

DISCUSSION PAPER SERIES

DP17133
(v. 6)

IDENTIFYING MONETARY POLICY SHOCKS: A NATURAL LANGUAGE APPROACH

Boragan Aruoba and Thomas Drechsel

**INTERNATIONAL MACROECONOMICS
AND FINANCE AND MONETARY
ECONOMICS AND FLUCTUATIONS**

CEPR

IDENTIFYING MONETARY POLICY SHOCKS: A NATURAL LANGUAGE APPROACH

Boragan Aruoba and Thomas Drechsel

Discussion Paper DP17133
First Published 24 March 2022
This Revision 13 January 2023

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- International Macroeconomics and Finance
- Monetary Economics and Fluctuations

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Boragan Aruoba and Thomas Drechsel

IDENTIFYING MONETARY POLICY SHOCKS: A NATURAL LANGUAGE APPROACH

Abstract

This paper proposes a novel method for the identification of monetary policy shocks. By applying natural language processing techniques to documents that staff economists at the Federal Reserve prepare for FOMC meetings, we capture the information set of the committee at the time of policy decisions. We verify econometrically that the language contains valuable information beyond what is incorporated in the staff's numerical forecasts. Using machine learning techniques, we then predict changes in the target interest rate conditional on the committee's information set and obtain a measure of monetary policy shocks as the residual. We find that the dynamic responses of macro variables to our identified shocks are consistent with the theoretical consensus. A real-time application of our method interprets the 2022 rate hikes as almost entirely systematic.

JEL Classification: C10, E31, E32, E52, E58

Keywords: N/A

Boragan Aruoba - aruoba@umd.edu
University Of Maryland

Thomas Drechsel - drechsel@umd.edu
University Of Maryland and CEPR

Identifying Monetary Policy Shocks: A Natural Language Approach*

S. Borağan Aruoba

Thomas Drechsel

University of Maryland

University of Maryland, CEPR

January 13, 2023

Abstract

This paper proposes a novel method for the identification of monetary policy shocks. By applying natural language processing techniques to documents that staff economists at the Federal Reserve prepare for FOMC meetings, we capture the information set of the committee at the time of policy decisions. We verify econometrically that the language contains valuable information beyond what is incorporated in the staff's numerical forecasts. Using machine learning techniques, we then predict changes in the target interest rate conditional on the committee's information set and obtain a measure of monetary policy shocks as the residual. We find that the dynamic responses of macro variables to our identified shocks are consistent with the theoretical consensus. A real-time application of our method interprets the 2022 rate hikes as almost entirely systematic.

Keywords: Monetary policy; Federal Reserve; Greenbook; Natural Language Processing; Machine learning.

JEL Classification: C10; E31; E32; E52; E58.

*We would like to thank Juan Antolin-Diaz, Michael Bauer, Pierre De Leo, Burcu Duygan-Bump, Marty Eichenbaum, Simon Freyaldenhoven, Friedrich Geiecke, Yuriy Gorodnichenko, Tarek Hassan, Matteo Iacoviello, Diego Kaenzig, Guido Kuersteiner, Michele Lenza, Kevin Lee, Xiang Li, Michael McMahon, Vitaly Meursault, Emi Nakamura, Ivan Petrella, Chris Redl, Frank Schorfheide, Minchul Shin, Lumi Stevens, Fabian Winkler, Chris Wolf and Jonathan Wright, as well as seminar and conference participants at the Federal Reserve Board, the Federal Reserve Bank of Philadelphia, the Federal Reserve Bank of Cleveland, the Bank of England, the Bank for International Settlements, the Reserve Bank of Australia, the Central Bank of Chile, FGV EESP Sao Paulo, the Glasgow Workshop on Recent Advances in Econometrics, the LSE Conference in Honor of Wouter Den Haan, the DC-MD-VA Econometrics Workshop at Georgetown University, the University of Hamburg, and the University of Maryland for helpful comments. Eugene Oue, Danny Roth, Mathias Vesperoni, Eric Youmans and Jialing Zhang provided excellent research assistance. Contact: Department of Economics, University of Maryland, Tydings Hall, College Park, MD 20742, USA; aruoba@umd.edu and drechsel@umd.edu. We make our estimated monetary policy shocks, the Fed sentiment time series, the FOMC composition data set, as well as the Greenbook forecast errors available to other researchers at <http://econweb.umd.edu/drechsel/research.html>.

1 Introduction

To study how monetary policy affects the economy, macroeconomists isolate changes in interest rates that are not systematic responses to economic conditions, but instead occur in a plausibly exogenous way. This paper proposes a novel method for the identification of such monetary policy shocks.

Our starting point is [Romer and Romer \(2004\)](#)'s seminal idea that exogenous movements in the Federal Funds Rate (FFR) are the difference between observed and systematic changes in the target FFR, where systematic changes are based on economic forecasts available to policy makers. These authors run a linear regression of the change in the FFR target on forecasts of inflation, output and unemployment contained in the "Greenbook" documents prepared by staff economists at the Federal Reserve for Federal Open Market Committee (FOMC) meetings. They retrieve a measure of monetary policy shocks as the residual from this regression.

We propose an identification approach that also exploits the information in documents prepared for the FOMC, but our methodology incorporates a larger set of information contained in these documents, including numerical forecasts and *human language*. We implement the approach with natural language processing and machine learning techniques, bringing the identification of monetary policy shocks into the growing literature that applies such techniques to Federal Reserve documents (see e.g. [Hansen, McMahon, and Prat, 2018](#)). Importantly, we show econometrically that the language in the documents produced by Fed staff contains valuable information beyond what is incorporated in the staff's numerical forecasts. We also develop a real-time version of our method to study recent Fed decisions, which is not possible with the original [Romer and Romer \(2004\)](#) procedure due to the forecast publication lag.

The identification procedure we propose estimates monetary policy shocks as the residuals from a prediction of changes in the FFR target using (i) all available numerical forecasts in the documents that Fed economists prepare for the FOMC; (ii) a comprehensive summary of the verbal information in the documents, including lagged information from documents prepared for previous meetings; and (iii) nonlinearities in (i) and (ii). (i) includes the original forecasts used by [Romer and Romer \(2004\)](#) but we expand the set to include additional variables that Fed economists provide forecasts for, such as industrial production, housing and government spending. To obtain (ii), we first identify the most commonly mentioned economic terms in the documents. This results in a set of 296 single or multi-word expressions, such as

“inflation,” “economic activity” or “labor force participation.” We then construct sentiment indicators that capture the degree to which these concepts are associated with positive or negative language, following work by [Hassan, Hollander, van Lent, and Tahoun \(2020\)](#). The documents we use are carefully crafted by Fed economists, with precise wording and consistency in language over time, so this type of natural language processing is particularly applicable. Our collection of sentiment time series paints a rich picture of the historical assessment of economic conditions by Fed staff.

A regression of FFR target changes on *(i)*, *(ii)* and *(iii)* is infeasible given that there are more regressors than observations. To overcome this issue, we resort to machine learning techniques. Specifically, we employ a ridge regression to predict changes in the FFR target using our large set of forecast and sentiment regressors. This choice is guided by recent insights about alternative types of machine learning methods for economic data ([Giannone, Lenza, and Primiceri, 2022](#)). A ridge regression minimizes the residual sum of squares plus an additional term that penalizes squared deviations of each coefficient from zero to achieve shrinkage. We select the ridge penalty parameter using k -fold cross-validation, a standard way in the machine learning literature to validate a model’s predictive ability in alternating subsets of the data.

Before applying our method to estimate monetary policy shocks and the associated macroeconomic responses, we assess the informational value added of our sentiment indicators. As pointed out by [Cochrane \(2004\)](#) in a conference discussion of [Romer and Romer \(2004\)](#), it would be enough to orthogonalize FFR changes with respect to the staff’s forecasts alone, if the forecast for a variable of interest incorporates all available information efficiently. While this argument is based on the Greenbook forecasts corresponding to a conditional mean expectation, we present evidence that they are better interpreted as modal forecasts. Moreover, to show that the language content of the documents reflects valuable information beyond what is incorporated in the numerical staff forecasts, we demonstrate that our sentiment indicators predict the errors of these forecasts. For example, the forecast error for the unemployment rate is predictable with one of the 296 sentiment indicators alone, across various horizons and to an economically significant degree. FFR target changes should thus be orthogonalized with respect to the additional verbal information in order to recover the response of the unemployment rate to a monetary policy shock.

We then apply our method and discuss five sets of findings. First, we examine the relative contribution of systematic and exogenous variation in the FFR target. A linear regression that contains only numerical forecasts for output, inflation and the

unemployment rate yields an R^2 of around 0.5, suggesting that half of the variation in the FFR target is attributed to systematic policy, while the other half is included in the monetary policy shock. The R^2 of our ridge regression is 0.94, implying that the exogenous component of FFR changes is reduced almost ten fold from 50% to 6%, when a larger set of forecasts, verbal sentiments, as well as nonlinearities are included. While our analysis of forecast error predictability already supports the view that more information about the systematic component of monetary policy should be included, a high R^2 is also economically appealing. Macroeconomists typically think of monetary policy decisions to be largely taken systematically, with a small role for exogenous shocks, as discussed for example by [Leeper, Sims, and Zha \(1996\)](#).

Second, we inspect the economic drivers of systematic changes in US monetary policy. We study which groups of variables have the highest economic and statistical significance in explaining FFR changes. We find numerical forecasts and sentiments surrounding real economic activity are important. We also identify Fed economists' sentiment about international economic developments as an important factor in explaining interest rate changes. There is some role, though limited, for sentiments surrounding inflation and credit conditions.

Third, we provide an interpretation of what estimated monetary policy shocks capture. We do so by closely analyzing the discussion that took place among the FOMC in meetings where the estimated shock is large in magnitude. It turns out that in these episodes the FOMC made decisions based on considerations not directly related to the staff's analysis, in particular long-run credibility concerns. For example, in the November 1994 meeting, the material prepared by the staff economists suggests a 50 basis point rate hike. However, in the FOMC meeting Chairman Alan Greenspan advocates a 75 basis point hike, arguing that "a mild surprise would be of significant value." Our procedure estimates almost the entire 25 basis point difference to be an unsystematic contractionary shift in policy.

Fourth, with our novel measure of monetary policy shocks at hand, we study impulse response functions (IRFs) of macro variables and compare them to canonical results in the literature. We estimate a state-of-the-art Bayesian vector autoregression (BVAR), in which our monetary policy shocks are included as an exogenous variable.¹ We find that a monetary policy tightening leads to a reduction in economic activity, a fall in the price level, an increase in bond premia and a decline in stock prices. These findings are in line with what economic theory predicts. Notably, following

¹We also use local projections ([Jordà, 2005](#)) as an alternative methodology and find similar results.

a tightening there is a relatively swift decline in real output, while the reduction in the price level builds up sluggishly over time. We also show that IRFs resulting from shocks computed using the original Romer-Romer methodology lead to responses not in line with the theoretical consensus, and discuss potential interpretations of what confounds this measure. Taken together, we conclude that natural language processing and machine learning deliver a cleanly identified estimate of monetary policy shocks.

Finally, we apply a real-time version our method to examine the FOMC's decisions in 2022. The Fed Tealbooks and associated forecasts are made available to the public only with a 5-year delay.² However, the Beigebooks are publicly available prior to every FOMC meeting. These summarize regional economic conditions in each Federal Reserve district, so extracting sentiment indicators from their text provides at least a limited proxy for the FOMC's information set. Leveraging the Beigebooks is not possible in the original [Romer and Romer \(2004\)](#) approach, as they do not contain numerical forecasts. We find that our method interprets the 2022 rate hikes as almost entirely systematic Fed policy. Over the course of that year, there are small contractionary shocks that cumulate to 19 basis points.

Literature. Our work contributes to three branches of research. The first is the literature that seeks to identify monetary policy shocks, most notably the seminal work of [Romer and Romer \(2004\)](#). Their method is still widely used, see [Tenreyro and Thwaites \(2016\)](#), [Coibion et al. \(2017\)](#) and [Wieland and Yang \(2020\)](#) for recent applications.³ There is a wide array of other approaches to identifying monetary policy shocks. One approach uses structural vector autoregressions (SVARs) identified in different ways.⁴ Another one is based on high-frequency (HF) strategies to elicit surprises in interest rates around FOMC announcements, e.g. [Gürkaynak, Sack, and Swanson \(2005\)](#) and [Gertler and Karadi \(2015\)](#).⁵ A survey of these different approaches

²"Tealbook" is a more recent labeling for the documents that the staff prepares for FOMC meetings. We sometimes use the terms Tealbook and Greenbook interchangeably, but in description of our method we precisely define which types of documents we process over which sample periods.

³[Bachmann, Gödl-Hanisch, and Sims \(2022\)](#) suggest summarizing the Fed's information set using forecast errors. The original Romer-Romer methodology has also been applied to other countries, e.g. [Cloyne and Hürtgen \(2016\)](#) use it for the UK and [Holm, Paul, and Tischbirek \(2021\)](#) for Norway.

⁴Identification in SVARs is obtained e.g. through zero restrictions ([Christiano, Eichenbaum, and Evans, 1999](#)), sign restrictions ([Uhlig, 2005](#)), or narrative sign restrictions ([Antolin-Diaz and Rubio-Ramirez, 2018](#)). [Coibion \(2012\)](#) compares SVAR approaches to that of [Romer and Romer \(2004\)](#).

⁵One challenge for HF approaches is the "Fed information effect", see e.g. [Romer and Romer \(2000\)](#), [Campbell et al. \(2012\)](#) and [Nakamura and Steinsson \(2018\)](#). [Jarocinski and Karadi \(2020\)](#) and [Miranda-Agrippino and Ricco \(2021\)](#) separate HF surprises in market interest rates between pure monetary and informational shocks. As our method does not rely on a HF identification strategy using market interest

is provided by [Ramey \(2016\)](#). We contribute to the literature on identifying monetary policy shocks by applying natural language processing and machine learning to achieve identification through a large set of information in economic data and text.⁶

The second branch of research we contribute to is a fast-growing literature that applies textual analysis or machine learning to documents produced by the Federal Reserve. [Hansen, McMahon, and Prat \(2018\)](#) show that communication in the FOMC changes after public transparency increased in the early 1990's. [Hansen and McMahon \(2016\)](#) investigate the impact of Fed communication on macroeconomic variables. Similar to us, [Sharpe, Sinha, and Hollrah \(2020\)](#) carry out sentiment analysis using documents produced by Fed economists and a pre-defined dictionary. Different from us, these authors construct a single sentiment index rather than sentiments for individual economic concepts (or 'aspect-based' sentiments). [Shapiro and Wilson \(2021\)](#) use sentiment analysis on FOMC transcripts, minutes, and speeches in order to draw inference about central bank objectives.⁷ [Cieslak and Vissing-Jorgensen \(2020\)](#) employ textual analysis on FOMC documents to understand if monetary policy reacts to stock prices.⁸ [Cieslak et al. \(2021\)](#) construct text-based measures of policy makers' uncertainty. None of the aforementioned studies identify monetary policy shocks, which is the goal of our methodology. To the best of our knowledge, two complementary papers use textual analysis on Fed documents for purposes similar to ours. [Handlan \(2020\)](#) applies textual analysis to FOMC statements and internal meeting materials to build a "text shock" that separates the difference between forward guidance and current assessment of the FOMC in driving FFR futures since 2005. We instead estimate a more conventional series of monetary policy shocks over several decades. [Ochs \(2021\)](#) uses sentiment analysis on publicly available FOMC documents to extract surprise changes in monetary policy from the point of view of private agents. We orthogonalize interest rates changes with respect to the central bank's information set as captured by the documents prepared internally for the FOMC. In that sense, our procedure is closer to the original [Romer and Romer \(2004\)](#) approach to estimating

rates, it is not contaminated by the Fed information effect.

⁶Our emphasis on a large information set has parallels to [Bernanke, Boivin, and Eliasziw \(2005\)](#) who incorporate many time series in a factor-augmented VAR (FAVAR), but do not consider text.

⁷[Acosta \(2022\)](#) studies how the FOMC responded to calls for transparency. Further papers on Fed language include [Lucca and Trebbi \(2009\)](#) and [Doh, Song, and Yang \(2022\)](#).

⁸Relatedly, [Peek, Rosengren, and Tootell \(2016\)](#) apply textual analysis to FOMC transcripts to understand how the FOMC reacts to financial stability concerns. Several others study the reverse, whether financial markets react to Fed language. [Gardner, Scotti, and Vega \(2021\)](#) study the response of equity prices to publicly released FOMC statement using sentiment analysis. [Gorodnichenko, Pham, and Talavera \(2021\)](#) use deep learning techniques to capture emotions in FOMC press conference.

monetary policy shocks. Natural language processing and machine learning enable us to capture the central bank’s information set in a comprehensive way.

The third branch of research we contribute to studies the Fed’s Greenbook forecasts. Important contributions include [Romer and Romer \(2000\)](#), [Faust and Wright \(2009\)](#), [Nakamura and Steinsson \(2018\)](#). This literature points to the high predictive quality of the Greenbook forecasts and the Fed’s informational advantage over private sector forecasts. Our new finding is that there is useful information, expressed verbally in the documents produced by staff, that can explain Greenbook forecast errors on average. This finding makes it critical to use that verbal information in the estimation of monetary policy shocks. We argue the Greenbook forecasts are best interpreted as modal, and sentiment indicators incorporate information about asymmetric risks.

Structure of the paper. Section 2 lays out our method. Section 3 assesses the information content of the sentiment indicators. Section 4 discusses the results of our identification procedure, including the contribution of systematic policy and our estimated shocks. Section 5 presents our results on the responses of macroeconomic variables to monetary policy shocks. In Section 6 we apply a real-time version of our method to interpret the interest rate hikes in 2022. Section 7 concludes.

2 A new method to identify monetary policy shocks

This section first provides the motivation for our approach, explains the relevant institutional setting, and lays out the main idea of our methodology. It then gives an in-depth description of the full shock identification procedure.

2.1 Motivation, institutional setting, and main idea

Definition of monetary policy shocks. The challenge of studying how monetary policy affects the economy is the fact that policy is set endogenously, by taking current economic conditions and the outlook for the economy into account. An influential literature has addressed this challenge by isolating monetary policy shocks, changes in monetary policy that are orthogonal to the information that policy makers react to. In this line of work, the central bank is typically assumed to set its policy instrument s_t , according to a rule

$$s_t = f(\Omega_t) + \varepsilon_t, \tag{1}$$

where Ω_t is the information set of the central bank, $f(\cdot)$ is the systematic component of monetary policy, and ε_t is the monetary policy shock. The systematic component of policy is endogenous, so the only way to understand the causal effect of monetary policy on the economy is to consider changes in ε_t . The formalization of the endogeneity challenge in equation (1) is the explicit or implicit starting point of most studies in the literature. There are different ways to estimate ε_t with data, e.g. using structural vector autoregressions (SVARs). A survey is provided by [Ramey \(2016\)](#).

The narrative approach. One approach to estimating monetary policy shocks, following the influential idea of [Romer and Romer \(2004\)](#), is to run a linear regression

$$\Delta i_t = \alpha + \beta i_{t-1} + \gamma \mathbf{X}_t + \varepsilon_t^{RR}, \quad (2)$$

where i_t is the FOMC’s FFR target, and \mathbf{X}_t contains the forecasts of the US economy that the Fed has at its disposal at time t .⁹ In their original work, these include forecasts of output growth, inflation, and the unemployment rate of various horizons, and enter in both levels and changes. Running regression (2) results in the residuals $\hat{\varepsilon}_t^{RR}$, which provide an empirical measure for ε_t in (1).

Two key assumptions underlie the above approach. First, the forecasts included in \mathbf{X}_t need to be a good proxy for the information set Ω_t that is relevant for the central bank’s decisions. The FOMC reviews a large amount of detailed information on the economic and financial conditions of the US economy, prepared by staff economists as part of different types of confidential documents. These documents contain numerical forecasts but also many pages of text. The numerical forecasts only provide a good proxy (or “sufficient statistic”) for the information set if they correspond to the FOMC’s mean expectation conditional on incorporating all this other information efficiently. The second assumption is that the mapping $f(\cdot)$ from the information to decisions is well captured by a linear relationship.

Main idea behind our approach. We revive the method championed by [Romer and Romer \(2004\)](#), and refine it along two dimensions. To do so, we exploit advances in natural language processing (NLP) and machine learning (ML) techniques.

The first dimension relates to the proxy for the information set Ω_t . The documents produced around FOMC meetings contain a vast amount of verbal information,

⁹In the notation we use in this paper, the time subscript t typically runs at the frequency of FOMC meetings. Where applicable, we transform the frequency of the shocks to monthly.

in addition to numerical forecasts. They are crafted by the Fed staff in a careful and analytical manner, with consistency in language over time, so natural language processing techniques are likely well suited for extracting valuable information from them. Our premise is that the human language in which Fed economists describe the subtleties around the economic outlook provides valuable information beyond what is contained in purely numerical predictions. We validate this premise also with a formal econometric test in Section 3. We then incorporate this information using NLP to fully capture systematic component of monetary policy.

The second dimension along which we refine the approach is through the potential presence of nonlinearities in $f(\cdot)$. We do so by including higher order terms in our econometric counterpart of (1). Since considering numerical forecasts, verbal information, as well as nonlinearities requires us to include a large number of variables on the right hand side of a regression model, we apply ML techniques to cope with the dimensionality of the problem. We then estimate monetary policy shocks as the residuals from a prediction of changes in the FFR using a large amount of numerical and verbal information.

2.2 Step-by-step description of our method

Our procedure to estimate monetary policy shocks consists of the following steps. First, we process the text of relevant FOMC meeting documents. Second, we identify frequently discussed economic concepts in these documents. Third, we construct sentiment indicators for each economic concept. Fourth, we run a regression that includes sentiment indicators and numerical forecasts, both linearly and nonlinearly.

Step 1: Process FOMC documents

In FOMC meetings, scheduled 8 times per year, the committee meets to discuss monetary policy decisions.¹⁰ We first retrieve historical documents associated with FOMC meetings from the website of the [Federal Reserve Board of Governors](#). We start with the meeting on October 5, 1982, in order to capture the period over which the Fed targeted the FFR as their main policy instrument, according to [Thornton \(2006\)](#).¹¹ Most of FOMC meeting documents are available with a 5-year lag, so the latest document

¹⁰There are also unscheduled meetings or conference calls during which the FOMC makes policy decisions. Most of these are excluded from our analysis because no new documents are prepared.

¹¹We vary the starting date for robustness, for example to include the entire Volcker period (starting in 1979) or to only begin with the Greenspan period (starting in 1987).

currently available is for the last FOMC meeting of 2016.¹² In our main analysis, we therefore process documents through to 2016, although in the regression of step 4 of our procedure, we use the period before the zero lower bound (ending with the FOMC meeting on October 29, 2008) to avoid running a regression with many zeros on the left hand side. In a real-time version of our method, we focus on a subset of documents that is available contemporaneously.

For each FOMC meeting, several different document types are available. We include the following documents: *Greenbook 1* and *Greenbook 2* (until June 2010), *Tealbook A* (after June 2010), *Redbook* (until 1983), *Beigebook* (after 1983).¹³ We focus on these documents to capture the Fed's information set at the onset of the meeting. In particular, we do not include the meeting minutes, transcripts or announcements because these might capture the decision process rather than the information set of policy makers going into an FOMC meeting. Our choice results in 772 pdf files for 276 meetings (630 files for 210 meetings before the zero lower bound), containing tens of thousands of pages of text and numbers.

For each document, we read its raw text into a computer and process it as follows. We remove stop words ("the", "is", "on", etc.); we remove numbers (that are not forecasts, e.g. dates, page numbers); we remove "erroneous" words. We then retrieve *singles*, *doubles* and *triples*. Singles are individual words. Doubles and triples are joint expressions that are not interrupted by stop words or sentence breaks. For example, "... consumer price inflation ..." is a triple, and also gives us two doubles ("consumer price" and "price inflation") and three singles ("consumer", "price" and "inflation"). "... inflation and economic activity ..." gives us three singles and one double. "... for inflation. Activity on the other hand..." only gives us three singles ("inflation", "activity" and "hand").¹⁴ For the 276 meetings there are roughly 18,000 singles, 450,000 doubles, and 600,000 triples. For comparison, the Oxford English dictionary has roughly 170,000 single words. We then calculate the frequency at which each single, double and triple occurs for each meeting date and each document.

¹²Our method exploits the information set available to the FOMC, and does not require the same information to be publicly available at the time of the policy decision.

¹³The *Greenbooks*, later replaced by the *Tealbooks*, contain staff analysis for the US economy. We exclude the *Bluebook* and *Tealbook B* as these contain different hypothetical scenario analyses, where outcomes conditional on alternative policy actions are described, and which we judged might obfuscate our sentiment extraction. The *Redbooks* (until 1983) / *Beigebooks* (from 1983) discuss economic conditions for each Federal Reserve district. We use the *Beigebooks* by themselves in our real-time application to recent Fed meetings in Section 6.

¹⁴We also added one quadruple: "money market mutual funds."

Figure 1: ECONOMIC CONCEPTS MENTIONED FREQUENTLY IN FOMC DOCUMENTS



Notes. Word cloud of the 75 most frequently mentioned economic concepts in documents prepared by Federal Reserve Board economists for FOMC meetings between 1982 and 2016. The size of concept reflects the frequency with which it occurs across the documents.

Step 2: Identify frequently used economic concepts

We rank all singles, doubles and triples from Step 1 by their total frequency of occurrence over the whole time period. We then start from the most frequent ones, move downwards and select those singles, doubles and triples that are economic concepts, such as “credit”, “output gap”, or “unit labor cost”.¹⁵ Sometimes there are economic concepts that overlap across singles, doubles and triples. For example, should “commercial real estate” be an economic concept or just “real estate” or both separately? To address this, we follow a precise selection algorithm that we describe in Appendix A. Our selection procedure results in 296 economic concepts. Figure 1 shows a word cloud for the 75 most frequent economic concepts, where the size of the concepts reflects its frequency across the documents.

¹⁵Both authors went through this selection independently and then discussed any disagreement case by case. When moving down the frequency ranking, we stop at a generous lower bound, for example one mention on average per meeting for triples. We discuss the general advantages of imposing some judgmental restrictions at the end of Section 2.

Table 1: EXAMPLES OF WORDS ASSOCIATED WITH POSITIVE AND NEGATIVE SENTIMENT

Positive sentiment	Negative sentiment
adequate	adversely
advantage	aggravate
benefit	bad
boost	burdensome
confident	collapse
conducive	concerning
desirable	decline
diligent	deficient
encouraging	eroded
excellent	exacerbate
...	...

Notes. Selected examples of words that are classified as expressing positive or negative sentiments in our improved version of the dictionary of [Loughran and McDonald \(2011\)](#). The total number of classified words is 2,882.

Step 3: Construct sentiment indicators for each economic concept

For each of the 296 individual economic concepts, we apply a method to capture the *sentiment* surrounding them, inspired by [Hassan, Hollander, van Lent, and Tahoun \(2020\)](#). For each occurrence of a concept in a document, we check whether any of the 10 words mentioned before and after the concept’s occurrence are associated with positive or negative sentiment.¹⁶ This classification builds on the dictionary of positive and negative terms in [Loughran and McDonald \(2011\)](#). This is a widely used dictionary in the literature, which is especially constructed for financial text, so it should already be reasonably suitable for the economic content discussed in the Fed documents. For our application we make several modifications to this dictionary.¹⁷ Based on our improved dictionary, each positive word then adds a score of +1 and each negative word a score of -1 towards the sentiment of the concept. Table 1 provides a few examples of

¹⁶The 10-word distance here refers to words after the pre-cleaning steps of the documents, and not words in a raw sentence. Further below, we explore robustness with sentiment indicators constructed using an alternative distance of 5 words. We also show that constructing sentiments based on positive and negative words within the same sentence, rather than inside a 10-word window, yields time series that are highly correlated with the ones we use. See Appendix C for two examples.

¹⁷We modify the [Loughran and McDonald \(2011\)](#) dictionary along two dimensions. First, we enhance the list of words by adding terms typical for Fed language, such as *tightening*. We also add some more variations of existing terms, for example the original dictionary contains *boom* and *booming*, and we add *booms* and *boomed*. Second, we remove some terms, either because they are among our selected economic concepts, such as *unemployment* and *unemployed*. Or because we think they should not necessarily be interpreted as positive or negative in the context of the Fed’s analysis, such as the term *unforeseen*.

positive and negative words. For each of our concepts, we then sum up the sentiment scores within the documents associated with an FOMC meeting, and scale by the total number of words in the documents to obtain a sentiment indicator. The final product of this procedure is a sentiment indicator time series for each economic concept, where the time variation is across FOMC meetings. For the purpose of entering these indicators in a regression, we also standardize all indicators to be mean zero and have standard deviation of one.

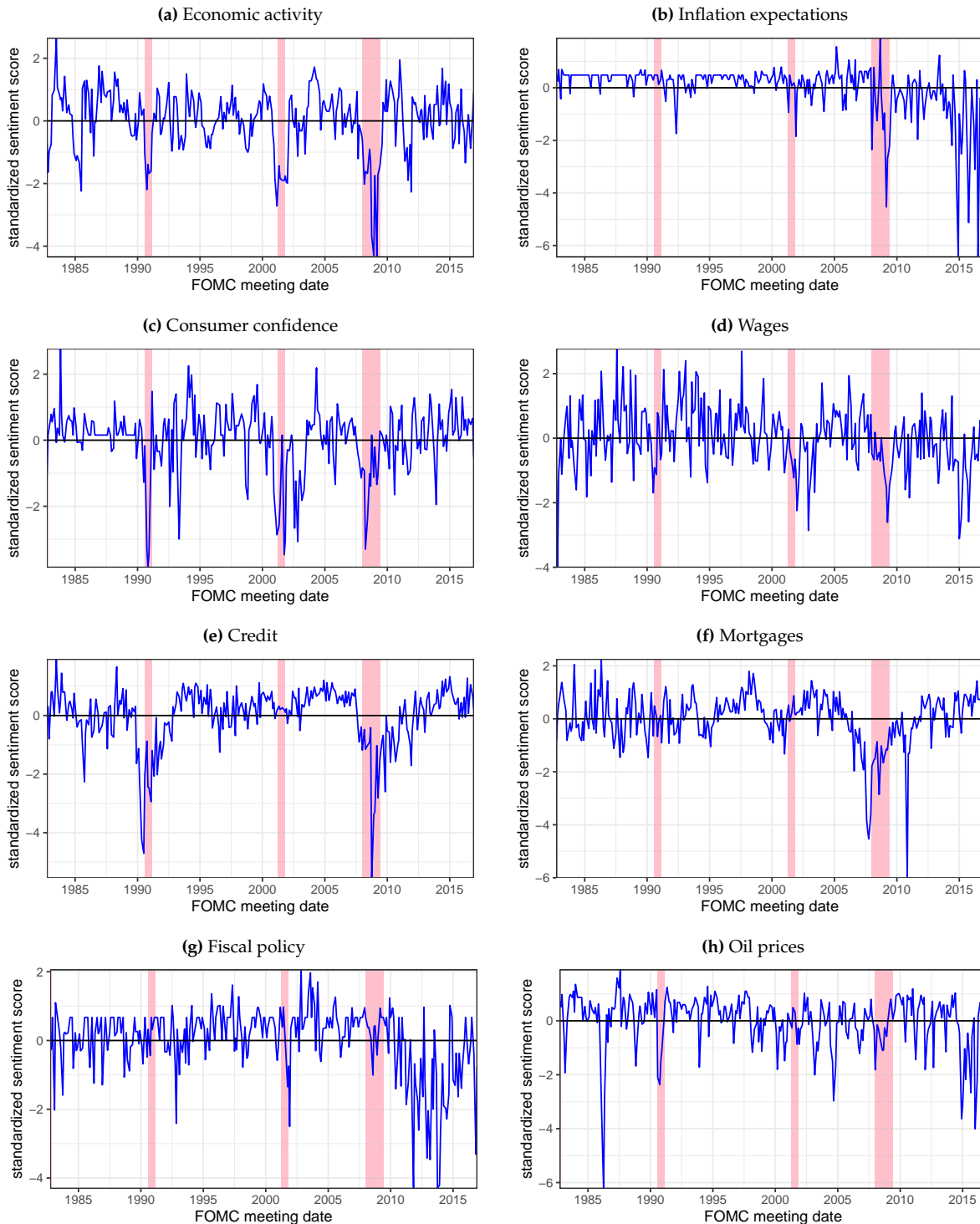
Figure 2 presents the sentiment indicators for some selected economic concepts. These indicators are standardized, but not otherwise smoothed or filtered after we compute them from the text. They clearly display meaningful variation. For example, Panel (a) shows that the sentiment surrounding “economic activity” falls sharply in recessions. Furthermore, comparisons across concepts reveal meaningful information about the Fed economists’ view on the nature of different recessions. For example, the sentiment around credit appears to fall both in the 1991 recession and the Great Recession of 2007-09, while negative sentiment surrounding mortgages plays a role primarily in the Great Recession and its aftermath (see Panels (e) and (f)). Another insight coming from the figure is that some concepts gain importance over time. For example, the sentiment around inflation expectations in Panel (b) moves relatively little for most of the sample, but displays larger volatility since the 2000’s. While we use the full set of 296 sentiment indicators in a multivariate econometric analysis, a by-product of our analysis is a rich descriptive picture of the Fed’s assessment of various aspects of the US economy over the last few decades. Appendix B contains sentiment plots for additional economic concepts.

Step 4: Specify and estimate the empirical model

Nonlinear specification using forecasts and sentiments. Our empirical counterpart of equation (1) includes the Fed’s policy instrument on the left hand side, and both numerical forecasts and sentiment indicators from FOMC documents on the right hand side. Both sets of variables can enter non-linearly. Formally, we define

$$\Delta i_t = \alpha + \beta i_{t-1} + \Gamma(\tilde{\mathbf{X}}_t, \mathbf{Z}_t) + \varepsilon_t^*. \quad (3)$$

Figure 2: SELECTED SENTIMENT INDICATORS



Notes. Sentiment indicators for a selection of economic concepts discussed in FOMC meeting documents, out of our full list of 296. The sentiments are constructed using the dictionary of positive and negative words in financial text of [Loughran and McDonald \(2011\)](#). Each indicator is standardized across the sample. Shaded areas represent NBER recessions.

Δi_t are changes in the FOMC’s FFR target, which for simplicity we mostly refer to as just the FFR.¹⁸ $\widetilde{\mathbf{X}}_t$ contains augmented set of forecasts that Fed economists produce, which includes additional production, investment, housing and government spending variables relative to \mathbf{X}_t in (2).¹⁹ Following [Romer and Romer \(2004\)](#)’s specification, we enter forecasts in levels and first differences, across several forecast horizons, which amounts to 132 forecast time series. \mathbf{Z}_t contains our 296 sentiment indicators. We also allow 4 lags of the sentiment indicators to enter, as the path of the economy, which includes recent historical performance, may have an influence on how the current state of the economy translates in to policy changes.²⁰ $\Gamma(\cdot)$ is a nonlinear mapping. In our main analysis, we specify this as a second-order polynomial. Together with the level of the FFR, i_{t-1} , which we also allow to enter quadratically, (3) includes 3,226 variables on the right hand side. We analyze different lag structures and alternative nonlinear specifications of $\Gamma(\cdot)$ for robustness.

Ridge regression. Our sample from October 1982 to October 2008 captures 210 FOMC meetings. Therefore, an ordinary least squares (OLS) regression with several thousands of regressors is infeasible. To overcome this issue, we resort to ML techniques. Specifically, we employ a ridge regression to estimate (3). The idea of a ridge regression, which was first introduced by [Hoerl and Kennard \(1970\)](#), is to minimize the residual sum of squares and an additional term that penalizes the squared deviations of each regression coefficient from zero. Formally, in the model $y_i = \gamma_1 x_{i1} + \dots + \gamma_k x_{ik} + \varepsilon_i$, the ridge minimizes $\sum_i \varepsilon_i^2 + \lambda \sum_j \gamma_j^2$. The Bayesian interpretation of a ridge regression is Bayesian OLS with a normal prior on each coefficient, centered around 0, with scale of the prior variance equal to λ . Unlike its close sibling, the LASSO regression, a ridge regression results in estimated coefficients for all regressors. An optimal λ (in a predictive sense) can be found using k -fold cross-validation. This is done as follows: randomly divide the sample into k subsamples, so-called folds, of equal size; use each subsample to evaluate the model when it is fit on the $k - 1$ other subsamples; in each case, compute a mean-squared error (MSE); compute an average MSE across these k MSEs; find the smallest average MSE by changing λ .

¹⁸In the part of our sample that overlaps with [Romer and Romer \(2004\)](#)’s sample, our left hand side is identical to theirs. Afterwards, we use the series constructed by [Thornton \(2005\)](#) and updated by FRED.

¹⁹This forecast data is conveniently made available by the Philadelphia Fed [here](#).

²⁰The staff documents are often formulated in a way that assumes knowledge of the previous meetings’ analysis, which calls for including lagged sentiment indicators in (3). For example, the language in the second Greenbook following the 9/11 terrorist attacks appears to take knowledge about the attacks and their impact on the economy as given, at least to some degree.

We follow this procedure using $k = 10$. Note that all variables that enter the ridge regression are standardized to be mean zero and have a standard deviation of one.

Discussion of NLP and ML choices. We conclude the step-by-step description of our method with two remarks. First, relative to the rich variety of modern NLP and ML methods, we opt for an approach with restrictions to reduce the complexity of the information. We carry out sentiment analysis for hand-selected economic concepts, an approach sometimes referred to as Aspect-Based Sentiment Analysis (Barbaglia, Consoli, and Manzan, 2022). One alternative to our Steps 2 and 3 would be to capture the entirety of the FOMC documents in (3), for example through term-document matrices in which rows correspond to documents, columns correspond to any English-language term, and entries in the matrix contain the frequency of each term.²¹ This alternative would potentially involve hundreds of thousands of regressors, and might be more suitable for less structured text. Instead, we build on the fact that the Greenbook documents we use contain very structured and carefully worded text with consistency through time. An advantage of our procedure is that the model retains interpretability and echoes the spirit of the original idea of Romer and Romer (2004).

Second, the ridge regression in Step 4 is one of several related ML techniques that could be applied here. One natural alternative would be the LASSO regression, which instead minimizes $\sum_i \varepsilon_i^2 + \lambda \sum_j^k |\gamma_j|$, or the elastic net, which is a mixture between ridge regression and LASSO. A key difference is that LASSO results in a *sparse* model that contains only a subset of the right-hand-side variables, while a ridge regression results in a *dense* model, containing all regressors and associated coefficients. In this sense, ridge regressions are more related to dynamic factor models and principal component analysis, which is often employed for macroeconomic data. We prefer ridge regression on the grounds that dense rather than sparse prediction techniques tend to be preferable for economic data, which typically consists of many correlated regressors with relatively small number of time series observations. This is confirmed in the in-depth analysis of Giannone, Lenza, and Primiceri (2022). These authors develop a Bayesian prior that allows for both shrinkage and variable selection, and find that including many predictors, rather than reducing the set of possible predictors, improves accuracy in several different economic applications. Although their study does not consider text data specifically, it shows that sparse method can become

²¹Kalamara et al. (2020) discuss and compare different prediction models based on high-dimensional text analysis methods in an application to newspaper text.

unstable in the presence of high collinearity between the predictors. This is clearly the case across the numerical forecasts and our sentiment measures based on text, and within both groups of variables.²²

3 Examining the information content of the sentiment indicators

Before we apply our method to estimate monetary policy shocks and the associated macroeconomic responses, this section presents an intermediate discussion and econometric validation exercise that assess the informational value added of our sentiment indicators. We examine the “sufficient statistic” argument laid out by [Cochrane \(2004\)](#) in a discussion of [Romer and Romer \(2004\)](#): suppose the numerical forecasts in the Greenbook efficiently incorporate all information that the FOMC has available about a variable of interest. In that case, it would not be necessary to include additional information on the right hand side of (2) to retrieve a shock measure that can be used to recover the true response of that variable to changes in monetary policy.

Mean vs. mode forecasts. [Cochrane \(2004\)](#)’s argument starts with the assumption that the Greenbook forecasts correspond to a conditional mean expectation. This assumption might not hold if the Greenbook forecasts are better interpreted as *modal* predictions. If the numerical prediction about an economic outcome corresponds the mode, then additional information in the text could alter the FOMC’s conditional mean expectation about that variable, for example by providing information about downside risks in the outlook for that variable. We systematically examine the transcripts of FOMC meetings and find ample evidence that support the view of the Greenbook forecasts representing modal predictions. For example, in the FOMC meeting on July 2-3, 1996, Michael Prell, the director of Research and Statistics at the time clarifies: *“I would characterize our forecasts over the years as an effort to present a meaningful, modal forecast of the most likely outcome. When we felt that there was some skewness to the probability distribution, we tried to identify it. In this instance, as we looked at the recent data, we felt that there was a greater thickness in the area of our probability distribution a little above our modal forecast.”* Appendix D provides numerous additional examples across our sample period. Indeed, the Tealbook A nowadays contains a “Risks and Uncertainties” section, where the asymmetric balance of risks around the numerical forecasts is described

²²In a macroeconomic forecasting context [Bianchi, Ludvigson, and Ma \(2022\)](#) find that elastic net methods, which weight ridge and LASSO penalties, perform best among various ML techniques including random forest techniques. They also emphasize the collinearity of macroeconomic data.

explicitly by the staff. This general insight is in line with some complementary research that alludes to the modal nature of Greenbook forecasts, for example by [Reifschneider and Tulip \(2019\)](#), and recently by [Cieslak et al. \(2021\)](#).²³

Forecast error predictability. Even in the presence of modal forecasts, [Cochrane \(2004\)](#)'s reasoning might still be valid when conditional mean and mode coincide. We therefore verify econometrically whether our sentiment indicators predict errors in the staff's numerical forecasts on average. If that is the case, then there is valuable information about the conditional mean available to the FOMC that is not captured by the forecasts, and thus should be removed from FFR variation to obtain valid monetary policy shocks. The tests presented here focus on the Greenbook unemployment rate forecast, which is particularly suitable because there is little definitional change over time and it is subject to only small data revisions. We provide analogous results for output growth and inflation forecasts in [Appendix E](#).

Table 2: GREENBOOK FORECAST ERROR PREDICTABILITY TESTS

	Left hand side: Greenbook unemployment rate forecast errors							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	current quarter	1-quarter ahead	1-year ahead	2-years ahead	current quarter	1-quarter ahead	1-year ahead	2-years ahead
First PC of all sentiments	-0.029*	-0.114**	-0.445**	-0.622**				
	[0.016]	[0.049]	[0.190]	[0.238]				
Economic activity sentiment					-0.026	-0.098**	-0.285*	-0.363**
					[0.016]	[0.048]	[0.165]	[0.171]
Constant	-0.019	-0.070**	-0.082	0.059	-0.019	-0.069**	-0.077	0.160
	[0.014]	[0.033]	[0.121]	[0.201]	[0.014]	[0.035]	[0.145]	[0.258]
R-squared	0.045	0.149	0.248	0.208	0.033	0.097	0.090	0.055
Obs	210	210	210	62	210	210	210	62

Notes. The forecast errors on the left hand side are constructed by subtracting the Greenbook forecast for the quarterly unemployment rate from the actual unemployment rate (final vintage). First sentiment PC is the first principal component of our 296 sentiment indicators. The regressions are run at FOMC meeting frequency over the October 1982 to October 2008 sample. Newey-West standard errors with optimal bandwidth are provided in brackets. *, **, *** indicate significance at the 10%, 5%, 1% level.

Table 2 presents estimates from different forecast error regressions, over the same sample period we use to estimate equation (3). The left hand side is the unemployment rate forecast error in percent (defined as final vintage – forecast). Columns (1)-(4)

²³[Reifschneider and Tulip \(2019\)](#) mainly focus on the FOMC's Summary of Economic Projections but also discuss the Greenbook/Tealbook forecasts. [Dimitriadis, Patton, and Schmidt \(2021\)](#) argue that Greenbooks forecasts can be rationalized as a mean forecast. They develop a test for rationality for a modal forecast and find that if it was modal then it could not be rational. However, it is not clear why one would start with the presumption that the Greenbook forecasts have to be rational.

include the first principal component (PC) of all 296 sentiment indicators on the right hand side. Columns (5)-(8) include one single sentiment indicator on the right hand side, the one for “economic activity” shown in Panel (a) of Figure 2. In both sets of regressions, we focus on the current quarter, 1-quarter, 1-year and 2-year ahead forecast errors. Note that for the 2-year horizon, the number of observations is lower because two-year ahead forecasts are not produced for all FOMC meetings.

The table reveals that Greenbook forecast errors are predictable with our text-based sentiment measures. The error in the Greenbook unemployment rate forecast is predictable with the first PC of our sentiments, and even with one of the 296 sentiment indicators alone, at various forecast horizons and to an economically significant degree. The economic significance increases with the forecast horizon, and the R^2 of the regression can be as high as 0.25. To give an example for how the magnitude of the coefficients should be interpreted, column (6) indicates that a one standard deviation increase in the sentiment around “economic activity” is associated with a 0.1 percentage point negative forecast error in the unemployment rate.²⁴

These estimates are in line with our argument that the staff construct modal forecasts that differ from the conditional mean. Assume for illustration that there is a well-calibrated distribution of unemployment rate forecasts with a lower bound of 4%, a mode of 6%, an upper bound of 8%, and a mean greater than 6%. That is, there is more mass to the right of the mode. If one computes a forecast error using the modal forecast, there would likely be a positive average forecast error, because on average the outcome should be greater than the mode forecast. This positive average error would – according to our regressions – be significantly negatively correlated with economic activity sentiment. This is consistent with negative economic activity sentiment capturing the upper tail of the unemployment rate distribution, and therefore predicting the positive forecast error on average. In other words, while the Fed staff provides a numerical modal forecast, through their narrative that accompany this forecast, they relay what is in this case an upside risk in unemployment, which, in turn, is captured by our sentiments. Note that this insight is in line with complementary work by [Sharpe, Sinha, and Hollrah \(2020\)](#), who find that language “tonality” surrounding forecasts predicts errors of both Fed and private

²⁴Appendix E provides results for real output growth and inflation. For these variables, forecast errors are predictable as well, though at a smaller subset of horizons. The same appendix also shows results based on using the first release instead of the final vintage in the construction of the forecast errors. The results are similar. Finally, note that we also tried including lags in the regressions, and found that the predictive power is mostly concentrated in the contemporaneous sentiment measures.

sector forecasts. We make clear that this means text-based information is crucial to inform the systematic component of policy when estimating monetary policy shocks.

4 Results of the identification procedure

This section discusses the estimation results for the empirical model represented by equation (3). The results include measures of fit, an analysis of what drives the systematic component of monetary policy, properties of the estimated shock time series, as well as an exploration of including further information.

4.1 Systematic vs. exogenous changes in the target interest rate

Table 3, column (1) presents the R^2 of alternative empirical specifications. First, as the simplest benchmark it includes equation (2), the restricted version of (3) where only the staff forecasts used in the original [Romer and Romer \(2004\)](#) specification enter in a linear OLS estimation. Second, a model that includes the expanded set of 132 forecast variables, and is estimated as a ridge rather than an OLS regression. Third, a ridge model where the augmented set of forecasts and our sentiment indicators are included, but function $\Gamma(\cdot)$ is still linear. Fourth, a ridge model with the same forecasts and sentiments variables entering linearly and quadratically. Fifth, the linear ridge model which also contains 4 lags of the sentiment indicators. Sixth, our main specification in which forecasts and sentiments enter linear, nonlinearly and with 4 lags.

We compare the goodness of fit between these alternative models to understand what they imply about the contribution of the systematic component of monetary policy. The first line in Table 3, column (1) shows that over the sample period October 1982 to October 2008 that we consider, the Romer-Romer OLS model implies an R^2 of 0.5. In other words, this empirical model attributes 50% of the variation in the FFR target to systematic policy, while 50% is attributed to monetary policy shocks. This seems undesirable – in the language of [Leeper, Sims, and Zha \(1996\)](#): *“Even the harshest critics of monetary authorities would not maintain policy decisions are unrelated to the economy.”*

The remaining lines in column (1) of the table reveal that expanding the information set in the empirical model increases the implied fit. Bear in mind that the ridge regression does not maximize fit, but instead optimizes predictive performance based on the 10-fold cross-validation. Nevertheless, each step of enriching the empirical model – going from OLS to ridge regression, including more numerical forecasts and

Table 3: R^2 ACROSS DIFFERENT SPECIFICATIONS

		(1)	(2)
	Number of	R^2 with	R^2 with
	regressors	10-word sentiment	5-word sentiment
		(main specification)	(robustness)
Romer-Romer original OLS with subset of forecasts	19		0.50
Ridge with extended set of forecasts	133		0.55
Ridge with all forecasts & sentiments (linear)	429	0.65	0.66
Ridge with all forecasts & sentiments (nonlinear)	858	0.75	0.77
Ridge with all forecasts & sentiments (linear with lags)	1,613	0.87	0.88
Ridge with all forecasts & sentiments (nonlinear with lags)	3,226	0.94	0.95

Notes. Implied goodness-of-fit, measured by R^2 , from estimating different empirical specifications of equation (3). For the first two specifications, sentiments are not included so the 10-word/5-word distinction does not apply. Our preferred specification is the last one presented in the table, with forecasts and sentiments entering nonlinearly and with 4 lags.

sentiment indicators, and allowing for nonlinearities and lagged sentiments – delivers some additional improvement in the fit of the model. Our preferred specification, the bottom line in Table 3 implies an R^2 of 0.94, suggesting that 94% of FFR variation is systematic, and 6% are explained by shocks. Relative to the Romer-Romer OLS model, this reduces the contribution of exogenous shocks almost ten fold.

Besides the economic appeal of these findings, our analysis of forecast error predictability in Section 3 already supported the view that more information about the systematic component of monetary policy should be included, i.e. that the R^2 should be higher than in a specification with forecasts only. Note that one downside of a higher R^2 and therefore less variation in the shocks could be low statistical power when studying the responses to the shocks in a finite sample of macroeconomic data. [Leeper et al. \(1996\)](#) describe this challenge quite pointedly, by saying “*This is what one would expect of good monetary policy, but it is also the reason why it is difficult to use the historical behavior of aggregate time series to uncover the effects of monetary policy.*” Our remaining results, in particular the IRFs we estimate in Section 5, show that lack of statistical power in the shock measure is not an issue in our application.

Further variations in the specification. We check the robustness of the results above along several dimensions. Column (2) of Table 3 focuses on empirical models in which our sentiment indicators are constructed using a 5-word instead of a 10-word window around economic concepts. By construction, the first two rows in each column remain unchanged, as these specifications do not incorporate sentiment indicators. The table shows that the meaningful increase in fit from expanding the information set remains present when we vary our way to construct sentiment indicators. We

verified that constructing sentiments based on positive and negative words within the same sentence, rather than inside a fixed word window, yields time series that are highly correlated with the ones we use, as shown in Appendix C for two examples. We also experimented with the lag structure of those specifications that include lags. We found that increasing the number of lags, starting at 0 lags, increased the R^2 for a given specification, but the increases becomes fairly small around 4 lags.²⁵ Furthermore, we constructed an auxiliary data set about the FOMC’s composition, in order to verify whether personal dynamics between FOMC members might drive FFR changes. We found that this was not the case: the R^2 from including this information in the ridge regression increased by less than 0.1%.²⁶ Finally, we tried alternative nonlinear specifications of $\Gamma(\cdot)$. We found that the model fit and time series of the residuals we obtained were similar to the quadratic version. For example, the residuals obtained with a cubic version were 99% correlated with the corresponding quadratic specification. A specification in which we added all possible interaction terms between all sentiment indicators and all forecasts – amounting to almost 40,000 variables on the right hand side of (3) – gave residuals that were 96% correlated with the corresponding quadratic version.

4.2 Inspecting the drivers of FFR changes

We now focus on our full nonlinear 10-fold validated ridge specification, which can account for 94% of the variation in FFR changes across FOMC meetings. The coefficient estimates corresponding to the different variables included in this empirical model convey information about which numerical forecasts and sentiment indicators are important for explaining interest rate decisions. For example, a positive, statistically significant and economically large coefficient associated with the sentiment indicator for “economic activity” implies that an improvement in the Fed economists’ sentiment around “economic activity” is associated with interest rate increases by the FOMC, all else equal. A variable-by-variable analysis of the coefficients is difficult given the long list of 3,226 regressors. Furthermore, comparisons between the coefficients associated with individual regressors are not easy to interpret, given that many of the variables closely overlap in terms economic content and are hence highly collinear. We

²⁵For our main ridge model, 0, 1, 2, 3, and 4 lags result in an R^2 of 0.75, 0.81, 0.90, 0.92 and 0.94.

²⁶The data set contains dummy variables that are 1 if a governor and regional bank representative attends a meeting and 0 otherwise. In addition, we collect information on voting status, the US presidents that have appointed given governors, as well as the number of female attendants. More details on the construction of this additional data set are provided in Appendix F.

therefore take a statistical approach to summarize systematically among which groups of variables the predictive power for FFR changes is most concentrated.

We proceed as follows. First, we extract the first 25 principal components (PCs) of the numerical forecast variables and the first 25 PCs of the sentiment variables.²⁷ Among the forecasts, the first 25 PCs explain 82% of the variation, with the first PC capturing 15%. Among the sentiment indicators, the first 25 PCs explain 49% of the variation, with the first PC capturing 10%. Second, we run an auxiliary LASSO model, where the left hand side is the change in the FFR and the right hand side contains these 50 PCs. In doing so we seek to eliminate PCs that are not particularly important in explaining FFR changes. Among the PCs, we then examine those that are associated with statistically significant and economically large coefficient estimates. It turns out that is the case for the first two sentiment PCs and the first numerical forecast PC. The coefficients (t-statistics) on these three regressors are 0.0206 (5.7195), 0.0159 (3.4511), and 0.0098 (2.2312), respectively.²⁸ Third, among these selected PCs we uncover which variables have the largest loadings in absolute value on a given PC. Examining those variables tells us what type of information from the documents prepared for the FOMC contributes meaningfully to explaining variation in the target interest rate.

Table 4: VARIABLES WITH HIGH LOADINGS ON PREDICTIVE PRINCIPAL COMPONENTS

Sentiment PC1		Sentiment PC2		Numerical forecast PC1	
economy	0.141	advanced foreign economies	-0.141	output growth (+1)	0.187
firms	0.139	merchandise	0.140	output growth (0)	0.175
economic activity	0.136	foreign economies	0.135	bus. fixed inv. growth (+2)	0.160
manufacturing activity	0.133	credit standards	-0.131	ind. prod. growth (+1)	0.160
commercial real estate	0.131	farm	0.127	output growth (+2)	0.158
manufacturing firms	0.130	cash	0.125	nominal output growth (+1)	0.153
labor market	0.125	core inflation	-0.124	housing starts (+1)	0.151
services	0.123	industrial production	0.123	housing starts (+2)	0.150
consumer confidence	0.118	trade deficit	0.121	housing starts (+3)	0.150
industries	0.117	developing countries	0.119	housing starts (0)	0.149

Notes. Variables and associated loadings on the principal components with largest economic significance in a LASSO regression with FFR changes on the left hand side. For numerical forecasts, the horizon in quarters is given in brackets.

Table 4 presents the 10 variables with the largest loadings in absolute value on each of the three selected PCs. A positive (negative) loading means that an increase in a given variable is associated with an increase (decrease) in the FFR. The first

²⁷For simplicity we consider the sentiments linearly and contemporaneously in this exercise. Including squared and lagged sentiment indicators, as in our preferred specification of (3), does not change the broad conclusions discussed in this section.

²⁸PCs are standardized, so the coefficients associated with them are comparable. The PC with the fourth largest coefficient has a much lower coefficient than the third largest one.

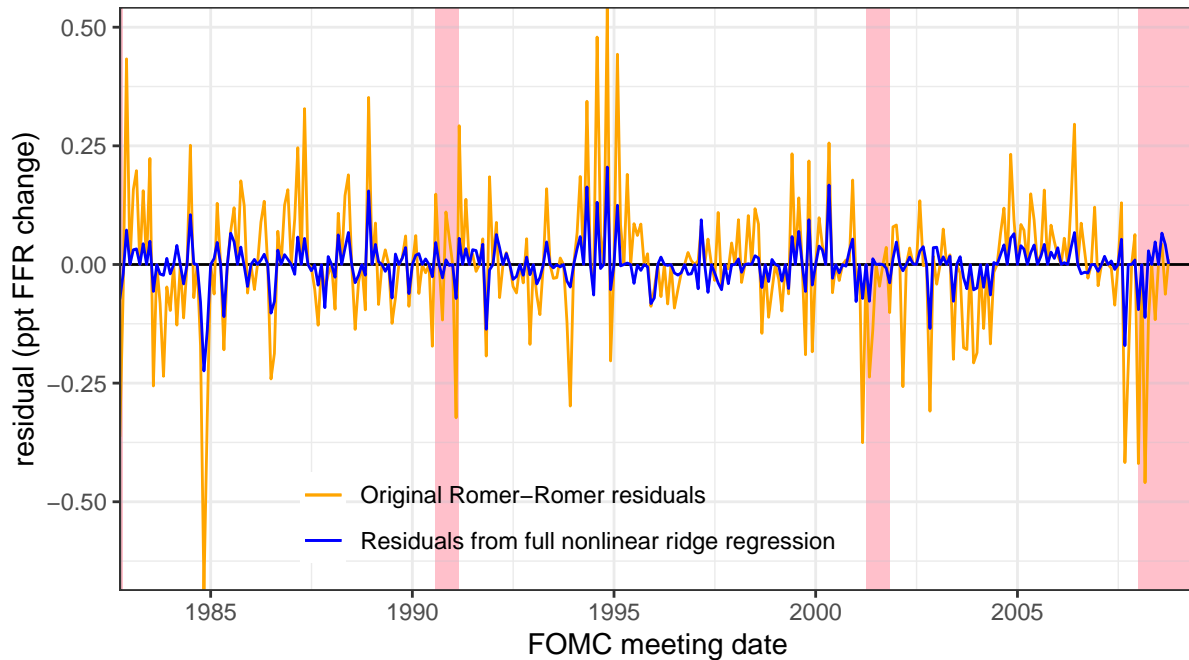
sentiment PC resembles a broad real activity sentiment factor, capturing the Fed economists' sentiments about activity, production, labor and housing markets, and consumer confidence.²⁹ Improvements in the sentiment around these concepts are associated with hikes in the FFR. Interestingly, the second sentiment PC mostly captures international economic developments, with the sentiments around advanced foreign economies, foreign economies, the trade deficit and developing countries featuring with high loadings. Improvements in the sentiment around these concepts translate into interest rate increases. The second sentiment PC also blends in some financial and price information, with sentiments surrounding both credit standards and core inflation appearing with a negative loading. Finally, the PC extracted from the numerical forecasts is again mostly related to real activity, with high loadings on output, production, investment and housing forecasts, at different horizons. Overall, it is perhaps somewhat surprising how little price and inflation information appears to drive the systematic variation in FFR changes according to our statistical analysis. Instead, systematic monetary policy appears to be mostly informed by the broader outlook around economic activity over the 1982 to 2008 period.

4.3 Inspecting the estimated shocks

The dark blue line in Figure 3 plots the estimated time series of monetary policy shocks, that is, the residuals $\hat{\varepsilon}_t^*$ from our preferred empirical specification which includes forecasts, sentiments and nonlinearities in a ridge model. The figure compares this with the estimated residuals from the Romer-Romer OLS model as the lighter orange line. The residuals have the same unit as that of the left hand side of the regression, so can be interpreted in percentage point changes in the FFR. Recall that the shocks represented by the blue line explain 6% of FFR variation while those represented by the orange line explain 50%. Related to the lower contribution in FFR variation, the figure shows that our measure of monetary policy shock displays a generally lower volatility. We also find it to display a lower degree of autocorrelation. It is also visible in the figure that our estimate of shocks is not simply a scaled-down version of the shocks implied by the Romer-Romer OLS model. In many instances, the orange line implies a larger shock in the same direction, while in others the shock measures go in different directions.

²⁹In the dynamic factor model literature, it is common that the first factor of macro data sets is associated with real activity, see e.g. [Giannone, Reichlin, and Sala \(2006\)](#).

Figure 3: ESTIMATED MONETARY POLICY SHOCKS



Notes. Time series of estimated monetary policy shocks. Dark blue: our preferred version, the residuals from predicting changes in the FFR based on numerical forecasts, sentiment indicators and nonlinearities in FOMC documents. Orange: benchmark version based on a specification that follows [Romer and Romer \(2004\)](#). Shaded areas represent NBER recessions.

What are monetary policy shocks? For those episodes in which the estimated shocks are particularly large in magnitude, we closely inspect the discussion that took place in the FOMC. Here we provide two examples, which shed light on what estimated monetary policy shocks capture.

The largest shock in absolute value is estimated for the November 7, 1984 FOMC meeting. The policy change is equivalent to a decline in the FFR of 75 basis points and our shock measure is minus 22 basis points, indicating that based on staff forecasts and sentiments, we predict a decline of 53 basis points. This is a period that has a mixed economic outlook: industrial production has declined for the first time in two years, employment shows the smallest rise since the expansion began at the end of 1982, yet investment and consumption show robust increases. The Fed staff concludes that the “slowdown may only be a pause in a recovery that has not run its full course” in the Beige Book. Accordingly, they forecast an increase of 3.5% in GDP in the last quarter of the year, compared to 2.75% in the previous quarter. Their inflation forecast is also flat relative to recent quarters and it is expected to pick up somewhat in 1985. When we read the transcript of the FOMC meeting, it becomes very clear that several participants

find the staff forecast too optimistic, and some of them consider the outlook to be more uncertain. As a result, at the end of the meeting, FOMC's policy actions are consistent with a sizable easing of policy, which is contrary to what one may have decided by simply reading the staff documents. In fact, one of the two policy options put forward by the staff involved no changes in policy. This episode is a good example of a situation where the FOMC participants' views about the economy are different from that of the Fed staffs', and the policy action appears to be quite far from what would be implied by the latter. It is important to emphasize that this is an unusual situation. If the disagreement happened more often, then our procedure would have picked it up as a systematic part of policy, and they would not show up in our monetary policy shocks.

Our second example is the November 15, 1994 meeting, where a 75 basis point FFR increase was decided, and our analysis shows 21 basis points of this was a monetary policy shock, our second largest in absolute terms. The staff analysis paints the picture of a very robust growth: they forecast an acceleration in output, in contrast to their prior forecast, final demand is high and banks are lending. They conclude that the economy is above its full capacity with the inflationary consequences not yet realized. The staff proposes two policy options: a no change option and one where the FFR increases by 50 basis points. In their forecasts, the staff uses an assumption of "appreciable further tightening" with a cumulative increase of 150 basis points in the following 6 months. During the meeting, Chairman Greenspan suggests that "they are behind the curve" and since the market already built in a significant rate hike "a mild surprise would be of significant value." He proposes a rate increase of 75 basis points to get "ahead of general expectations." Most of the participants agree with this proposal, with several participants emphasizing the credibility of keeping inflation under control. Once again this is a situation where the FOMC decided on an action not simply based on the current economic outlook but also other considerations, and our procedure therefore implies that this reflects a monetary policy shock. Indeed, the difference between the 75 basis point decision and the staff's suggested 50 basis point option almost exactly matches the 21 basis point contractionary shock we estimate.

One might argue that credibility concerns such as the ones motivating the strong hike in November 1994, might in some way be a feature of the Fed's policy rule and should therefore not constitute a shock. However, the types of decisions that imply monetary policy shocks in our procedure must occur in a completely unsystematic way, not in response to changes in the Fed's information set. Otherwise they would be picked up by the ridge regression as a feature of policy. Systematic credibility concerns,

that arise based on available information, will be part of policy. Monetary policy shocks are instead *unsystematic* FFR changes, orthogonal to the information set.

Shocks vs. surprises. An alternative branch of research, described in our review of the literature, identifies shocks from HF surprise movements in market interest rates. We do not rely on HF techniques, as our left hand side variable is the *target* FFR. The discussion above reveals that one could still give a surprise interpretation to our shock measure, which might be best interpreted as FFR target decisions by the FOMC that constitute *surprises to the Fed staff*. While a systematic comparison of alternative methodologies is beyond the scope of this paper, we would expect that surprises to the Fed staff should also be a surprise to markets, but not vice versa.

5 The effects of monetary policy shocks on the economy

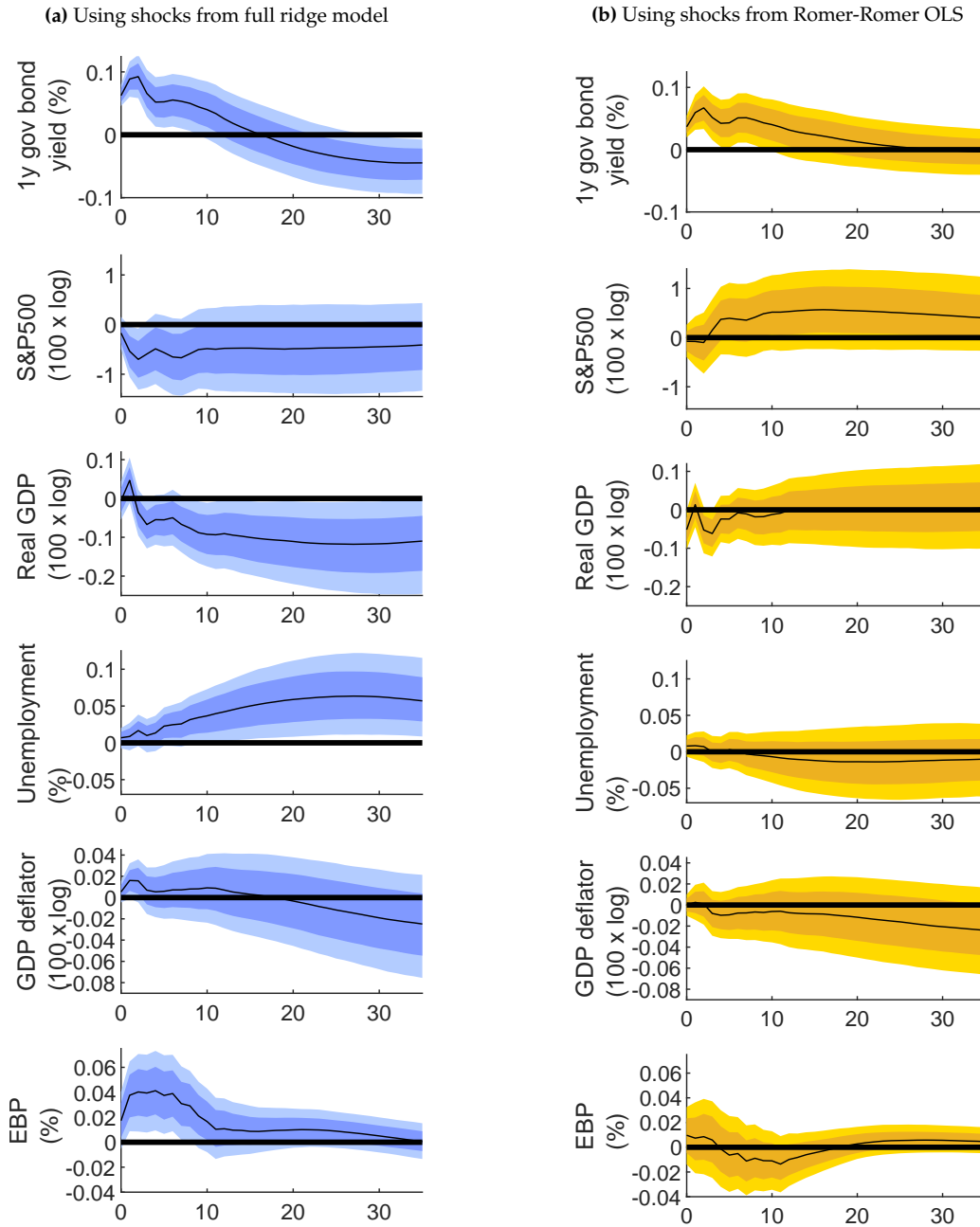
This section uses our new shock measure to study the effects of changes in monetary policy on the US economy in a state-of-the-art BVAR model, following [Jarocinski and Karadi \(2020\)](#). Alternatively, we also study the responses to our shocks using a frequentist local projections approach following [Jordà \(2005\)](#).

The BVAR system is estimated at monthly frequency, and includes the 1-year Treasury yield, the log of the S&P500, log real GDP, the log GDP deflator, the unemployment rate, and the excess bond premium (EBP) of [Gilchrist and Zakrajšek \(2012\)](#). Our time series of monetary policy shocks enters first in a Cholesky ordering. This yields asymptotically identical results to using the shock series as an external instrument ([Plagborg-Møller and Wolf, 2021](#)).³⁰ While our shock series spans the period 1982:10-2008:10, applying it as an external instrument allows us to estimate the system over a longer sample. The 1-year yield is included as it is mostly free to move while the target FFR is stuck at the zero lower bound for part of the sample. GDP and its deflator are included to capture the effect of monetary policy on activity and prices. We use their monthly versions, interpolated using the Kalman filter. We include the unemployment rate, given that we found the Greenbook forecast error predictability to be particularly strong for this variable. The S&P500 and EBP are included as forward-looking financial variables. For comparability reasons, we use the same sample period (1984:02-2016:12), settings and priors as in [Jarocinski and Karadi \(2020\)](#).³¹

³⁰For more on external instruments see [Mertens and Ravn \(2013\)](#), [Stock and Watson \(2018\)](#).

³¹We thank these authors for making their Gibbs sampler codes available. Their sample starts in February 1984 which [Bernanke and Mihov \(1998\)](#) identify as the end of the Volcker disinflation period.

Figure 4: IRFS TO DIFFERENT MONETARY POLICY SHOCK MEASURES



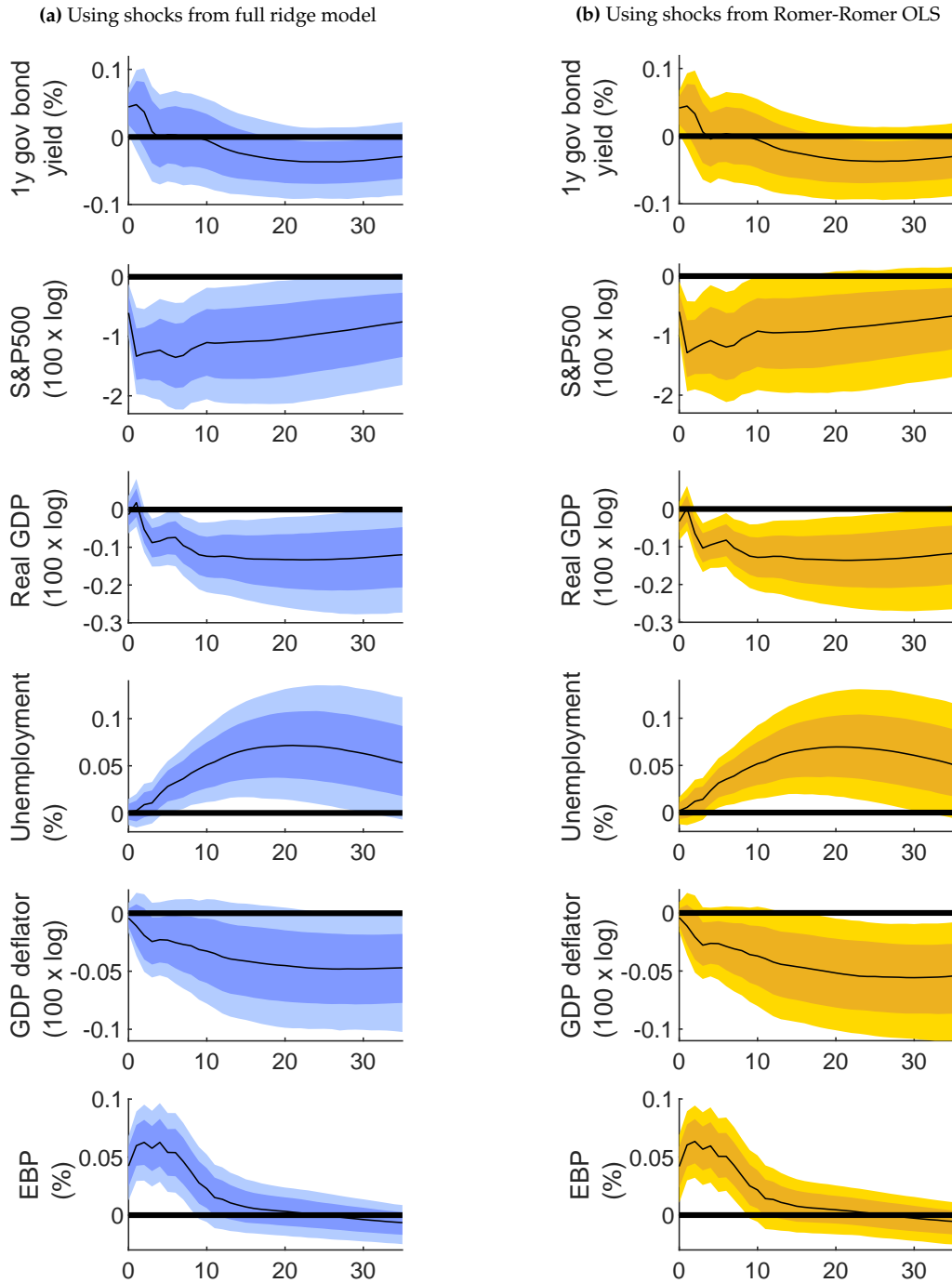
Notes. IRFs to different estimated monetary policy shocks in BVAR model (without additional sign restrictions imposed). Panel (a) uses our proposed measure of monetary policy shocks, estimated using the full nonlinear ridge model on the extended set of numerical forecasts and our sentiment indicators from FOMC documents. Panel (b) shows the analogous IRFs when a simpler empirical specification is used to estimate the shocks, which includes only the original set of numerical forecasts in a Romer-Romer OLS regression. The solid line represents the median, the 16th and 84th percentiles are represented by the darker bands, and the 5th and 95th percentiles by the lighter bands. The sample period to estimate the shocks is 1982:10-2008:10. The sample used to estimate the IRFs is 1984:02-2016:12.

Figure 4, Panel (a) presents IRFs of macroeconomic variables to our preferred measure of monetary policy shocks. We find that a monetary policy tightening is characterized by a relatively persistent increase in yields, lasting for about 20 months. The rise in interest rates leads to a reduction in real economic activity and a fall in the price level, directly in line with what standard economic theory predicts. The reduction in real output and the increase in unemployment take about a year to materialize and are very persistent. The price level response displays a very mild version of a “price puzzle” (Sims, 1992) in the first months, but is persistently negative thereafter. It takes about 18 months for the point estimate to be visibly negative, and 30 months for the response to be significantly negative. Panel (a) also shows that bond premia increase sharply and significantly after a monetary policy tightening, a finding in line with standard models of monetary policy and external finance premia (e.g. Bernanke, Gertler, and Gilchrist, 1999). Furthermore, our identified monetary policy shocks imply a fall in stock prices following a tightening in monetary policy, consistent with theory (Jarocinski and Karadi, 2020).

These results contrast with Panel (b) of Figure 4, which presents IRFs to monetary policy shocks constructed using the original Romer-Romer OLS specification, in which a only handful of numerical forecasts are used to predict the systematic component of monetary policy. While the shock induces a similar path for market interest rates, as well as a comparable reduction in the price level, a monetary policy tightening appears to have little effect on output and unemployment. The effect on real GDP and the unemployment rate are completely flat. This is different from the IRFs in the original Romer and Romer (2004) paper, using the 1969-1996 sample, where output is significantly reduced after a tightening. This contrast connects to earlier findings on the fact that the IRFs to their original shocks give results at odds with standard theory in more recent samples. Ramey (2016) and Barakchian and Crowe (2013) provide discussions. Moreover, the shocks computed using the original Romer-Romer methodology imply an insignificant response of the EBP to a policy tightening, and positive comovement between the S&P500 and interest rates conditional on a monetary policy shocks, both of which are inconsistent with standard theory.

The differences between Panels (a) and (b) may suggest that some systematic policy variation is still present in the shock measure only based on numerical forecasts, whereas our measure based on a larger information set is more plausibly exogenous. In particular the unemployment response is markedly different, consistent with the evidence in Section 3 that text-based sentiment is predictive of Greenbook

Figure 5: IRFS CONSTRUCTED WITH ADDITIONAL SIGN RESTRICTIONS



Notes. IRFs to different estimated monetary policy shocks. The two panels correspond to those in Figure 4, but impose the additional sign restrictions suggested by [Jarocinski and Karadi \(2020\)](#) to separate monetary policy shocks from central bank information shocks. Specifically, the IRFs shown here are constructed under the restriction that stock prices fall in response to a monetary policy tightening. The solid line represents the median, the 16th and 84th percentiles are represented by the darker bands, and the 5th and 95th percentiles by the lighter bands. The sample period to estimate the shocks is 1982:10-2008:10. The sample used to estimate the IRFs is 1984:02-2016:12.

unemployment rate forecasts errors. Given the modal nature of Greenbook forecasts, this could indicate that the Romer-Romer OLS regression does not fully reflect the effects of asymmetric changes in the balance of risks around forecasts on the systematic conduct of policy but that our ridge regression with sentiments can capture these.

Another interpretation, in light of the difference in response of stock prices, is that the Fed systematically reacts to equity markets, e.g. lowers the FFR after contractions in stock prices (Cieslak and Vissing-Jorgensen, 2020). If orthogonalizing the FFR only with respect to a small set of numerical forecasts does not control for this systematic feature of monetary policy, then the implied residuals might spuriously pick up a positive correlation between stock prices and the FFR, as observed in Panel (b). Instead, our sentiment indicators might reflect the relevant information about financial market developments that the FOMC considers. To support this interpretation, Figure 5 shows the same IRFs but additionally restricts stock prices to fall in response to a monetary policy tightening, as suggested by Jarocinski and Karadi (2020). Panel (a) of Figure 5 looks similar to its counterpart in Figure 4, making clear that our preferred shock measure already satisfies the additional sign restrictions.³² On the contrary, the sign restrictions alter the IRFs in Panel (b) quite drastically. Imposing a negative comovement between interest rates and stock prices also “corrects” the activity, price and bond premia responses, which are now very similar to our preferred measure, and what economic theory would predict.

Additional results. Appendix G presents IRFs constructed with shocks from intermediate specifications of (3) (refer to Table 3 for an overview). One noteworthy observation is that monetary policy shocks retrieved using the extended set of numerical forecasts, but without including sentiments, already render the IRFs to be in line with theory, that is, are sufficient to correct the endogeneity problem evident in Panel (b) of Figure 4. We emphasize that IRFs in line with consensus of the economic literature should be a necessary, but not sufficient criterion for a good measure of monetary policy shocks. As shown in Section 3, errors from numerical forecasts are predictable using our sentiment indicators, a strong argument for including them when retrieving a monetary policy shock measure. Furthermore, Section 4 shows that the Romer-Romer ridge specification has an R^2 of only 0.55, as opposed to 0.94 in our preferred specification, implying an unappealingly strong contribution of shocks to

³²One exception is that the mild price puzzle is eliminated when the sign restrictions are imposed. Another difference is that the negative response of the S&P500 is clearer with the sign restrictions.

variation in the FFR target when only numerical forecasts are included.

Appendix H presents results analogous to Figure 4, but instead constructed using local projections. As one would expect without the shrinkage imposed by the BVAR, the IRFs are generally noisier, but we confirm the general results using this alternative approach. Most notably, the shocks from the original Romer-Romer specification again result in responses of real activity and stock prices that are not in line with theory.

Taken together, the results across the different parts of our analysis lead us to prefer our estimate of monetary policy shocks, based on including a wide set of information contained in numerical forecasts, sentiments reflected in natural language used by Federal Reserve economists, and nonlinearities.

6 Interpreting the 2022 hikes with a real-time version of our method

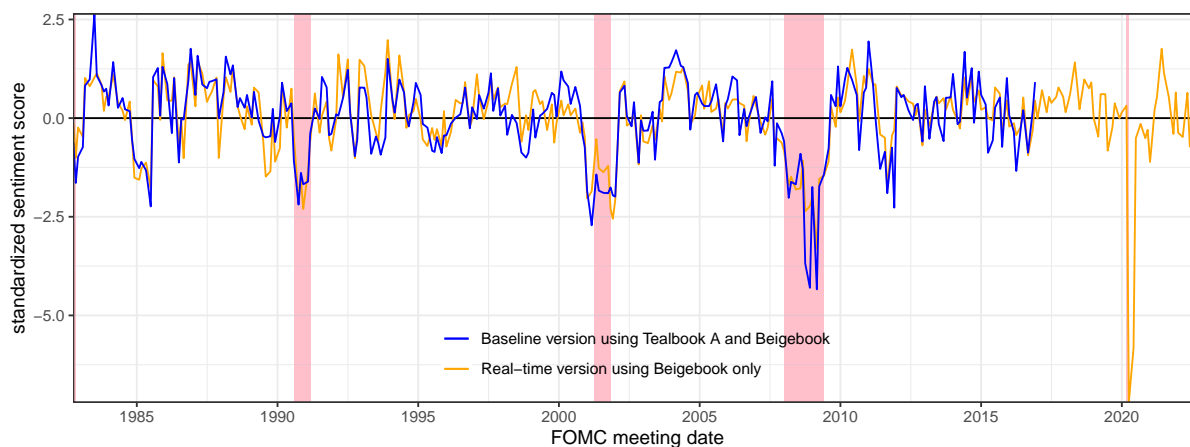
In this section, we present a real-time version of our method and use it to examine the FOMC's decisions in 2022. While the Fed Tealbooks are made available to the public only with a 5-year delay, the Beigebooks are publicly available prior to every FOMC meeting. These summarize regional economic conditions for each individual Federal Reserve district. The idea behind the real-time version of our method is that constructing our sentiment indicators only from the Beigebook text provides at least a limited proxy for the FOMC's information set. We verify how well this proxy works: while in our main analysis we use both the Tealbook A and the Beigebook, we find that using only the Beigebook over our main 1982 to 2008 sample gives us strongly correlated sentiment indicators, as illustrated for the example of "economic activity" in Figure 6.³³ Leveraging the Beigebooks is not possible in the original Romer and Romer (2004) approach, as the Beigebooks do not contain any numerical forecasts. This highlights a further advantage of our methodology based on natural language.

We run the Beigebook-only version of our main ridge regression, with 4 lags and squared terms, over the period 2015 to 2022, after the 2008 to 2015 zero lower bound period.³⁴ In the last year of this period our procedure measures sharp changes in the sentiment indicators around various economic concepts in the Beigebooks. For example, the sentiment around "inflation" drops massively, with a reduction of more

³³Over that period, the R^2 from using only Beigebook sentiments to estimate (3) is 0.68, compared to 0.92 with information from Tealbooks and Beigebooks combined. The resulting shocks have a correlation of 0.90 with each other. We confirmed that the BVAR IRFs we study in the previous section look qualitatively similar for the shocks constructed using only the Beigebook sentiments.

³⁴When estimating equation (3) in that sample, we exclude observations corresponding to the second zero lower bound period between March 2020 and December 2021.

Figure 6: SENTIMENT SURROUNDING ECONOMIC ACTIVITY: BASELINE VS. REAL-TIME



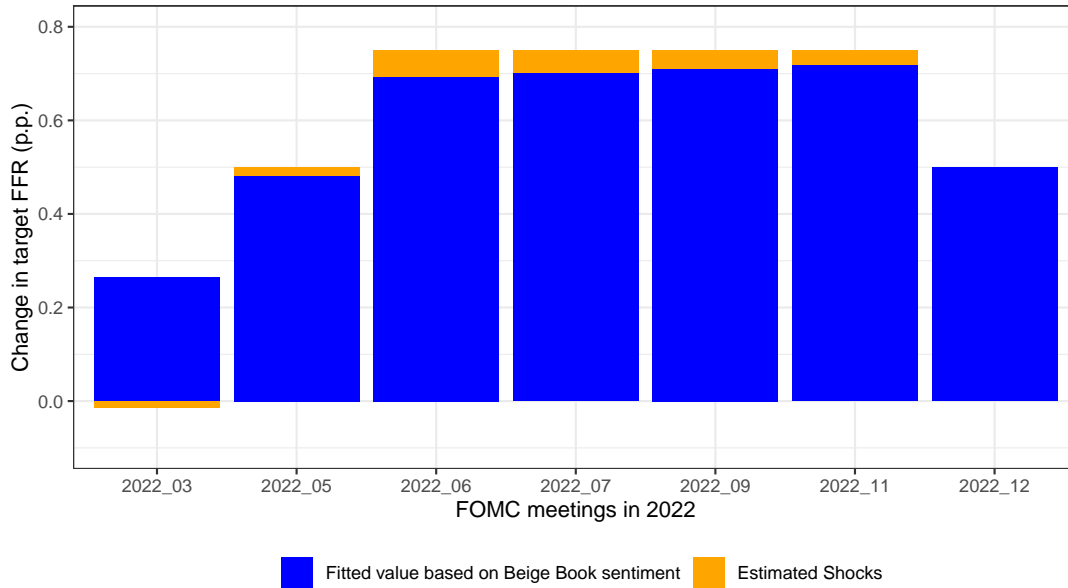
Notes. Sentiment around economic activity over time. Dark blue: indicator used for our main analysis based on Tealbook A and Beigebook. Orange: real-time version based on Beigebook only. The 5-year period after the blue line stops corresponds to the publication lag of the Tealbook and associated forecasts. Shaded areas represent NBER recessions.

than 6 standard deviations (in terms of its 1982 to 2022 variability). One of the main contributors to this pattern is a sharp increase in the use of the negatively connotated word “concern” from the [Loughran and McDonald \(2011\)](#) in proximity to inflation. Other concepts around which the sentiment expressed in the Beigebook text deteriorates strongly into negative territory are “recession”, “fuel”, and “China”.

We find that the R^2 from estimating the real-time version of (3) over the 2015 to 2022 sample is 0.98, suggesting that monetary policy is very systematic over that period, with only a small role for monetary policy shocks. Recall that this is the case despite the fact that we can only include the Beigebook sentiments on the right hand side, without using the information from Tealbook sentiments and numerical forecasts which we find add significant predictive power in the 1982 to 2008 period.

Figure 7 decomposes the Fed’s series of interest rate hikes in 2022, displayed as bars, into systematic component (blue) and monetary policy shocks (orange). The FOMC’s decisions in 2022 are estimated to be almost entirely systematic, that is, to be almost fully explained by the variation in sentiment in the Beigebooks. While the total increase in the FFR target amounted to 425 basis points, the shock component cumulates to around 19 basis points over the course of the year 2022. In other words, our method implies that the tightening in 2022 was almost perfectly in line what our model estimates as the Fed’s historical systematic policy making.

Figure 7: DECOMPOSITION OF 2022 INCREASES IN THE TARGET FED FUNDS RATE



Notes. The bars represent the seven consecutive FOMC meetings during which the FFR was increased in 2022. The height of each bar corresponds to the percentage point change in the target FFR. The blue area corresponds to the estimated systematic component (the fitted values of the ridge regression). The orange area corresponds to the estimate monetary policy shock, where an area in the positive (negative) part of the y-axis corresponds to a contractionary (easing) shock.

7 Conclusion

This paper develops a method for the identification of monetary policy shocks using natural language processing and machine learning. We extract sentiment indicators for 296 economic concepts that are discussed by Fed economists in the documents they prepare for FOMC meetings. We include those indicators, alongside the economists' numerical forecasts of macroeconomic variables, in a ridge regression to predict systematic changes in the target interest rate. Since the sentiment indicators can predict errors from the numerical forecasts, their inclusion is crucial to summarize the Fed's information set. The residuals of our ridge regression represent our new measure of monetary policy shocks. We find that economic activity and prices decline, bond premia rise, and stock prices fall after a monetary policy tightening, in line with theoretical predictions. Our analysis as a whole shows that the novel procedure proposed in this paper delivers a cleanly estimated series of monetary policy shocks. To some degree, our procedure can be applied in real time and allows us to interpret the Fed hikes in 2022 as mostly systematic changes in monetary policy.

References

- ACOSTA, M. (2022): "A New Measure of Central Bank Transparency and Implications for the Effectiveness of Monetary Policy," *Working Paper*.
- ANTOLIN-DIAZ, J. AND J. F. RUBIO-RAMIREZ (2018): "Narrative Sign Restrictions for SVARs," *American Economic Review*, 108, 2802–29.
- BACHMANN, R., I. GÖDL-HANISCH, AND E. R. SIMS (2022): "Identifying monetary policy shocks using the central bank's information set," *Journal of Economic Dynamics and Control*, 145, 104555.
- BARAKCHIAN, S. M. AND C. CROWE (2013): "Monetary policy matters: Evidence from new shocks data," *Journal of Monetary Economics*, 60, 950–966.
- BARBAGLIA, L., S. CONSOLI, AND S. MANZAN (2022): "Forecasting with Economic News," *Journal of Business & Economic Statistics*, 0, 1–12.
- BERNANKE, B. S., J. BOIVIN, AND P. ELIASZ (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach*," *The Quarterly Journal of Economics*, 120, 387–422.
- BERNANKE, B. S., M. GERTLER, AND S. GILCHRIST (1999): "The financial accelerator in a quantitative business cycle framework," *Handbook of macroeconomics*, 1, 1341–1393.
- BERNANKE, B. S. AND I. MIHOV (1998): "Measuring monetary policy," *The quarterly journal of economics*, 113, 869–902.
- BIANCHI, F., S. C. LUDVIGSON, AND S. MA (2022): "Belief distortions and macroeconomic fluctuations," *American Economic Review*, 112, 2269–2315.
- CAMPBELL, J. R., C. L. EVANS, J. D. FISHER, AND A. JUSTINIANO (2012): "Macroeconomic effects of federal reserve forward guidance [with comments and discussion]," *Brookings papers on economic activity*, 1–80.
- CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (1999): "Chapter 2 Monetary policy shocks: What have we learned and to what end?" Elsevier, vol. 1 of *Handbook of Macroeconomics*, 65–148.
- CIESLAK, A., S. HANSEN, M. MCMAHON, AND S. XIAO (2021): "Policymakers' Uncertainty," *Available at SSRN 3936999*.
- CIESLAK, A. AND A. VISSING-JORGENSEN (2020): "The Economics of the Fed Put," *The Review of Financial Studies*, 34, 4045–4089.
- CLOYNE, J. AND P. HÜRTGEN (2016): "The Macroeconomic Effects of Monetary Policy: A New Measure for the United Kingdom," *American Economic Journal: Macroeconomics*, 8, 75–102.

- COCHRANE, J. (2004): “Comments on ‘A new measure of monetary shocks: Derivation and implications’ by Christina Romer and David Romer,” *July 17, 2004, presented at NBER EFG Meeting*.
- COIBION, O. (2012): “Are the Effects of Monetary Policy Shocks Big or Small?” *American Economic Journal: Macroeconomics*, 4, 1–32.
- COIBION, O., Y. GORODNICHENKO, L. KUENG, AND J. SILVIA (2017): “Innocent Bystanders? Monetary policy and inequality,” *Journal of Monetary Economics*, 88, 70–89.
- DIMITRIADIS, T., A. J. PATTON, AND P. W. SCHMIDT (2021): “Testing forecast rationality for measures of central tendency,” *Working paper*.
- DOH, T., D. SONG, AND S.-K. YANG (2022): “Deciphering federal reserve communication via text analysis of alternative fomc statements,” *Federal Reserve Bank of Kansas City Working Paper Forthcoming*.
- FAUST, J. AND J. H. WRIGHT (2009): “Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset,” *Journal of Business & Economic Statistics*, 27, 468–479.
- GARDNER, B., C. SCOTTI, AND C. VEGA (2021): “Words speak as loudly as actions: Central bank communication and the response of equity prices to macroeconomic announcements,” *Journal of Econometrics*.
- GERTLER, M. AND P. KARADI (2015): “Monetary Policy Surprises, Credit Costs, and Economic Activity,” *American Economic Journal: Macroeconomics*, 7, 44–76.
- GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2022): “Economic predictions with big data: The illusion of sparsity,” *Econometrica*.
- GIANNONE, D., L. REICHLIN, AND L. SALA (2006): “VARs, common factors and the empirical validation of equilibrium business cycle models,” *Journal of Econometrics*, 132, 257–279, common Features.
- GILCHRIST, S. AND E. ZAKRAJŠEK (2012): “Credit Spreads and Business Cycle Fluctuations,” *American Economic Review*, 102, 1692–1720.
- GORODNICHENKO, Y., T. PHAM, AND O. TALAVERA (2021): “The voice of monetary policy,” Tech. rep., National Bureau of Economic Research.
- GÜRKAYNAK, R. S., B. SACK, AND E. SWANSON (2005): “The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models,” *American Economic Review*, 95, 425–436.
- HANDLAN, A. (2020): “Text Shocks and Monetary Surprises: Text Analysis of FOMC Statements with Machine Learning,” .

- HANSEN, S. AND M. MCMAHON (2016): “Shocking language: Understanding the macroeconomic effects of central bank communication,” *Journal of International Economics*, 99, S114–S133.
- HANSEN, S., M. MCMAHON, AND A. PRAT (2018): “Transparency and deliberation within the FOMC: a computational linguistics approach,” *The Quarterly Journal of Economics*, 133, 801–870.
- HASSAN, T. A., S. HOLLANDER, L. VAN LENT, AND A. TAHOUN (2020): “The Global Impact of Brexit Uncertainty,” Tech. rep., National Bureau of Economic Research.
- HOERL, A. E. AND R. W. KENNARD (1970): “Ridge Regression: Biased Estimation for Nonorthogonal Problems,” *Technometrics*, 12, 55–67.
- HOLM, M. B., P. PAUL, AND A. TISCHBIREK (2021): “The Transmission of Monetary Policy under the Microscope,” *Journal of Political Economy*, 129, 2861–2904.
- JAROCINSKI, M. AND P. KARADI (2020): “Deconstructing Monetary Policy Surprises—The Role of Information Shocks,” *American Economic Journal: Macroeconomics*, 12, 1–43.
- JORDÀ, O. (2005): “Estimation and Inference of Impulse Responses by Local Projections,” *American Economic Review*, 95, 161–182.
- KALAMARA, E., A. TURRELL, C. REDL, G. KAPETANIOS, AND S. KAPADIA (2020): “Making text count: economic forecasting using newspaper text,” .
- LEEPER, E. M., C. A. SIMS, AND T. ZHA (1996): “What does monetary policy do?” *Brookings papers on economic activity*, 1996, 1–78.
- LOUGHRAN, T. AND B. MCDONALD (2011): “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks,” *The Journal of Finance*, 66, 35–65.
- LUCCA, D. O. AND F. TREBBI (2009): “Measuring Central Bank Communication: An Automated Approach with Application to FOMC Statements,” Working Paper 15367, National Bureau of Economic Research.
- MERTENS, K. AND M. O. RAVN (2013): “The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States,” *American Economic Review*, 103, 1212–47.
- MIRANDA-AGRIPPINO, S. AND G. RICCO (2021): “The Transmission of Monetary Policy Shocks,” *American Economic Journal: Macroeconomics*, 13, 74–107.
- NAKAMURA, E. AND J. STEINSSON (2018): “High-frequency identification of monetary non-neutrality: the information effect,” *The Quarterly Journal of Economics*, 133, 1283–1330.

- OCHS, A. C. (2021): "A New Monetary Policy Shock with Text Analysis," *Faculty of Economics, University of Cambridge*.
- PEEK, J., E. S. ROSENGREN, AND G. TOOTELL (2016): "Does Fed policy reveal a ternary mandate?" .
- PLAGBORG-MOLLER, M. AND C. K. WOLF (2021): "Local Projections and VARs Estimate the Same Impulse Responses," *Econometrica*, 89, 955–980.
- RAMEY, V. A. (2016): "Macroeconomic shocks and their propagation," *Handbook of macroeconomics*, 2, 71–162.
- REIFSCHNEIDER, D. AND P. TULIP (2019): "Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve's approach," *International Journal of Forecasting*, 35, 1564–1582.
- ROMER, C. D. AND D. H. ROMER (2000): "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 90, 429–457.
- (2004): "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 94, 1055–1084.
- SHAPIRO, A. H. AND D. WILSON (2021): "Taking the fed at its word: A new approach to estimating central bank objectives using text analysis," Federal Reserve Bank of San Francisco.
- SHARPE, S. A., N. R. SINHA, AND C. A. HOLLRAH (2020): "The Power of Narratives in Economic Forecasts," *Working Paper*.
- SIMS, C. A. (1992): "Interpreting the macroeconomic time series facts: The effects of monetary policy," *European Economic Review*, 36, 975–1000.
- STOCK, J. H. AND M. W. WATSON (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 128, 917–948.
- TENREYRO, S. AND G. THWAITES (2016): "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions," *American Economic Journal: Macroeconomics*, 8, 43–74.
- THORNTON, D. L. (2005): "A new federal funds rate target series: September 27, 1982–december 31, 1993," *FRB of St. Louis Working Paper No.*
- (2006): "When Did the FOMC Begin Targeting the Federal Funds Rate? What the Verbatim Transcripts Tell Us," *Journal of Money, Credit and Banking*, 38, 2039–2071.
- UHLIG, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.
- WIELAND, J. F. AND M.-J. YANG (2020): "Financial dampening," *Journal of Money, Credit and Banking*, 52, 79–113.

ONLINE APPENDIX TO
**Identifying Monetary Policy Shocks:
A Natural Language Approach**

by S. Borağan Aruoba and Thomas Drechsel

A Algorithm to combine and exclude concepts

The below algorithm describes how we deal with overlapping economic concepts in Step 2 of our procedure, which is described in Section 2 of the main text.

1. Start with triples. Go through the list of triples that have at least 250 mentions (around one per meeting on average). Select triples that are economic concepts (based on judgment).
- 2.a) Go through the list of doubles that have at least 500 mentions. Select doubles that are economic concepts (based on judgment).
- 2.b) **IF** a selected double is a subset of one or several triples:
 - Unselect the double and keep the triple(s) **IF**
[*Criterion 1*] the triples close to add up to the double **AND**
[*Criterion 2*] the triples are sufficiently different concepts
OR
[*Criterion 3*] the double by itself is too ambiguous
 - **ELSE:** keep the double and unselect the triple(s)
- 3.a) Go through the list of singles that have at least 2000 mentions. Select singles that are economic concepts (based on judgment).
- 3.b) **IF** a selected single is a subset of one or several doubles:
 - Unselect the single and keep the double(s) **IF**
[*Criterion 1*] the doubles close to add up to the single **AND**
[*Criterion 2*] the doubles are sufficiently different concepts
OR
[*Criterion 3*] the single by itself is too ambiguous
 - **ELSE** Keep the single and unselect the double(s)

END

An example of *Criterion 1* and *Criterion 2* being satisfied is for: “commercial real estate” and “residential real estate”. The occurrences of these two triples almost exactly add up to the occurrences of the double “real estate”. Since they are also sufficiently different concepts (e.g. capture meaningfully different markets and thus span richer information), we kept the two triples.

An example *Criterion 1* not being satisfied and *Criterion 3* not being satisfied is for the single “credit”. While there are doubles such as “consumer credit” and “bank credit”, the overall occurrence of credit is much larger than the associated doubles. So we decided to keep credit.

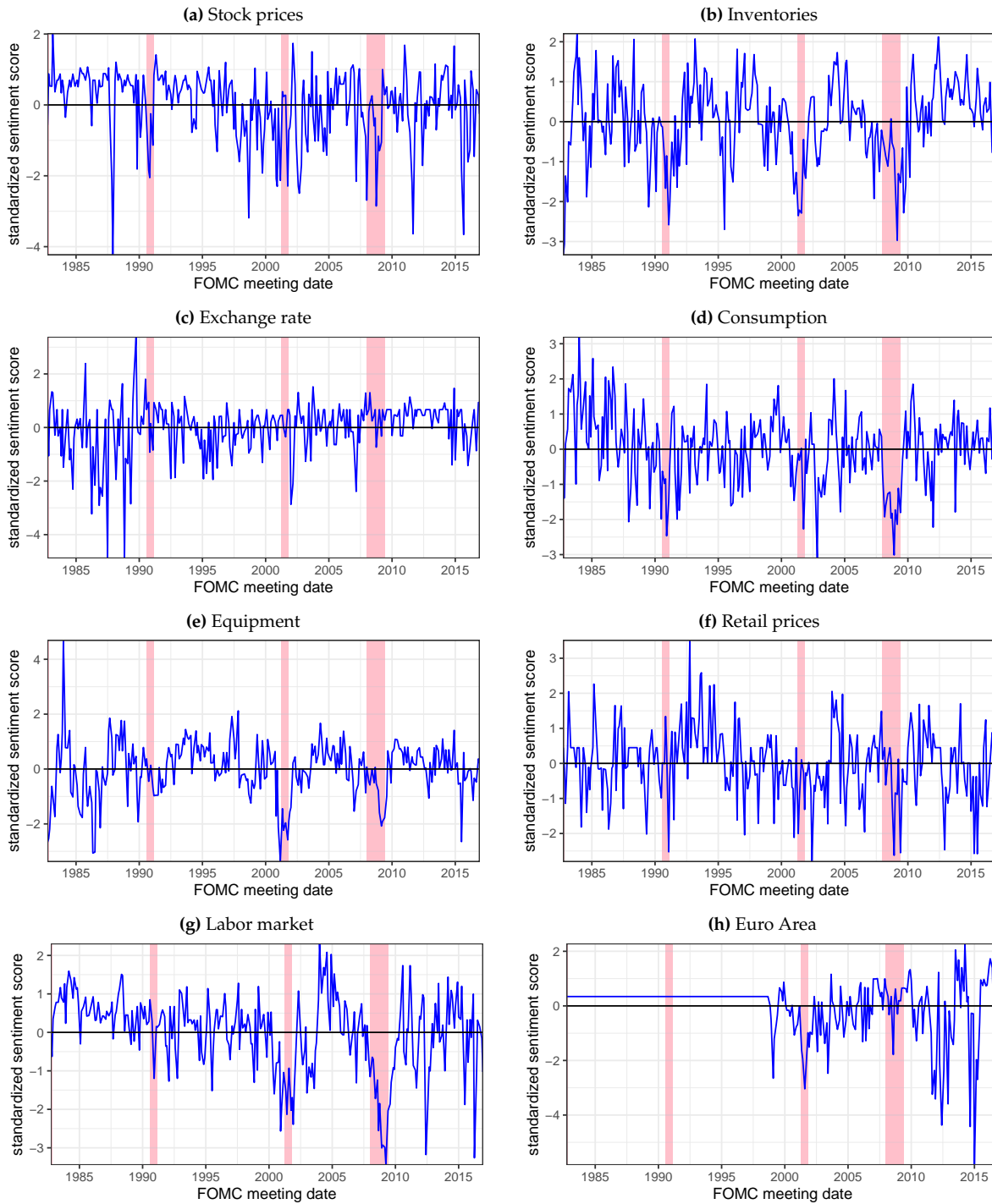
An example *Criterion 1* not being satisfied and *Criterion 3* satisfied is for the single “expenditures”. Unlike credit, this single by itself is too vague based on our judgment (as “capital expenditures” and “government expenditures” are quite different). We therefore selected the doubles, even though their added-up occurrence is well below the one of “expenditures” by itself.

After going through algorithm, we also applied to following additional steps to clean up the list:

- Sometimes a concept occurred as a singular and a plural, for example “oil price” and “oil prices”. In this case, we add them up.
- Sometimes the algorithm produced different concepts that are quite similar, which we unified. For example “stock prices” and “equity prices”. We add them up.
- In a few instances we selected singles and doubles separately for the same single. For example “employment” and “employment cost”.

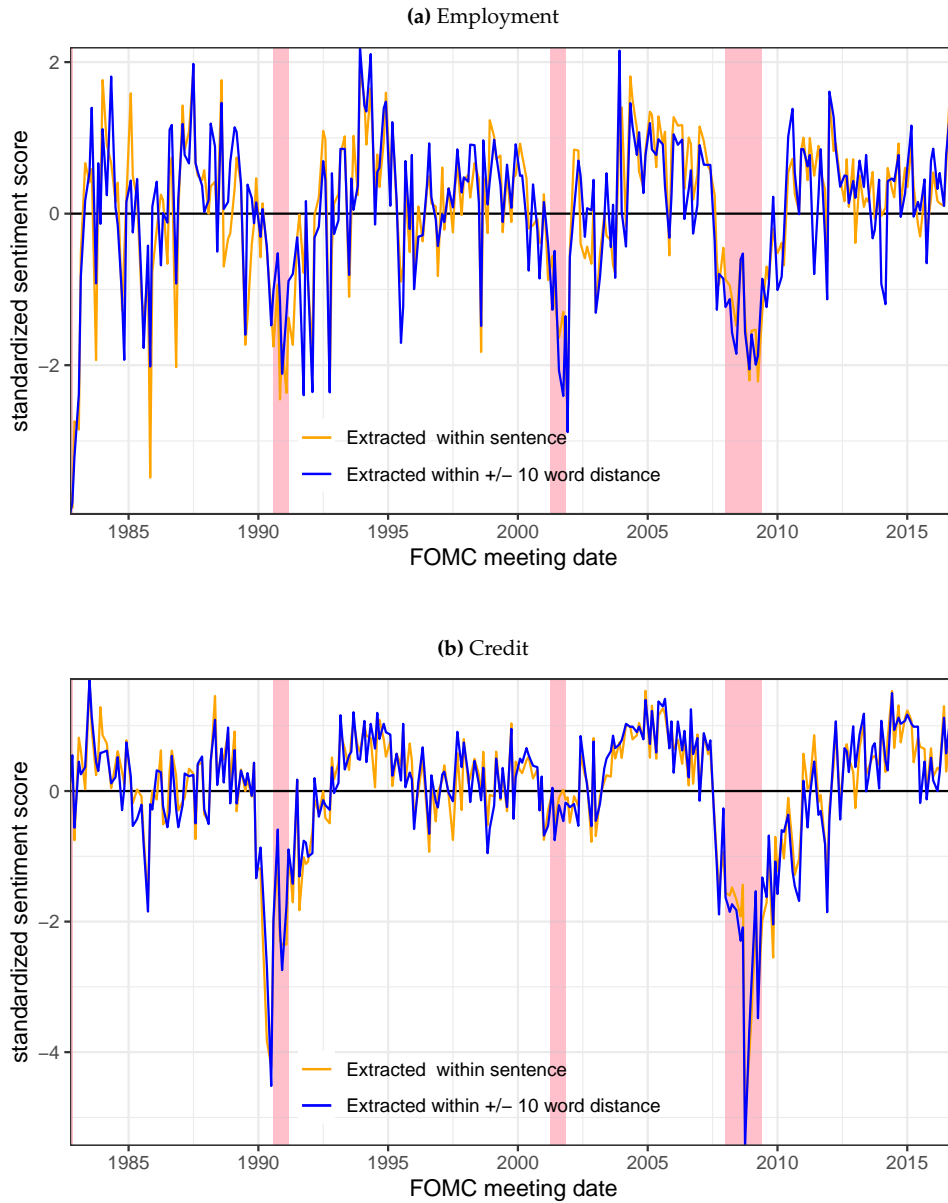
B Additional sentiment indicators

Figure B.1: SELECTED SENTIMENT INDICATORS



C Sentiments in +/- 10 word distance vs. in sentences

Figure C.1: SENTIMENT INDICATORS CONSTRUCTED IN ALTERNATIVE WAYS



Notes. Two examples of sentiment indicators constructed based on positive and negative words within +/- 10 word window vs. based on positive and negative words within the same sentence. See discussion in Section 2.2. For the sentiment surrounding employment the correlation across the two alternative indicators is 0.875. For the case of credit sentiment, the correlation is 0.959. Shaded areas represent NBER recessions.

D Evidence for the modal nature of Greenbook forecasts

We systematically check the transcripts of the FOMC meetings in our sample period 1982 to 2016 for mentions of the terms “modal” and “modal forecasts” and then read the discussions around those instances. Below we provide several examples, spanning all decades over our sample period, that indicate that the staff and members of the FOMC interpret the Greenbook forecasts as modal in nature.

- In the **February 1985** meeting, Governor Wallich asks the staff “Could I ask a question on that? The greater probability is the number on a skewed distribution. Presumably, the probability distribution of inflation is that it can’t go much below zero but it can go up quite far; it has a long right hand tail. Are you thinking in terms of the mode—the most likely single value—or the mean, including the tail?”

The director of Research and Statistics James L. Kichline responds “We have alleged for years that we have a modal forecast. I would say that it’s very difficult, but basically, if we use the model and try to come out with confidence intervals, the model comes out with substantially lower rates of inflation. In fact, if you put a 70 percent confidence interval around our deflator estimate, a couple of times we drift out of that range on the high side. So with the same policy assumptions for 1985, the model forecast, for whatever it’s worth, is a rate of increase in the deflator one percentage point less than in the staff forecast. I view that information as saying that the risks tend to be skewed on the down side. We think 3-1/2 percent is the most likely outcome; but if we’re wrong, I’d say we’re probably too high rather than too low.”

- In the **July 1996** meeting Michael Prell, the director of Research and Statistics clarifies: “I think there have been some occasions when we have indicated that the risks in our outlook were asymmetric. I would characterize our forecasts over the years as an effort to present a meaningful, modal forecast of the most likely outcome. When we felt that there was some skewness to the probability distribution, we tried to identify it. In this instance, as we looked at the recent data, we felt that there was a greater thickness in the area of our probability distribution a little above our modal forecast.”

[This is the example we provide in the main text.]

- In the **November 2001** meeting, Governor Meyer states, in reference to the 9/11 terrorist attacks that “The Greenbook, like most forecasts, seems to assume a one-time terrorist attack with a near-term effect on confidence that dissipates over time. That might be appropriate for a modal forecast. But relative to this assumption, there seems to be significant asymmetric downside risks, specifically of further terrorist attacks that affect

confidence in the economy or perhaps for other reasons as well. The forecast for the first state of the world is therefore likely to be biased in an optimistic direction though, as David Stockton noted, we would be hard pressed to parameterize the downside risks associated with the second state of the world. Still this analysis suggests that the mean of the forecast might be interpreted as being below the mode in this case. So the question is how policy should respond to this type of uncertainty and whether policy should be set to err on the side of ease relative to the modal forecast.”

- In the **March 2005** meeting, President of the Federal Reserve Bank of San Francisco Janet Yellen states that “While the Greenbook expectation of a relatively flat path for bond rates through the end of next year may be a reasonable modal forecast, I don’t think the risks here are balanced.”
- In the **June 2009** meeting, FOMC secretary Brian Madigan lays out different policy options, with reference to the forecasts: “With both a modal outlook for weak growth and low inflation, and downside risks around the outlook for activity, macroeconomic considerations would seem to argue for providing additional monetary policy stimulus at this juncture. However, with the federal funds rate at the zero bound, the Committee has limited policy options at its disposal.”
- In the **June 2011** meeting, President of the Federal Reserve Bank of San Francisco John Williams explains “Furthermore, despite the deep cuts to the output projection, the Tealbook has also shifted to a downside skew to the risks of the growth outlook. This combination of a downward modal revision to the growth forecast and downside risk assessment is a truly sobering development, but it’s consistent with what we see in financial markets.”
- In the **December 2016** meeting, Vice Chairman Dudley says “I guess my view of the risks to the forecast is that you have a modal forecast and then you ask, where is the skew of the distribution? It’s not about where the lower bound lies relative to the funds rate. So I guess I interpret the balance of the risks differently (...).”

E More results on forecast error predictability

E.1 Additional results for output and inflation forecasts

Table E.1: ADDITIONAL GREENBOOK FORECAST ERROR PREDICTABILITY TESTS

Panel (a): unemployment rate forecast errors on LHS								
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	-0.029* [0.016]	-0.114** [0.049]	-0.445** [0.190]	-0.622** [0.238]				
Economic activity sentiment					-0.026 [0.016]	-0.098** [0.048]	-0.285* [0.165]	-0.363** [0.171]
Constant	-0.019 [0.014]	-0.070** [0.033]	-0.082 [0.121]	0.059 [0.201]	-0.019 [0.014]	-0.069** [0.035]	-0.077 [0.145]	0.160 [0.258]
R-squared	0.045	0.149	0.248	0.208	0.033	0.097	0.090	0.055
Obs	210	210	210	62	210	210	210	62
Panel (b): output forecast errors on LHS								
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	0.121 [0.220]	0.411 [0.325]	0.540* [0.310]	-0.171 [0.402]				
Economic activity sentiment					0.036 [0.228]	0.146 [0.272]	0.079 [0.251]	-0.485 [0.403]
Constant	0.300* [0.167]	0.139 [0.276]	-0.252 [0.340]	-0.380 [0.750]	0.298* [0.163]	0.131 [0.299]	-0.268 [0.374]	0.442 [0.717]
R-squared	0.005	0.030	0.049	0.003	0.000	0.003	0.001	0.021
Obs	206	204	198	54	206	204	198	54
Panel (c): inflation forecast errors on LHS								
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	0.148 [0.101]	0.170 [0.133]	0.142 [0.173]	-0.011 [0.164]				
Economic activity sentiment					0.263*** [0.092]	0.222* [0.126]	0.236* [0.141]	0.013 [0.214]
Constant	-0.163 [0.109]	-0.136 [0.167]	-0.267 [0.208]	0.056 [0.216]	-0.167 [0.103]	-0.140 [0.160]	-0.271 [0.201]	-0.019 [0.207]
R-squared	0.029	0.032	0.017	0.013	0.081	0.049	0.041	0.000
Obs	210	210	210	62	210	210	210	62

Notes. Panel (a) repeats Table 2 from the main text. Panels (b) and (c) show analogous results for real output growth and inflation forecasts.

E.2 Results for first release instead of final vintage

Table E.2: GREENBOOK FORECAST ERROR PREDICTABILITY TESTS FOR FIRST RELEASE

Panel (a): unemployment rate forecast errors on LHS								
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	-0.025* [0.013]	-0.104** [0.045]	-0.433** [0.189]	-0.637** [0.242]				
Economic activity sentiment					-0.020 [0.014]	-0.089* [0.044]	-0.272 [0.166]	-0.376** [0.173]
Constant	-0.032*** [0.011]	-0.084*** [0.031]	-0.097 [0.119]	0.048 [0.240]	-0.032*** [0.011]	-0.083* [0.032]	-0.093 [0.142]	0.150 [0.260]
R-squared	0.038	0.129	0.240	0.214	0.020	0.084	0.084	0.058
Obs	210	210	210	62	210	210	210	62
Panel (b): output forecast errors on LHS								
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	-0.093 [0.125]	0.172 [0.256]	0.327 [0.282]	-0.291 [0.245]				
Economic activity sentiment					-0.144 [0.131]	0.052 [0.235]	-0.069 [0.228]	-0.551* [0.318]
Constant	0.214** [0.103]	0.070 [0.192]	-0.236 [0.256]	-0.348 [0.568]	0.218** [0.106]	0.067 [0.200]	-0.241 [0.283]	-0.374 [0.535]
R-squared	0.006	0.009	0.024	0.015	0.012	0.001	0.001	0.045
Obs	206	204	198	54	206	204	198	54
Panel (c): inflation forecast errors on LHS								
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	0.104 [0.091]	0.049 [0.093]	0.116 [0.126]	0.062 [0.155]				
Economic activity sentiment					0.201** [0.087]	0.098 [0.093]	0.232** [0.115]	0.130 [0.196]
Constant	-0.167** [0.079]	-0.133 [0.123]	-0.281* [0.155]	-0.483** [0.214]	-0.170** [0.073]	-0.135 [0.120]	-0.285** [0.143]	-0.470** [0.212]
R-squared	0.018	0.003	0.013	0.004	0.059	0.010	0.046	0.012
Obs	210	210	210	62	210	210	210	62

Notes. This table repeats Table E.1, based on the outcome being the first release (constructed from ALFRED) rather than the final vintage of each variable.

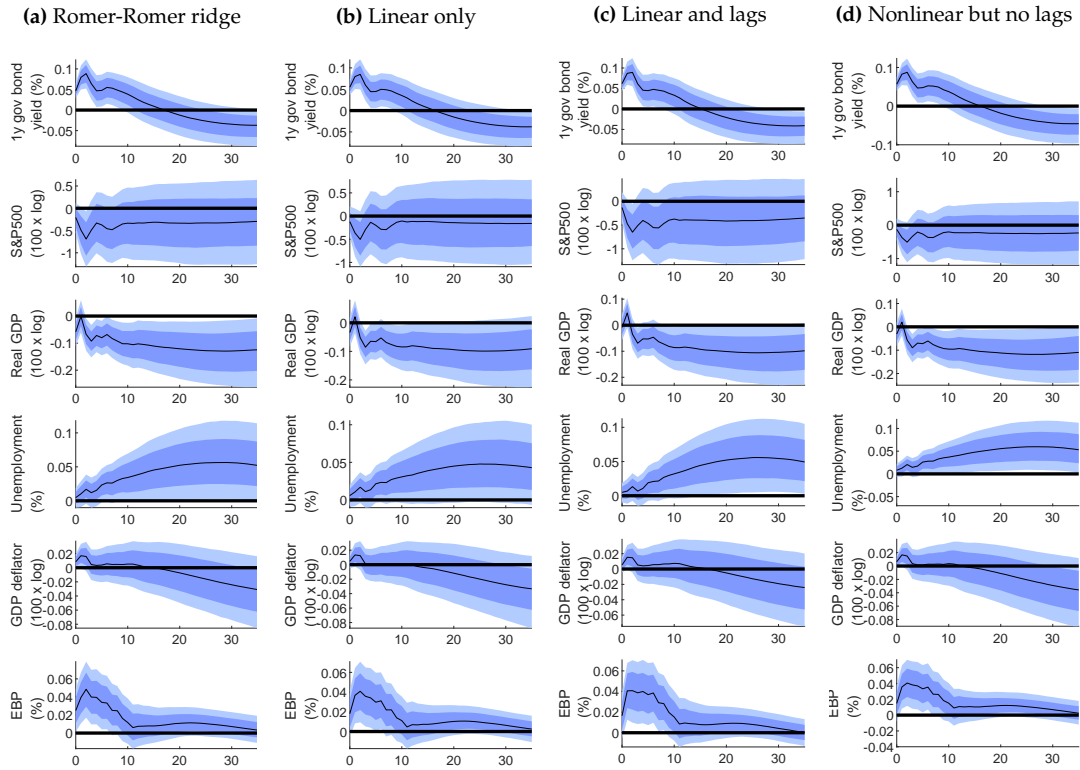
F Construction of committee composition variables

The additional data set that captures information on the composition of the FOMC in each meeting, which we use for robustness, is constructed as follows. For each FOMC meeting, we record the list of participants. This list consists of the governors at the board as well as the representatives from each regional bank. Typically, regional bank representatives are their respective presidents, except in cases where there is an interim president. We classify the participants by their voting status: they are either voting members, alternate members, or non-voting members. The governors always vote and the regional bank presidents alternate between the three roles. For each governor, we create a dummy variable that equals 1 if he/she attended a given meeting and 0 otherwise. We record the attendance of each regional bank representative in a similar way. Here we create three sets of dummy variables. The first set of variables are constructed at the participant-position-voting status level, meaning for example that we distinguish between Mr. Boehne (president of the FRB of Philadelphia) when he is attending as a voting member and when he is attending as a non-voting member. The second set of variables are constructed only at the participant-position level, without regard to their voting statuses. The last set of variables recorded whether a regional bank's representative voted during the meeting for each of the 12 banks. For governors, we also record information on who appointed them. We tally the total number of governors in attendance by the US president who made the appointment, as well as the number of governors appointed by a Republican and Democratic administration respectively.¹ In addition to attendance, for each meeting we record the number of motions voted upon and the results of each vote. Indicator variables are constructed for whether there is only one vote during the meeting, whether there is not a vote at all, and in the case that there is one vote, whether the voting result was unanimous. Lastly, we tally the total number of female participants in attendance at each meeting. Over the sample period 1982:10 to 2008:10, this results in 298 variables.

¹In the case that a governor served multiple tenures appointed by different US presidents, we make that distinction. For example, Janet Yellen was appointed by Bill Clinton to serve as a governor in 1994 and then by Barack Obama in 2010 – and these are recorded separately.

G Additional IRFs from BVAR

Figure G.1: IRFS ESTIMATED FROM INTERMEDIATE SHOCK VERSIONS

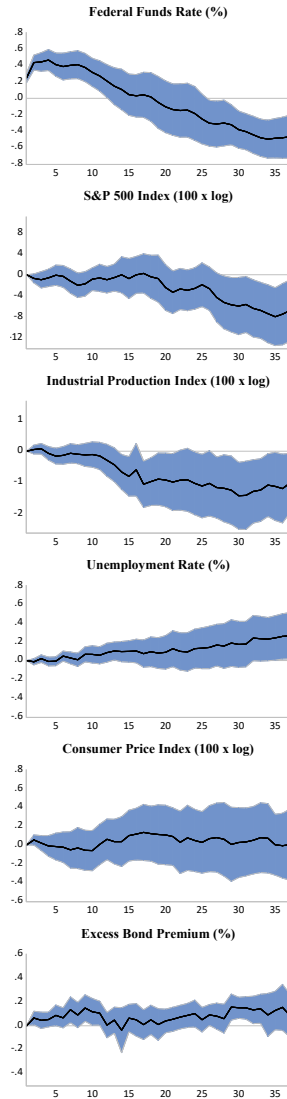


Notes. IRFs to different intermediate versions of the estimated monetary policy shocks, computed from the BVAR model. Panel (a) shows the IRFs to the shocks from an empirical specification where only the extended set of forecasts are used in a ridge regression. Panel (b) uses the measure of monetary policy shocks retrieved from a linear instead of nonlinear ridge model using the extended set of numerical forecasts and sentiment indicators, but where no lags or squared sentiment indicators are included. Panel (c) is similar to Panel (b) but the specification to estimate the shocks also adds lagged sentiments. Panel (d) is similar to Panel (b) but the specification to estimate the shocks also adds squared terms. The sample period to estimate the shocks is 1982:10-2008:10. The solid line represents the median, the 16th and 84th percentiles are represented by the darker bands, and the 5th and 95th percentiles by the lighter bands. The sample used to estimate the IRFs is 1984:02-2016:12.

H Alternative IRFs using local projections

Figure H.1: IRFS TO DIFFERENT MONETARY POLICY SHOCKS USING LOCAL PROJECTIONS

(a) Using shocks from full ridge model



(b) Using shocks from Romer-Romer OLS



Notes. IRFs to different estimated monetary policy shocks based on local projections approach (Jordà, 2005). Panel (a) uses our proposed measure of monetary policy shocks, estimated using the full nonlinear ridge model on the extended set of numerical forecasts and our sentiment indicators from FOMC documents. Panel (b) shows the analogous IRFs when a simpler empirical specification is used to estimate the shocks, which includes only the original set of numerical forecasts in a Romer-Romer OLS regression. The solid line represents the median, and the 5th and 95th percentiles are captured by the bands. The sample used to estimate the IRFs is 1984:02-2008:10.