Monetary Communication Rules*

Laura Gáti† and Amy Handlan‡

September 7, 2023

Abstract

Does the Federal Reserve follow a communication rule? We propose a simple framework to estimate communication rules, which we conceptualize as a systematic mapping between the Fed's expectations of macroeconomic variables and the words it uses to talk about the economy. Using text analysis and regularized regressions, we find strong evidence for systematic communication rules that vary over time, with changes in the rule often being associated with changes in the economic environment. Our method is general and can be applied to investigate systematic communication in a wide variety of settings.

Keywords: communication, expectations, monetary policy, NLP, text analysis
JEL codes: E52, E58, C49

*We thank our discussant Taeyoung Doh for his feedback. We are also grateful for the helpful comments and suggestions from Fernando Duarte, Michael Ehrmann, John Friedman, Peter Karadi, Soonwoo Kwon, David Lagakos, Bartosz Maćkowiak, Michael McMahon, Vitaly Meursault, Sarah Mouabbi, Emi Nakamura, Davide Romelli, Jon Roth and David Weil. We also thank seminar and conference participants for their feedback. Finally, we thank Haoyu Sheng and Frank Chiu for their research assistance. The views expressed herein are our own and do not necessarily reflect those of the ECB or the Eurosystem.

†European Central Bank, Directorate General Research, Sonnemannstraße 20, 60314 Frankfurt am Main, Germany. Email: laura_veronika.gati@ecb.europa.eu.

‡Brown University, Department of Economics, 64 Waterman Street, Providence, Rhode Island, USA. Email: amy_handlan@brown.edu.

1 Introduction

The conventional wisdom of economic policy is that it is rule-based. For example, macroeconomists think of a central bank's policy rate as being determined through a mapping from economic fundamentals to the interest rate. However, these policy instruments are not necessarily numerical. We consider policy announcements as a tool for policymakers to communicate new information and to interact with the public's expectations of the economy and policy. These announcements may include new information about policymaker preferences, beliefs over economic variables, or the path of future policy. Accordingly, we argue that we should think of the text in policy announcements in a similar rule-based framework as numerical policy.

This paper estimates the systematic mappings from what central bankers think about the economy to the words of their policy announcements. We call these estimated policy functions the **monetary communication rules**. Using the post-meeting statements from the Federal Open Market Committee (FOMC), we represent the Federal Reserve's (Fed's) language with a bag-of-words approach.[1] We use ridge regressions to estimate the communication rules, that is, to connect the FOMC statements to internal Fed forecasts. We find that the FOMC systematically communicates most of their internal forecasts and policy decisions through their post-meeting policy statements except for forecasts about short-run inflation. With this framework, we also provide a simple procedure to estimate time-varying communication rules and to detect when the rules shift. We find changes in the communication rules for macroeconomic forecasts, indicating the Fed is changing how they reveal their beliefs to the public. We also find changes to the communication rule for the target policy rate and forward guidance throughout the zero-lower-bound period. This suggests that language in

[1] Bag-of-words models use measures of word frequency, or the frequency of sequences of words called n-grams, while largely abstracting from the order in which those words or sequences occur throughout the document.

statements are a separate policy instrument; the Fed can change their communication when they are unable to change interest rates. Our methodology for quantifying time-varying language mappings is tractable and flexible enough to be applied to other contexts that want to connect text data to numerical data over time.

Our finding that Fed communication is exceedingly systematic has broad economic implications. Central bankers have long stated the importance of managing public expectations about the economy and policy. At the same time, the public pays close attention to the Fed's words. The result that the Fed communicates systematically is consistent with a model where the Fed is transparent about their beliefs and it is thus rational for the public to closely watch the Fed's announcements. But to eventually explore the public's evolving understanding of the Fed's use of communication as a *policy tool*, we first need a way to measure communication quantitatively. We thus see our main contribution as providing an empirical strategy to capture communication as a policy tool and measure its variation over time.

The advantage of our measurement strategy is threefold: objectivity, flexibility, and tractability. First, we let the connection between words and economic variables of interest be determined econometrically. Conventional methods for text analysis in economics, including narrative and dictionary approaches,[2] are valuable for settings with specialized language. However, they are not easily standardized as the measure of the text is dependent on the coder's interpretation of the text or their priors on the meanings of particular words. Implicitly these methods impose a priori structure on the text measure which may not be consistent with patterns in the data. But, when we can use observables to link to text, as is the case in our setting, then we can let the correlations with observables guide how meaning

[2]Narrative approaches involve researchers assigning labels directly to text to create a measure of the text. Dictionary methods are where researchers assign labels to words by creating keyword lists. Then they count how often words from the dictionary list occur in their text to create measures. Section 7 compares our baseline to a dictionary approach.

is ascribed to words. For example, we say particular words are indicative of higher economic growth because those words are correlated with higher growth forecasts and not because we assume it. In this sense, we argue the objectivity of our approach comes from allowing the data to guide our language model instead of using the researcher's potentially subjective interpretation of words.

Second, using a regression model allows us to flexibly capture both direct and indirect associations between text and other variables. For example, our approach nonparametrically associates language about high economic growth with forecasts of higher inflation because the Fed often references increased economic activity in announcements when they have higher inflation expectations. That is, we are not restricted to phrases that explicitly mention "inflation" to learn about inflation forecasts. On the other hand, a dictionary approach would require the researcher to identify a list of words that directly map to a concept of interest, such as inflation, and are limited only to direct references based on the list of key words. Thus we see this flexibility as another advantage.

Finally, although these first two points would hold for any supervised, machine learning approach with text, our method also has the added benefit of being tractable. At this time, state-of-the-art large language models are predominately trained on English text and therefore are not implementable for researchers with specialized language, such as non-English text or settings with unique vocabulary. Furthermore, time-varying analysis requires adjusting the sample throughout estimation. This is not feasible with frontier models - like Bert or ChatGPT - which are trained on huge datasets that incorporate information from a variety of different time periods in a black-box fashion. Fine-tuning these models can help adapt general language models for cross-section analysis, but this does not side-step the issue for time series analysis where the timing of training data matters. Meanwhile, a penalized regression strategy can be trained on a much smaller sample, and therefore can be estimated from scratch allowing the researcher full control over sample restrictions and parameter estimation.

Accordingly, the objective, flexible, and tractable methods in this paper can more broadly be used for any environment where researchers are interested in quantifying systematic language related to quantitative measures. Using the language-ridge regression allows the researcher to test for correlations between language and other variables. For example, it can be used to analyze text-based survey responses, to identify political affiliation based on speeches,[3] or to test for research language that increases the chance of acceptance into a top journal. Again, our methodology also can be applied to non-English communication, making it promising for applications in international and development economics as well.

There are also many opportunities for time-varying applications in economics. For instance, one could extend Baker et al. (2016) and Kalamara et al. (2022) by identifying when newspapers change their coverage of macroeconomic variables or economic policy. Or, one could build on Hassan et al. (2019) and Liang et al. (2022) by testing when companies change their communication strategies for their earnings calls.[4] Our framework for thinking of communication as an estimable policy function opens up a new area of research where economists may measure communication rules and document their changes over time in a variety of settings.

1.1 Related Literature

This paper sets out to measure systematic communication policy. In this effort, it touches base with a number of literatures that do related, but distinct things. The first is the growing literature that uses text analysis methods to study monetary policy. Existing works in this space use text to quantify policy shocks and surprises (Acosta, 2023; Aruoba and Drechsel,

[3]One could think of an exercise similar to Gentzkow and Shapiro (2010) but with the less structural ridge or Lasso regression design instead of their probability language model.

[4]These papers focus on measurement and prediction tasks. See Gentzkow et al. (2019) for a summary of related prediction exercises with text data.

2023; Doh et al., 2022; Gorodnichenko et al., 2023; Handlan, 2020b; Hansen and McMahon, 2016; Romer and Romer, 2004), to measure transparency and preferences of central banks (Doh et al., 2022; Hansen et al., 2018; Romelli and Bennani, 2021; Shapiro and Wilson, 2021), to understand how the public responds (Calomiris et al., 2022; Campbell et al., 2012; Gardner et al., 2021; Gnan et al., 2022; Handlan, 2020a; Husted et al., 2020; Lunsford, 2020; Lüdering and Tillmann, 2020), or to think about how policymakers craft their policy statements (Ashwin, 2021; Byrne et al., 2021; Cieslak et al., 2021; Ehrmann and Fratzscher, 2005, 2007; Ericsson, 2017; Stekler and Symington, 2016). Our work most closely relates to those in the last category, as we are estimating a policy function for the Fed's communication in their post-meeting statements. So far, most papers in this literature either rely on dictionary approaches – using lists of positive/negative words to create sentiment measures – or on embedding approaches – using machine learning techniques that are relatively more difficult to interpret and are not time-varying. Overall, our contribution to this vast literature is to use a minimal set of assumptions on the form of communication so that we can measure how the information set of the Fed relates to their announcements directly from textual data in an objective, flexible, and tractable way.

A second literature explores empirical methods for textual data to estimate time-varying relationships. Regularized regressions, also called penalized regressions, are valuable tools for estimation where there are overfitting concerns. This class of regressions, including Lasso (Tibshirani, 1996), ridge (Hoerl and Kennard, 1970), and elastic-net regressions (Zou and Hastie, 2005), impose a trade-off between within-sample fit and generalizability by penalizing coefficient size. For applications to text data, Gentzkow et al. (2019) provide a general discussion of text-penalized regressions. Nevertheless, when considering time-varying parameters, even for numerical data, there is the additional consideration about how we allow those parameters to evolve (Goulet Coulombe, 2023; Kapetanios and Zikes, 2018). Currently, most applications of regularized regressions with text are forecasting or prediction exercises.

For example, Kalamara et al. (2022) forecast macroeconomic variables using newspaper text, and Liang et al. (2022) use a sliding-window penalized regression to predict returns from earnings call text. In our setting, we leverage the time-varying function to study when the mapping from text to numerical variables shifts. This is more in line with the literature on break detection, going back to Chow (1960) and Brown et al. (1975). Our application brings this idea to time-varying regressions with text data.

The third and last literature studies public communication more broadly in monetary economics. A set of theory papers in this large literature explore the value of public communication (Amador and Weill, 2010; Angeletos and La'O, 2013; Angeletos and Pavan, 2007; Basu et al., 1990; Gáti, 2023; Morris and Shin, 2002; Ou et al., 2022). The common thread in this group is to investigate the properties of optimal signals or signal structures, or to postulate a communication structure and derive optimal parameter values given the structure. Another set of papers uses DSGE, SVAR or purely empirical frameworks to look at how the private sector responds to public information. Here, public information is conceptualized either as news shocks (Barsky and Sims, 2011; Beaudry and Portier, 2006), as monetary policy announcements and surprises (Bauer and Swanson, 2023a,b; Gertler and Karadi, 2015; Gürkaynak et al., 2005; Jarociński and Karadi, 2020; Kuttner, 2001; Lewis et al., 2020; McQueen and Roley, 1993; Nakamura and Steinsson, 2018; Piazzesi, 2005; Swanson and Jayawickrema, 2023), or as information treatments in randomized control trials (Coibion et al., 2022). Generally, the strategy for measuring information provision is to impose structure on belief formation and signalling and then use responses of other variables to back out communication effects. We instead impose minimal structure in order to measure systematic communication directly from announcement text.

The paper is structured as follows. Section 2 presents a simple framework that forms the backbone of our estimation strategy. Section 3 provides an overview of the data and the data cleaning. Section 4 spells out the estimation strategy in detail, while Section 5 estimates the

Fed's communication rule under the assumption that the Fed kept their communication rule unchanged over time. Section 6 performs the estimation allowing that communication rules were time-varying, and presents our indicator for shifts in the communication rule. Section 7 performs robustness checks, and Section 8 concludes.

2 A Model of Communication Rules

In this section, we lay out a simple model of monetary communication rules that guides our empirical work. Throughout the paper, we use the terms "communication rule" and "communication policy function" interchangeably to capture the notion of an empirical association between the Fed's beliefs and their word use. In this paper, this is not a structural object, as the first step is to measure the empirical regularity directly. Future research can investigate structural communication rules.

We assume that the central bank (the Fed) communicates about the economy in the following way. Upon seeing data on various economic variables, the Fed forms expectations of these economic variables and their own policy decisions. The set of variables the Fed makes expectations over is Y, and a variable in the set is called $y \in Y$. Throughout the paper we will also use y to index parameters pertaining to variable y. Some y represent contemporaneous policy decisions at time t, y_t, and others represent forecasts up to k quarters in the future, y_{t+k}. We denote the Fed's expectation as $\mathbb{E}_t^{Fed}[y]$.[5]

Given their expectations about the economy and policy, the Fed then sends the private sector a message, $m_t \in M$, about their forecasts. M is the general message space. We think of the message, and thus the information of the message, as a linear combination of tokens, w. Depending on the specification, tokens may be individual words or sequences of

[5]For example, in our dataset, some y are contemporaneous policy variables such as the current target fed funds rate, while others are forecasts k quarters ahead. Examples for the latter are one- or four-quarter-ahead output growth forecasts.

several words. In our baseline, we use overlapping sequences of four words, called quadgrams or 4-grams. In this formulation, the message m_t corresponds to an FOMC statement at a meeting t and we assume we can represent that message using a weighted sum of tokens, as formalized in the following assumption.[6]

Assumption 1 *Message as combination of tokens.*

$$m_t = \sum_j \beta_j w_{j,t}. \tag{1}$$

For tractability we will make another assumption that the overall FOMC statement is made up of sub-messages on the different variables in the Fed's information set, Y. The sub-message m_t^y is the information in the statement that corresponds directly or indirectly with the Fed's belief over variable $y \in Y$ at time t. Then the sub-message is also made up of a linear combination of tokens in the statement, but with potentially different weights β^y. The overall message is the union of sub-messages in that information about one element of Y may be informative about other elements. That is, the messages may be overlapping and need not be disjoint. For example, information about Fed forecasts of unemployment may have information on what the target interest rate may be in the future.

Assumption 2 *Messages are made of sub-messages for each variable $y \in Y$.*

$$m_t = \bigcup_{y \in Y} m_t^y \tag{2}$$

$$m_t^y = \sum_j \beta_j^y w_{j,t}. \tag{3}$$

The novelty of our paper is to entertain the idea that there may be a systematic mapping between the Fed's expectations of economic variables, $\mathbb{E}_t^{Fed}[y]$, and the message it sends, m_t^y.

[6] An alternative would be to think of the *individual* tokens of the statement as *separate* messages. Conceptually this is inconsistent with our idea that combinations of words are needed to convey information. Accordingly, we follow the majority of the central bank communication literature in conceiving of the full FOMC statement as a single message.

That is, we are treating the message itself like an "interest rate rule." This object is what we refer to as the *communication rule* and what we estimate in the Section 4. To obtain a relation that we can take to the data, we make the following assumption.

Assumption 3 *The Federal Reserve is mean truthtelling.*
We assume that the Fed chooses m_t^y such that on average

$$m_t^y = \mathbb{E}_t^{Fed}[y]. \tag{4}$$

for each variable $y \in Y$.

The interpretation is that the Fed selects the message that corresponds to their expectation on average. In other words, the Fed has a goal of transparency. In a companion paper, we show that the Fed does not have a strategic incentive to systematically mislead the public (Gáti and Handlan, 2023).[7] Observationally, we think that this is a reasonable assumption given the recent push in central banking towards increased transparency. Over time, the Fed has increased transparency by releasing internal meeting materials, minutes, and meeting transcripts. These materials, while informative about the inner workings of the FOMC, are released on a lag. The Fed has also increased transparency for policy by increasing the modes of communication on policy meeting days. Policy statements occur after every FOMC meeting since 1999, the Summary of Economic Projections (SEP) is released after every other FOMC meeting since 2007, and post-meeting press conferences now occur after every FOMC meeting.

We now use this simple model to guide our quantification of the Fed's communication rules. We derive our estimating equations from Equation 3 and Equation 4. These equations

[7] In Gáti and Handlan (2023), we find that the central bank in equilibrium will release mean truthtelling statements that are imprecise or noisy to balance a trade-off between moving public expectations today and maintaining a reputation of providing informative forward guidance.

imply a relationship between Fed expectations and the words of the form

$$\mathbb{E}_t^{Fed}[y] = \sum_j \beta_j^y w_{j,t}, \tag{5}$$

allowing us to estimate the unknowns weights β^y by regressing data on the Fed's expectations of various economic variables on measures of tokens from FOMC statements. The predictive relationship given by the estimates $\hat{\beta}_j^y$ is our notion of a communication rule. In Section 4 we will spell out the regression specification in more detail after introducing our data in Section 3.

3 Data

For the empirical analysis we use a variety of data from the Federal Reserve Bank System. We analyze the FOMC post-meeting statements as the mode of monetary communication for this paper. We also work with data on Federal Reserve forecasts of macroeconomic variables: unemployment, inflation, and output growth. The remaining variables in the paper are the target federal funds rate, the 10-year Treasury less the federal funds rate, the shadow rate, and the total assets of the Federal Reserve. The remainder of the section provides a more detailed discussion of the data.

For our text data, each FOMC statement is downloaded from the Federal Reserve Board's website and the text is extracted and cleaned to remove any URLs, the release time, and the FOMC member voting records. We use a bag-of-words approach to represent the text of the FOMC statements. This means that we are looking at the counts of different words or sequences of words, which we call tokens or n-grams. The main specification of the paper uses term-frequency-inverse-document-frequency (TFIDF) weighted counts of tokens. This weighting scheme down-weights the more frequent tokens across the corpus that may be too common to be informative. The other cleaning of the text includes the removal of numbers, punctuation, and stop words - that is, words that are so common that they are not informative about variation in information across documents in the sample such as "the" or "a". We also combine common multi-word concepts into a single term, e.g. "funds rate" becomes "fundrate", "basis point" becomes "bp", and "Federal Open Market Committee" becomes "fomc". The full lists and processing procedure is available in Appendix B.

For the baseline model, we represent tokens as sequences of four words, called quadgrams or 4-grams. A major advantage of a bag-of-words approach is that in abstracting from token order, it reduces the amount of information we need to keep track of in measuring text. That is, we only keep track of occurrences of tokens rather than a more complex measure of word order and joint occurrences. However, to capture a sense of context we look at tokens that are longer sequences of words. That is, when phrases of four words in a row are used repeatedly in association with particular expectations of the central bank, we see this as evidence of systematic communication. We run robustness exercises in varying the word-sequence length - trigrams/3-grams, quintgrams/5-grams - in Appendix C and find similar results.

We then can represent each FOMC statement as a vector of quadgram counts. The length of the vector is the number of unique tokens across all FOMC statements in the sample, also known as the vocabulary. As part of our processing, we drop tokens that occur in less than 5% of statements to remove outlier tokens that are used too infrequently to measure correlations with economic variables. This removes tokens that are used in fewer than 10 statements, which covers just over one year of FOMC statements. In robustness exercises, we vary this threshold, and our results still go through if we remove no tokens or if we remove tokens that occur in fewer than 10% of statements as well. In dropping extremely infrequent tokens we are able to better focus on systematic communication. With this vector representation of text, we can look at how quadgram usage correlates with macroeconomic variables.

For forecasts of macroeconomic variables at each FOMC meeting, we use the Tealbook Data Sets (formerly the Greenbook Data Sets) from the Federal Reserve Bank of Philadelphia's website. These forecasts and all information in the Tealbooks are released to the public with a five-year lag. The forecasts are quarterly forecasts made by the staff at the Federal Reserve for macroeconomic variables up to nine quarters in the future, but not for policy variables. We work with forecasts for real output growth, inflation, and unemployment.[8] For each, we use the forecasts for next quarter and for next year to compare short-run and medium-run expectations, respectively. Although these forecasts constructed by the Federal Reserve staff and not the FOMC members themselves, we think about the communication strategy of the Fed as being one at the institutional level and thus use these forecasts as the Fed's expectations.

In addition to the forecast variables, we also look at realized policy variables of the Fed. We start with the target fed funds rate from the Federal Reserve Bank of St. Louis's FRED website.[9] We also work with a range of variables capturing different aspects of Fed policy, such as the realized fed funds rate a year from now, the shadow policy rate, longer maturity treasuries, and the Fed's total assets. We source data on the shadow rate from Wu and Xia (2020). This allows us to pick up on additional language variation during the zero-lower-bound periods following the Great Recession and the coronavirus pandemic. Additionally, we work with 10-year Treasury less the federal funds rate to capture language that may be correlated with longer run expectations about interest rates. The total assets of the Fed capture the asset-purchasing behavior of the Fed associated with quantitative easing and other large scale asset purchasing (LSAP) policy decisions. This measure allows us to test for communication rules related to this newer policy instrument.

[8]Output growth and inflation are annualized quarter-on-quarter growth rates. For inflation, we work with change in headline CPI.

[9]For periods when the Federal Reserve has a target range, we look at the midpoint of the range.

4 Estimating Monetary Communication Rules

To estimate the communication rules from the Fed's perspective, we use a ridge regression with the tokens of text from the FOMC statements as inputs and macroeconomic variables as the output of different specifications. To allow for potentially time-varying communication rules, we estimate the regression parameters for different time windows h that are either expanding or rolling. For the fixed communication rules, we assume there is one time window that includes the whole sample. For a given window h and a given output variable y, we estimate the penalty parameter $\alpha^{h,y,*}$ and vector of coefficients $\hat{\boldsymbol{\beta}}^{h,y}$ using a two-step estimation procedure.

The first step in our estimation is to find the optimal penalty term, $\alpha^{h,y,*}$, for the ridge regression. This is the regularization parameter that best controls, in an out-of-sample accuracy sense, for potential overfitting of the regression. We find $\alpha^{h,y,*}$ using stratified, k-fold cross validation. The data is split into five subsets (or folds) such that each has a similar representation of statements associated with rate changes and Fed chair. For example, each of the five folds has a similar number of observations associated with rate hikes under Bernanke. We perform cross validation where one of the folds is used as a validation set and the other four folds are used to estimate the $\boldsymbol{\beta}^{h,y}$ coefficients for a given α penalty term. We fit different $\boldsymbol{\beta}^{h,y}$ as we go over a grid of candidate α parameters and calculate the prediction error on the validation set using the mean square error between the fitted values and the actual output variable. We then repeat this procedure four more times such that each of the folds is the validation set once. The $\alpha^{h,y,*}$ is the α associated with the lowest average out-of-sample mean squared error, where the average is taken across the five different cross validation splits, for macroeconomic variable y and the h-th window.

Given the optimal penalty parameter $\alpha^{h,y,*}$, we estimate the vector of coefficients $\hat{\boldsymbol{\beta}}^{h,y}$

using ridge regression such that:

$$\hat{\boldsymbol{\beta}}^{h,y} = \underset{\beta}{\operatorname{argmin}} \sum_t (\mathbb{E}_t^{Fed}[y] - \sum_j \beta_j^{h,y} w_{j,t})^2 + \alpha^{h,y,*} \sum_j (\beta_j^{h,y})^2 \tag{6}$$

where $w_{j,t}$ represents the term-frequency-inverse-document-frequency (TFIDF) weighted count of quadgram j from FOMC statement at meeting t. Quadgrams from the text of the FOMC statement are sequences of four words in a row. We can then represent the FOMC statement as a vector of token counts, as stated in Equation 3. Again, the length of the vector is the number of unique tokens across all FOMC statements in the sample plus one to account for the constant term in the regression. The reason we use quadgrams instead of uni-, bi-, tri-, or quint-grams is that we found sequences of four words to have enough context within each observation without being so specific that they did not commonly occur throughout statements in our sample. The TFIDF weighting is standard in computational text analysis to downweight term frequencies for tokens that are more common across documents in the corpus and therefore are not as valuable in differentiating information content across documents.[10] We conduct robustness to how we represent text by changing the size of n-grams, the text-weighting schemes, and other cleaning procedures which do not alter the qualitative results of the paper.

The $\mathbb{E}_t^{Fed}[y]$ represents the Federal Reserve Board's expectation of macroeconomic variables from the meeting materials given *before* the FOMC meeting actually occurs. Each FOMC meeting is indexed by t and each macroeconomic variable that makes up the Fed's information set is indexed by y. The macroeconomic variables include realizations of the target federal funds rate, changes in the target rate, the shadow rate, the Fed's total assets,

[10] We implement the baseline TFIDF technique from sklearn package in python with 4-grams. Term-frequency (TF) is the number of times token is used in the current document, divided by the total number of tokens in the document. For inverse-document-frequency (IDF), we divide by the fraction of documents in the corpus that contain the token. In sklearn, there are smoothing transformations to the measure, such as taking the log and adding one to TF and IDF.

10-year Treasuries, but also Federal Reserve forecasts of next quarter's and next year's real GDP growth, inflation, and unemployment as specified in the Tealbooks. In this way, the set of macroeconomic variables Y includes both contemporaneous realizations, such as the current policy rate target, as well as expectations at different horizons.

We estimate different regressions for each macroeconomic output variable, indexed by y. This makes an implicit assumption that the communication rules that map expectations of particular macroeconomic variables to words can be separately measured from the rules for other macroeconomic variables, even if their information content interacts. This is a strong assumption that we make in order to simplify the estimation procedure and get a first pass notion of communication rules in the data. In future work, we will relax the assumption to allow for joint communication rules. Equation 3 and Equation 4 show the mathematical representation of this assumption.

5 Fixed Communication Rule

We first consider a case where the Federal Reserve has the same communication rule over our entire sample for each macroeconomic variable. This means that there is a single, stable mapping from expectations over macroeconomic variables and policy to words. Since we have around 750 regressors, instead of showing the regression coefficients, we plot the fitted values from our regressions.[11] In other words, for each y variable, we plot the realized value in black against the fitted value in red. The fitted value, $f^{h,y}$, is constructed as the vector of estimated coefficients times the regressor

$$f_t^{h,y} = \sum_j \hat{\beta}_j^{h,y} w_{j,t}. \tag{7}$$

[11] In Appendix E, we show the tokens with the 15 largest positive and 15 largest negative coefficients for all the y variables. These are the tokens that are most predictive of higher and lower values of each y variable, respectively.

That is, $f^{h,y}$ is a vector of variable y implied by the communication rule estimated on window h and variable y.

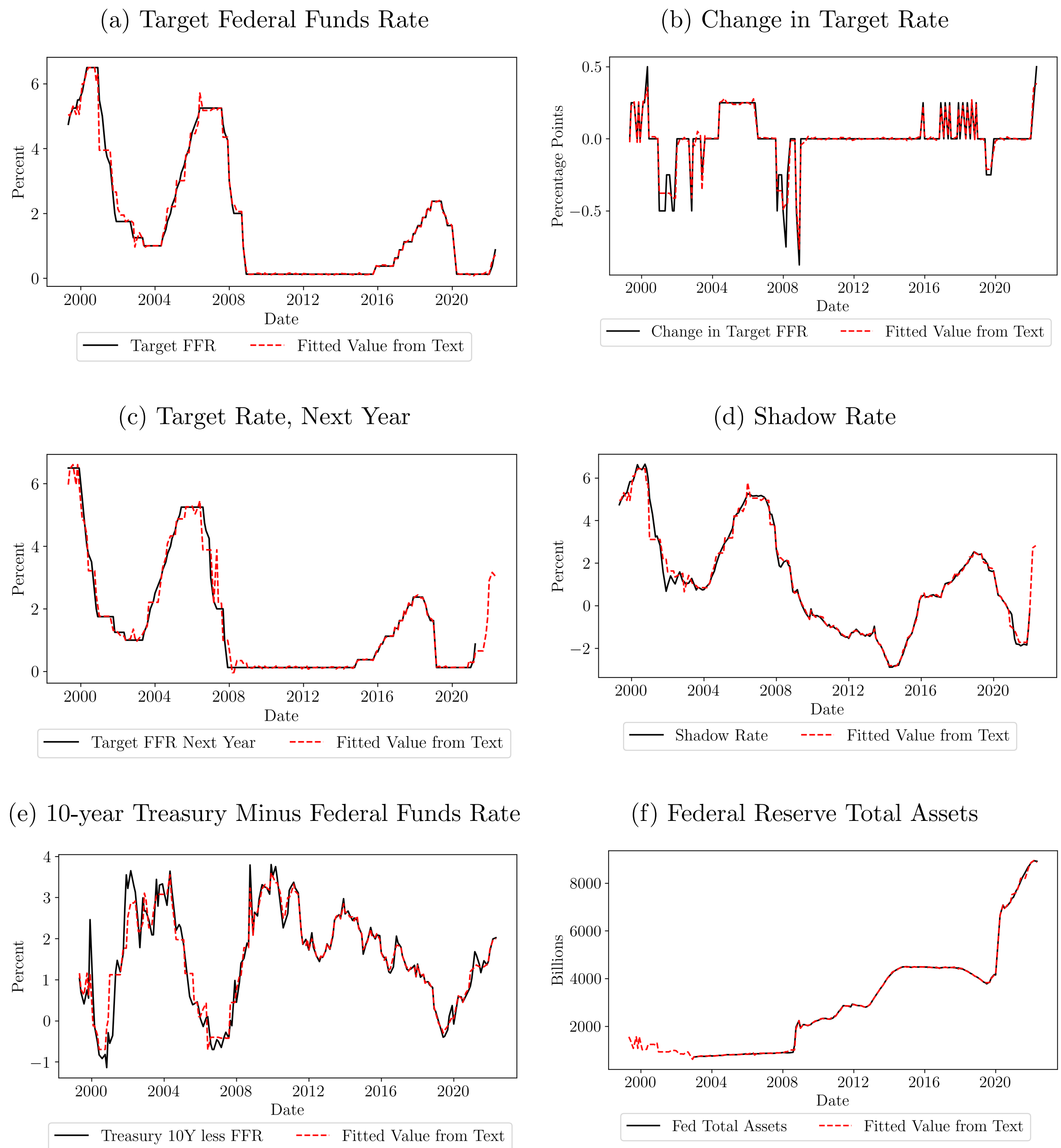
Figure 1 shows the fitted values for the policy variables, while Figure 2 depicts them for the forecast variables. The first observation is that the red dashed lines lie almost completely on top of the black ones for all the policy variables. In other words, we find a very good fit for the target fed funds rate, the change in the target rate, the one-year-ahead target rate, the shadow rate, the spread between the 10-year Treasury and the fed funds rate, and the total assets of the Fed. The interpretation of this is that Fed communication regarding these variables is extremely systematic. Taking the one-year-ahead target rate as an example, if one knows the communication rule (has access to the $\hat{\beta}_j^{h,y}$ from our estimation), then reading the FOMC statement at a particular time allows one to back out next year's target rate almost perfectly from the wording of the statement.

When considering Figure 1, it is important to remember that we have removed all numbers from the FOMC statements. Thus the predictive power of the FOMC statements does not come from numerical information, but just from the information content of the Fed's words alone. This means that the Fed chooses their words extremely consistently; movements in the policy variables up or down are consistently associated with the same sets of words. In other words, the Fed's communication is very transparent.

When it comes to the Fed's expectations in Figure 2, the picture remains largely the same. The fit is excellent for one-quarter-ahead and one-year-ahead unemployment expectations, real GDP growth expectations, and one-year-ahead inflation expectations of the Fed. That is, the Fed communicates just as transparently about their expectations as it does about their policy variables. This transparency allows one to infer the Fed staff forecasts in real time, even though the Tealbooks are only published with a five-year lag.

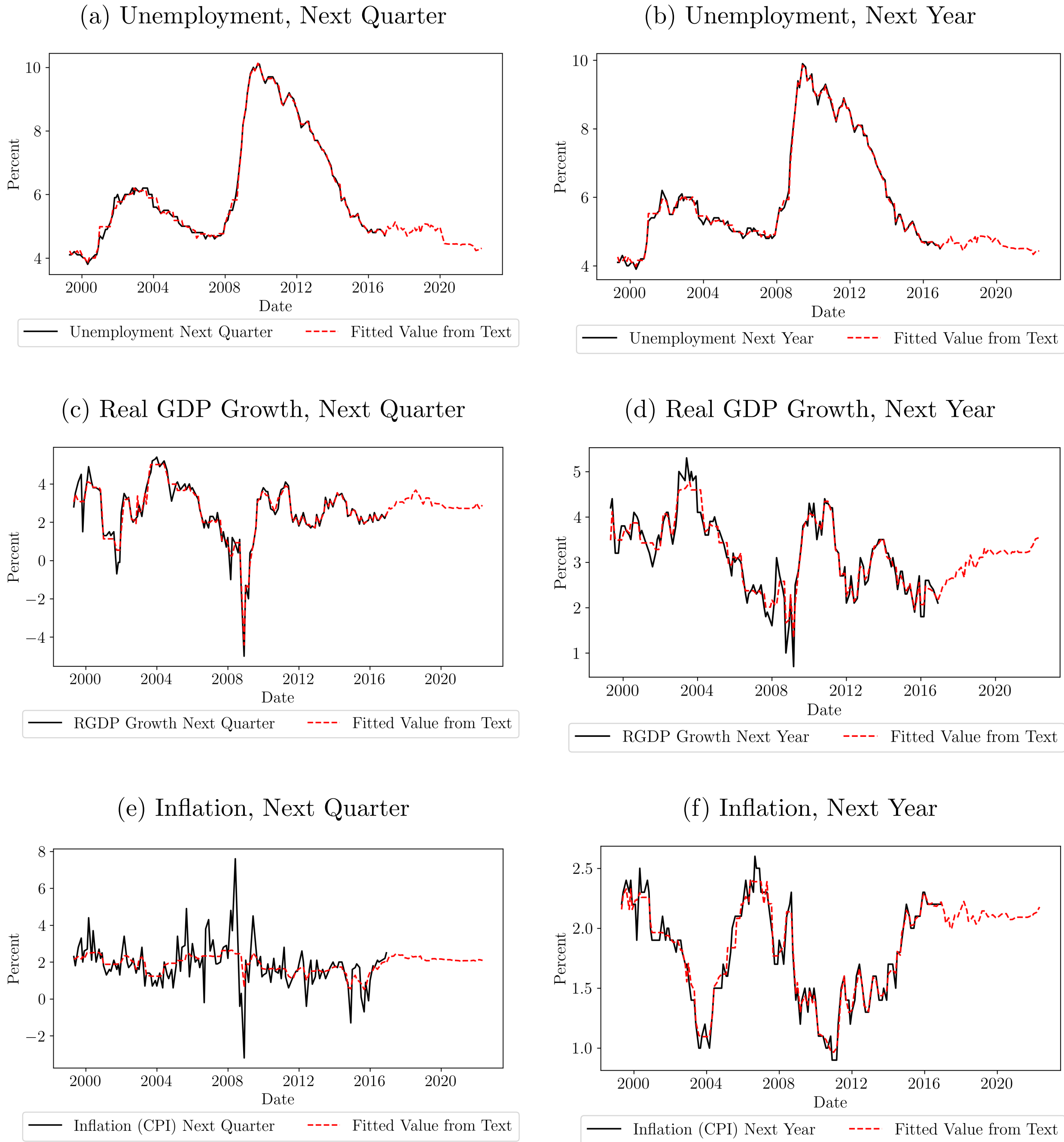
The only variable where the fit completely deteriorates is the one-quarter-ahead inflation expectation of the Fed. This means that the Fed's word-use does not systematically vary

Figure 1: Fixed Communication Rules for Monetary Policy



Note: The red dashed lines are the fitted value from communication rules for the corresponding policy and macroeconomic variables, assuming the communication rule is stable or <u>fixed</u> over the entire sample. The target federal funds rate is the midpoint of the target rate when a range is stated as policy. The Target Rate, Next Year is the realized target federal funds rate one year in the future. This acts as a proxy measure for communication about forward guidance. The Shadow Rate is as constructed by Wu and Xia (2020).

Figure 2: Fixed Communication Rules for Macroeconomic Forecasts



Note: The red dashed lines are the fitted value from communication rules for the corresponding forecasts of macroeconomic variables from the Tealbooks, assuming the communication rule is stable or <u>*fixed*</u> *over the entire sample. Real GDP growth is the quarter-over-quarter growth rate in annualized percentage points. Inflation is measured as the quarter-over-quarter growth in headline CPI in annualized percentage points. The next quarter forecasts are the Federal Reserve forecasts for next quarter and the next year forecasts are the forecasts for four quarters into the future.*

with their short-run inflation expectations. The fact that the fit recovers completely for the one-year-ahead inflation expectation suggests that instead of reflecting their short-run inflation expectations, the Fed's words reflect their longer-run expectations. Given that the Fed's mandate is formulated in terms long-run inflation, and that policymakers usually assume that monetary policy operates with lags, it is not surprising that the Fed's words convey what they expect inflation to be over a longer horizon.

These results indicate that the Fed communicates very systematically about their policy and expectations. The Fed's communication exhibits transparency to a degree that far exceeds what our mean truthtelling assumption of Equation 4 requires, namely that the Fed should reveal their beliefs on average. We see this as evidence that the Fed employs communication as a *policy tool*, systematically picking their words to provide information about their policy and expectations.

6 Time-Varying Communication Rules

The previous section showed evidence for highly systematic Fed communication when looking at the whole sample cross-sectionally. However, the relationship between Fed expectations, policy, and their word-use is not necessarily fixed. In fact, we find evidence that the communication rule changes over time and we provide a simple indicator to track the changes in this mapping over time.

6.1 Detecting Shifts in Communication Rules

In this section we therefore consider the possibility that the coefficients β of the communication rule are only fixed within subsets of the data but can vary over time. We reintroduce the superscript h to denote windows of length H, and estimate $(\beta^{h,y}, \alpha^{h,y})$ separately for each window. In our baseline, we estimate regressions over an expanding window. Our initial window includes scheduled FOMC meetings over 8 years, providing us with 64 observations. Each expanded window adds one more FOMC meeting into the sample, and we repeat the estimation procedure described in Section 4. This provides us with coefficients on quadgrams that vary from one meeting to the next as we expand the window incrementally.

To detect changes in the estimated model over windows, we look at the differences in the fitted values between sequential windows. Recall the notation of $f^{h,y}$ as the fitted value from the ridge regression estimated on window h for macroeconomic variable y. The correlation between $f^{h,y}$ and $f^{h+1,y}$ captures the similarity of communication rules estimated on window h versus $h+1$. We look at one less the correlation as our indicator of changes in the communication rule.

$$\text{Shift Indicator}^{h,h+1,y} \equiv 1 - Corr\left(f^{h,y}, f^{h+1,y}\right) \tag{8}$$

This way, our indicator takes on the value of zero if the correlation between the fitted values is perfect, and, by contrast, a value of one means that there is such a dramatic change that there is no correlation whatsoever. Thus, an uptick in the indicator suggests that a change in the communication rule occurred.

For clarity, we provide confidence intervals and normalize the shift indicator. We estimate bootstrapped confidence intervals for the shift indicator by re-sampling 2500 times for each variable and window h at the 95% confidence level. This gives us a sense that the spikes in the indicator are statistically significant. To compare indicators across variables, we normalize each indicator to a scale of zero to one.[12] We produce graphs without normalization in Section H.1.

The meetings with shifts in the communication rule are different from meetings that just change the words in the statement. Changes in words from one FOMC statement to the next

[12]To account for the confidence intervals (s^b, s^u) we use the following to normalize the shift indicator: $s_h^n = [s_h - min(s^b)]/[max(s^u) - min(s^b)]$.

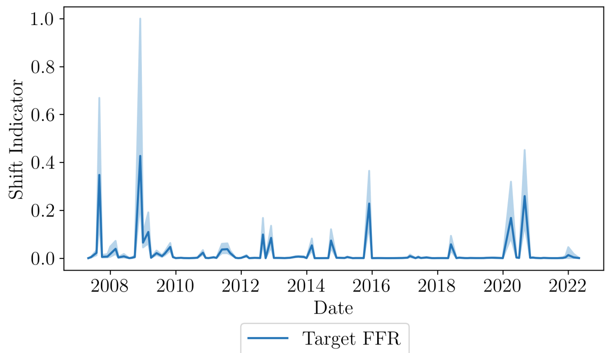
meeting's statement can reflect changes in the state of the economy and still have the same mapping from information to words; that is, the same communication rule. However, when we measure changes in the communication rule, this states that the relationship between word-use and economic variables - such as our policy variables and macroeconomic forecasts - has changed. So there could be a situation where the language from one meeting to the next does not change much, but if there are big shifts in internal forecasts then this would show up as a change in the communication rule. In this sense, we isolate a sufficient statistic for measuring changes in the communication policy function.

Figure 3 depicts our indicator for the Fed's policy variables, while Figure 4 shows the indicator for the Fed forecasts. At the end of 2008, all communication rules dramatically changed due to the incorporation of new policy language surrounding LSAPs and the ZLB. The communication rules for forecasts of real GDP growth and unemployment also change dramatically at this point with a shift indicator that is at least three times larger than in any other period in our sample. The other side of this result is that communication from the FOMC on these variables otherwise are very stable. We can see that as the the shift indicator remains very low at meetings besides the end of 2008. The remaining variables – the target federal funds rate, the target rate one year in the future, and the Fed's total assets – also have a large spike at the December 2008 meeting, but they exhibit greater variation in the shift indicator which suggests additional changes in the communication rules.

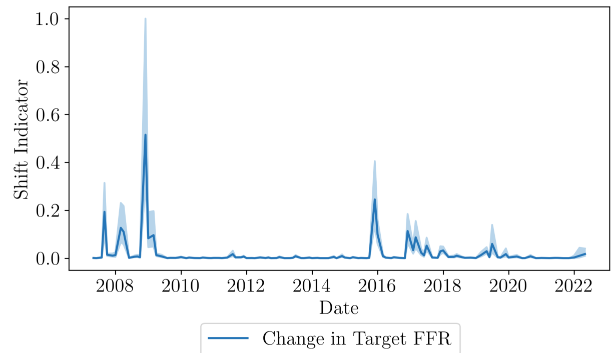
When looking at how the FOMC communicates about their policy tools, the largest changes are leading up to or following periods at the ZLB. September 2007 was the first target rate cut in a series of rate cuts that eventually led to the zero lower bound at the end of 2008. This meeting signalled a shift in language with lower rates due to the nature of the forthcoming recession and housing market crash. We also see a spike in December 2015, likely associated with the eventual lift-off from the ZLB. The communication rule shifted here as a result of forward guidance communicating raising rates, not because the economy

Figure 3: Shifts in Communication Rules for Monetary Policy

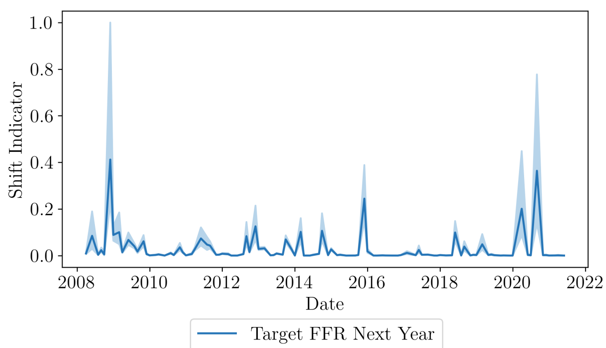
(a) Target Federal Funds Rate



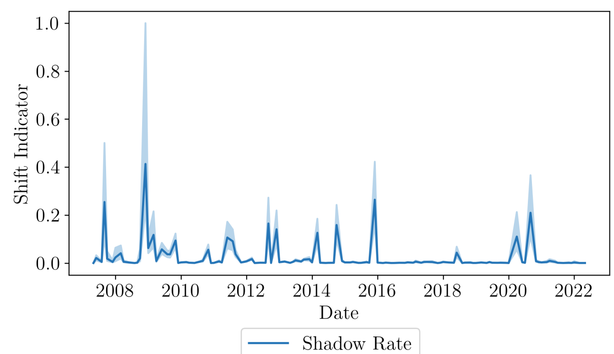
(b) Change in Target Rate



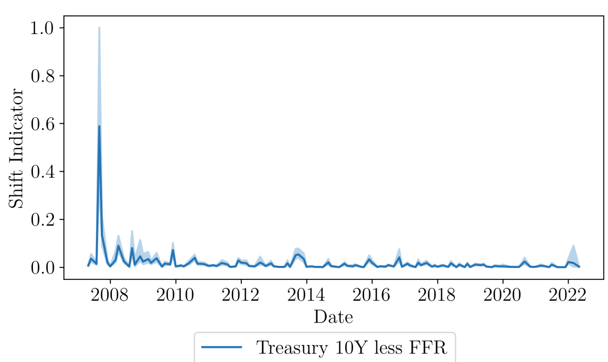
(c) Target Rate, Next Year



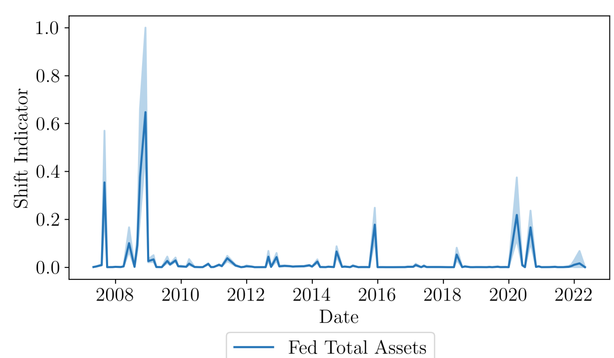
(d) Shadow Rate



(e) 10-year Treasury Minus Federal Funds Rate

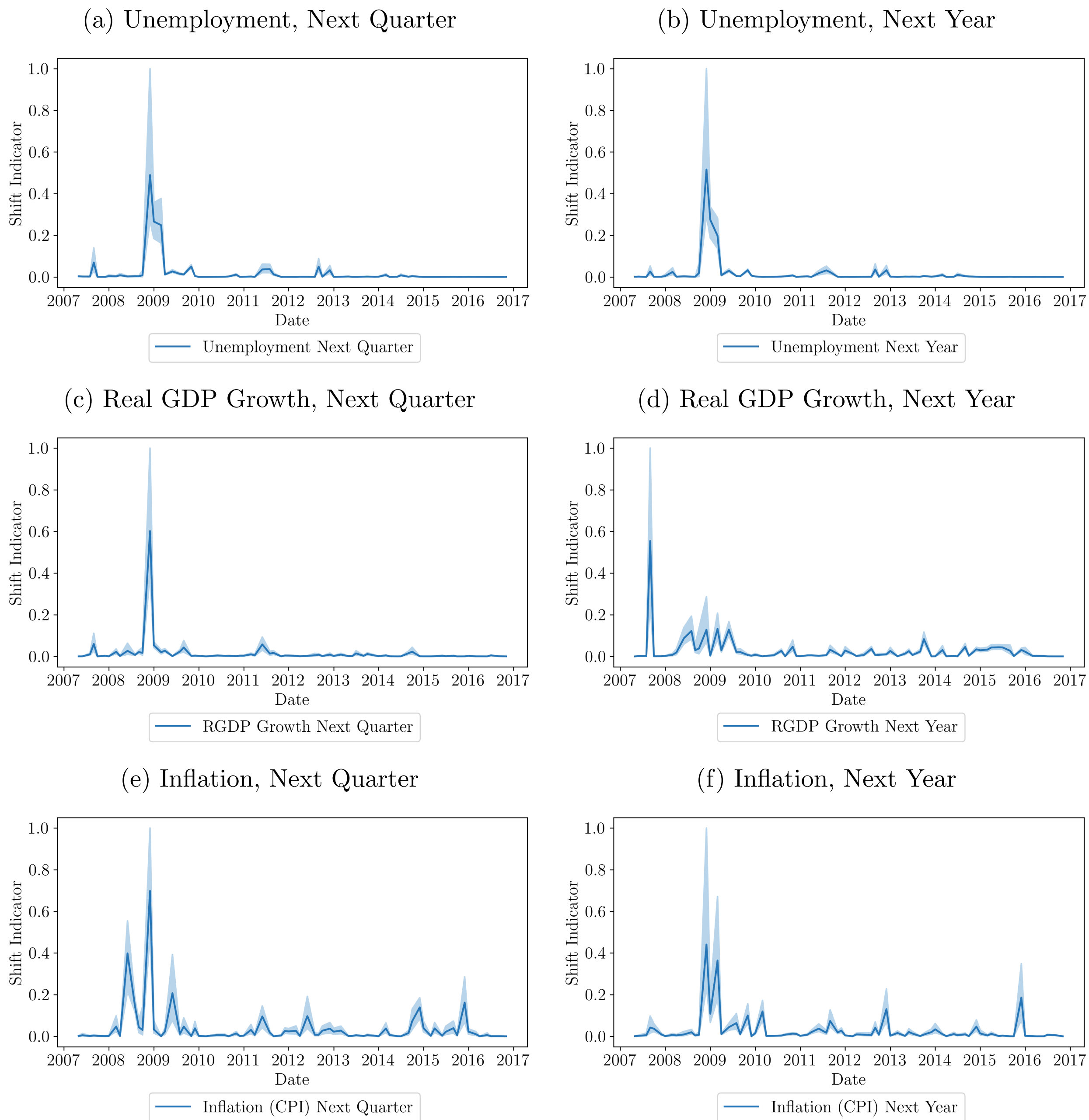


(f) Federal Reserve Total Assets



Note: The shift indicator is one minus the correlation between the fitted values from communication rules estimated on expanding windows ending with meeting t and meeting $t-1$. The higher this indicator, the more the communication rule changed from meeting $t-1$ to t. 95% confidence intervals are plotted around the shift indicator point estimates. We scale plots to have a minimum of zero and maximum of one for comparison across variables. Non-scaled figures are in Appendix H.

Figure 4: Shift in Communication Rules for Macroeconomic Forecasts



Note: The shift indicator is one minus the correlation between the fitted values from communication rules estimated on <u>*expanding*</u> *windows ending with meeting t and meeting $t-1$. The higher this indicator, the more the communication rule changed from meeting $t-1$ to t. 95% confidence intervals are plotted around the shift indicator point estimates. We scale plots to have a minimum of zero and maximum of one for comparison across variables. Non-scaled figures are in Appendix H.*

was overheating, but as a way to "return to normal." Then other rule shifts occur in 2020 due to new language about the pandemic, the ZLB, and how they relate to ongoing policy decisions.

Besides the ZLB, we also see spikes in 2014 and 2018 when Janet Yellen and Jerome Powell, respectively, became the new Fed chair. This indicates that the communication rule, the mapping from information of the Fed to the language in the policy statements, seems to be influenced by the chair. In Appendix G, we estimate communication rules for Greenspan, Bernanke, and Yellen based on the statements and variables when they were chair. Interestingly, we find that Bernanke's communication rules are the most similar to the full-sample fixed communication rule fitted values despite much of his tenure being during the ZLB.

We also find empirical evidence of changes in the forward guidance strategy that aligns with qualitative analysis of shifts. Ben Bernanke describes the change in forward guidance strategy under his tenure as Fed chair and the changes in strategy align with shifts in the forward guidance communication rule (Bernanke, 2020). First, the spikes in 2009 are picking up changes in the relationship between policy and statement language as forward guidance was increasingly used to interact with public expectations during the Great Recession at the ZLB. There is a subsequent spike in 2011 when the guidance strategy changed from being relatively imprecise – or "qualitative" as Bernanke calls it – to more explicit guidance that linked lift-off to specific dates around mid-2013. Once this period of commitment to the ZLB ended in 2013, we also see another shift to the forward guidance communication rule towards conditional statements around lift-off from the ZLB. Overall, we see that the shifts in the communication rule come from a combination of changes in the economy and changes in Fed priorities.

Figure 4 shows the shifts in communication rules for the Tealbook macroeconomic forecasts. The way the FOMC communicates about next-quarter forecasts of macroeconomic variables – unemployment, output growth, and inflation – is fairly stable except for the change in communication at the beginning of the Great Recession. However, the shift-indicator for forecasts of macroeconomic variables one year in the future suggest changes that match up with changes to forward guidance strategy.

In addition to the expanding window, we also estimate time-varying communication rules with a rolling window. Qualitatively, we find similar dates with communication rule shifts, such as the end of 2008 and the end of 2015 with the beginning and end of the ZLB period. However, with the rolling window it is more difficult to attribute changes in the estimated rules to the addition of new meetings to the in-sample window or the dropping of old meetings from the training data. Accordingly, we focus on the expanding window as our main specification and include the results for the rolling window in Section H.3 of the online appendix.

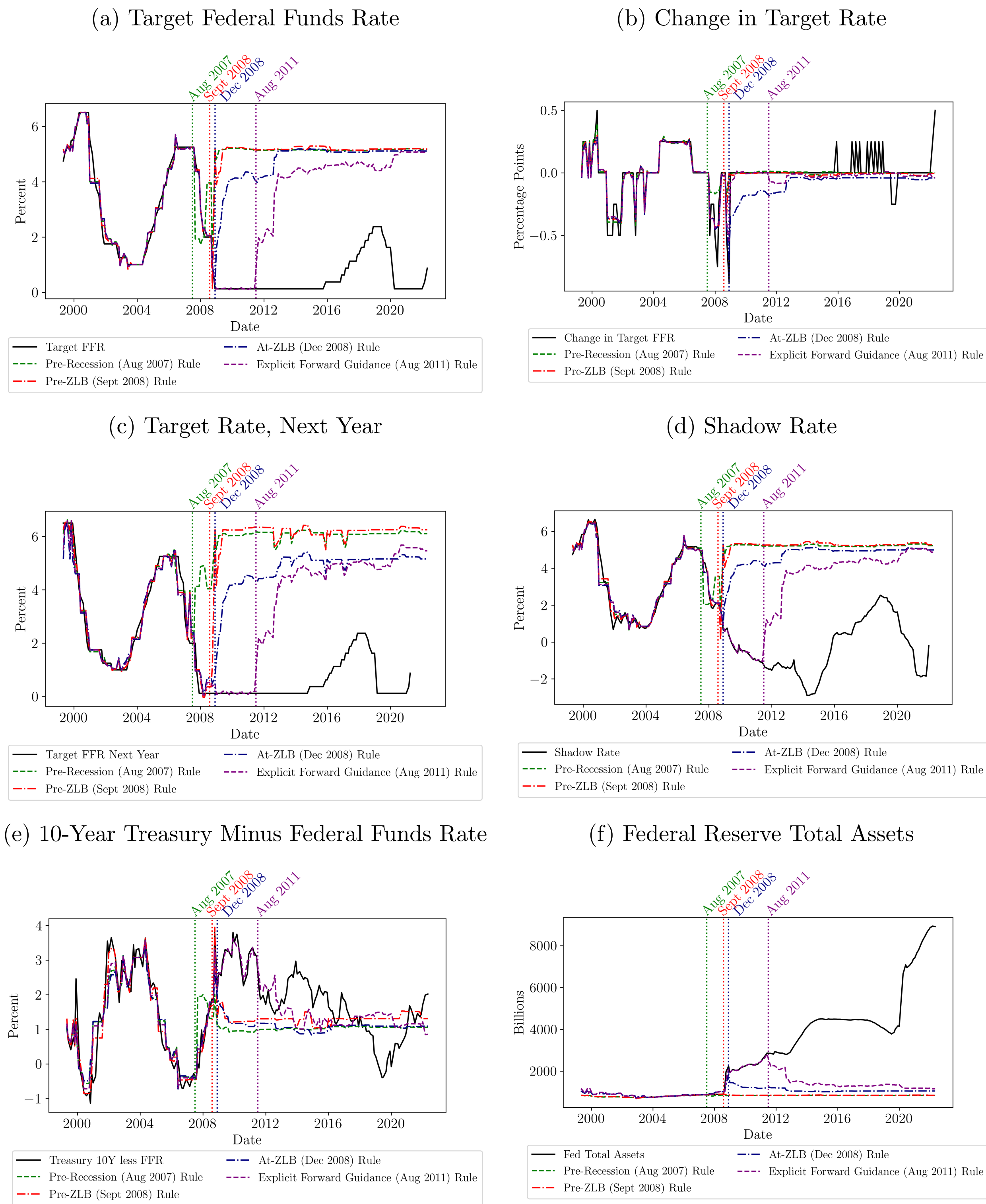
6.2 Communication Rules around the 2008 Crisis

In the previous section, we saw major shifts in the communication rules for policy and macroeconomic forecast variables around the start of the 2008 Financial Crisis. Here, we investigate the changes in the communication rule in more detail. We plot the fitted values for four different windows that end at key dates from the onset of the crisis in 2007, to the start of the zero-lower bound (ZLB) in 2008, to a change in forward guidance language in 2011.[13]

Throughout the panels of Figure 5 and 6, the black lines correspond to the realizations of the y-variables, and the colored, dashed lines show the fitted values coming from communication rules estimated from various windows. Note that these fitted values are for a mix of in- and out-of-sample observations. The vertical, color-coded lines show the stopping times for each window, such that left of the line is in-sample and right of the line is out-of-sample.

[13] Appendix G repeats this exercise for subsets of data that correspond to tenures of various Fed chairs.

Figure 5: Time-Varying Communication Rules for Monetary Policy



Note: The dashed lines are the fitted value from communication rules for the corresponding policy and macroeconomic variables estimated on different subperiods in our sample. We use an <u>*expanding*</u> *window to capture time variation in the communication rule.*

The green lines correspond to the shortest window which ends in August 2007, right before the Great Recession. The next window, in red, ends in September 2008, and thus already captures some of the Great Recession, but not yet the ZLB. The blue window ends just three months later, in December 2008, right as the economy hit the ZLB. The longest window, in purple, ends in August 2011 to capture some explicit forward guidance announcements of the Fed in addition to many months at the ZLB.

Looking first at Figure 5, we see that the in-sample forecasts continue to exhibit the good fit from the fixed rules estimated in Section 5. The out-of-sample fit degenerates, however, for all communication rules. This indicates that the Fed either introduced new language in this period, or changed the way it used existing language (or both). The objective here is not to provide a method that perfectly predicts economic variables from past text of the Fed. It is to understand how the FOMC's information connects with the language they use in their policy statements, and how that relationship changes over time.

It is interesting to note that even though the out-of-sample forecasts coming from all windows are off, they are all off in a similar fashion. There seems to be a level shift in them early on, reflecting that longer windows incorporate more data, but out-of-sample forecasts from the different rules converge to one another towards the end of the sample. This means that while some changes to the Fed's language use occurred around the onset of the Great Recession and the ZLB, the Fed made a concerted effort to communicate consistently over time overall. When getting far enough out-of-sample, though, the new language of newer rules has replaced old language to the extent that the only thing that predicts the left-hand-side variable is the intercept.[14]

How exactly did the Fed's language change early on in the crisis? To investigate this, let us focus on the target fed funds rate in Figure 5a. We see that all out-of-sample forecasts

[14] It is interesting to note that the estimated constants seem to reflect meaningful long-run concepts, such as the inflation target or the natural rate of unemployment.

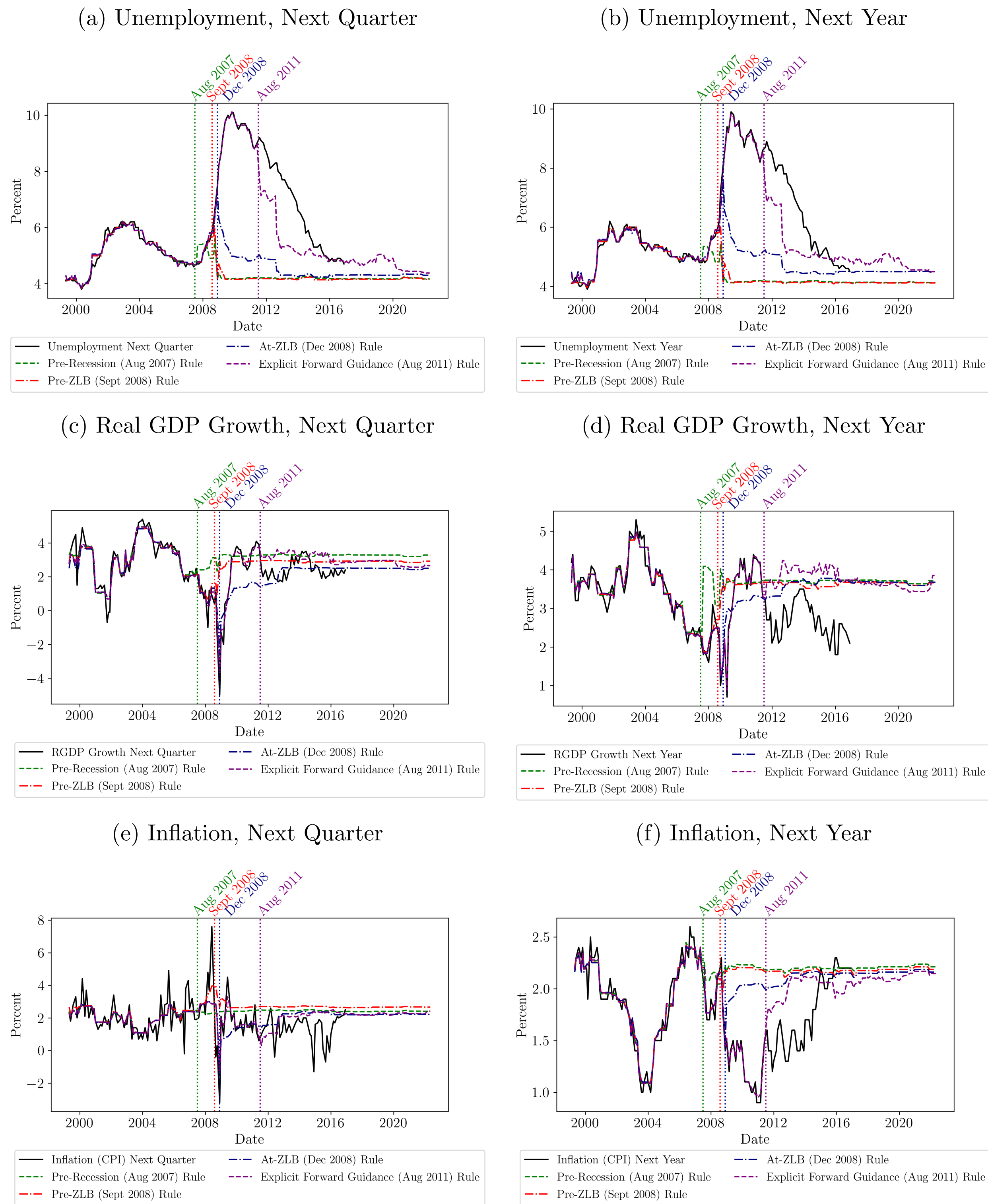
overestimate the target fed funds rate. This has the interpretation that if one was reading an FOMC statement for example in 2013 through the lens communication rules trained on the green, red or blue windows, one would have expected the Fed's current target rate to be around 5%, while in truth the target rate was stuck at the ZLB. Even the longest window with a stopping time of August 2011, in purple, would have had one expect a 4% target rate from the wording of the 2013 FOMC statement. One way to read this is as an early signalling of lift-off.

This means that the FOMC statements after 2008 continued to use words that were previously associated with a higher target fed funds rate. The Fed continued to use words throughout the crisis that reflected stronger economic fundamentals than what the near-zero target rate suggested. Indeed, looking at the other policy variables on the other panels of Figure 5, we see that the out-of-sample forecasts suggest much more contractionary paths for policy than what was realized. For example, for the total assets of the Fed on Figure 5f, the out-of-sample forecasts all suggests that the Fed will decrease their balance sheet in the decade following the crisis, whereas the Fed had only just gotten started increasing their balance sheet in 2012.

The fitted values for communication rules about Fed forecasts in Figure 6 underscore this interpretation by and large. Out-of-sample, the communication rules underpredict unemployment and overpredict GDP growth and inflation compared to the Fed's actual forecasts. Also here, we see that the Fed used words during the recession that reflected a more optimistic economic outlook (lower unemployment, higher GDP growth and inflation closer to target) in the old communication rules.[15] This is evidence of the Fed trying to support the economic recovery by systematically using language that in previous communication was associated with a stronger economy. Again, this suggests the Fed relying on communication

[15] In the case of the Fed's one-year-ahead unemployment expectation, the August 2011 rule is an exception to this, presenting a more pessimistic unemployment outlook starting about in 2015.

Figure 6: Time-Varying Communication Rules for Macroeconomic Forecasts



Note: The dashed lines are the fitted value from communication rules for the corresponding forecasts of macroeconomic variables estimated on different subperiods in our sample. We use an <u>*expanding*</u> *window. Real GDP growth and headline CPI inflation are the quarter-over-quarter growth rate in annualized percentage points.*

as a policy tool, which became increasingly relevant when the Fed's standard policy tool, the fed funds rate, was stuck at the ZLB.

7 Robustness

We conduct many robustness exercises to validate the results from the paper. In this section, we discuss selected exercises, delegating the rest to the appendix.[16] The first tests the degree of overfitting for text data in ridge regressions. The second compares our method for estimating communication rules with alternative text analysis methods, like a dictionary approach or a generative large-language model.

7.1 Shuffled Timing of Communication

One may be concerned that the large number of regressors in our communication rule estimation – from the fact that each quadgram enters as its own variable – may make it possible to fit any body of text to predict the output variables of interest. We test this hypothesis by shuffling the dates of FOMC meeting statements so that the the output variable is no longer from the same FOMC meeting as the quadgram input variables. For simplicity, assume that there is one window.

We estimate optimal penalty weights and coefficients in the same manner as in the fixed communication rule framework, but where the timing is shuffled:

$$\hat{\boldsymbol{\beta}}^y = \underset{\beta}{\operatorname{argmin}} \sum_t (\mathbb{E}_t^{Fed}[y] - \sum_j \beta_j^y w_{j,g(t)})^2 + \alpha^{y,*} \sum_j (\beta_j^y)^2 \tag{9}$$

where $g(t)$ is the date of a random FOMC meeting such that $g(t) \neq t$. If the fitted values

[16] We consider robustness for different representations of text in Appendix B, for penalty parameter selection in Appendix D and different specifications for estimating time-varying communication rules in Appendix H.

from this shuffled exercise are of similar quality as those from Section 5, then that would be evidence that our text analysis approach would produce high in-sample fit mechanically from having many regressors. However, as seen on Figure 7, we find that the shuffled exercise produces fitted values that are poor fits with very high penalty parameters. This indicates that the correlations we estimate are not just spurious. In other words, this provides evidence that the estimated communication rules are indeed indicative of highly systematic Fed communication. Figures for the shuffled communication rules for our other variables are in Appendix F.

7.2 Alternative Text Analysis Methods

We compare our method for measuring communication rules with other popular methods for text analysis in economics: a custom dictionary approach and a frontier generative language model (ChatGPT/GPT-3.5). These other methods provide different ways of measuring variation in the FOMC statement text that we can then relate to the Fed's forecasts and policy decisions. As before, we still discuss robustness in the context of a fixed communication rule. Overall, we find that our approach of penalized regression with quadgrams provides the ideal balance of simplicity and flexibility.

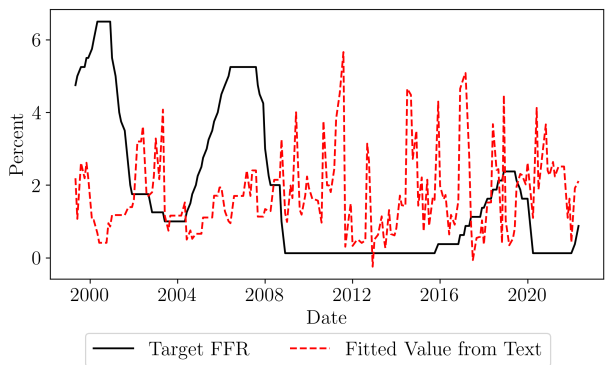
7.2.1 Dictionary Approach

Dictionary methods are currently the most popular strategy for text analysis in economics research. The method involves using a list or multiple lists of words that are associated with a value. For example, we could have an "uncertainty" dictionary that is a list of words indicative of uncertainty. To produce a text-based measure, the researcher then counts the occurrences of words or the number of sentences with key words from their dictionary in the text.

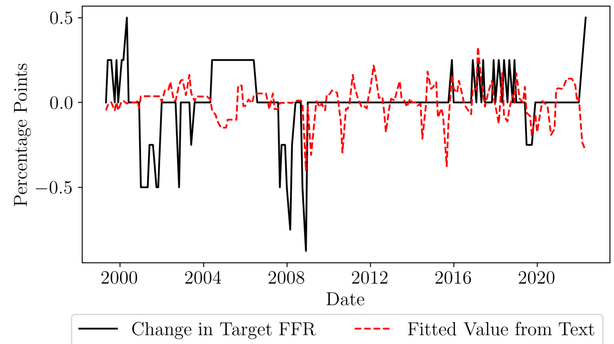
For this comparison exercise, we create two types of word lists – words indicative of

Figure 7: Shuffled Communication Rules

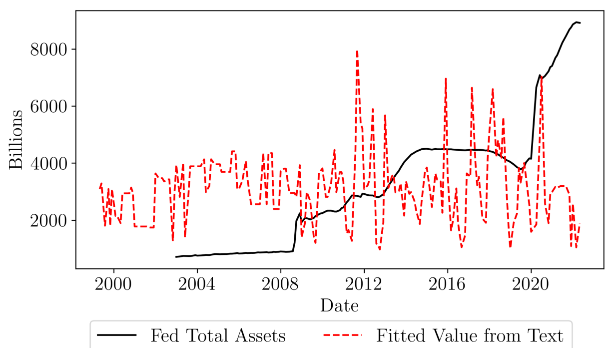
(a) Target Federal Funds Rate



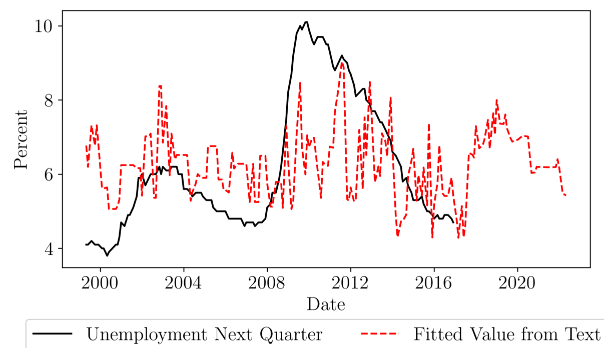
(b) Change in Target Rate



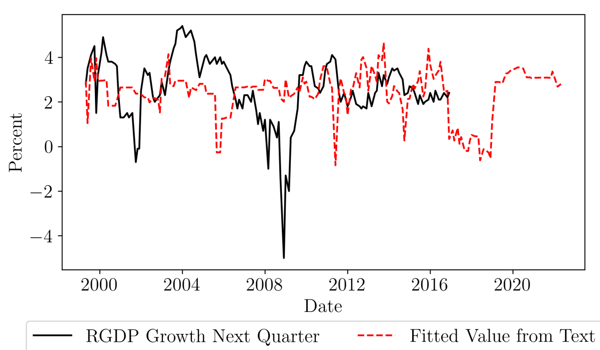
(c) Federal Reserve Total Assets



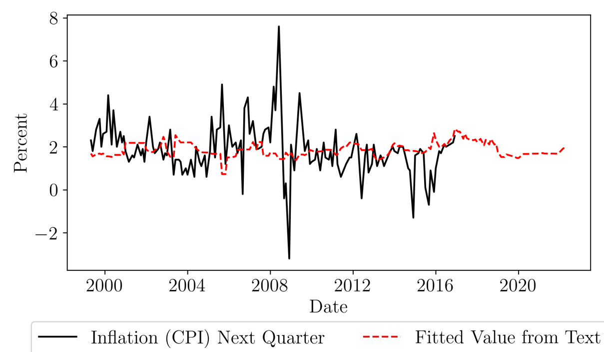
(d) Unemployment, Next Quarter



(e) Real GDP Growth, Next Quarter



(f) Inflation, Next Quarter



Note: The red dashed lines are the fitted value from shuffled communication rules for the corresponding policy and macroeconomic variables, assuming the communication rule is stable over the entire sample and the statement text is randomly assigned to a different date. That is, we have shuffled our observations. The target federal funds rate is the midpoint of the target rate when a range is stated as policy. Real GDP growth is the quarter-over-quarter growth rate in annualized percentage points. Inflation is measured as the quarter-over-quarter growth in headline CPI in annualized percentage points.

macroeconomic variables and words indicative of value or direction. We then count sentences with occurrences of words on the lists to create text-based measures of whether macroeconomic variables are increasing or decreasing. For each sentence in an FOMC statement, we first assign it a topic based on the presence of words from the "unemployment" list, "inflation" list, "economic growth" list, "policy rate" list, and "assets" list.[17] Then we tally the number of "increasing" words less the number of "decreasing" words that occur in a 10-word neighborhood around a topic word. We use "increasing" and "decreasing" but the words in the lists also cover high level values and low level values, respectively. Finally, we also account for negation handling; if "not" shows up before an increasing word, then it is counted as a decreasing word.[18]

Figure 8 shows the dictionary implied measures from the FOMC statements. We can think of the dictionary approach as a restricted version of our approach. In both, there are coefficients associated with words to indicate higher and lower output variables. But rather than estimating coefficients to uncover how words relate to that output variable, the dictionary method requires the researcher to assign coefficient values manually – plus one for increasing words and minus one for decreasing words. Understandably, the dictionary method has a worse fit compared to the fixed communication rule.

The dictionary approach also is not able to distinguish between different forecast horizons. Although the language in the FOMC statements does contain temporal information, it is not specific enough to build a dictionary of "next quarter" and "next year" words that would effectively capture those forecasts.[19] Accordingly, Figure 8 plots both the next

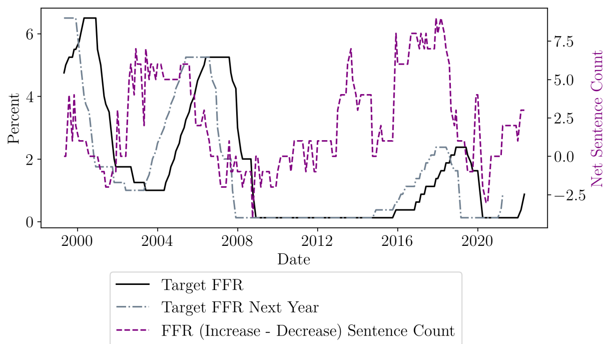
[17]The FOMC sometimes uses long sentences joined with conjunctions that can make this neighborhood approach difficult. Thus, we also do this exercise when looking at sub-sentence phrases and the dictionary measures are fairly similar. We provide this additional specification in the appendix.

[18]We also apply this for the unemployment measure. That is, when an increasing word occurs near "employment" then this is counted as a decreasing word for unemployment.

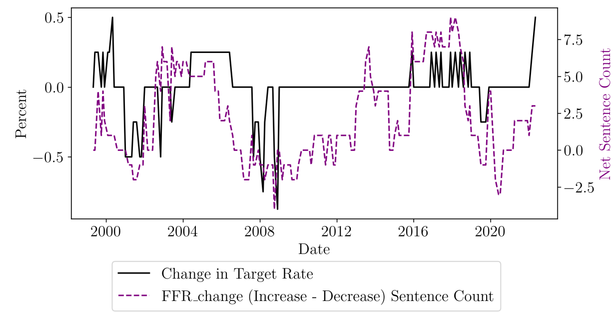
[19]Byrne et al. (2021) have created an algorithm to measure the temporal dimension of central bank communication. The approach sorts information from text into backward-looking or forward-looking categories. Although they incorporate numerical date references from the text into their measure, it is still a reality of central bank communication that they use words that do not always explicitly distinguish between different horizons into the future.

Figure 8: Dictionary-Based Communication Rules

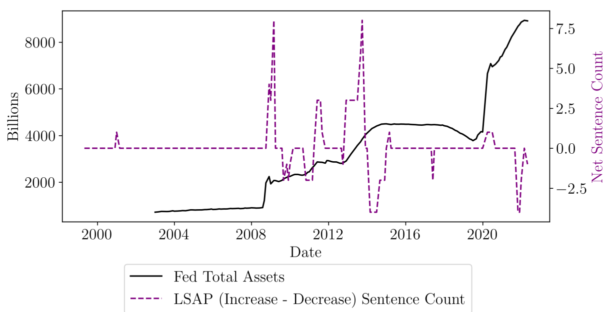
(a) Target Federal Funds Rate



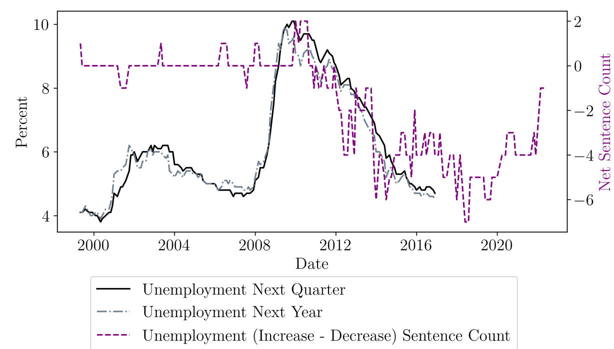
(b) Change in Target Rate



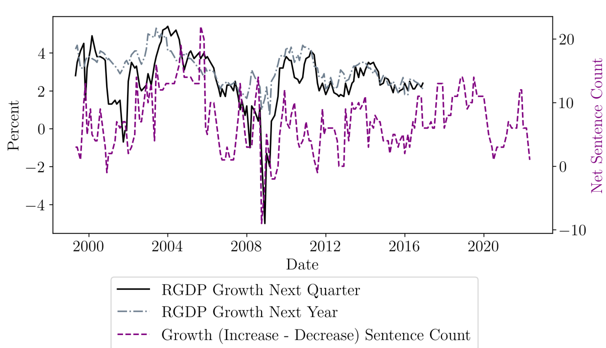
(c) Federal Reserve Total Assets



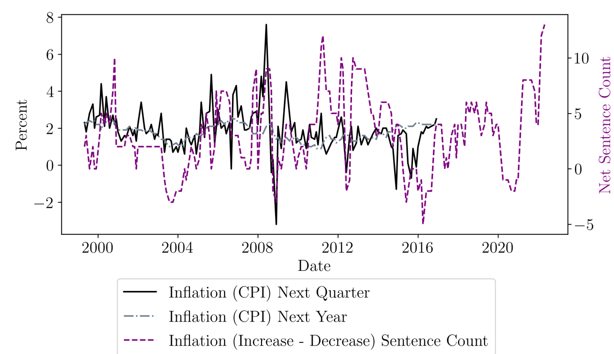
(d) Unemployment



(e) Real GDP Growth



(f) Inflation



Note: The purple line is the dictionary-based measure of the policy variable or macroeconomic forecast. The list of increasing and decreasing words are in the appendix. We do not have a horizon distinction here because it is not feasible to distinguish between next quarter and next year with the dictionary method. Implicitly, by using a fixed dictionary we assume the communication rule is stable over the entire sample. However, we are limited in our measure of direction to occurrences of increasing/decreasing words near our variable-relevant words.

quarter and next year macroeconomic forecasts on the same plot for each variable with different horizons.[20] Indeed, we see that the dictionary approach cannot distinguish between, for example, unemployment next quarter or next year. Instead, our regularized regression strategy is flexible enough to work with any numerical output variable, so we can separately estimate the relationship between variables at different horizons and the words in the policy statements.

Another drawback from the dictionary approach is that the researcher only identifies direct discussions of macroeconomic variables where specific topic words are present. For example, the dictionary approach would not connect discussion of increased concerns over inflation to the FOMC's expectations for higher interest rates. A major advantage of the ridge regression with text data is that we allow patterns in the data to indicate word meanings and associations. In fact, the word lists from estimating the communication rule with our proposed methodology (Appendix E) show that language associated with higher interest rates includes discussions of inflation.

7.2.2 ChatGPT Approach

On the other extreme, we ask a generative large language model – specifically, GPT-3.5-Turbo which is the foundation of the current version of ChatGPT (hereafter called GPT) – to back out Fed forecasts from FOMC statements by asking it to guess an exact number. With GPT, the researcher needs to make a prompt which asks a question and possibly provides select examples to guide the text generation. We implement a few-shot learning strategy where we include three examples of statements and their corresponding policy or macroeconomic forecast variable in our code prompt.

The general phrasing of our prompt asks,

[20] For a parallel reason, we are not able to create word lists indicative of shadow rates or 10-year treasury rates separately from the federal funds rate. Accordingly, we do not plot dictionary based measures for those variables.

"Based on the following FOMC statement, what is your best guess of the `<measure>` *the Federal Reserve thinks the* `<variable>` *will be* `<horizon>`*? FOMC statement:* `<statement>`*"*

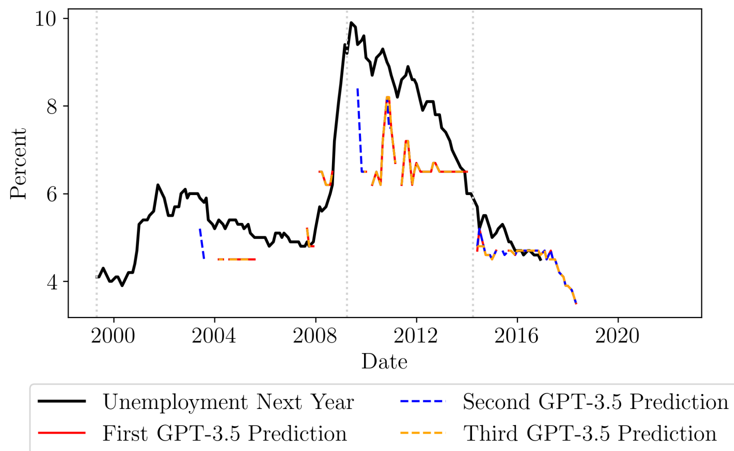
where `<measure>` refers to the units of our variable, `<variable>` refers to the variable name, `<horizon>` indicates whether the variable refers to current, next quarter, or next year, and the final entry, `<statement>`, is for the actual statement itself. The full prompt and description is in Appendix I. With limited training examples, GPT is able to leverage similarities in the text to make predictions.

However, three clear problems arise when using GPT to estimate communication rules: replicability, missing values, and the inability to capture timing. Figure 9 plots the GPT fitted values from repeated prompt submissions. Despite the same prompts, the model produces slightly different responses. This result in variation is useful for a model trying to produce human sounding text, but it is less useful for predicting the same number consistently.

Second, there is a much higher percentage of statements where it says there is insufficient information to provide a numerical prediction. This leads to many missing values in the GPT communication rules. To address these first to problems, we recommend researchers use an average of responses from multiple GPT submissions for regression, classification, or imputation tasks. This approach could then be similar to a researcher averaging over responses from human coders assigning labels to text data.

Another problem is that GPT struggles with distinguishing between different timing or horizons of variables. In Figure 10a, we can see that when GPT is able to provide a response it is correlated with the actual values for the target policy rate. However, in Figure 10b GPT predictions for the target rate next year are the same as those for the current target rate. Unlike for the ridge regressions, for this example we did not remove numbers from the FOMC statements before analysis. It is from those numbers that GPT is able to provide

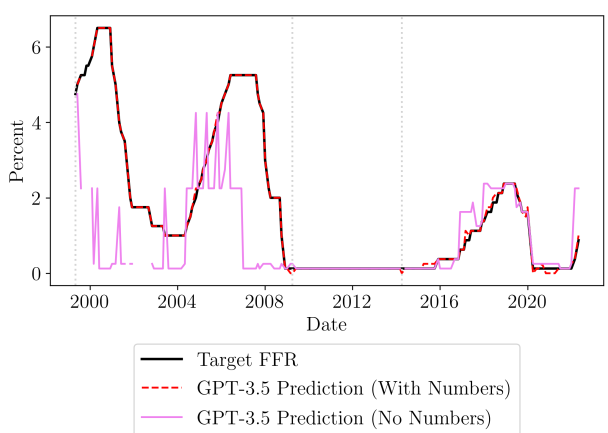
Figure 9: ChatGPT Communication Rule for Unemployment



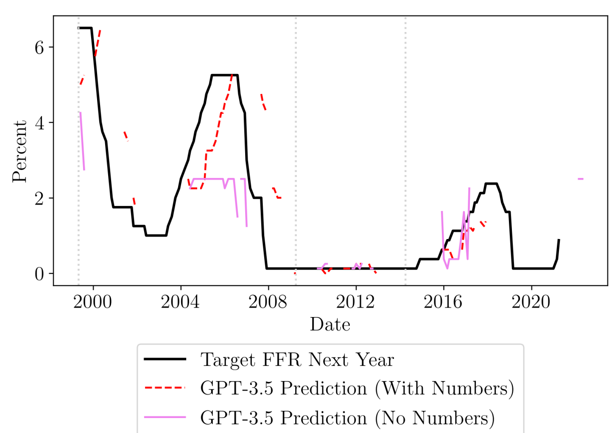
Note: This plots the predicted values from GPT-3.5. We run the prompt multiple times to produce the multiple GPT-3.5 predictions. Even with the same prompt and temperature set to zero, we get different responses. Additionally note that there are many missing values where GPT-3.5 does not provide a number.

Figure 10: Timing and ChatGPT Communication Rules

(a) Target Federal Funds Rate



(b) Target Rate, Next Year



Note: This plots the predicted values from GPT-3.5. The red, dashed line is the prediction for FOMC statements with numbers, and the violet line is the prediction for FOMC statements with numbers removed. With numbers, provides almost identical predictions for the current target rate and the target rate in a year. This indicates the GPT-3.5 model is over weighting the request for information on the target federal funds rate and ignoring the request for different horizons. Without numbers, the fit is much worse.

accurate predictions for the current target rate, as shown with the red, dashed line. After removing numbers, the GPT prediction (violet line) is very inaccurate for this variable as well.

There are two high-level reasons for this lower accuracy. First, we are technically using a generative model for a task where we do not need to generate text. This introduces a mismatch between the model's objective and that of the researcher. Second, is that GPT it is currently limited in how much "training" it can receive in the prompt. Even though it is a strength that GPT is able to keep track of conversations and learn from past text, the memory is still limited. Ultimately, this is evidence of the importance of training and fine-tuning language models specific for the research task and research domain. While some frontier models, like GPT, are powerful, they are not the best tool for every job.

Finally, a disadvantage shared by both of these alternative approaches relative to our own is that there is no time-varying counterpart. Therefore, they are not able to detect changes in communication rules. The dictionary method requires the researcher to pre-specify word meaning prior to analysis and thus implicitly assumes a fixed meaning for words. In other contexts, this may be a desirable feature, but fixing the meaning of words a priori makes it impossible to detect changes in communication rules. A model like GPT is unable to capture a time-varying communication rule for a different reason. Large generative models, like GPT-3.5 and ChatGPT, are implicitly trained on language from over a large span of time that we cannot specify or restrict. That is, language from news and policy statements may be used to train the language model even though they would be "out-of-sample" given a specified window for estimating the model. For all the reasons stated above, we find that our approach strikes the right balance between simplicity and flexibility.

8 Conclusion

Is there a systematic way that the Fed maps their expectations of macroeconomic variables into FOMC statements? To answer this question, we propose a simple procedure based on text analysis and regularized regression to estimate systematic monetary communication rules. We first estimate such rules under the assumption that the Fed has a fixed communication rule for the full sample, and then we reestimate communication rules for various subsamples with a sufficient statistic to detect shifts in the communication rules over time.

Two main results emerge. First, while the language contained in the FOMC statements allows one to back out the Fed's expectation on real variables very well, it does not provide a good fit to the Fed's short-run inflation expectations. This may reflect the notion that the Fed talks more about long-run inflation expectations because that is the horizon where it thinks it can achieve their price stability objective.

Second, while communication rules tend to be stable over time, there is strong evidence for occasional time-variation in the Fed's communication rule. Oftentimes, this is driven by changes in policy that necessitate new language (such as unconventional monetary policy and quantitative easing). Our approach provides a consistent procedure to detect changes in this mapping beyond narrative approaches.

Lastly, we emphasize that the procedure we lay out is general and flexible. As such, it is easy to adopt to a wide variety of settings in which systematic communication may play a role. Corporate earnings calls, political campaigns or announcements of judges all form examples of environments where one can use our method to estimate and study communication rules. Furthermore, our approach is not limited to environments where data is written in English. For example, a researcher could study the monetary policy statements for the Banco Central do Brasil in the original Portuguese without translation, or evaluate survey responses from households in developing countries directly from the original.

References

Acosta, Miguel (2023) "The Perceived Causes of Monetary Policy Surprises," *Working Paper*.

Amador, Manuel and Pierre-Olivier Weill (2010) "Learning from Prices: Public Communication and Welfare," *Journal of Political Economy*, 118 (5), 866–907.

Angeletos, George-Marios and Jennifer La'O (2013) "Sentiments," *Econometrica*, 81 (2), 739–779.

Angeletos, George-Marios and Alessandro Pavan (2007) "Efficient Use of Information and Social Value of Information," *Econometrica*, 75 (4), 1103–1142.

Aruoba, Borağan and Thomas Drechsel (2023) "Identifying Monetary Policy Shocks: A Natural Language Approach," *Working Paper*.

Ashwin, Julian (2021) "The Shifting Focus of Central Bankers," in *Essays in Macroeconomics and Machine Learning*, Chap. 4, 98–139: University of Oxford.

Baker, Scott, Nicholas Bloom, and Steven Davis (2016) "Measuring Economic Policy Uncertainty," *Quarterly Journal of Economics*, 131 (4), 1593–1636.

Barsky, Robert and Eric Sims (2011) "News Shocks and Business Cycles," *Journal of Monetary Economics*, 58 (3), 273–289.

Basu, Susanto, Miles Kimball, Gregory Mankiw, and David Weil (1990) "Optimal Advice for Monetary Policy," *Journal of Money, Credit and Banking*.

Bauer, Michael and Eric Swanson (2023a) "An Alternative Explanation for the "Fed Information Effect"," *American Economic Review*, 113 (3), 664–700.

——— (2023b) "A Reassessment of Monetary Policy Surprises and High-Frequency Identification," *NBER Macroeconomics Annual*, 37, 87–155, Publisher: The University of Chicago Press.

Beaudry, Paul and Franck Portier (2006) "Stock Prices, News, and Economic Fluctuations," *American Economic Review*, 96 (4), 1293–1307.

Bernanke, Ben (2020) "The New Tools of Monetary Policy," *American Economic Association Presidential Address*.

Brown, R. L., J. Durbin, and J. M. Evans (1975) "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society. Series B (Methodological)*, 37 (2), 149–192.

Byrne, David, Robert Goodhead, Michael McMahon, and Conor Parle (2021) "The Central Bank Crystal Ball: Temporal Information in Monetary Policy Communication," *Working Paper*.

Calomiris, Charles, Joanna Harris, Harry Mamaysky, and Cristina Tessari (2022) "Fed Implied Market Prices and Risk Premia," *NBER Working Paper*.

Campbell, Jeffrey, Charles Evans, Jonas Fisher, and Alejandro Justiniano (2012) "Macroeconomic Effects of Federal Reserve Forward Guidance," *Brookings Papers on Economic Activity*, 43 (1), 1–80.

Chow, Gregory (1960) "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 28 (3), 591–605.

Cieslak, Anna, Stephen Hansen, Michael McMahon, and Song Xiao (2021) "Policymakers' Uncertainty," *Working Paper*.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber (2022) "Monetary Policy Communications and Their Effects on Household Inflation Expectations," *Journal of Political Economy*, 130 (6), 1537–1584.

Doh, Taeyoung, Dongho Song, and Shu-Kuei Yang (2022) "Deciphering Federal Reserve Communication via Text Analysis of Alternative FOMC Statements," *FRB Kansas City Working Paper*.

Ehrmann, Michael and Marcel Fratzscher (2005) "How Should Central Banks Communicate?" *ECB Working Paper*.

——— (2007) "Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness?" *Journal of Money, Credit and Banking*, 39 (2-3), 509–541.

Ericsson, Neil (2017) "Predicting Fed forecasts," *Journal of Reviews on Global Economics*.

Gardner, Ben, Chiara Scotti, and Clara Vega (2021) "Words Speak as Loudly as Actions: Central Bank Communication and the Response of Equity Prices to Macroeconomic Announcements," *Journal of Econometrics*, 231 (2), 387–409.

Gáti, Laura (2023) "Talking Over Time – Dynamic Central Bank Communication," *Journal of Money, Credit and Banking*, 55 (5), 1147–1176.

Gáti, Laura and Amy Handlan (2023) "Reputation for Competence," *Working Paper*.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019) "Text as Data," *Journal of Economic Literature*, 57 (3), 535–74.

Gentzkow, Matthew and Jesse Shapiro (2010) "What Drives Media Slant? Evidence from U.S. Daily Newspapers," *Econometrica*, 78 (1), 35–71.

Gertler, Mark and Peter Karadi (2015) "Monetary Policy Surprises, Credit Costs, and Economic Activity," *American Economic Journal: Macroeconomics*, 7 (1), 44–76.

Gnan, Phillipp, Maximilian Schleritzko, Maik Schmeling, and Christian Wagner (2022) "Deciphering Monetary Policy Shocks," *CEPR Working Paper*.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera (2023) "The Voice of Monetary Policy," *American Economic Review*, 113 (2), 548–584.

Goulet Coulombe, Phillippe (2023) "Time-Varying Parameters as Ridge Regressions," *Working Paper*.

Gürkaynak, Refet, Brian Sack, and Eric Swanson (2005) "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements," *International Journal of Central Banking*, 1 (1), 55–93.

Handlan, Amy (2020a) "FedSpeak Matters: FOMC Statements and Monetary Policy Expectations," *Working Paper*.

——— (2020b) "Text Shock and Monetary Suprises: Text Analysis of FOMC Statements with Machine Learning," *Working Paper*.

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

Hansen, Stephen, Michael McMahon, and Andrea Prat (2018) "Transparency and deliberation within the FOMC: a computational linguistics approach," *The Quarterly Journal of Economics*, 133 (2), 801–870.

Hassan, Tarek, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun (2019) "Firm-Level Political Risk: Measurement and Effects," *Quarterly Journal of Economics*, 134 (4), 2135–2022.

Hoerl, Arthur and Robert Kennard (1970) "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12 (1), 55–67.

Husted, Lucas, John Rogers, and Bo Sun (2020) "Monetary Policy Uncertainty," *Journal of Monetary Economics*, 115 (1215), 20–36.

Jarociński, Marek and Peter Karadi (2020) "Deconstructing Monetary Policy Surprises—The Role of Information Shocks," *American Economic Journal: Macroeconomics*, 12 (2), 1–43.

Kalamara, Eleni, Arthur Turrell, Chris Redl, George Kapetanios, and Sujit Kapadia (2022) "Making Text Count: Economic Forecasting Using Newspaper Text," *Journal of Applied Econometrics*, 37 (5), 896–919.

Kapetanios, George and Filip Zikes (2018) "Time-Varying Lasso," *Economics Letters*, 169, 1–6.

Kuttner, Kenneth (2001) "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market," *Journal of Monetary Economics*, 47 (3), 523–544.

Lewis, Daniel, Christos Makridis, and Karel Mertens (2020) "Do Monetary Policy Announcements Shift Household Expectations?" *FRB Dallas Working Paper*.

Liang, Pierre, Vitaly Meursault, Bryan Routledge, and Madeline Scanlon (2022) "PEAD.txt: Post-Earnings Announcement Drift Using Text," *Working Paper*.

Lunsford, Kurt (2020) "Policy Language and Information Effects in the Early Days of Federal Reserve Forward Guidance," *American Economic Review*, 110 (9), 2899–2934.

Lüdering, Jochen and Peter Tillmann (2020) "Monetary Policy on Twitter and Asset Prices: Evidence from Computational Text Analysis," *The North American Journal of Economics and Finance*, 51, 100875.

McQueen, Grant and Vance Roley (1993) "Stock Prices, News, and Business Conditions," *The Review of Financial Studies*, 6 (3), 683–707.

Morris, Stephen and Hyun Song Shin (2002) "Social Value of Public Information," *American Economic Review*, 92 (5), 1521–1534.

Nakamura, Emi and Jón Steinsson (2018) "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect," *Quarterly Journal of Economics*, 133 (3), 1283–1330.

Ou, Shengliang, Donghai Zhang, and Renbin Zhang (2022) "The Return of Greenspan: Mumbling with Great Incoherence," *Working Paper*.

Piazzesi, Monika (2005) "Bond Yields and the Federal Reserve," *Journal of Political Economy*, 113 (2), 311–344.

Romelli, Davide and Hamza Bennani (2021) "Disagreement inside the FOMC: New Insights from Tone Analysis," *Working Paper*.

Romer, Christina and David Romer (2004) "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94 (4), 43.

Shapiro, Adam and Daniel Wilson (2021) "Taking the Fed at its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis," *Review of Economic Studies*, 89 (5), 2768–2805.

Stekler, Herman and Hilary Symington (2016) "Evaluating Qualitative Forecasts: The FOMC Minutes, 2006–2010," *International Journal of Forecasting*, 32 (2), 559–570.

Swanson, Eric and Vishuddhi Jayawickrema (2023) "Speeches by the Fed Chair Are More Important Than FOMC Announcements: An Improved High-Frequency Measure of U.S. Monetary Policy Shocks," *Working Paper*.

Tibshirani, Robert (1996) "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58 (1), 267–288.

Wu, Jing Cynthia and Fan Dora Xia (2020) "Negative Interest Rate Policy and the Yield Curve," *Journal of Applied Econometrics*, 35 (6), 653–672.

Zou, Hui and Trevor Hastie (2005) "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67 (2), 301–320.