Identifying Monetary Policy Shocks: A Natural Language Approach*

S. Borağan Aruoba — University of Maryland

Thomas Drechsel — University of Maryland, CEPR

July 8, 2022

Abstract

We propose a novel method for the identification of monetary policy shocks. By applying natural language processing techniques to documents that economists at the Federal Reserve prepare for Federal Open Market Committee meetings, we capture the information set available to the committee at the time of policy decisions. Using machine learning techniques, we then predict changes in the target interest rate conditional on this information set, and obtain a measure of monetary policy shocks as the residual. An appealing feature of our procedure is that only a small fraction of interest rate changes is attributed to exogenous shocks. We find that the dynamic responses of macroeconomic variables to our identified shocks are consistent with the theoretical consensus.

Keywords: Monetary policy; Federal Reserve; 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, Tarek Hassan, Diego Kaenzig, Guido Kuersteiner, Michele Lenza, Michael McMahon, Vitaliy Mersault, Ivan Petrella, Chris Redl, Frank Schorfheide, Minchul Shin, and Lumi Stevens, as well as seminar participants and the Federal Reserve Bank of Philadelphia, the Central Bank of Chile, the Glasgow Workshop on Recent Advances in Econometrics, the Reserve Bank of Australia, the University of Hamburg, and the University of Maryland for helpful comments. Eugene Oue, Danny Roth, Mathias Vesperoni and Jialing Zhang provided excellent research assistance. Contact: aruoba@umd.edu and drechsel@umd.edu. *We make the Fed sentiment time series, the FOMC composition data set, as well as our estimated monetary policy shocks 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 a systematic response to economic conditions, but instead occur exogenously. 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 FFR, where systematic changes are based on information and forecasts about the economy 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 Federal Reserve Board economists for Federal Open Market Committee (FOMC) meetings. They then retrieve a measure of monetary policy shocks as the residual from this regression. We propose an identification approach that follows the idea of exploiting the information in documents prepared for the FOMC. Our methodology incorporates *more* 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).

We estimate monetary policy shocks as the residuals from a prediction of changes in the FOMC's FFR target using *(i)* all available numerical forecasts in the documents that Federal Reserve Board economists prepare for the FOMC; *(ii)* a comprehensive summary of the verbal information in the documents; 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). Our collection of 296 sentiment time series paints a rich picture of the historical assessment of economic conditions by Fed economists.

A regression with FFR changes on the left hand side and *(i)*, *(ii)* and *(iii)* on the right hand side is infeasible given that there are many 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 regressors. This choice is guided by the recent insight 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 regression coefficient from zero to achieve shrinkage. To select the ridge penalty parameter, we suggest two alternative options. The first is k-fold cross-validation, a standard way in the machine learning literature to validate a model's ability to perform out-of-sample in alternating subsets of the data. The second is a restriction on the implied share of FFR variation that can be attributed to systematic changes in monetary policy. 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).

We discuss four sets of findings. First, we examine the relative contribution of systematic and exogenous variation in interest rates implied by our cross-validated ridge regression, compared with benchmark specifications. A linear regression that contains 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 systematic policy, while the other half is exogenous. The R^2 of our ridge regression is 0.75, implying that the systematic component is 25 percentage points more important when a larger set of forecasts, Fed economists' sentiments, as well as nonlinearities are included. In other words, exogenous shocks are much less important in explaining interest rate changes when constructed with our new identification method.

Second, we inspect the economic drivers of systematic monetary policy, and provide an interpretation of what estimated monetary policy shocks capture. In terms of systematic 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 broader 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. In terms of the interpretation of monetary policy shocks, we closely analyze the discussion that took place in the FOMC for 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 current economic outlook, such as long-run credibility concerns. For example, in the November 1994 meeting, the material prepared by the staff economists indicated that the market had already built in a rate hike. However Chairman Alan Greenspan advocated a larger hike, arguing that "a mild surprise would be of significant value."

Third, we verify whether including information in our ridge regression that is not contained in the documents prepared by the Fed staff alters our measure of shocks. For this purpose, we construct two additional sets of regressors. One consists of sentiment indicators constructed from FOMC meeting transcripts. These should reflect information that arrives between the time the staff documents are completed and the committee meets. The other set of regressors captures the composition of the FOMC, which includes a dummy for each committee member as well as several personal characteristics. These regressors measure dynamics and meeting interactions not captured in the information provided by staff economists. We show that neither of these sets of variables improves the fit of our ridge regression. This indicates that our measure of shocks is not explained by information beyond that made available to FOMC members by the Fed economists at the beginning of a meeting.

Fourth, with our novel measure of monetary policy shocks at hand, we study impulse response functions (IRFs) of macro variables to changes in monetary policy, 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. While the shocks span the period 1982:10-2008:10, we can study their impact across a sample that includes more recent observations. 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 a tightening there is a swift decline in real GDP after only a few months, while the reduction in the price level builds up sluggishly over time. We also show that IRFs resulting from benchmark shock measures constructed based on a smaller information set lead to responses not in line with the theoretical consensus, and discuss potential interpretations of what confounds these measures. Taken together, we conclude that natural language processing and machine learning deliver a cleanly identified estimate of monetary policy shocks.

Literature. Our work contributes to two 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.[1] There is also a wide array of other approaches to identifying monetary policy shocks. One approach uses SVARs identified in different ways.[2] 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).[3] A survey of these different approaches 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

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

[2] Identification in SVARs can be obtained for example through zero restrictions (Christiano, Eichenbaum, and Evans, 1999), sign restrictions (Uhlig, 2005), or narrative sign restrictions (Antolin-Diaz and Rubio-Ramirez, 2018). Coibion (2012) systematically compares SVAR approaches to that of Romer and Romer (2004).

[3] One 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 rates, it is not contaminated by the Fed information effect.

through a large set of information contained in economic data and text.[4]

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 make inference about central bank objectives.[5] A large set of papers in this branch of research focuses specifically on the interaction between the Fed and financial markets. For example, Cieslak and Vissing-Jorgensen (2020) employ textual analysis on FOMC documents to understand if monetary policy reacts to stock prices.[6] 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 fed funds futures prices 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

[4]Our emphasis on a large information set has parallels to Bernanke, Boivin, and Eliasz (2005) who incorporate many time series in a factor-augmented VAR (FAVAR).

[5]Further papers analyzing Fed language include Acosta (2015) who studies in transcripts how the FOMC's responded to calls for transparency, and Cieslak et al. (2021) who construct text-based measures of uncertainty from FOMC transcripts.

[6]Peek, Rosengren, and Tootell (2016) apply textual analysis to FOMC transcripts to understand to what degree 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 conferences, and then study how these affect markets.

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

Structure of the paper. Section 2 introduces our method to identify monetary policy shocks. Section 3 discusses the results of our identification procedure, including the contribution of systematic policy and our estimated shocks. Section 4 presents our results on the responses of macroeconomic variables to monetary policy shocks. Section 5 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, that is, 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, for example 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

$$\Delta i_t = \alpha + \beta i_{t-1} + \boldsymbol{\gamma} \boldsymbol{X}_t + \varepsilon_t^{RR}, \tag{2}$$

where i_t is the FOMC's FFR target, $\boldsymbol{X}_t$ contains the forecasts of the US economy that the Fed has at its disposal at time t. 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 $\boldsymbol{X}_t$ need to be a good proxy for the whole information set Ω_t that is relevant for the central bank's decisions. Second, the mapping $f(\cdot)$ from the information to decisions is well captured by a linear relationship.

The FOMC reviews a large amount of detailed information on the economic and financial conditions of the US economy. This information is prepared by staff economists as part of different types of confidential documents. Part of the information composed by the Fed's economists consists of numerical forecasts of key macroeconomic variables. Romer and Romer (2004) exploit these numerical forecasts as a proxy for $\boldsymbol{X}_t$.

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, in addition to numerical forecasts. 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 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 $f(.)$. 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 model 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.[7] 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).[8] FOMC meeting documents are available with a 5-year lag, so the latest document currently available is for the last FOMC meeting of 2016.[9] We process documents through to

[7]There are also unscheduled meetings or conference calls during which the FOMC makes policy decisions. Most of these are excluded from the estimation of (2) because usually no new documents are prepared for unscheduled meetings.

[8]We 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).

[9]Our 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.

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 hands side. For each FOMC meeting, a number of 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).[10] 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. In a separate analysis, we do explore using information from the meeting transcripts to study whether they contain additional information. Our choice results in 772 pdf files for 276 meetings (630 files for 210 meeting before the zero lower bound), containing tens of thousands of pages of text and numbers.

For each document, we read its raw textual content into a computer and process it as follows. We remove stop words (such as *the*, *is*, *on*, ...); we remove numbers (that are not separately recorded as forecasts, e.g. dates, page numbers); we remove "erroneous" words. After processing the raw text, we 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").[11] 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.

[10]The *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 by Federal Reserve district.

[11]We also added one quadruple: "money market mutual funds."

Step 2: Identify frequently used economic concepts

We now 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*.[12] 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.

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.

[12] Both 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.

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.[13] This classification builds on the dictionary of positive and negative terms in Loughran and McDonald (2011). Their dictionary is especially constructed for financial text, so should be reasonably suitable for the economic content discussed in the Fed documents. We slightly enhance the dictionary by adding terms specific to Fed language, such as "tightening."[14] 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 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.

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 at business cycle frequency. 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

[13] The 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.

[14] We also remove some terms from this dictionary, such *unemployment* and *unemployed*, because these are among our selected economic concepts.

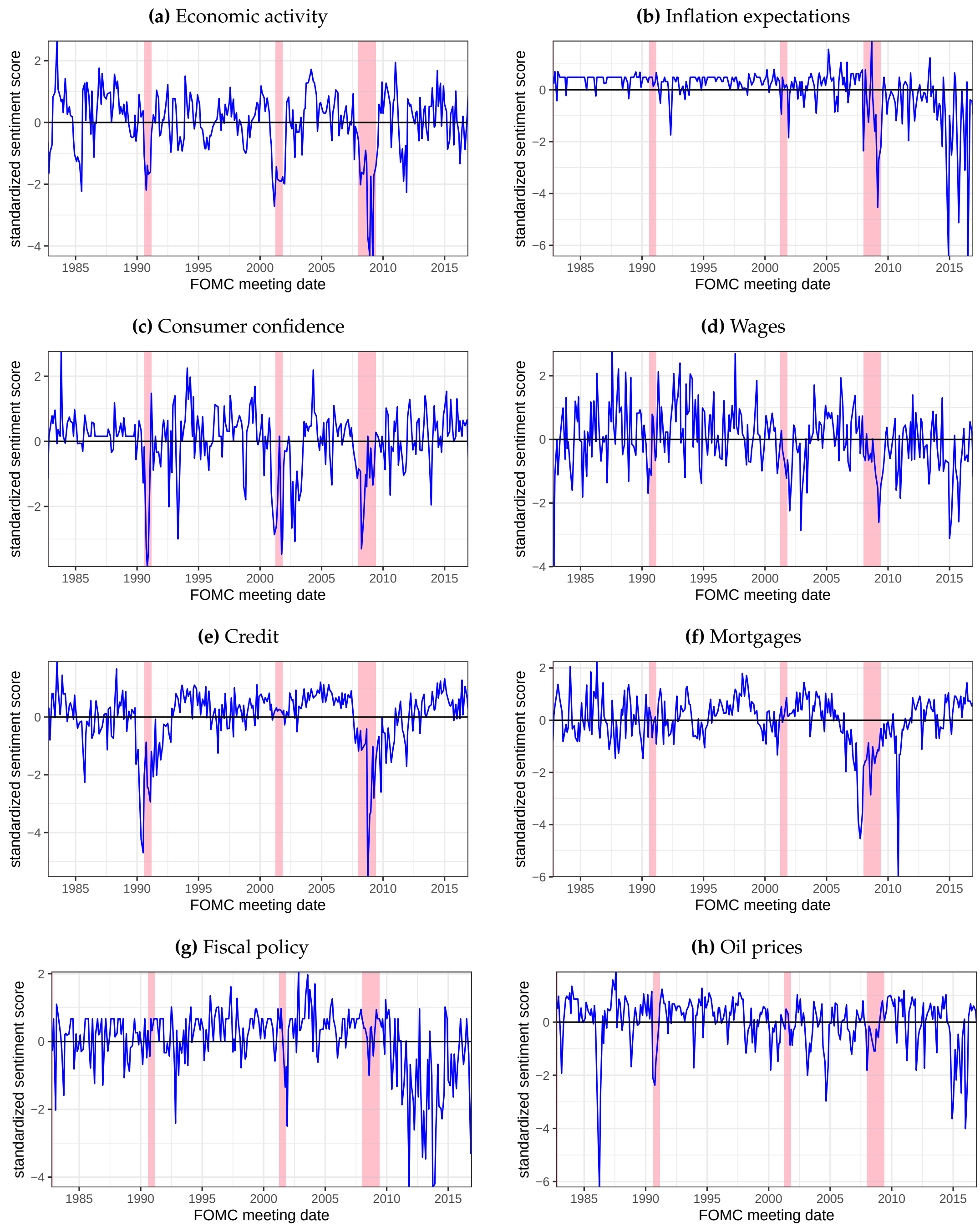
mortgages plays a role primarily in the Great Recession and its aftermath (see Panels (e) and (f)). Another insight coming from the figure is that some concepts gain importance over time. For example, the sentiment around inflation expectations in Panel (b) moves relatively little for most of the sample, 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.

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

Positive sentiment	**Negative sentiment**
able	abandon
best	bad
charitable	calamities
delight	damage
easier	egregious
fantastic	fail
gain	grievances
happiest	halt
ideal	idle
leadership	jeopardize
meritorious	lack
opportunities	malfeasance
perfect	negative
...	...

Notes. Examples of words classified as expressing positive or negative sentiments in the dictionary of Loughran and McDonald (2011). The total number of classified words is 2,885.

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.

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

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

Δi_t are changes in the FOMC's FFR target, which for simplicity we mostly refer to as just the FFR.[15] $\widetilde{\boldsymbol{X}}_t$ contains augmented set of forecasts that Fed economists produce, which includes additional production, investment, housing and government spending variables relative to $\boldsymbol{X}_t$ in (2).[16] Following Romer and Romer (2004)'s specification, we enter forecasts in levels and first differences, across several forecast horizons. Using the available variables and horizons in levels and differences amounts to 132 forecast time series. $\boldsymbol{Z}_t$ contains our 296 sentiment indicators. $\Gamma(\cdot)$ 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 858 variables on the right hand side.

Estimation as ridge regression. Our sample from 1982:10 to 2008:10 captures 210 FOMC meetings. Therefore, an ordinary least squares (OLS) regression with our 858 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 a regression model $y_i = \gamma_1 x_{i1} + \cdots + \gamma_k x_{ik} + \varepsilon_i$, the ridge regression minimizes $\sum_i \varepsilon_i^2 + \lambda \sum_j^k \gamma_j^2$. The Bayesian interpretation of a ridge regression

[15] In the part of our sample that overlaps with the original Romer and Romer (2004) sample, our left hand side is identical to theirs. Afterwards, we use the series constructed by Thornton (2005), updated by FRED through to the end of our sample in 2008.

[16] This forecast data is conveniently made available by the Philadelphia Fed here.

is Bayesian OLS with a prior on each coefficient that is normally distributed, centered around 0, and the scale of the prior variance is equal to λ. Unlike its close sibling, the LASSO regression, which we discuss further below, a ridge regression results in estimated coefficients for all 858 regressors. Importantly, there are different ways to choose λ. We propose two alternative options:

- **Option 1: Optimal tuning parameter.** 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 of equal size; use each subsample to fit model on the $k-1$ other subsamples; in each case, compute a mean-squared error (MSE); compute an average MSE across the k MSEs; find the smallest average MSE by changing λ. We follow this procedure using $k = 10$.

- **Option 2: Impose a restriction about the contribution of systematic policy.** An alternative way to proceed is to impose a restriction on the share of FFR variation that can be attributed to systematic changes in monetary policy. Macroeconomists would typically think of monetary policy decisions to be largely made in a systematic fashion, with a small role for exogenous shocks (see e.g. the discussion in Leeper, Sims, and Zha, 1996). Our suggested restriction is a 90% share of FFR variation attributed to systematic changes and a 10% share explained by shocks.[17]

In short, option 1 selects a value for λ based on maximizing the out-of-sample performance of the model, while option 2 selects λ based on a desired in-sample fit of the model. We implement both options in our application.

Discussion of our NLP and ML choices. We conclude the step-by-step description of our method with two important remarks. First, relative to the rich variety of methods that modern NLP and ML techniques provide, we opt for an approach in which we impose a fair amount of manual restrictions. In particular, we carry out a sentiment analysis for a hand-selected, finite number of economic concepts, an approach sometimes referred to as Aspect-Based

[17] In the same spirit as our second option, Giannone, Reichlin, and Lenza (2008) set priors for a VAR that imposes a high in-sample fit for the interest rate equation of the system.

Sentiment Analysis (Barbaglia, Consoli, and Manzan, 2021). One natural 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.[18] This alternative would involve many tens of thousands of regressors. Instead, we select economic concepts using judgment, reducing the dimensionality of the problem. We prefer this procedure as 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. A 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, according to 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 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. Nevertheless, we also use LASSO regressions for robustness.

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

3 Results of the identification procedure

This section discusses the estimation results for different versions of 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.

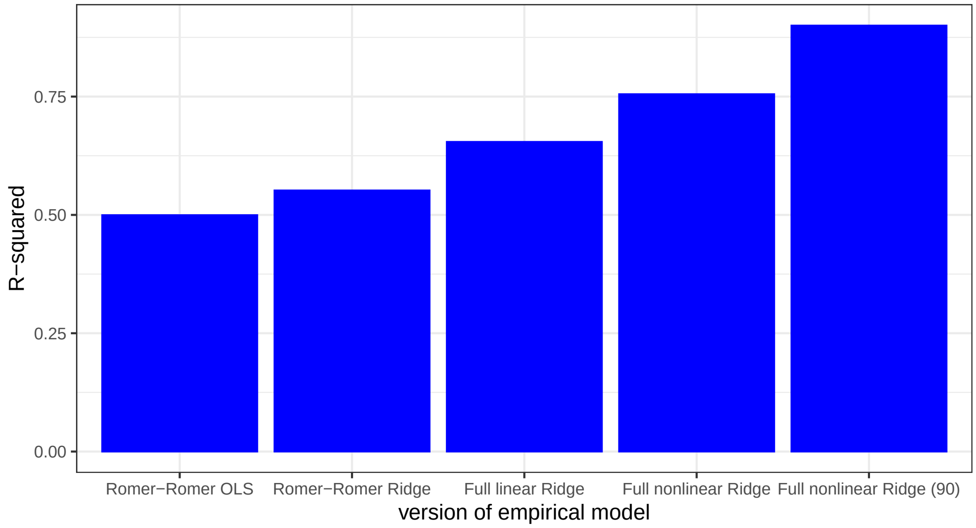
3.1 Systematic vs. exogenous changes in interest rates

Figure 3 presents measures of R^2 across empirical specifications. First, as the simplest benchmark it includes equation (2), a restricted version of (3) where only the forecasts used by Romer and Romer (2004) enter in a linear OLS estimation (results are labeled 'Romer-Romer OLS'). Second, a model that includes the expanded set of forecasts, and is estimated as a ridge regression ('Romer-Romer Ridge'). Third, the figure contains a ridge model where our augmented set of forecasts and sentiments are included, but function $\Gamma(\cdot)$ is linear ('Full linear Ridge'). Fourth, our main model with all 858 variables entering linearly and quadratically ('Full nonlinear Ridge'). In all ridge models so far, the ridge penalty parameter λ is estimated based on an optimal average MSE. The fifth model in Figure 3 corresponds again to the full nonlinear ridge regression, but restricts R^2 to be 0.9 by adjusting the ridge penalty parameter.

We compare the goodness of fit between these alternative models to understand what different elements of our approach imply about the contribution of the systematic component of monetary policy. The first bar in Figure 3 shows that over the sample period 1982:10-2008:10 we consider, the Romer-Romer OLS model implies an R^2 of 0.5. In other words, this empirical model implies that 50% of the variation in the FFR target is systematic, while 50% is attributed to shocks. This seems undesirable, given that macroeconomists typically think of monetary policy decisions to be largely made in a systematic fashion, with a small role for exogenous shocks. 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 bars of the figure reveal that expanding the information

Figure 3: FIT OF ALTERNATIVE EMPIRICAL SPECIFICATIONS



Notes. R^2 implied by the estimation of different versions of the empirical model specified in equation (3), over the sample period 1982:10 to 2008:10. *Romer-Romer OLS*: set of variables used by Romer and Romer (2004), estimated with OLS; *Romer-Romer Ridge*: same set of variables, estimated with ridge regression; *Full linear Ridge*: augmented set of forecasts and sentiment indicators, estimated with ridge regression; *Full nonlinear Ridge*: augmented set of forecasts, sentiment indicators, and quadratic terms in these variables, estimated with ridge regression; *Full nonlinear Ridge (90)*: same model, but ridge penalty parameter chosen based on restriction that systematic policy contributes to 90% of the variation in the FFR.

set in the empirical model increases the implied fit. Each step of enriching the empirical model – going from OLS to ridge regression, including more numerical forecasts and sentiment indicators, and allowing for nonlinearities – delivers some additional improvement in the fit of the model. This is not a purely mechanical effect, as the ridge regression does not maximize fit, but instead optimizes out-of-sample performance in the k-fold cross-validation. Our preferred specification, the fourth bar in Figure 3 implies an R^2 of 0.75, suggesting that 75% of FFR variation is systematic, and 25% are explained by shocks. Relative to the Romer-Romer OLS model, this reduces the contribution of exogenous shocks by about half. The last bar, showing an R^2 of 0.9 is an illustration of the fact that our proposed procedure can incorporate a restriction on the R^2, reflecting a researcher's a priori view about the relative contribution of the systematic and exogenous component in FFR variation.

We check the robustness of these results in Table 2. The first column repeats the results from Figure 3. The remaining columns vary our procedure along

Table 2: R^2 FOR DIFFERENT MACHINE LEARNING TECHNIQUES AND SENTIMENT VERSIONS

	(1) 10-word sentiment Ridge regression	(2) 5-word sentiment Ridge regression	(3) 10-word sentiment LASSO regression	(4) 5-word sentiment LASSO regression
Romer-Romer OLS	0.50	0.50	0.50	0.50
Romer-Romer ML	0.55	0.55	0.57	0.57
Full linear ML	0.65	0.66	0.55	0.63
Full nonlinear ML	0.75	0.77	0.81	0.72
Full nonlinear ML (90)	0.90	0.90	0.90	0.90

Notes. Implied R^2 from estimating different empirical specifications of equation (3). Column (1) is our preferred model. Column (2) uses a 5-word rather than 10-word distance in Step 3 of our procedure. Columns (3) and (4) employ LASSO rather than ridge regression in Step 4.

two dimensions. First, in column (2) we show the corresponding measures of R^2 of empirical models in which our sentiment indicators are constructed using a 5-word instead of a 10-word window around economic concepts (see discussion under Step 3 above). Second, in columns (3) and (4) we apply LASSO rather than ridge regressions. This illustrates that other ML techniques can be used in Step 4 of our methodology without a substantive change in results. By construction, the first and the last row in each column remain unchanged, as the first row does not incorporate sentiments and uses OLS, and the row fixes the R^2 based on an a priori restriction. The table shows that the increase in fit from expanding the information set, which is not a mechanical relation in the case of cross-validation techniques, remains present when we vary our method along the two dimensions.

3.2 Inspecting the drivers of FFR changes

We now focus on our full nonlinear 10-fold validated ridge specification, which can account for 75% of the variation in FFR changes across FOMC meetings. The coefficient estimates corresponding to the different variables included on the right hand side of 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 systematically associated with interest rate increases by the FOMC. A variable-by-variable analysis of the coefficients is difficult given the long list of 858 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 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. Among the numerical 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 ridge model, where the left hand side is the change in the FFR and the right hand side contains these 50 PCs. 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.[19] Third, among these selected PCs we uncover which variables have the largest loadings in absolute value on a given PC. Examining those variables is then informative about what type of information from the documents prepared for the FOMC contributes meaningfully to explaining variation in the Fed's target interest rate.

Table 3 presents the 10 variables with the largest loadings in absolute value on each of the three PCs selected by our procedure. A positive (negative) loading means that an increase in a given variable is associated with an increase (decrease) in the FFR. The first sentiment PC looks very much like a broad real activity sentiment factor, capturing the Fed economists' sentiments about activity, production, labor and housing markets, and consumer confidence.[20] Improvements in the sentiment around these concepts are associated with hikes in the FFR. Interestingly, the second sentiment PC

[19] PCs 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.

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

Table 3: 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

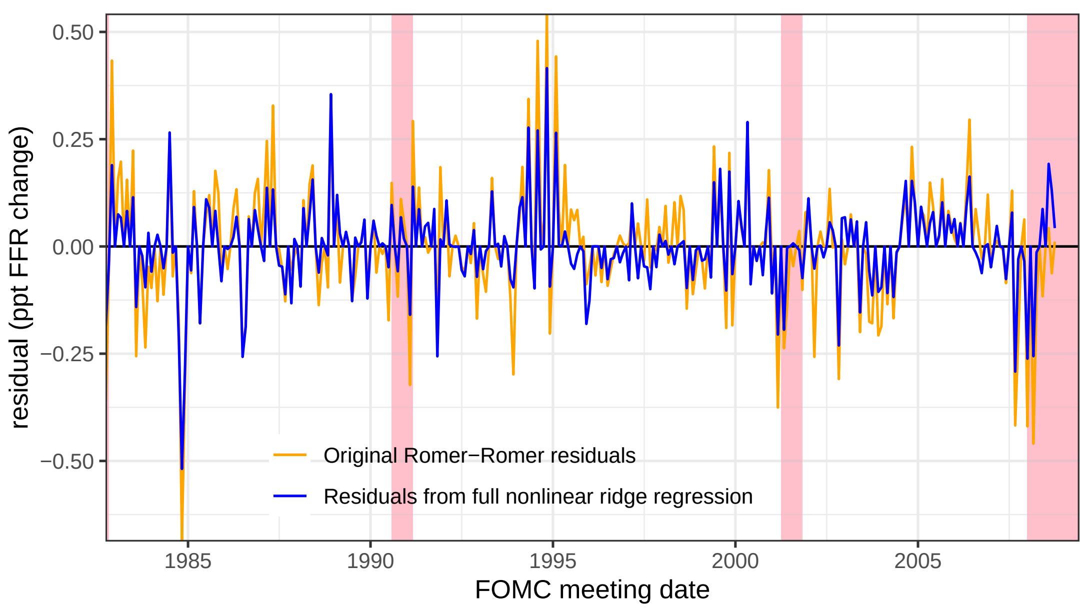
Notes. Variables and associated loadings on the principal components with largest economic significant in a ridge regression with FFR changes on the left hand side. For numerical forecasts, the horizon in quarters is given in brackets.

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.

3.3 Inspecting the shocks

The dark blue line in Figure 4 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 25% 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 absolute terms, while in various others larger shocks are visible in the blue line.

Figure 4: 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. In this meeting, the policy change is equivalent to a decline in the FFR of 75 basis points and our shock measure is minus 52 basis points, indicating that based on staff forecasts and sentiments, we predict a decline of

23 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 there was a 75 basis point increase in the FFR and our analysis shows 43 basis point of this was a monetary policy shock, our second largest in absolute terms. The Fed staff documents paint 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 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.

3.4 Is there information that is omitted in the identification?

Our approach includes 858 linear and nonlinear forecast and sentiment variables to predict changes in the FFR. It could be the case that this might still not be sufficient to capture the full information set available to the FOMC when policy decisions are made. In this case, our measure of monetary policy shocks would not be completely exogenous but instead include some remaining endogenous variation in interest rates. We aim to investigate whether this issue is relevant for our estimate of monetary policy shocks by further expanding the information set, and then verifying whether the contribution of exogenous shocks to FFR variation decreases further. In particular, we include two additional sets of variables into equation (3).

Transcript sentiment indicators. We carry out the same sentiment analysis for the same 296 economic concepts described in Steps 2 and 3 of our procedure, but do so also on the FOMC meeting transcripts. While the documents we use in our main analysis are prepared by Federal Reserve Board economists prior to meetings, the transcripts describe the actual discussion that take place during FOMC meetings. We do not include these variables in our preferred specification because they might capture information about the decision process, rather than about the information set available to policy makers. Yet we include them as an extra set of regressors here to see if they actually do provide additional information about FFR variation.

Committee composition variables. To further capture information about the FOMC meetings, we construct a separate data set that captures the composition of the FOMC for each meeting. This is composed of dummy variables that are 1 if a specific member attends a meeting and 0 otherwise, for each governor and regional bank representative that has ever served on the committee over the sample period 1982:10-2008:10. In addition to attendance dummies, we collect information on voting status (not all regional bank representatives vote in each meeting), the US presidents that have appointed given governors, information on the unanimity of votes, as well as the number of female attendants. In total this results in 298 additional variables that capture the composition of the FOMC. More details on the construction of this additional data set are provided in Appendix D.

Table 4: ADDING FURTHER INFORMATION

Specification	R^2
Full nonlinear Ridge	0.7505
Adding transcript sentiments and committee composition	0.7516
Difference	0.0011

Notes. Comparison of the systematic policy contribution with and without including additional information in the form of sentiment indicators from FOMC meetings transcripts and variables capturing information about the FOMC members.

Table 4 shows the R^2 from our preferred specification in comparison to the model that includes both the transcript sentiment indicators, as well as the committee composition variables, linearly and non-linearly, in addition to everything else. That expanded set of information amounts to a list of 1,585 variables. It is evident from the table that the R^2 hardly increases when the additional information is included, leading us to conclude that the estimate of the systematic component does not change meaningfully.[21] This also increases our confidence in our shocks being truly exogenous to the FOMC's information set. Our analysis in the next section provides further evidence that our identified shock series captures exogenous changes in monetary policy, and is not confounded by information effects.

[21] The same R^2 across two specifications could still imply a different sequence of estimated shocks. We verified that the shocks from the two estimations are close to identical. As an alternative, we tried to predict the residuals from our full nonlinear ridge regression using the additional information. The fit of this "second stage" regression was near 0.

4 The effects of monetary policy shocks

The findings above indicate that our estimated shocks capture changes in the target interest rate that are orthogonal to the systematic conduct of monetary policy by the Federal Reserve. To further validate our shock measure, we use it to study the effects of monetary policy shocks on the US economy in a state-of-the-art BVAR model. Following Jarocinski and Karadi (2020), the 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, and the excess bond premium (EBP) of Gilchrist and Zakrajšek (2012). Our time series of shocks enters first in a Cholesky ordering. This yields asymptotically identical results to using the shock series as an external instrument (Plagborg-Moller and Wolf, 2021).[22] 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, through to 2016. 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. The S&P500 and EBP are included as forward-looking financial variables. We use the same sample period, settings and priors as in Jarocinski and Karadi (2020).[23]

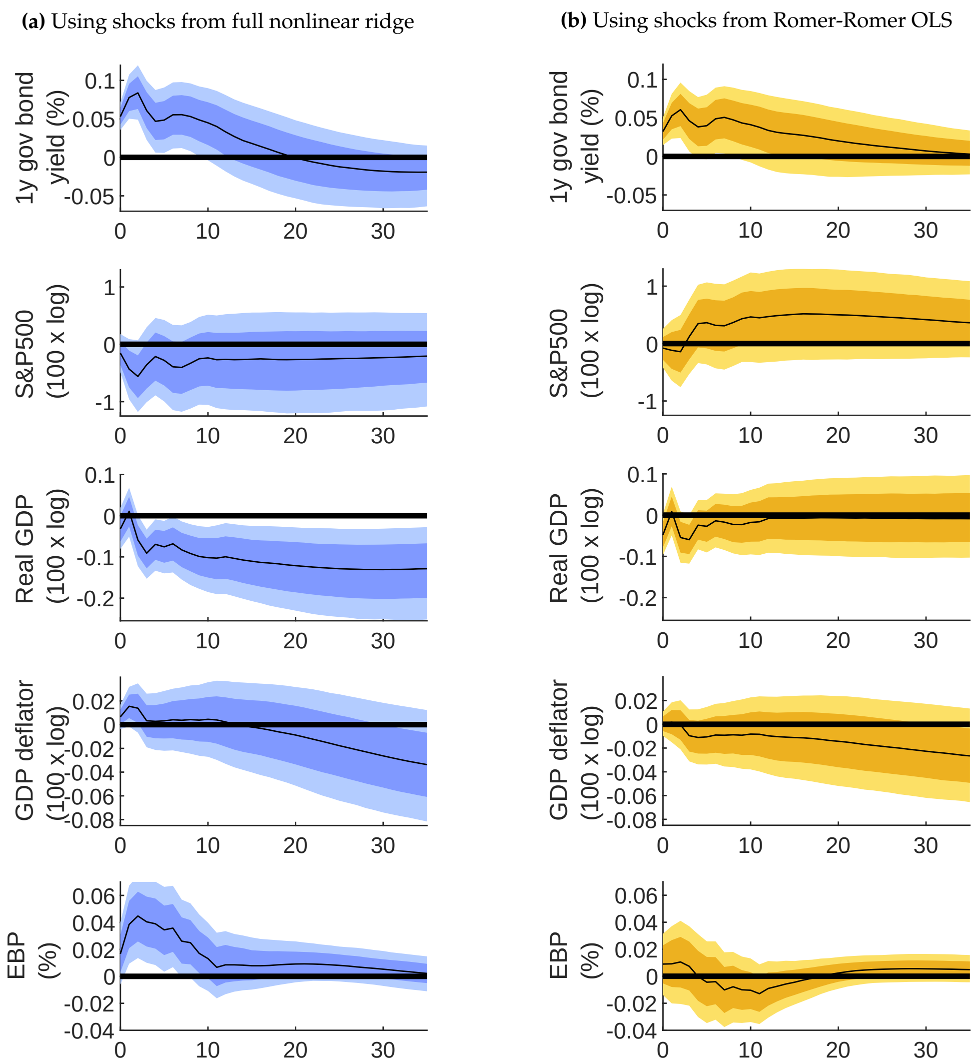
Figure 5, Panel (a) presents IRFs of macroeconomic variables to our preferred measure of monetary policy shocks.[24] We find that a monetary policy tightening is characterized by a relatively persistent increase in yields, lasting for about 20 months. The rise in interest rates leads to a reduction in real economic activity and a fall in the price level, directly in line with what standard economic theory predicts. The reduction in real output is quite immediate and very persistent. The price level response displays a very mild version of a "price puzzle" (Sims, 1992) in the first two months,

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

[23]We thank these authors for making their Gibbs sampler codes available online.

[24]These IRFs are very similar to the ones based on the shock measure estimated from imposing an R^2 of 0.9 in the ridge regression, which we present in Appendix E. Hence the estimated effect of monetary policy do not depend on which option we choose for the ridge regression, as long as a large information set is included.

Figure 5: 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.

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 result in 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 5, which presents IRFs to monetary policy shocks constructed using the 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. 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 Romer-Romer shocks, which are extracted based on a smaller information set than ours, imply an insignificant response of the EBP to a monetary policy tightening, and positive comovement between the S&P500 and interest rates conditional on a monetary policy shocks. Both responses are inconsistent with standard theory.

The differences between IRFs in Panels (a) and (b) indicate that some systematic policy variation may still be present in the shock measure only based on numerical forecasts, whereas our preferred measure based on a larger information set is more plausibly exogenous. One potential interpretation, given the difference in response of stock prices, is that the Fed systematically reacts to equity markets, for example lowers the FFR after contractions in stock prices (see e.g. 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.

This could explain the patterns observed in Panel (b). Instead, our sentiment indicators plausibly reflects the relevant information about financial market developments that the FOMC considers, and therefore the corresponding systematic response of monetary policy.

To support this interpretation, Figure 6 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 6 looks broadly similar to its counterpart in Figure 5, making clear that our preferred shock measure already satisfies the additional sign restrictions and thus delivers responses in line with theory.[25] On the contrary, the additional sign restrictions significantly alter the IRFs in Panel (b). Comparing them to their counterpart in Figure 5, it is evident that 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. This lends support to the conjecture that our inclusion of a larger information set contains relevant information about how monetary policy reacts to stock market developments.

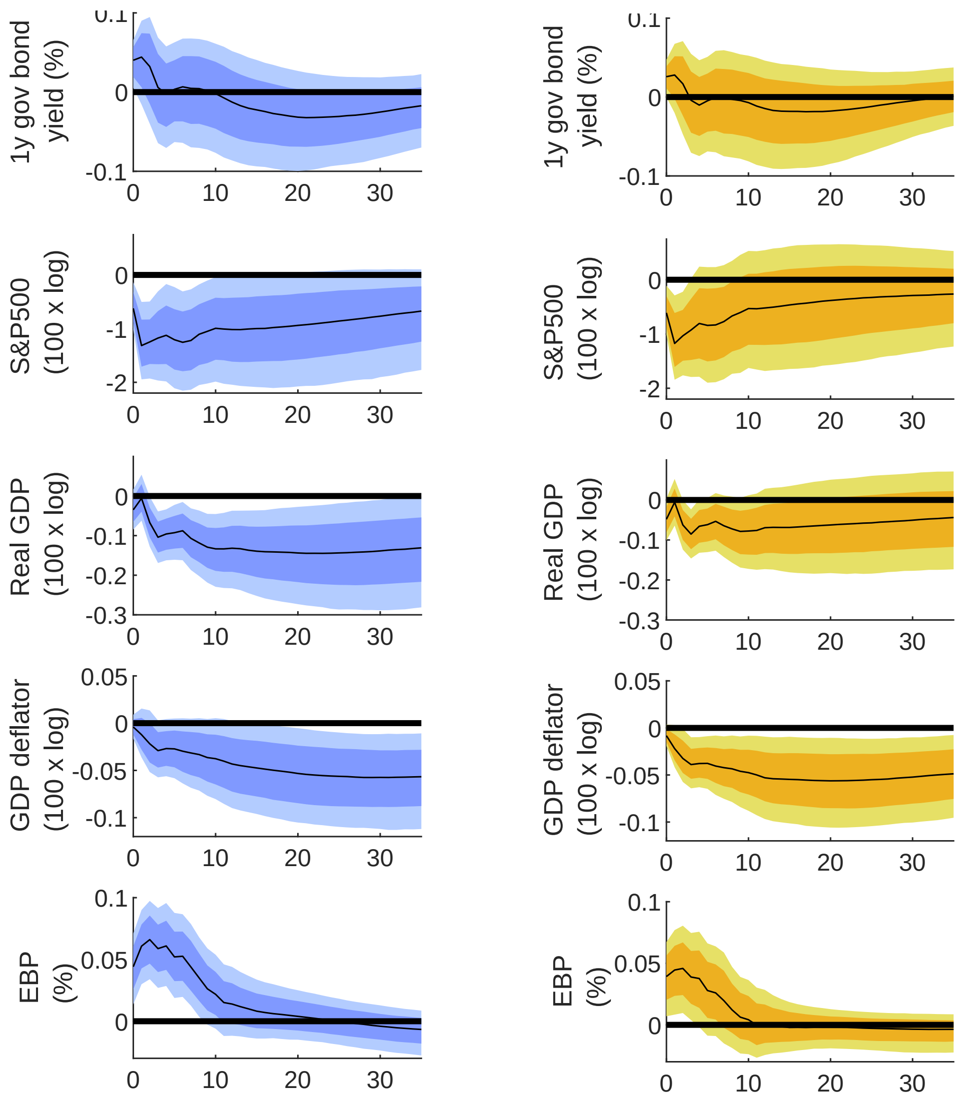
Additional results. In Appendix E we present IRFs constructed with shocks that are estimated using alternative specifications of the underlying regression (3). This includes the Romer-Romer ridge, the full linear ridge, as well as the ridge regression where the penalty parameter is based on the a priori restriction that systematic monetary policy captures 90% of FFR variation.

One noteworthy observation among these additional results is that monetary policy shocks retrieved from a ridge regression using the extended set of numerical forecasts, but without including sentiment indicators, also implies IRFs that are in line with theoretical predictions. This indicates that the additional information contained in a wider set of forecasts relative to the original Romer-Romer OLS specification already helps to correct the endogeneity problem evident in Panel (b) of Figure 5. We emphasize, however, that IRFs in line with consensus of the economic literature should be a

[25]One exception is that the mild price puzzle is eliminated from when the sign restrictions are imposed. Another difference is that the negative response of the S&P500 becomes much clearer through the sign restrictions. So while the sign restrictions are not required for our shock, they do help to sharpen the significance of the IRFs further.

Figure 6: IRFS CONSTRUCTED WITH ADDITIONAL SIGN RESTRICTIONS

(a) Using shocks from full nonlinear ridge **(b)** Using shocks from Romer-Romer OLS



Notes. IRFs to different estimated monetary policy shocks. The two panels correspond to those in Figure 5, 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. In the appendix, we also provide the IRFs to a central bank surprise shock, which is identified based on the restriction that stock prices increase in response to a rate hike. 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.

necessary, but not sufficient criterion for a good measure of monetary policy shocks. As shown in Section 3, the Romer-Romer ridge specification has an R^2 of only 0.55, as opposed to 0.75 in our preferred specification, implying an unappealingly strong contribution of shocks to variation in the FFR target when only numerical forecasts are included. Taken together, the results across Sections 3 and 4 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.

5 Conclusion

This paper develops a method to identify 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. The residual of this regression is our new measure of monetary policy shocks. An appealing feature of this measure is that only a small fraction of interest rate changes is attributed to exogenous 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.

References

ACOSTA, M. (2015): "Fomc responses to calls for transparency," *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 (2021): "Identifying

Monetary Policy Shocks Using the Central Bank's Information Set," Tech. rep., National Bureau of Economic Research.

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 (2021): "Forecasting with economic news," *Available at SSRN 3698121*.

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.

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.

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.

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 M. Lenza (2008): "Explaining the Great Moderation: It Is Not the Shocks," *Journal of the European Economic Association*, 6, 621–633.

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, 38th Annual NBER International Seminar on Macroeconomics.

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.

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.

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," .

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.

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," .

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 Machine Learning 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.

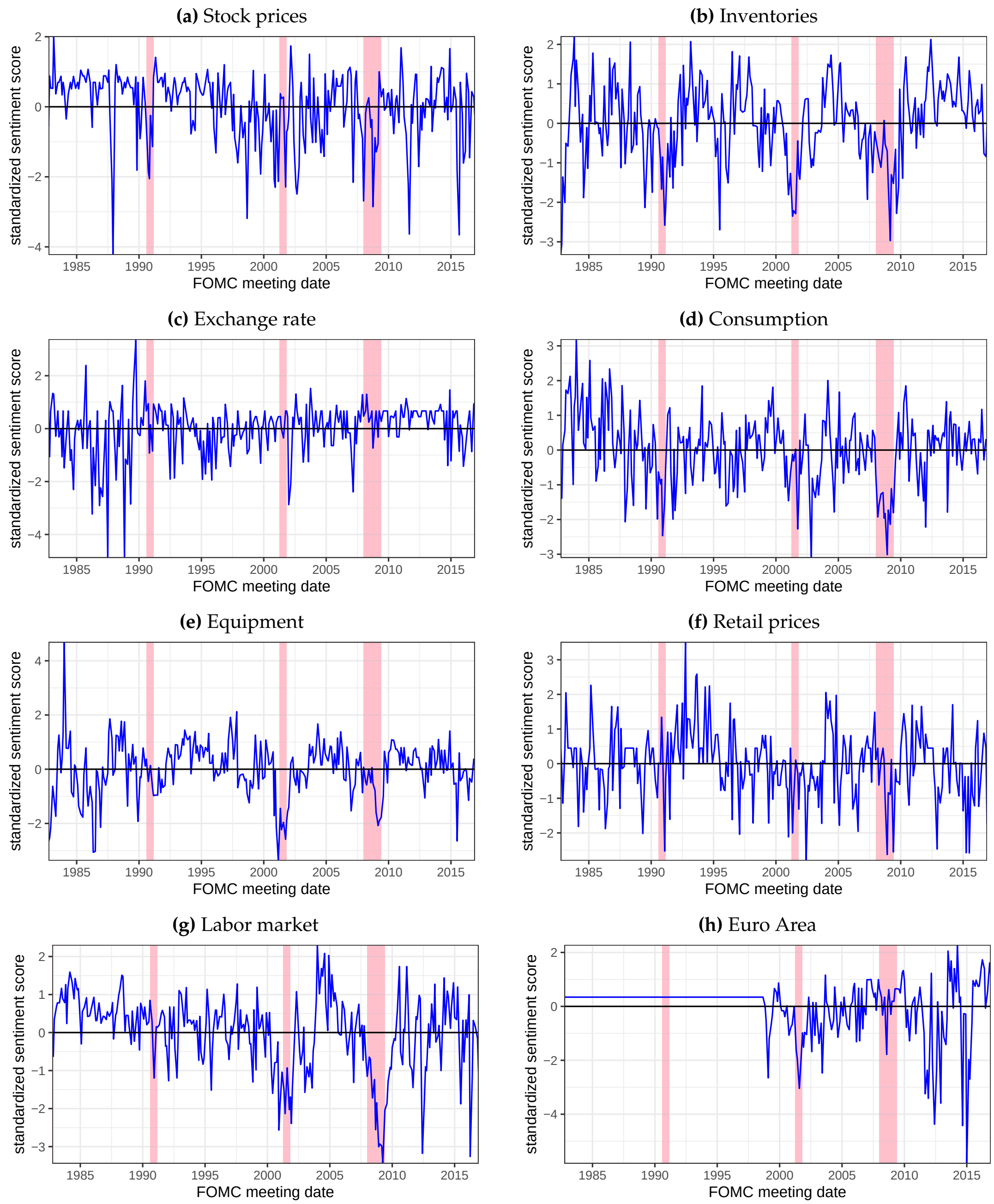
An example ***Criterion 1*** not being satisfied and ***Criterion 3*** satisfied is for the single "expenditures". Unlike credit, this single by itself is too vague based on our judgment (as "capital expenditures" and "government expenditures" are quite different). We therefore selected the doubles, 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 More sentiment indicators

Figure B.1: SELECTED SENTIMENT INDICATORS

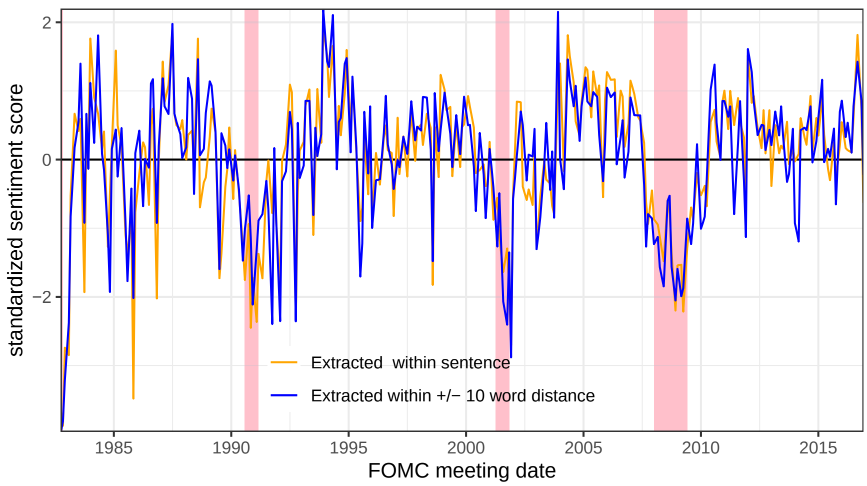


Notes. Sentiment indicators for a selection of economic concepts discussed in FOMC meeting documents, out of our full list of 296. The sentiments are constructed using the dictionary of positive and negative words in financial text of Loughran and McDonald (2011). Each indicator is standardized across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

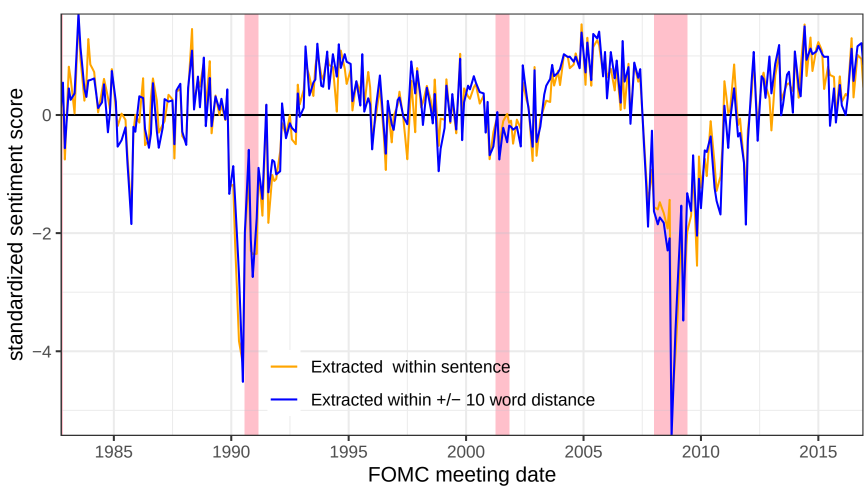
C Sentiments in +/- 10 word distance vs. in sentences

Figure C.1: SENTIMENT INDICATORS CONSTRUCTED IN ALTERNATIVE WAYS

(a) Employment



(b) Credit



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 Construction of committee composition variables

The additional data set that captures information on the composition of the FOMC in each meeting, which we use in Section 3.4 of the main text, 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.

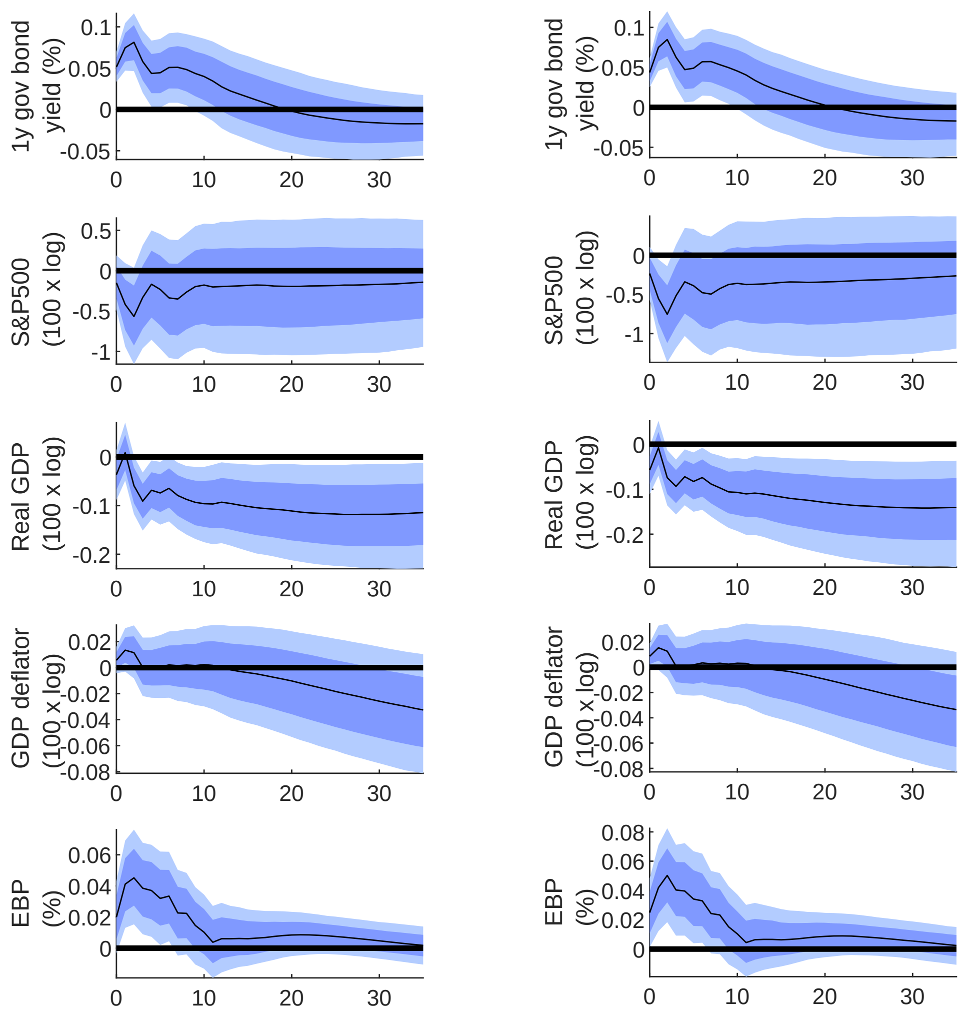
[1] In the case that a governor served multiple tenures appointed by different US presidents, we make that distinction. For example, Janet Yellen was appointed by Bill Clinton to serve as a governor in 1994 and then by Barack Obama in 2010 – and these are recorded separately.

E Additional IRFs to monetary policy shocks

Figure E.1: IRFS TO MONETARY SHOCKS ESTIMATED FROM INTERMEDIATE MODELS

(a) Using shocks from full linear ridge

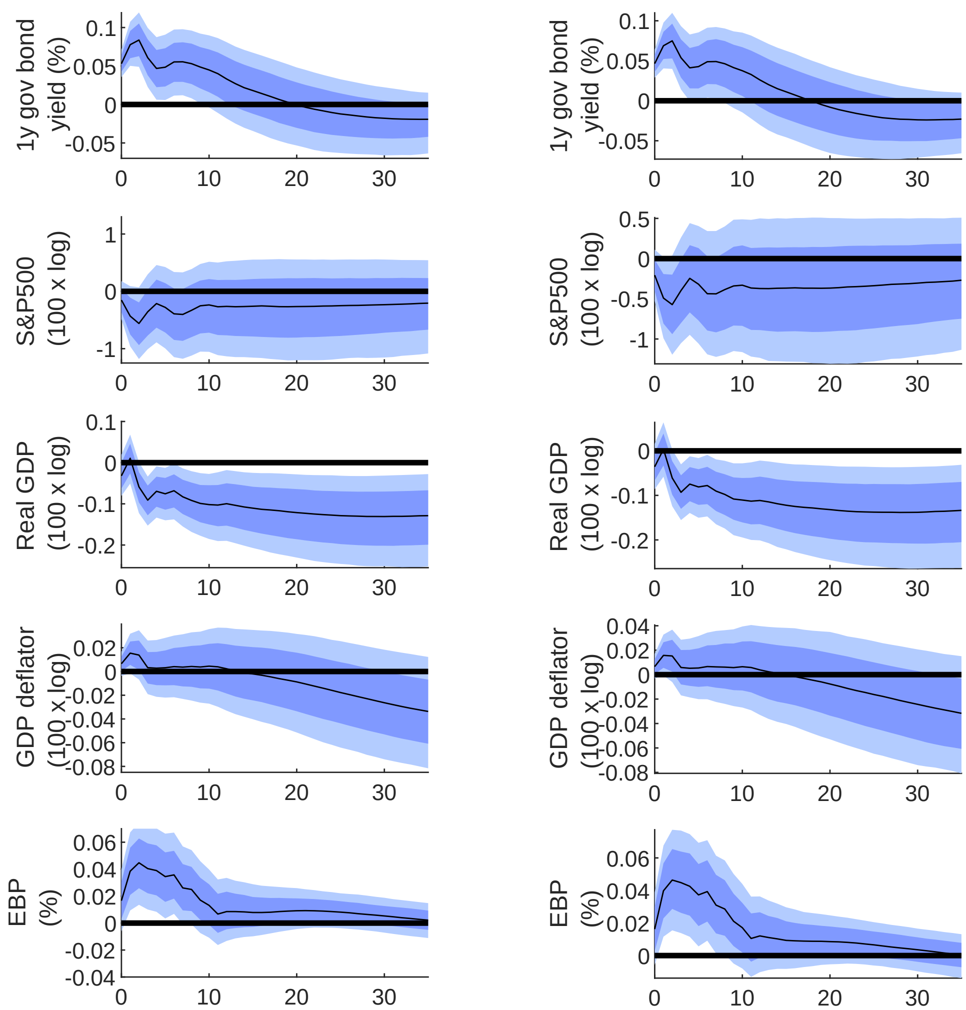
(b) Using shocks from Romer-Romer ridge



Notes. IRFs to different estimated monetary policy shocks in BVAR model. Panel (a) 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. Panel (b) shows the analogous IRFs from an empirical specification where the extended set of forecasts are used in a ridge regression. 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.

Figure E.2: IRFS TO MONETARY SHOCKS ESTIMATED WITH DIFFERENT RIDGE PENALTY PARAMETERS

(a) Full nonlinear ridge, Option 1: optimal λ **(b)** Full nonlinear ridge, Option 2: λ based on restriction



Notes. IRFs to different estimated monetary policy shocks in BVAR model. Panel (a) repeats Figure 5 Panel (a) from the main text, corresponding to 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 from the same empirical model where the ridge penalty parameter λ is set based on a restriction about a contribution of the systematic component of monetary policy of 90%. 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.