic growth with price stability. At a macro level, we fulfill this goal through monetary policy. But, monetary policy is a blunt tool—it doesn't give us the ability to focus on improving conditions in specific regions and sectors of the United States economy. When we at the Chicago Fed want to focus on local economies, we have to turn to approaches other than monetary policy.
October 2, 2015	Stanley Fischer	Given these considerations, how should monetary policy be deployed to foster financial stability? This topic is a matter for further research, some of which will look similar to the analysis in an earlier time of whether and how monetary policy should react to rapidly rising asset prices. That discussion reached the conclusion that monetary policy should be deployed to deal with errant asset prices (assuming, of course, that they could be identified) only to the extent that not doing so would result in a worse outcome for current and future output and inflation.
July 18, 1985	Paul Volcker	Our monetary policy actions need to be conducted with a clear vision of the continuing longer-term goals – a financial environment in which we as a nation can enhance prospects for sustained growth in a framework of greater stability. To succeed fully in that effort, monetary policy will need to be complemented by action elsewhere.
March 9, 2018	Charles Evans	Financial stability is an important goal of the Federal Reserve. Indeed, the Fed was established to provide an elastic currency that supports credit intermediation. As we were all too aware during the crisis, a breakdown in financial intermediation can have severe consequences for the real economy. So we must ask if some alternative monetary policy frameworks might be more (or less) prone to generating financial instability risks.
January 10, 2002	Anthony Santomero	As you know, the goal of monetary policy is to create financial conditions that foster maximum sustainable growth. The Federal Reserve, as the nation's central bank, makes two important contributions in this regard. First, we provide essential price stability. Second, we lean against the wind, offsetting as best we can, shifts in aggregate demand that push the economy away from its potential.

Notes: The table provides examples of statements that had the highest cosine similarity scores with the statement "monetary policy should be used to achieve financial stability."

Table A12: Text Feature Examples: Past, Present, and Future

Date	Speaker	Tense	Passage
May 24, 1996	Jack Guynn	Past	Another lesson I learned came from the period during which I headed our Supervision and Regulation Division. In 1982, you may recall, a chain of Tennessee banks controlled by the Butcher brothers failed. You will remember that this debacle turned a page of the history books–from the 1930s model of banking, with regulation and stability as the watchwords, to an early glimpse of the go-go 1980s.
May 25, 2022	Lael Brainard	Past	And you came to the right place to do so. SAIS was founded in 1943—during World War II. It was established one year before the Bretton Woods Conference, two years before the United Nations was established, and four years before the National Security Council first met. SAIS was established during a crisis to enable its graduates to make a better world. Sound familiar?
March 20, 2006	Ben Bernanke	Past	Given the global nature of the decline in yields, an explanation less centered on the United States might be required. About a year ago, I offered the thesis that a "global saving glut"–an excess, at historically normal real interest rates, of desired global saving over desired global investment–was ...
March 9, 2010	Charles Evans	Past	We can gain some insight into this dynamic from earlier periods. Figure 3 highlights the path of the unemployment rate and duration during the last two severe recessions. As you'd expect, both measures rise in tandem during the recession. But during the recovery phase, unemployment duration remains persistently high for quite some time ...
May 25, 2017	Charles Evans	Past	So, now that I've given you this executive summary, let's start with some familiar background. Financial strains began to emerge intensely in the summer of 2007, and the FOMC initiated its first policy rate cut in September of 2007. That, by the way, was my first FOMC meeting as the Chicago Fed president.
January 22, 1992	John Laware	Present	Each year, information about the persons who apply for and receive home loans is provided by the institutions covered by HMDA to the Federal Financial Institutions Examination Council (FFIEC) in Washington, D.C., through their respective supervisory agencies. The Federal Reserve compiles the data, on behalf of ...
April 26, 1968	George Mitchell	Present	... we are making progress in winnowing out the more stable and important relationships from a monetary point of view. As this work has progressed our ability to forecast the consequences of alternative policy actions has improved. While progress has been slow, it has become clearer and clearer that in a dynamic economy, with flexible and adaptable financial markets, no one aspect or variable is an ...
February 28, 2007	Timothy Geithner	Present	In this sense, liquidity is like confidence. And, like confidence, liquidity plays a critical role both in establishing the conditions than can lead to a financial shock, and in determining whether that shock becomes acute, threatening broader dam age to the functioning of financial and credit markets.
April 11, 2002	William Poole	Present	... the gap as the difference between the equilibrium and actual rates of unemployment; others use the gap between actual and high-employment real GDP. When the unemployment rate is used, some like to call the equilibrium rate the "natural rate" and some like to call it the "non-accelerating inflation rate of unemployment" or "NAIRU."
September 19, 2000	Thomas Hoenig	Present	Let me conclude my discussion with some general observations about monetary policy in a changing world. The theme of my remarks this evening is that, while monetary policy is always challenging, it is especially so when there are important structural changes occurring in the economy or in financial markets.
December 18, 2020	Lael Brainard	Future	In the global development context, see Lael Brainard, Abigail Jones, and Nigel Purvis, Climate Change and Global Poverty: a Billion Lives in the Balance (Washington, DC: Brookings Institution Press, 2009). See also Department of Treasury, "United States Takes a Significant Step Toward a Clean Energy Future," news ...
October 9, 1982	Paul Volcker	Future	No. I think the implication of a de-emphasis on M1 has to be that. It's not new, except in a matter of degree. I say in the statement: We'll look at M2 and M3 to see what information we get from that–because we always do look to them–but we expect those to be less distorted by this technical change because almost all the changes I anticipate will be ...
February 19, 2004	Michael Moskow	Future	The big question going forward is whether the most recent surge in demand will sustain itself and lead to substantial job growth, or whether this one will falter like the spurts we saw in 2002. While there are always uncertainties, I think there are reasons to be optimistic that economic growth will remain solid and lead to an acceleration in employment.
October 2, 2002	Jack Guynn	Future	One final risk is the uncertainty associated with developments in the Middle East, Latin America and Asia. Any one of these hot spots has the potential to deliver a large negative shock to the United States economy; the uncertainty and anxiety caused by geopolitical developments in these regions is having an effect on financial markets and the economy even now.
February 14, 1995	Robert Forrestal	Future	There is, however, another view of the world that argues that the economy is less prone to inflation than it has been. There are many reasons for this new perspective, including changes in the labor force, business investment, and general expectations about inflation. I will not go into them here. Suffice it to say that, according to this point of view, the Federal Reserve should ...

Notes: The table above provides example speech passages along with their tense classifications. We select five passages for each tense. The tense classification is determined by selecting passages that have a classification score of over 0.75 for the tense of interest and a classification score of under 0.25 for the other two tenses. While many passages in speeches contain a mixture of tenses, we focus on passages that have a dominant tense for the purpose of illustration.

Recent Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

The Macroeconomic Effects of Trade Tariffs: Revisiting the Lerner Symmetry Result *by Jesper Lindé and Andrea Pescatori*	2019:363
Biased Forecasts to Affect Voting Decisions? The Brexit Case *by Davide Cipullo and André Reslow*	2019:364
The Interaction Between Fiscal and Monetary Policies: Evidence from Sweden *by Sebastian Ankargren and Hovick Shahnazarian*	2019:365
Designing a Simple Loss Function for Central Banks: Does a Dual Mandate Make Sense? *by Davide Debortoli, Jinill Kim and Jesper Lindé*	2019:366
Gains from Wage Flexibility and the Zero Lower Bound *by Roberto M. Billi and Jordi Galí*	2019:367
Fixed Wage Contracts and Monetary Non-Neutrality *by Maria Björklund, Mikael Carlsson and Oskar Nordström Skans*	2019:368
The Consequences of Uncertainty: Climate Sensitivity and Economic Sensitivity to the Climate *by John Hassler, Per Krusell and Conny Olovsson*	2019:369
Does Inflation Targeting Reduce the Dispersion of Price Setters' Inflation Expectations? *by Charlotte Paulie*	2019:370
Subsampling Sequential Monte Carlo for Static Bayesian Models *by David Gunawan, Khue-Dung Dang, Matias Quiroz, Robert Kohn and Minh-Ngoc Tran*	2019:371
Hamiltonian Monte Carlo with Energy Conserving Subsampling *by Khue-Dung Dang, Matias Quiroz, Robert Kohn, Minh-Ngoc Tran and Mattias Villani*	2019:372
Institutional Investors and Corporate Investment *by Cristina Cella*	2019:373
The Impact of Local Taxes and Public Services on Property Values *by Anna Grodecka and Isaiah Hull*	2019:374
Directed technical change as a response to natural-resource scarcity *by John Hassler, Per Krusell and Conny Olovsson*	2019:375
A Tale of Two Countries: Cash Demand in Canada and Sweden *by Walter Engert, Ben Fung and Björn Segendorf*	2019:376
Tax and spending shocks in the open economy: are the deficits twins? *by Mathias Klein and Ludger Linnemann*	2019:377
Mind the gap! Stylized dynamic facts and structural models *by Fabio Canova and Filippo Ferroni*	2019:378
Financial Buffers, Unemployment Duration and Replacement Labor Income *by Mats Levander*	2019:379
Inefficient Use of Competitors' Forecasts? *by André Reslow*	2019:380
How Much Information Do Monetary Policy Committees Disclose? Evidence from the FOMC's Minutes and Transcripts *by Mikael Apel, Marianna Blix Grimaldi and Isaiah Hull*	2019:381
Risk endogeneity at the lender/investor-of-last-resort *by Diego Caballero, André Lucas, Bernd Schwaab and Xin Zhang*	2019:382
Heterogeneity in Households' Expectations of Housing Prices – Evidence from Micro Data *by Erik Hjalmarsson and Pär Österholm*	2019:383
Big Broad Banks: How Does Cross-Selling A Affect Lending? *by Yingjie Qi*	2020:384
Unemployment Fluctuations and Nominal GDP Targeting *by Roberto Billi*	2020:385
FAQ: How do I extract the output gap? *by Fabio Canova*	2020:386

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se