Identifying Monetary Policy Shocks:
A Machine Learning Approach*

S. Borağan Aruoba
University of Maryland

Thomas Drechsel
University of Maryland, CEPR

February 28, 2022
(preliminary draft)

Abstract

We propose a novel method to identify monetary policy shocks. By applying natural language processing techniques to documents that economists at the Federal Reserve Board 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 from this prediction. An appealing feature of our procedure is that only a small fraction of interest rate variation is attributed to exogenous shocks. We find that the dynamic responses of macroeconomic variables to our identified shock measure are consistent with the theoretical consensus. We also demonstrate that our estimated shocks are not contaminated by the "Fed information effect."

Keywords: Monetary policy; Federal Reserve; Machine learning; Natural Language Processing; Fed Information Effect.

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

1 Introduction

To study how monetary policy affects the economy, macroeconomists isolate changes in interest rates that are not a response to economic conditions, but instead occur exogenously. This paper proposes a novel method to identify such monetary policy shocks. Our starting point is Romer and Romer (2004)'s influential idea that exogenous movements in the Federal Funds Rate (FFR) are the difference between observed and intended changes in the FFR. Intended changes are based on information and forecasts about the economy available to policy makers at the time of their decisions. Romer and Romer (2004) run a linear regression of the change in the FFR on numerical 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 approach that follows the idea of exploiting the information in documents prepared for the FOMC, but aims to include *all* information contained in these documents, including numerical forecasts, and also human language. We implement this approach with natural language processing and machine learning methods.

We estimate monetary policy shocks as the residuals from a prediction of changes in the FFR 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" or "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 intended changes in the FFR

using our large set of regressors. The idea of a ridge regression is to minimize the residual sum of squares and an additional term that penalizes squared deviations of each regression coefficient from zero.[1] To select the ridge penalty parameter, we suggest two alternative options. The first is to use 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 to formulate a prior 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 (see for example the discussion in Leeper, Sims, and Zha, 1996). Our baseline prior for this second way to implement the ridge is a 90% share of FFR variation attributed to systematic changes and a 10% share explained by shocks.

We discuss five sets of findings. First, we examine the role of systematic and exogenous variation in interest rates implied by our cross-validated ridge regression, in comparison with benchmark specifications. A linear regression that is exactly specified as in Romer and Romer (2004), containing numerical forecasts for output, inflation and the unemployment rate, implies an R^2 of around 0.5, suggesting that 50% of the variation in the FFR is intended by policy makers, while 50% is exogenous. The R^2 of our ridge regression is 0.76, implying that the systematic component is 26 percentage points more important when a larger set of forecasts, Fed economists' sentiments, as well as nonlinearities are taken into account. In other words, exogenous shocks are much less important in explaining observed interest rate changes when constructed with our new method.[2]

Second, we analyze which predictors contribute to systematic changes in monetary policy in important ways. Our ridge regression contains hundreds of linear and nonlinear regressors, which we can group into different economic categories and then rank them based on their normalized ridge coefficients, that is, the FFR change that is induced on average when the sentiment of a group of related variables increases by one standard deviation. We show that there is a relatively balanced contribution of variables capturing real activity, prices, financial markets, the foreign sector and fiscal policy.

Third, we verify whether including additional information in our ridge

---

[1]There are obvious alternatives to a ridge regression, such as a LASSO, which we explore for robustness. We prefer ridge on the grounds that *dense* rather than a *sparse* prediction techniques tend to be preferable for economic data, as recently shown by Giannone, Lenza, and Primiceri (2022).

[2]Our alternative procedure to select the ridge penalty parameter allows to reduce this importance of shocks even further if desired based on a researcher's ex ante beliefs about the policy process.

regression alters our measure of shocks. We construct two additional sets of regressors. One set consists of sentiment indicators constructed from FOMC meeting transcripts rather than documents prepared by staff economists. These should reflect information that arrives between the time the staff documents are written and the committee comes together. The other set of regressors consists of variables that capture the personal composition of the FOMC, which includes a dummy for each committee member as well as several personal characteristics. These regressors should capture personal dynamics and meeting interactions not captured in the information provided by staff economists. We show that neither of these sets variables can predict the residuals from 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 staff at the beginning of a meeting.[3]

Fourth, with our novel measure of monetary policy shocks at hand, we study impulse response functions (IRFs) of macro variables to monetary policy shocks, and compare them to canonical results in the literature. We estimate a state-of-the-art Bayesian vector autoregression (BVAR) proposed by Jarocinski and Karadi (2020), in which our shock measure is included as an exogenous variable. While our shock series spans the period 1982:10-2008:10, Kalman filtering techniques allow us to study the impact of monetary policy shocks for a longer period, including the zero lower bound (ZLB) period. We find that a monetary tightening leads to a reduction in production activity and a fall in the price level, in line with what economic theory predicts. This contrasts with IRFs to the shocks constructed from the original Romer and Romer (2004) specification, where a monetary tightening appears to have no significant effect on economic activity. This issue is not present in their original paper using the 1969-1996 sample, suggesting that in more recent periods, some systematic policy variation may still be present in shock measures constructed purely based on numerical forecasts. Our findings indicate that the method we suggest overcomes this problem by including a larger information set with sentiments and nonlinearities.

Fifth, we demonstrate that our shock measure does not appear to be subject to the "Fed information effect" (Nakamura and Steinsson, 2018), by which measures of monetary surprises from high-frequency (HF) identification techniques contain information both about monetary policy shocks and the central bank's changed economic outlook. Jarocinski and Karadi (2020) argue that a monetary tightening should raise interest rates and reduce stock prices, while the confounding positive

---

[3]Our shock series also displays a lower autocorrelation than benchmark series.

central bank information shock increases both. They impose additional sign restrictions to isolate these two forces. Using the same data and specification, we show that our shock results in an interest rates increase and a fall in stock prices without imposing any additional sign restrictions. We conclude that natural language processing and machine learning are useful to deliver a cleanly identified estimate of monetary policy shocks.

Literature. Our work contributes to three branches of research. First, we fit into 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 applications.[4,5] Apart from Romer and Romer (2004), there is a wide array of approaches to identifying monetary policy shocks. A survey is provided by Ramey (2016).[6] We contribute to this literature by applying natural language processing and machine learning to identify monetary policy shocks. Our findings regarding the differences in IRFs implied by alternative empirical specifications, in particular the fact that the original Romer and Romer (2004) shocks give different IRFs in more recent samples, relate to earlier findings of Barakchian and Crowe (2013).[7]

Second, our work speaks to the discussion around the Fed information effect, see e.g. Romer and Romer (2000), Campbell et al. (2012) and Nakamura and Steinsson (2018).[8] Jarocinski and Karadi (2020) and Miranda-Agrippino and Ricco (2021), among others, aim to separate HF surprises in market interest rates between pure monetary shocks and informational shocks. Our method to estimate monetary policy shocks does not rely on a HF identification strategy. We show that, just like HF surprises, our shock series can be included in a BVAR alongside market instruments, an approach similar to using the shock series as an external instrument (Plagborg-Moller and Wolf, 2021). The estimated IRFs suggest that our identified

---

[4]In a recent paper, Bachmann, Gödl-Hanisch, and Sims (2021) suggest an interesting new spin on the idea of Romer and Romer (2004), by summarizing the Fed's information set using forecast *errors*.

[5]The method has also been applied to other countries: Cloyne and Hürtgen (2016) use it for the UK and Holm, Paul, and Tischbirek (2021) for Norway.

[6]This literature includes SVARs identified in different ways, e.g. with zero restrictions (Christiano, Eichenbaum, and Evans, 1999), sign restrictions (Uhlig, 2005), narrative sign restrictions (Antolin-Diaz and Rubio-Ramirez, 2018). Coibion (2012) compares SVAR approaches to that of Romer and Romer (2004). It also includes HF strategies to elicit surprises in interest rates around FOMC announcements, e.g. Gürkaynak, Sack, and Swanson (2005) and Gertler and Karadi (2015).

[7]Barakchian and Crowe (2013) also show that including more information is crucial for estimating IRFs more in line with theoretical predictions. They use fed funds futures contracts to do so. See Rudebusch (1998), Kuttner (2001), Thapar (2008) for related approaches.

[8]See also Bauer and Swanson (2021) for a recent perspective.

shock series is not contaminated by the Fed information effect.

The third branch of research we contribute to is a fast growing literature that applies textual analysis or machine learning to documents produced by or related to the Federal Reserve. 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 aspect-based sentiments for individual economics concepts. Shapiro and Wilson (2021) use sentiment analysis on FOMC transcripts, minutes, and speeches in order to make inference about central bank objectives, such as the inflation target.[9] 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.[10] None of the aforementioned studies identify monetary policy shocks. To the best of our knowledge, the only exception is Handlan (2020) who uses textual analysis of both FOMC statements and internal meeting materials to build a "text shock" that isolates the difference between forward guidance and current assessment of the FOMC in driving fed funds futures prices since 2005. We estimate a more conventional series of monetary policy shocks over several decades, in the tradition of Romer and Romer (2004).

Structure of the paper. Section 2 introduces our method to identify monetary policy shocks using natural language processing and machine learning. Section 3 discusses implications of our method and the resulting series of shocks, such as the contribution of systematic vs. exogenous changes in policy and the role of information. Section 4 presents our results on the responses of macroeconomic variables to monetary policy shocks. Section 5 concludes.

---

[9]Further papers analyzing Fed language include Acosta (2015) who studies in transcripts how the FOMC's responded to calls for transparency, and Hansen, McMahon, and Prat (2018) who show that communication in the FOMC changes after public transparency around was increased in the early 1990's. Cieslak et al. (2021) construct text-based measures of uncertainty from FOMC transcripts.

[10]Peek, Rosengren, and Tootell (2016) apply textual analysis to FOMC meeting transcripts to understand to what degree the FOMC reacts to financial stability concerns. Several other paper study the reverse, whether financial markets react to Fed text and language. Hansen and McMahon (2016) investigate the impact of Fed communication on asset prices as well as macroeconomic variables. 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.

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 that we propose.

2.1 Motivation, institutional setting, and main idea

Definition of monetary policy shocks. When studying how monetary policy affects the economy, macroeconomists are challenged by 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. A formalization of the endogeneity challenge in the spirit of equation (1) is the explicit or implicit starting point of most studies in the literature. For example, it is explicitly emphasized in the Handbook Chapter of Christiano, Eichenbaum, and Evans (1999).

Estimating monetary policy shocks. There are different ways to estimate ε_t with data, for example using structural vector autoregressions (SVARs). A survey of different methodologies is provided by Ramey (2016). One approach, 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 captures the Federal Funds Rate (FFR). $\boldsymbol{X}_t$ contains the forecasts of the US economy that the central bank has at its disposal at time t. In the original work of Romer and Romer (2004), these include forecasts of output growth, inflation, and the unemployment rate, and are entered both levels and changes for different forecast

horizons. Running regression (2) results in the residuals $\hat{\varepsilon}_t^{RR}$, which provide an empirical measure for ε_t in (1). Two key assumption 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.

Using information in Fed staff forecasts. Romer and Romer (2004) retrieve the forecasts contained in $\boldsymbol{X}_t$ from documents that economists of the Federal Reserve Board prepare for each FOMC meeting. In FOMC meetings, scheduled 8 times per year, the committee meets to discuss monetary policy decisions. The committee reviews a large amount of detailed information on the economic and financial conditions in US economy. This information is prepared by staff economists as part of different types of confidential documents, in particular the so-called "Greenbook" (later "Tealbook"). These documents are made available to the public with a 5-year delay and can be used by researchers. Part of the information composed by the Fed's economists consists of numerical forecasts of key macroeconomic variables. These have shown to be superior, or at least comparable to formal econometric models (Faust and Wright, 2009; Antolin-Diaz, Drechsel, and Petrella, 2021), indicating that the Fed might have an informational advantage over the private sector (Romer and Romer, 2000; Nakamura and Steinsson, 2018). Romer and Romer (2004) exploit these numerical forecasts as a proxy for the FOMC's information set.

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 capture this information using NLP to fully capture systematic component of monetary policy. The second dimension along which want to 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 amount of variables on the right hand side of a regression model,

8

we apply ML techniques to cope with the dimensionality of the problem. Using these techniques, we then estimate monetary policy shocks as the residuals from a prediction of changes in the FFR using are large amount of numerical, verbal, and nonlinear 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 inspired by Romer and Romer (2004), which includes sentiment indicators and numerical forecasts, both linearly and nonlinearly.

Step 1: Process FOMC documents

We first retrieve historical pdf 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] FOMC meeting documents are available with a 5-year lag, so the latest document currently available is for the last FOMC meeting of 2017. We process documents through to 2017, although in the regression for step 4 of our estimation procedure, we limit ourselves to the time before the zero lower bound, ending with the FOMC meeting on October 29, 2008. For each FOMC meeting, a number of 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).[12] We focus on these documents to capture the Fed's information set *at the onset of the meeting*, in the same spirit as Romer and Romer (2004). In particular, we do not include the meeting minutes and transcripts because these might capture the decision process rather than the information set. We do explore using information

---

[11]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).

[12]The *Greenbooks*, later replaced by the *Tealbooks*, contain staff analysis and outlook for US Economy. We exclude the Bluebook and the Tealbook B because these contain different hypothetical scenario analyses which we judged might obfuscate our sentiment extraction. The *Reedbooks* (until 1983) / *Beigebooks* (from 1983) discuss economic conditions by Federal Reserve district. An overview on the different documents in provided here.

from the meeting transcripts to study whether they contain additional information. Our choice results in 772 pdf files for 267 meetings (630 files for 210 meeting before the ZLB), containing 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").[13] For the 267 meetings there are roughly 18,000 singles, 450,000 doubles, and 600,000 triples (note that the Oxford English dictionary has roughly 170,000 single words). We the calculate the frequency at which each single, double and triple occurs for each meeting date and each document.

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*.[14] 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.

---

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

[14]Both of us went through this selection independently and then discussed any disagreement case by case. When moving down along the frequency ranking we stop at a very 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.

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 2017. The size of concept reflects the frequency with which it occurs across the documents.

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 each 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.[15] This classification is based on the dictionary of positive and negative terms in Loughran and McDonald (2011), which is a dictionary especially constructed for financial text.[16] Each positive word then gives a score of +1 and each negative word of -1. 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 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.

---

[15]The 10 word distance here refers to words after the pre-cleaning steps of the documents, and not words in a raw sentence. We explore robustness with a an alternative distance of 5 words.

[16]We remove some terms from this dictionary, such *unemployment* and *unemployed*, because these are among our selected economic concepts. We also slightly enhanced the dictionary with some terms specific to Fed language, such as "tightening."

Figure 2 presents the sentiment indicators for some selected economic concepts. These indicators 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 mortgages played a role primarily in the Great Recession and its aftermath (see Panels (c) and (d)). 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.[17] 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, taken from the dictionary of Loughran and McDonald (2011). The total number of classified words is 2,885.

---

[17]We are planning to make these sentiment indicators available for other researchers.

Figure 2: SELECTED SENTIMENT INDICATORS
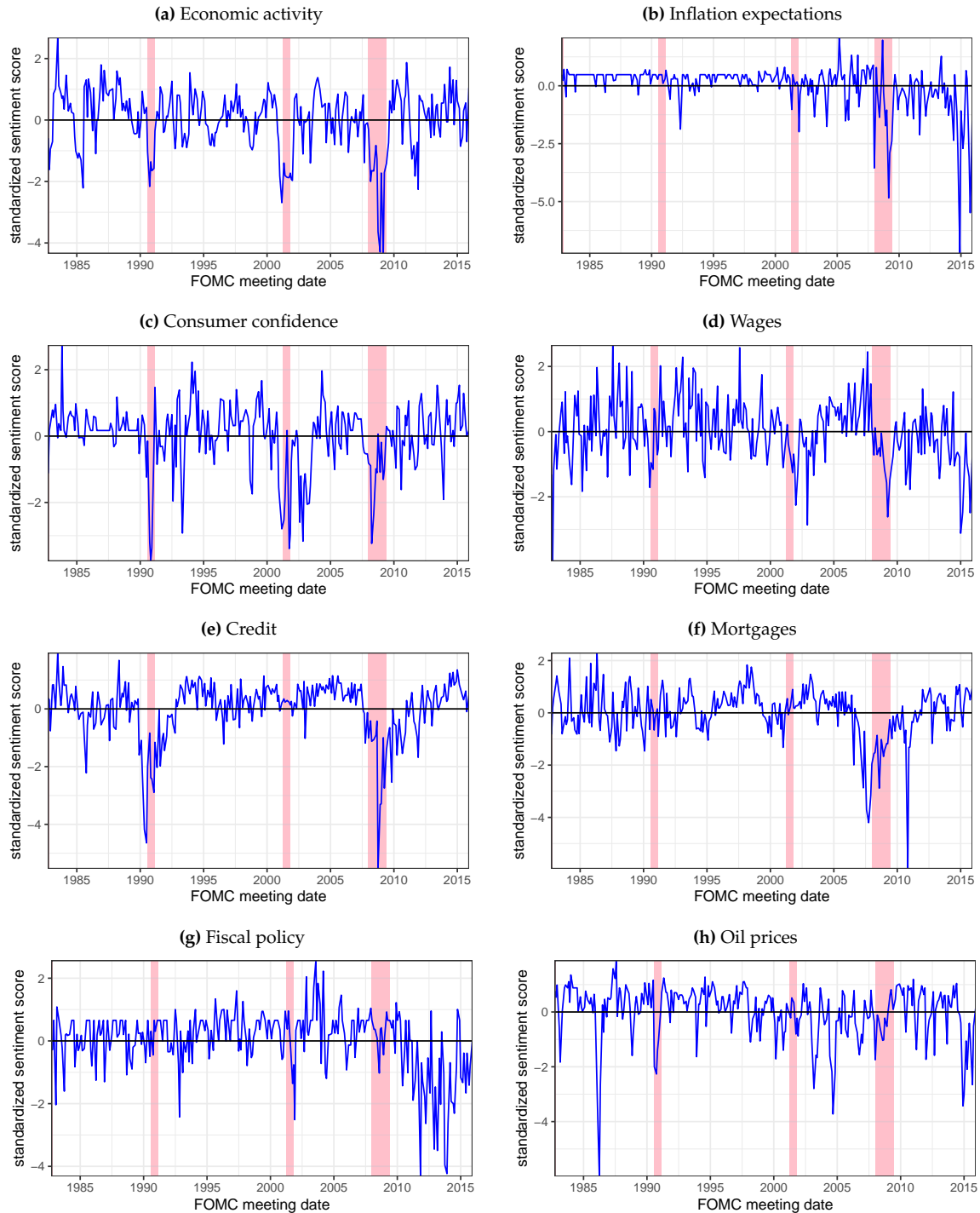
Notes. Sentiment indicators for a selection of economic concepts discussed in FOMC meeting documents, out of our full list of 296. The sentiments are constructed using the dictionary of positive and negative words in financial text of Loughran and McDonald (2011). Each indicator is standardized across 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 nonlinearly. 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 FFR. $\widetilde{\boldsymbol{X}}_t$ contains numerical forecasts. Following Romer and Romer (2004)'s specification of $\boldsymbol{X}_t$ in (2), we enter forecasts in levels and first differences, across several forecast horizons. Relative to their original paper, we use an augmented set of forecasts that Fed economists produce since the starting point of our sample in 1982, which includes additional production, investment, housing and government spending variables.[18] 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 linear-quadratic function. 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 FFR changes. Therefore, an ordinary least squares (OLS) regression with our 858 regressors is infeasible. To overcome this issue, we resort to machine learning techniques. Specifically, we employ a ridge 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 minimizes $\sum_i \varepsilon_i^2 + \lambda \sum_j^k \gamma_j^2$.[19] Unlike its close sibling, the LASSO regression, which we discuss further below, a ridge results in estimated coefficients for all 858 regressors. Importantly, there are different ways to choose λ. We propose two alternatives:

- **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

---

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

[19]The Bayesian interpretation of ridge 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 λ.

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: Set a prior about the contribution of systematic policy.** An alternative way to proceed is to formulate a prior 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 baseline prior for this second way to implement the ridge is a 90% share of FFR variation attributed to systematic changes and a 10% share explained by shocks.

In short, option 1 selects a value for λ based on maximizing the out-of-sample performance of the model, while option 1 selects a value for 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, we note that 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 amount of economic concepts, an approach sometimes referred to as Aspect-Based 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.[20] This alternative would involve tens or hundreds of thousands of regressors. Instead, we select economic concepts using judgment, reducing the dimensionality of the problem to 868 regressors. We prefer this procedure because 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 machine learning techniques that could be applied here. A natural alternative would be the LASSO regression, which instead minimizes $\sum_i \varepsilon_i^2 + \lambda \sum_j^k |\gamma_j|$. A key difference is that

---

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

LASSO results in a *sparse* model that contains only a subset of the right-hand-side variables, while ridge results in a *dense* model, containing all regressors and associated coefficients. In this sense, ridge is more related to dynamic factor models and principal component analysis, which is often employed for macroeconomic data. While we do not have a strong view, we prefer ridge 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). Given its close relationship to ridge, we exploit LASSO regressions for robustness.

3 Implications of our method

This section discusses implication of our methodology. It reviews estimation results for different versions of the empirical model represented by equation (3). The results include measures of fit, coefficient estimates, 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 different empirical specifications. First, as the simplest benchmark it includes equation (2), a restricted version of (3) where only the original forecasts used by Romer and Romer (2004) enter linearly in an OLS estimation (results are labeled 'Romer-Romer OLS'). Second, a model that includes the same smaller set of forecasts, but is instead estimated as a ridge regression ('Romer-Romer Ridge'). Third, the figure contains 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 aforementioned ridge models the ridge penalty parameter λ is estimated based on an optimal average MSE. The fifth and last model in Figure 3 corresponds again to the full nonlinear ridge, but features an R^2 of 0.9 by construction, as we solve for the ridge penalty that achieves a contribution of the systematic component of policy on 90%.

We compare the goodness of fit between these alternative models to understand in discrete steps 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, estimating an equation as specified in the original study of Romer and Romer (2004) implies an

R^2 of 0.5. In other words, this empirical model implies that 50% of the variation in the FFR 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.*"

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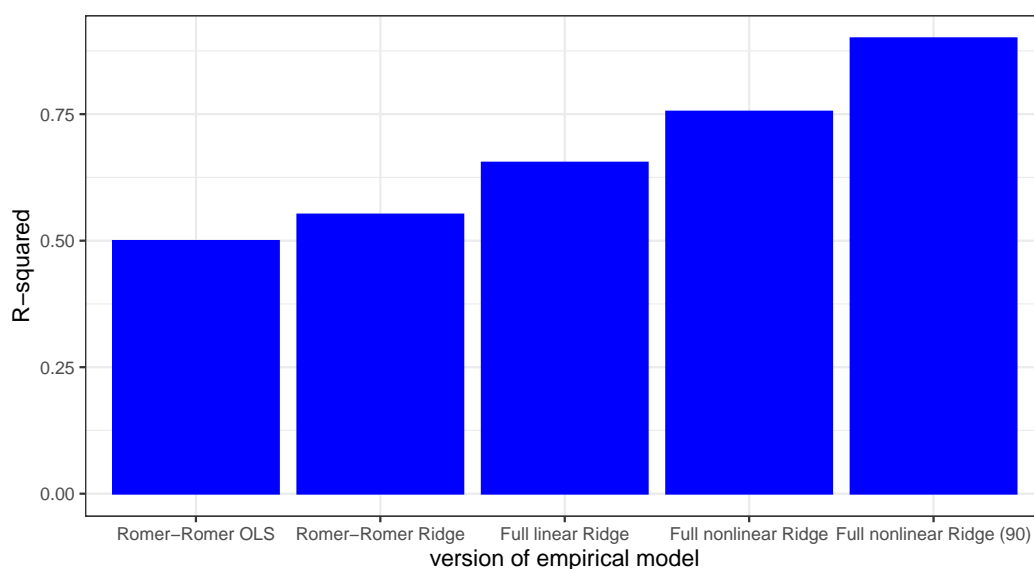
Notes. R^2 implied by the estimation of different versions of the empirical model specified in equation (3), over the sample period 1982:10 to 2008:10. *Romer-Romer OLS*: set of variables used by Romer and Romer (2004), estimated with OLS; *Romer-Romer Ridge*: same set of variables, estimated with ridge;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

The remaining bars of the figure reveal that expanding the information set in the empirical model increases the implied fit. Each of the ways in which the right hands side of the model is enriched – going from OLS to ridge, including more numerical forecasts and sentiment indicators, and allowing for nonlinearities – delivers some additional improvement in the fit of the model. Note that 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 0.76, suggesting that 76% of FFR variation is systematic, and 24% . 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 purely an illustration of the fact that our proposed procedure can incorporate a researcher's prior about the relative contribution of the systematic and exogenous component in FFR variation.

Table 2: R^2 ACROSS 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	0.56	0.61
Full nonlinear ML	0.76	0.77	0.80	0.69
Full nonlinear ML (90)	0.90	0.90	0.90	0.90

Notes. R^2 measures implied 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 in Step 4 of our procedure.

We provide some robustness for these results in Table 2. The first column simply repeats the results from Figure 3 for reference. The remaining columns vary our procedure along two dimensions. First, in column (2) we show the corresponding measures of R^2 of empirical models in which our sentiments indicators are constructed based using a 5-word window instead of a 10-word window around economic concepts (see discussion under Step 3 above). Second, in columns (3) and (4) we apply LASSO regressions rather than ridge regressions. This illustrates that other machine learning techniques can be used in Step 4 of our methodology. 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 a prior. 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 predictors

We now focus on our preferred full nonlinear ridge specification. This empirical model explains 76% of the variation in the FFR across FOMC meetings. To some degree, the coefficient estimates corresponding to different variables that are included on the right hand side of this model can tell us something about which forecasts and sentiment indicators are particularly important for predicting the FFR. In the ridge regression, the coefficients are normalized, and for each variable a linear and squared term are included. This means that it is possible to calculate,

the average percentage point (ppt) change in the FFR, when a given variable increases, say by one standard deviation from its mean.[21] A variable-by-variable comparison is difficult given the long list of 858 regressors. We therefore group the forecasts and sentiment indicators into different economic categories, and then rank the contribution of these groups of variables as a whole to variation in the FFR. Specifically, we assign variables to one of the following categories: real activity, prices, financial markets, the foreign sector, fiscal policy. We then compute the change in the FFR when every variable in a group were to increase by one standard deviation from its mean. *[To be completed.]*

Figure 4: ESTIMATED MONETARY POLICY SHOCKS
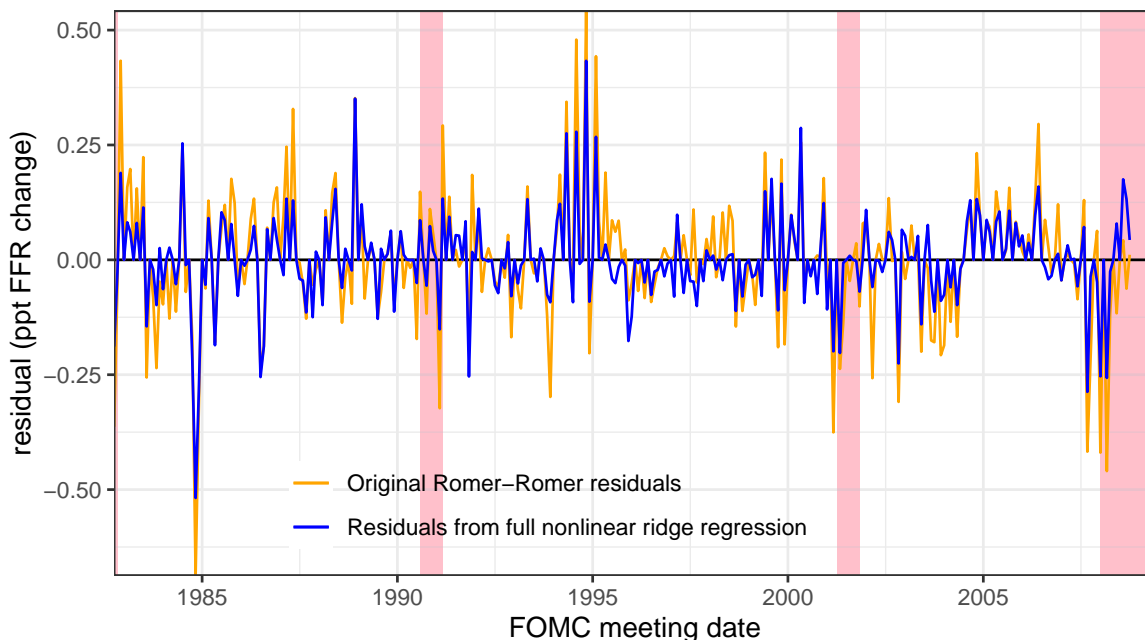
Notes. Time series of estimated monetary policy shocks. Dark blue: our preferred version, the residuals from predicting changes in the FFR based on numerical forecasts, sentiment indicators and nonlinearities in FOMC documents. Orange: benchmark version based on a specification that follows Romer and Romer (2004). Shaded areas represent NBER recessions.

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

---

[21]In Appendix C we provide more detailed calculations on how to compute such contributions with standardized coefficients of linear and squared terms.

figure compares this with the simplest benchmark, the estimated residuals from the Romer-Romer OLS model as the lighter orange line. The unit corresponds to 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 24% 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 original Romer and Romer (2004) method. 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.

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 decision 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 baseline our prepared by Federal Reserve Board economists priors 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 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 a given governors, information on the unanimity of votes, as well as the number of female attendants. In total this results in 298 variables that capture the composition of the FOMC. More details on the construction of this additional data set are provided in Appendix D.

Table 3: ADDING FURTHER INFORMATION

Specification	R^2
Full nonlinear Ridge	0.7550
Adding transcript sentiments and committee composition	0.7572
Difference	0.0022

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 3 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 nonlinearly, 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.[22] 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.

---

[22]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 using the additional information. The fit of this "second stage" regression was near 0.

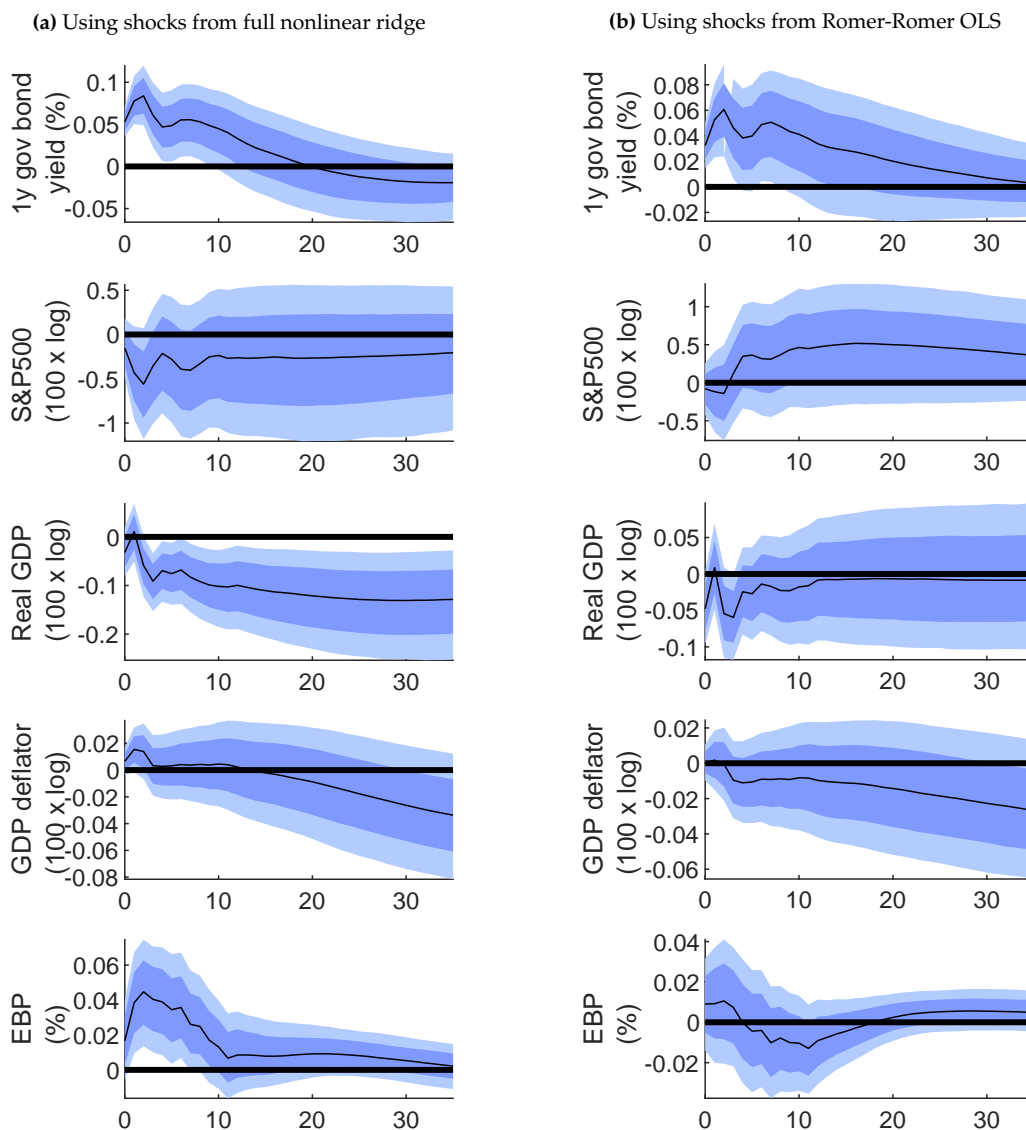# 4 The effects of monetary policy shocks

We use our estimated shock series to study the effects of monetary policy shocks on the US economy in a state-of-the-art BVAR model estimated at monthly frequency, following Jarocinski and Karadi (2020). The system 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 identified monetary policy shocks is ordered first in a Choleski identification scheme. This yields asymptotically identical results to using the shock series as an external instrument (Plagborg-Moller and Wolf, 2021).[23] GDP and its deflator are included to capture the effect of monetary policy on activity and prices. The Kalman filter is applied to interpolate these quarterly series to monthly frequency. The use of the Kalman filter also allows us to estimate the system through to 2016, a period that includes the ZLB, while our shock series spans the period 1982:10-2008:10. The inclusion of the S&P500 allows us to use the logic of Jarocinski and Karadi (2020) to investigate the presence of the Fed information effect in our identified monetary policy shock series. Finally, the EBP is included as a forward-looking financial variable, as is widely done in the literature. We use the same settings and priors as in Jarocinski and Karadi (2020).[24]

The impact on activity, inflation and bond spreads. Figure 5 compares the IRFs of macroeconomic variables to two versions of estimated monetary policy shocks. Panel (a) is constructed using the measure of monetary policy shocks that we propose in this paper. These shocks are estimated using the full nonlinear ridge model on the extended set of numerical forecasts and our constructed sentiment indicators. Based on these identified shocks, we find that a monetary tightening is characterized by a relatively persistent increase in yields, lasting for about 20 months. The increase in rates leads to a reduction in real production activity and a fall in the price level, directly in line with what economic theory predicts. The reduction in output is quite immediate and very persistent. The price level response displays somewhat of a "price puzzle" (Sims, 1992) in the first two months, but is persistently negative thereafter. It takes about 30 months for the response to be significantly negative. In line with previous findings, the EBP increases sharply and significantly after a monetary tightening.

---

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

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

Figure 5: IMPULSE RESPONSE FUNCTIONS TO DIFFERENT MONETARY POLICY SHOCK MEASURES

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

Notes. IRFs to different estimated monetary policy shocks in the BVAR of Jarocinski and Karadi (2020). 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 an OLS regression, as in Romer and Romer (2004). 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.

These results contrast with those in Panel (b), which presents IRFs to monetary policy shocks constructed using the original Romer and Romer (2004) specification, in which a 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 tightening appears to have little effect on output. After an initial reduction, the effect is essentially flat. Another important difference to the responses based on our shock measure is the insignificant response of the EBP. These conclusions are different from the IRFs in the original Romer and Romer (2004) paper, using the 1969-1996 sample, where output is significantly reduced after a monetary tightening. This suggests that in more recent periods some systematic policy variation may still be present in a shock measure only based on numerical forecasts. The results suggest that our method overcomes this problem by including a larger information set with sentiment indicators and nonlinearities.

The impact on stock prices and the Fed information effect.　The Fed information effect has often been argued to affect monetary surprises constructed from HF identification techniques. The worry is that movements in market interest in narrow windows around FOMC announcements contain information both about monetary policy shocks as well as about the central bank's changed economic outlook. Jarocinski and Karadi (2020) argue that a monetary tightening in the traditional textbook sense should raise interest rates and reduce stock prices, while the confounding positive central bank information shock increases both. The latter happens because the market may interpret an interest rate hike as a positive assessment about the economy by the Fed, and therefore as good news about the stock market. Figure 5, Panel (a) makes clear that using our identified shocks results in an increase in interest rates and a fall in stock prices. It is worth noting that this is the case without imposing additional sign restrictions, as suggested by Jarocinski and Karadi (2020). In Panel (b), however, we see that if only numerical forecasts are used to identify the shock, a positive comovement of stock prices and interest rates arises. This is indicative of the fact that those shocks are not fully exogenous. While they are not estimated based on market interest rate movements around FOMC announcements, some anticipation effects may be present in the error term of equation (2) if the Fed's information set is not appropriately controlled for on the right hand side. We conclude that our method, which applies natural language processing and machine learning techniques to extract information from FOMC

documents, delivers a cleanly identified estimate of monetary policy shocks, not subject to the Fed information effect.

Additional results. In Appendix E we provide additional results. First, we show the IRFs for the same variables as in Figure 5, but using the other versions of the shocks, estimated using the intermediate empirical specifications (Romer-Romer ridge, and the full linear Ridge), as well as the alternative option where a prior about the R^2 is imposed. Second we show the results of other outcomes variables (to be completed.) Third, we show the impact of additionally imposing the sign restrictions of Jarocinski and Karadi (2020). *[To be completed.]*

5 Conclusion

This paper develops a novel 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 Fed Funds Rate. The residual of this regression is our new measure of monetary policy shocks. We find that activity and prices fall after a monetary tightening, in line with theoretical predictions. The negative response of stock prices suggest that we have identified monetary policy shock and not an informational shocks.

References

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

ANTOLIN-DIAZ, J., T. DRECHSEL, AND I. PETRELLA (2021): "Advances in nowcasting economic activity: Secular trends, large shocks and new data," *CEPR Discussion Paper No. DP15926*.

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*.

BAUER, M. D. AND E. T. SWANSON (2021): "The Fed's response to economic news explains the" Fed information effect"," Tech. rep., Working paper.

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.

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*.

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," .

KUTTNER, K. N. (2001): "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market," *Journal of Monetary Economics*, 47, 523–544.

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.

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.

RUDEBUSCH, G. D. (1998): "Do measures of monetary policy in a VAR make sense?" *International economic review*, 907–931.

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.

THAPAR, A. (2008): "Using private forecasts to estimate the effects of monetary policy," *Journal of Monetary Economics*, 55, 806–824.

THORNTON, D. L. (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.

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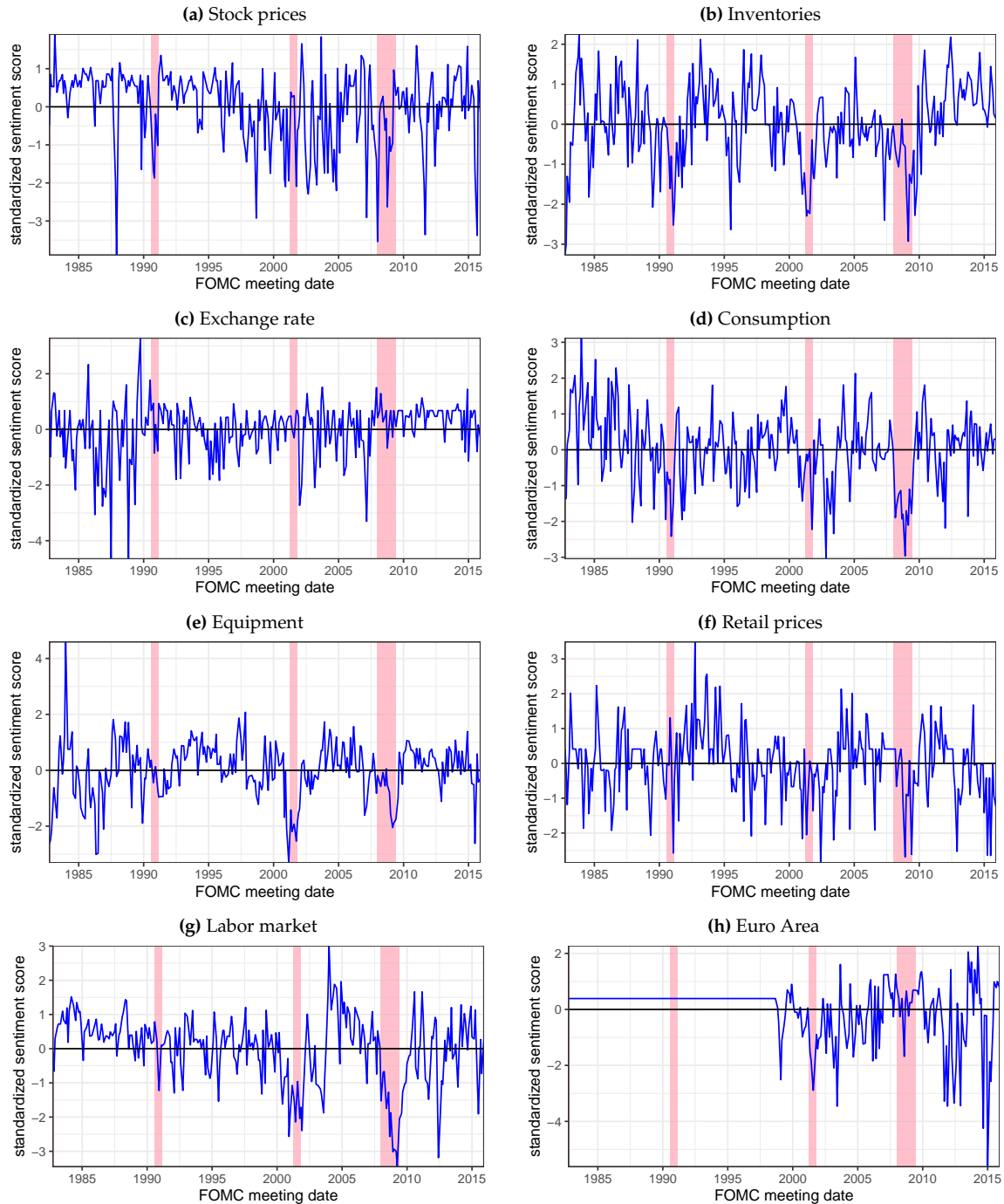
Notes. Sentiment indicators for a selection of economic concepts discussed in FOMC meeting documents, out of our full list of 296. The sentiments are constructed using the dictionary of positive and negative words in financial text of Loughran and McDonald (2011). Each indicator is standardized across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

3

C Calculation of coefficients with squared terms

Denote x the original variable, which is not standardized. Suppose x and x^2 are included in the ridge regression. The ridge command in *R* standardizes all variables, so what effectively enters the regression are two variables $z_1 = (x - \mu_x)/\sigma_x$ and $z_2 = (x^2 - \mu_{(x^2)})/\sigma_{(x^2)}$. The regression spits out coefficients α_1 and α_2.

Now suppose we want to know by how much the left hand side changes if the original x goes up by m standard deviations. This should be calculated as follows. We first derive dy as a function of dx:

$$dy = \frac{\partial}{\partial x} \left\{ \alpha_1 \left(\frac{x - \mu_x}{\sigma_x} \right) + \alpha_2 \left(\frac{x^2 - \mu_{(x^2)}}{\sigma_{(x^2)}} \right) \right\} dx \tag{4}$$

$$= \left\{ \frac{\alpha_1}{\sigma_x} + \frac{2\alpha_2 x}{\sigma_{(x^2)}} \right\} dx \tag{5}$$

Now we can plug in $dx = m\sigma_x$. For x we can plug in a 'location' of choice, for example the mean μ_x or $\mu_x + \sigma_x$ or $\mu_x - \sigma_x$.

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 a total of 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
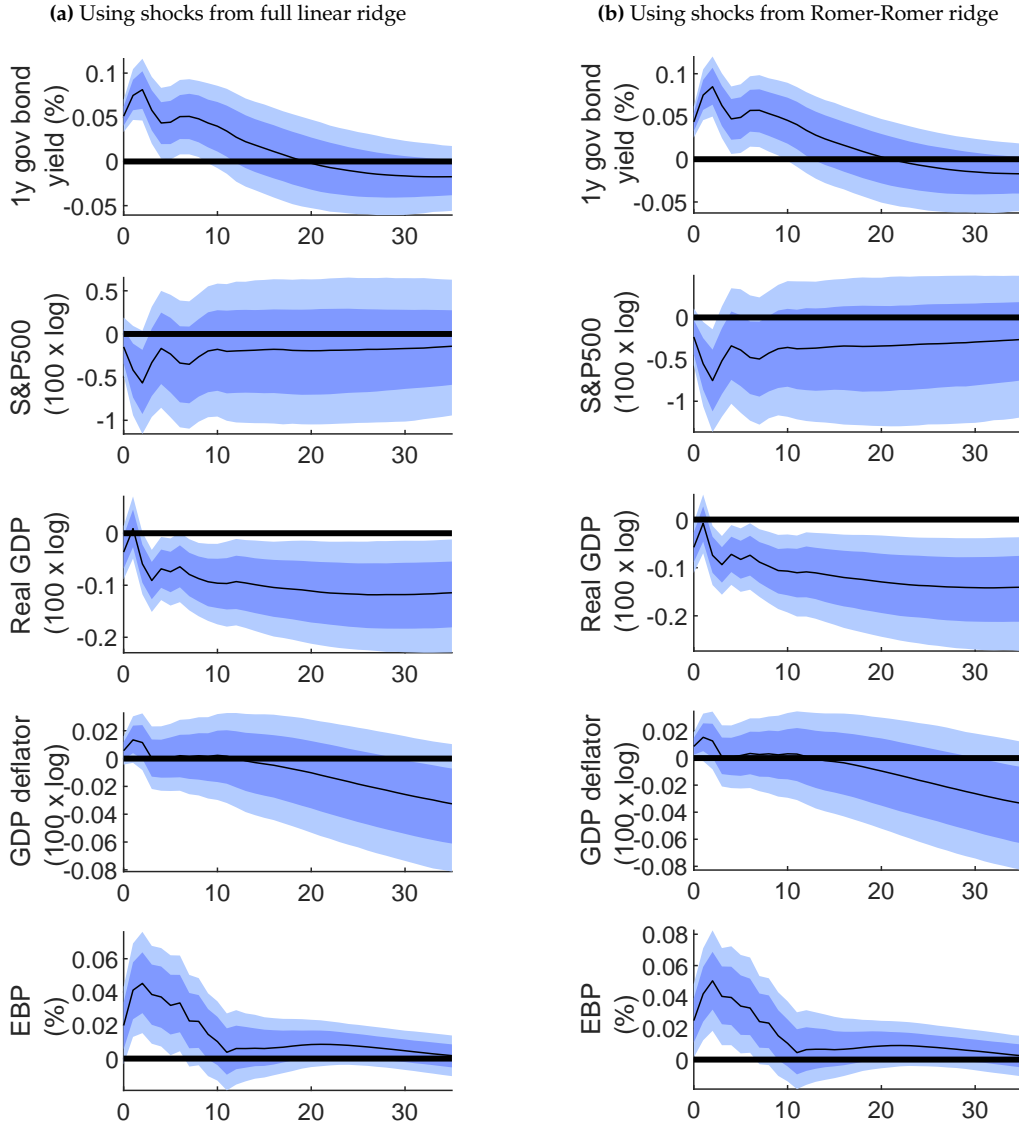
(a) Using shocks from full linear ridge   **(b)** Using shocks from Romer-Romer ridge

Notes. IRFs to different estimated monetary policy shocks in the BVAR of Jarocinski and Karadi (2020). 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
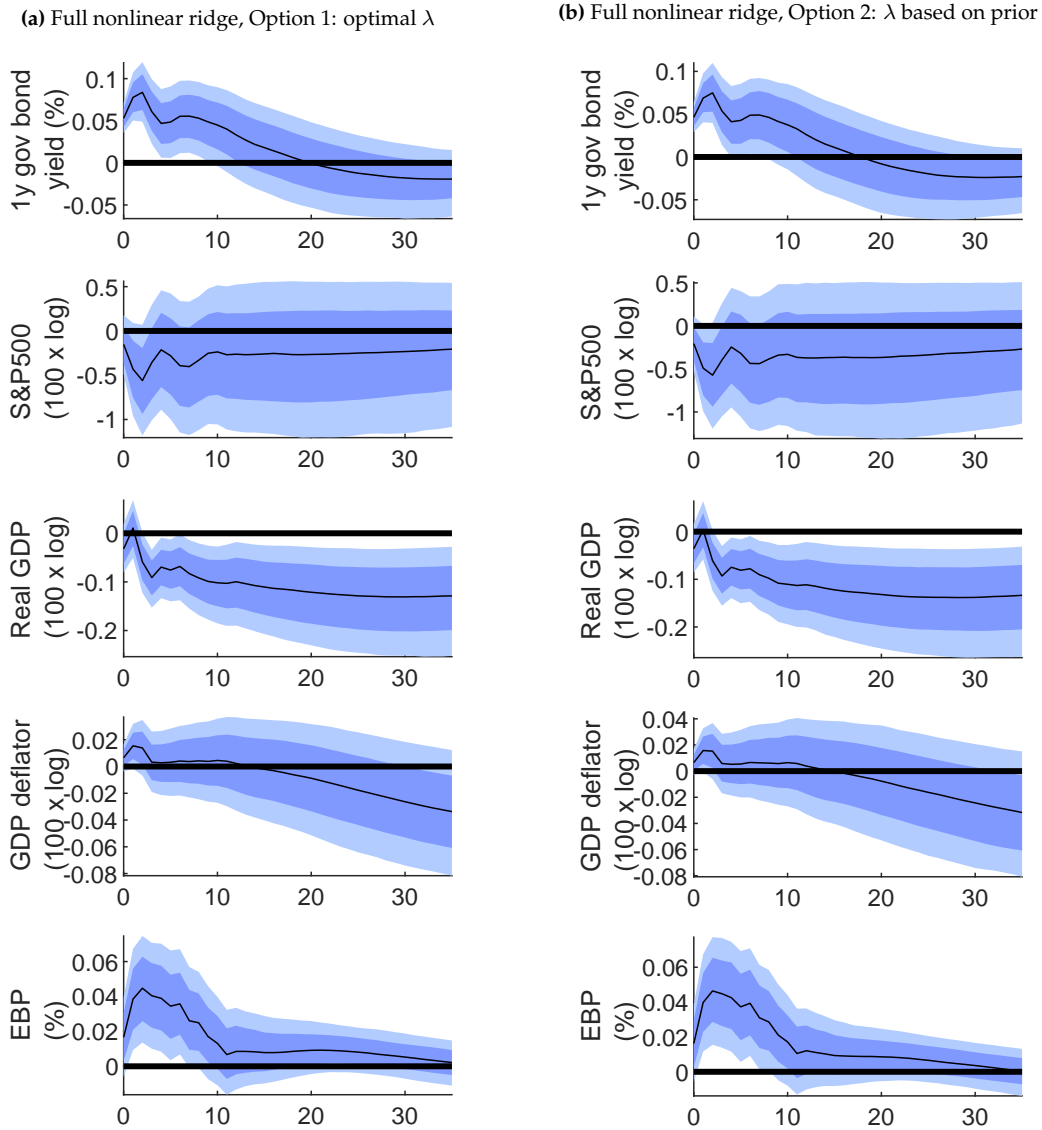
(a) Full nonlinear ridge, Option 1: optimal λ

(b) Full nonlinear ridge, Option 2: λ based on prior

Notes. IRFs to different estimated monetary policy shocks in the BVAR of Jarocinski and Karadi (2020). 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 prior 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.