ONLINE APPENDIX TO

Identifying Monetary Policy Shocks: A Natural Language Approach

by S. Borağan Aruoba and Thomas Drechsel

A Algorithm to combine and exclude concepts

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

1. Start with triples. Go through the list of triples that have at least 250 mentions (around one per meeting on average). Select triples that are economic concepts (based on judgment).

2.a) Go through the list of doubles that have at least 500 mentions. Select doubles that are economic concepts (based on judgment).

2.b) **IF** a selected double is a subset of one or several triples:

- Unselect the double and keep the triple(s) **IF**
  [***Criterion 1***] the triples close to add up to the double **AND**
  [***Criterion 2***] the triples are sufficiently different concepts
  OR
  [***Criterion 3***] the double by itself is too ambiguous
- **ELSE**: keep the double and unselect the triple(s)

3.a) Go through the list of singles that have at least 2000 mentions. Select singles that are economic concepts (based on judgment).

3.b) **IF** a selected single is a subset of one or several doubles:

- Unselect the single and keep the double(s) **IF**
  [***Criterion 1***] the doubles close to add up to the single **AND**
  [***Criterion 2***] the doubles are sufficiently different concepts
  OR
  [***Criterion 3***] the single by itself is too ambiguous
- **ELSE** Keep the single and unselect the double(s)

END

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

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

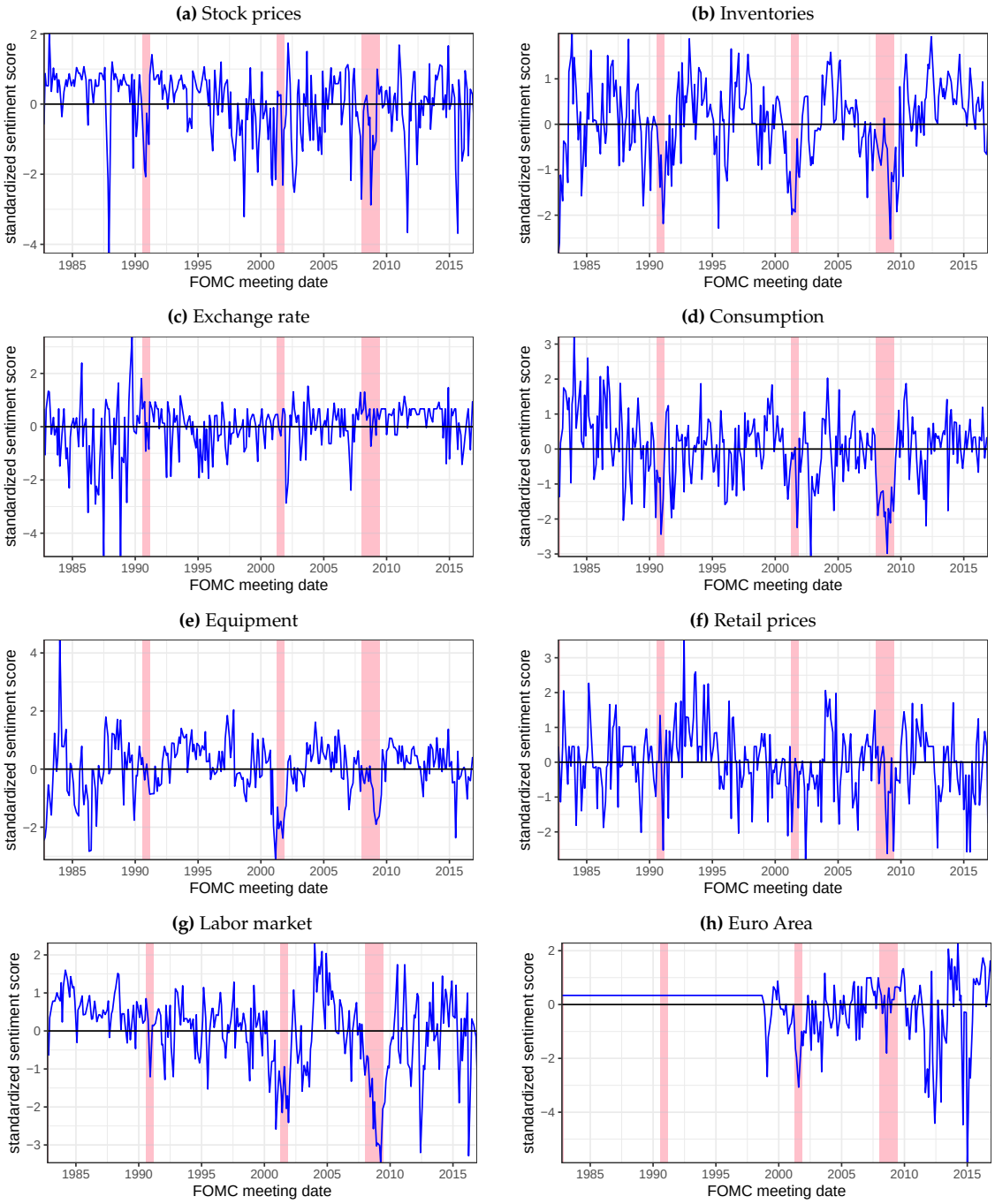
An example ***Criterion 1*** not being satisfied and ***Criterion 3*** satisfied is for the single "expenditures". Unlike credit, this single by itself is too vague based on our judgment (as "capital expenditures" and "government expenditures" are quite different). We therefore selected the doubles,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

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

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

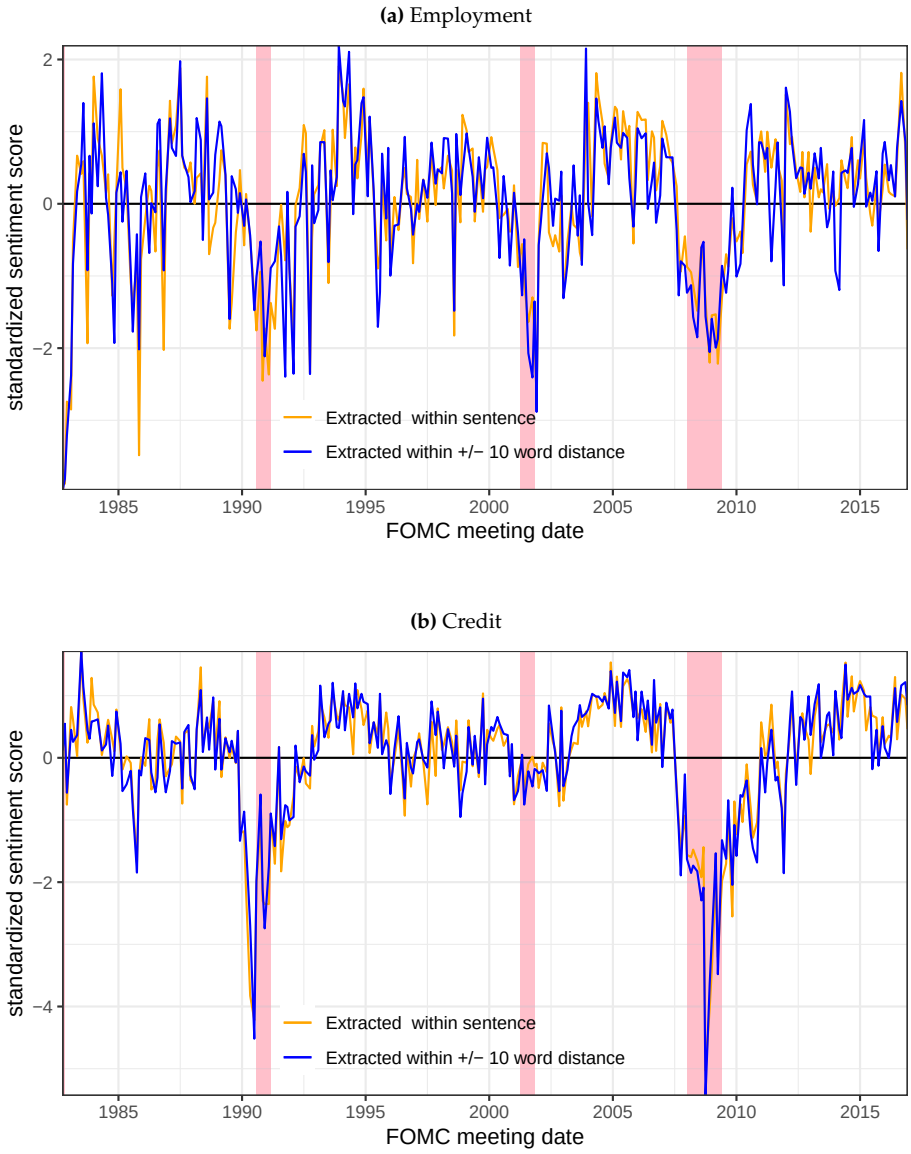
B Additional sentiment indicators

Figure B.1: SELECTED SENTIMENT INDICATORS



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

Figure C.1: SENTIMENT INDICATORS CONSTRUCTED IN ALTERNATIVE WAYS



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

D Alternative machine learning approaches

Tables D.1 and D.2 repeat the analysis of goodness-of-fit across different model specifications from the main text using LASSO and elastic net regressions instead of ridge. In the case of the elastic net, the hyperparameter that governs how the ridge and LASSO penalties are weighted (α) is chosen optimally and the optimal α^* is also reported.[1] Note that different from ridge, in LASSO and elastic net regressions the number of regressors can differ between the 10-word and 5-word version, so the 'Number of regressors' column includes two numbers.

Table D.1: GOODNESS-OF-FIT ACROSS DIFFERENT SPECIFICATIONS – LASSO VERSION

	Selected number of regressors	(1) Variance explained with 10-word sentiment (main specification)	(2) Variance explained with 5-word sentiment (robustness)
Romer-Romer original OLS with subset of forecasts	18	0.50	
LASSO with extended set of forecasts	29	0.57	
LASSO with extended set of forecasts (nonlinear)	22	0.49	
LASSO with all forecasts & sentiments (linear)	24 / 41	0.55	0.62
LASSO with all forecasts & sentiments (nonlinear)	80 / 63	0.81	0.72
LASSO with all forecasts & sentiments (linear with lags)	41 / 39	0.64	0.63
LASSO with all forecasts & sentiments (nonlinear with lags)	37 / 49	0.59	0.66

Notes. Implied goodness-of-fit, measured by R^2 (OLS) and deviance ratio (ridge), from estimating different empirical specifications of equation (3). For the first two specifications, sentiments are not included so the 10-word/5-word distinction does not apply. Different from ridge, in the LASSO regressions the number of regressors can differ between the 10-word and 5-word version, so the 'Number of regressors' column includes two numbers.

A key observation about the LASSO regressions in Table D.1 is that both the number of regressors as well as the explained variance of the left-hand-side variable change highly nonmonotonically as the number of available regressors increases. For example, moving from a set of regressors with linear and nonlinear forecasts and sentiments to the same set of regressors with lags, the LASSO prefers fewer variable (37 instead of 80) and the explained variance drops from 81% to 59%. This reflects the fact that with many strongly correlated regressors, the cross-validation with LASSO picks very different regressors as the specific set of regressors changes.[2] This echoes insights of Giannone, Lenza, and Primiceri (2022) and the reasoning for our choice of ridge as a dense rather than sparse technique, which extracts at least some information from any available regressor.

[1] We chose the weight optimally by defining a grid of α values and then using a two-layered 10-fold cross-validation procedure where the average MSE is minimized over both α and λ We used a grid of 21 grid points to generate the results in Table D.2.

[2] We found that this is even true when the set of regressors only changes marginally.

Table D.2: GOODNESS-OF-FIT ACROSS DIFFERENT SPECIFICATIONS – ELASTIC NET VERSION

	Number of regressors	(1) Variance explained with 10-word sentiment (main specification)	(2) Variance explained with 5-word sentiment (robustness)
Romer-Romer original OLS with subset of forecasts	18	0.50	
Elastic net with extended set of forecasts	42	0.57 [$\alpha^* = 0.2$]	
Elastic net with extended set of forecasts (nonlinear)	49	0.66 [$\alpha^* = 0.7$]	
Elastic net with all forecasts & sentiments (linear)	429 / 429	0.65 [$\alpha^* = 0$]	0.66 [$\alpha^* = 0$]
Elastic net with all forecasts & sentiments (nonlinear)	69 / 858	0.78 [$\alpha^* = 0.6$]	0.77 [$\alpha^* = 0$]
Elastic net with all forecasts & sentiments (linear with lags)	129 / 1,613	0.91 [$\alpha^* = 0.95$]	0.88 [$\alpha^* = 0$]
Elastic net with all forecasts & sentiments (nonlinear with lags)	3,226 / 3,226	0.94 [$\alpha^* = 0$]	0.95 [$\alpha^* = 0$]

Notes. Implied goodness-of-fit, measured by R^2 (OLS) and deviance ratio (ridge), from estimating different empirical specifications of equation (3). For the first two specifications, sentiments are not included so the 10-word/5-word distinction does not apply. Different from ridge, in the elastic net regressions the number of regressors can differ between the 10-word and 5-word version, so the 'Number of regressors' column includes two numbers. In square brackets, the optimally chosen weight between the LASSO and ridge penalties (α^*) is reported. $\alpha = 0$ corresponds to a pure ridge model and $\alpha = 1$ corresponds to a pure LASSO model.

Table D.2 provides further support for our choice of ridge. It shows that with an elastic net regression, which weighs the penalties of a ridge and a LASSO, the cross-validation procedure in many cases prefers a pure ridge model, i.e. $\alpha^* = 0$. Most notably, this is the case for our preferred specification with all forecasts, sentiments and nonlinear terms and lags, in the last line of the table.

In addition to these insights about how the alternative ML methodologies work in terms of fit, we found that the resulting monetary policy shocks that one gets in each case are not drastically different, as long as sufficient information is included. For example, the correlation between the shocks from our main ridge specification and the analogous LASSO specification is 0.93. We also constructed IRFs in the BVAR and found those to be broadly similar between shocks based on our main ridge model, when using LASSO instead of a ridge, and when using the richest elastic net model in which the optimal α did not select a ridge. Importantly, all of these shocks are very different from the original Romer-Romer shocks in terms of the IRFs they imply.

E Evidence for the modal nature of Greenbook forecasts

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

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

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

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

- In the **February 1994** meeting, Chairman Greenspan explains "Watching the market behave in the long end since our move just reinforces what Joan was discussing. I'm not certain that we can say at this stage that the modal forecast for growth in the first quarter has changed materially. But the probability that the growth rate in the first quarter will be significantly higher than previously expected may be higher while the probability that growth in the first quarter will be significantly below the expected modal forecast is clearly much lower. As a consequence, the average expectation for the first quarter clearly has increased."

- In the **July 1996** meeting Michael Prell, the director of Research and Statistics clarifies: "I think there have been some occasions when we have indicated that the risks in our outlook were asymmetric. I would characterize our forecasts over the

years as an effort to present a meaningful, modal forecast of the most likely outcome. When we felt that there was some skewness to the probability distribution, we tried to identify it. In this instance, as we looked at the recent data, we felt that there was a greater thickness in the area of our probability distribution a little above our modal forecast."

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

- In the **November 2001** meeting, Governor Meyer states, in reference to the 9/11 terrorist attacks that "The Greenbook, like most forecasts, seems to assume a one-time terrorist attack with a near-term effect on confidence that dissipates over time. That might be appropriate for a modal forecast. But relative to this assumption, there seems to be significant asymmetric downside risks, specifically of further terrorist attacks that affect confidence in the economy or perhaps for other reasons as well. The forecast for the first state of the world is therefore likely to be biased in an optimistic direction though, as David Stockton noted, we would be hard pressed to parameterize the downside risks associated with the second state of the world. Still this analysis suggests that the mean of the forecast might be interpreted as being below the mode in this case. So the question is how policy should respond to this type of uncertainty and whether policy should be set to err on the side of ease relative to the modal forecast."

- In the **March 2005** meeting, President of the Federal Reserve Bank of San Francisco Janet Yellen states that "While the Greenbook expectation of a relatively flat path for bond rates through the end of next year may be a reasonable modal forecast, I don't think the risks here are balanced."

- In the **June 2009** meeting, FOMC secretary Brian Madigan lays out different policy options, with reference to the forecasts: "With both a modal outlook for weak growth and low inflation, and downside risks around the outlook for activity, macroeconomic considerations would seem to argue for providing additional monetary policy stimulus at this juncture. However, with the federal funds rate at the zero bound, the Committee has limited policy options at its disposal."

- In the **June 2011** meeting, President of the Federal Reserve Bank of San Francisco John Williams explains "Furthermore, despite the deep cuts to the output projection, the Tealbook has also shifted to a downside skew to the risks of the growth outlook. This combination of a downward modal revision to the growth forecast and downside risk assessment is a truly sobering development, but it's consistent with what we see in financial markets."

- In the **December 2016** meeting, Vice Chairman Dudley says "I guess my view of the risks to the forecast is that you have a modal forecast and then you ask, where is the skew of the distribution? It's not about where the lower bound lies relative to the funds rate. So I guess I interpret the balance of the risks differently (...)."

F More results on forecast error predictability

F.1 Additional results for output and inflation forecasts

Table F.1: ADDITIONAL GREENBOOK FORECAST ERROR PREDICTABILITY TESTS

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

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

F.2 Results for first release instead of final vintage

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

	Panel (a): unemployment rate forecast errors on LHS							
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	-0.025*	-0.104**	-0.433**	-0.637**				
	[0.013]	[0.045]	[0.189]	[0.242]				
Economic activity sentiment					-0.020	-0.089*	-0.272	-0.376**
					[0.014]	[0.044]	[0.166]	[0.173]
Constant	-0.032***	-0.084***	-0.097	0.048	-0.032***	-0.083*	-0.093	0.150
	[0.011]	[0.031]	[0.119]	[0.201]	[0.011]	[0.032]	[0.142]	[0.260]
R^2	0.038	0.129	0.240	0.214	0.020	0.084	0.084	0.058
Number of observations	210	210	210	62	210	210	210	62
	Panel (b): output forecast errors on LHS							
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	-0.093	0.172	0.327	-0.291				
	[0.125]	[0.256]	[0.282]	[0.345]				
Economic activity sentiment					-0.144	0.052	-0.069	-0.551*
					[0.131]	[0.235]	[0.228]	[0.318]
Constant	0.214**	0.070	-0.236	-0.348	0.218**	0.067	-0.241	-0.374
	[0.103]	[0.192]	[0.256]	[0.568]	[0.106]	[0.200]	[0.283]	[0.535]
R^2	0.006	0.009	0.024	0.015	0.012	0.001	0.001	0.045
Number of observations	208	206	200	54	208	206	200	54
	Panel (c): inflation forecast errors on LHS							
	current quarter	1 quarter ahead	1 year ahead	2 years ahead	current quarter	1 quarter ahead	1 year ahead	2 years ahead
First PC of all sentiments	0.104	0.049	0.116	0.062				
	[0.091]	[0.093]	[0.126]	[0.155]				
Economic activity sentiment					0.201**	0.098	0.232**	0.130
					[0.087]	[0.093]	[0.115]	[0.196]
Constant	-0.167**	-0.133	-0.281*	-0.483**	-0.170**	-0.135	-0.285**	-0.470**
	[0.079]	[0.123]	[0.155]	[0.214]	[0.073]	[0.120]	[0.143]	[0.212]
R^2	0.018	0.003	0.013	0.004	0.059	0.010	0.046	0.012
Number of observations	210	210	210	62	210	210	210	62

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

G Construction of committee composition variables

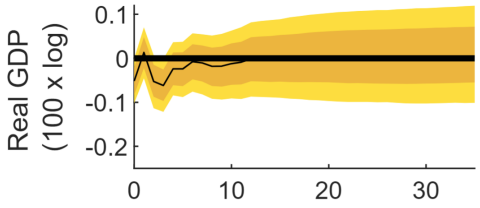
The additional data set that captures information on the composition of the FOMC in each meeting, which we use for robustness, is constructed as follows. For each FOMC meeting, we record the list of participants. This list consists of the governors at the board as well as the representatives from each regional bank. Typically, regional bank representatives are their respective presidents, except in cases where there is an interim president. We classify the participants by their voting status: they are either voting members, alternate members, or non-voting members. The governors always vote and the regional bank presidents alternate between the three roles. For each governor, we create a dummy variable that equals 1 if he/she attended a given meeting and 0 otherwise. We record the attendance of each regional bank representative in a similar way. Here we create three sets of dummy variables. The first set of variables are constructed at the participant-position-voting status level, meaning for example that we distinguish between Mr. Boehne (president of the FRB of Philadelphia) when he is attending as a voting member and when he is attending as a non-voting member. The second set of variables are constructed only at the participant-position level, without regard to their voting statuses. The last set of variables recorded whether a regional bank's representative voted during the meeting for each of the 12 banks. For governors, we also record information on who appointed them. We tally the total number of governors in attendance by the US president who made the appointment, as well as the number of governors appointed by a Republican and Democratic administration respectively.[3] 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.

[3] 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.

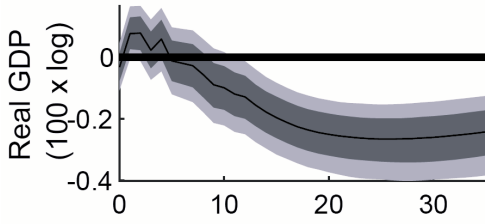
H Additional results

Figure H.1: OUTPUT IRF TO SHOCKS FROM ROMER-ROMER OLS IN DIFFERENT SAMPLES

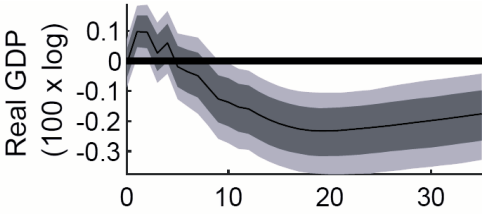
(a) Our main 1982–2008 sample (BVAR includes all variables)



(b) Romer-Romer 1969–1996 sample (BVAR excludes EBP)

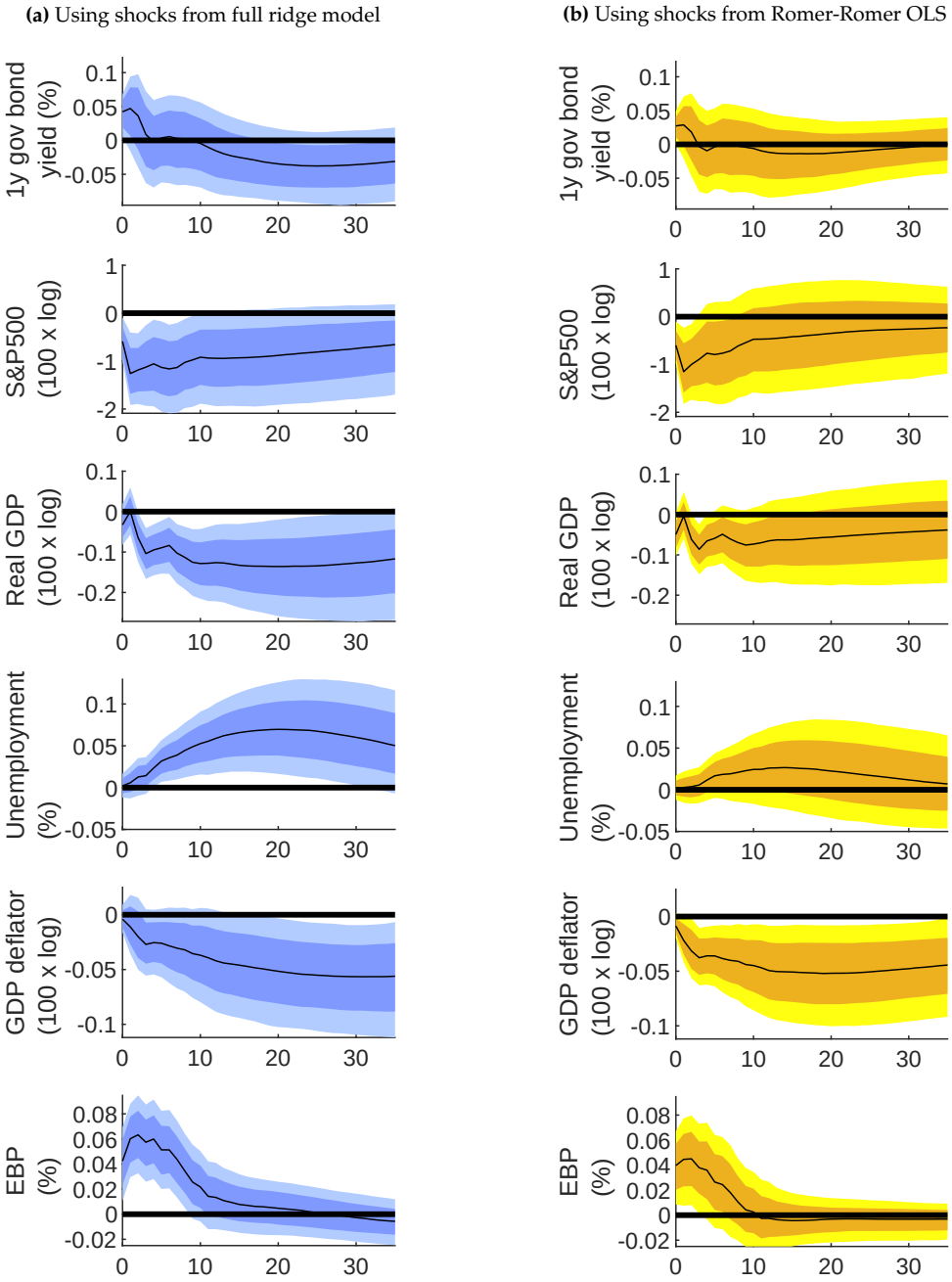


(c) 1973–1996 sample (BVAR includes all variables)



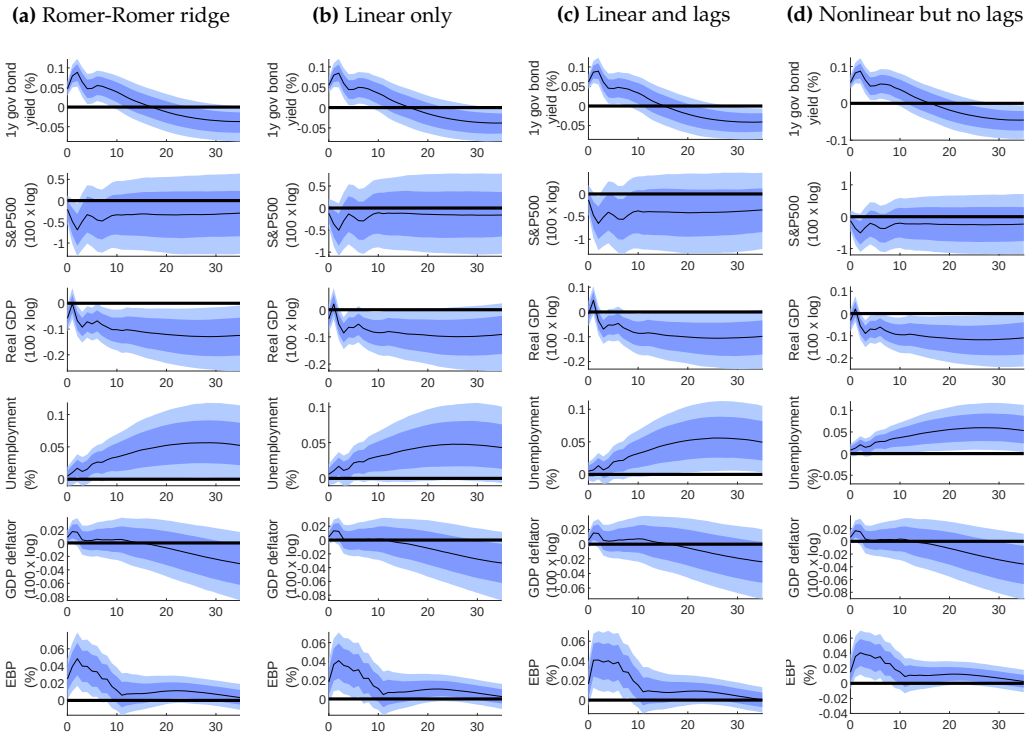
Notes. Panel (a) repeats the output IRF from Figure 6, Panel (b), which is based on our replication of Romer and Romer (2004) and our estimated BVAR. Panel (b) shows the output response to the same shock and in the same BVAR but for the original Romer-Romer sample from 1969 to 1996. That sample excludes the EBP from the BVAR due to data availability. Panel (c) shows the results for the 1973 to 1996 sample, which has the biggest overlap with the original Romer-Romer sample that can feature all variables in our BVAR, including the EBP. In both Panel (b) and (c) real GDP falls after a monetary policy tightening. This finding makes clear that it is the sample choice that drives the lack of an effect on activity in Figure 6, Panel (b), and not the fact that we use a different method from Romer and Romer (2004) to construct IRFs.

Figure H.2: IRFS CONSTRUCTED WITH ADDITIONAL SIGN RESTRICTIONS



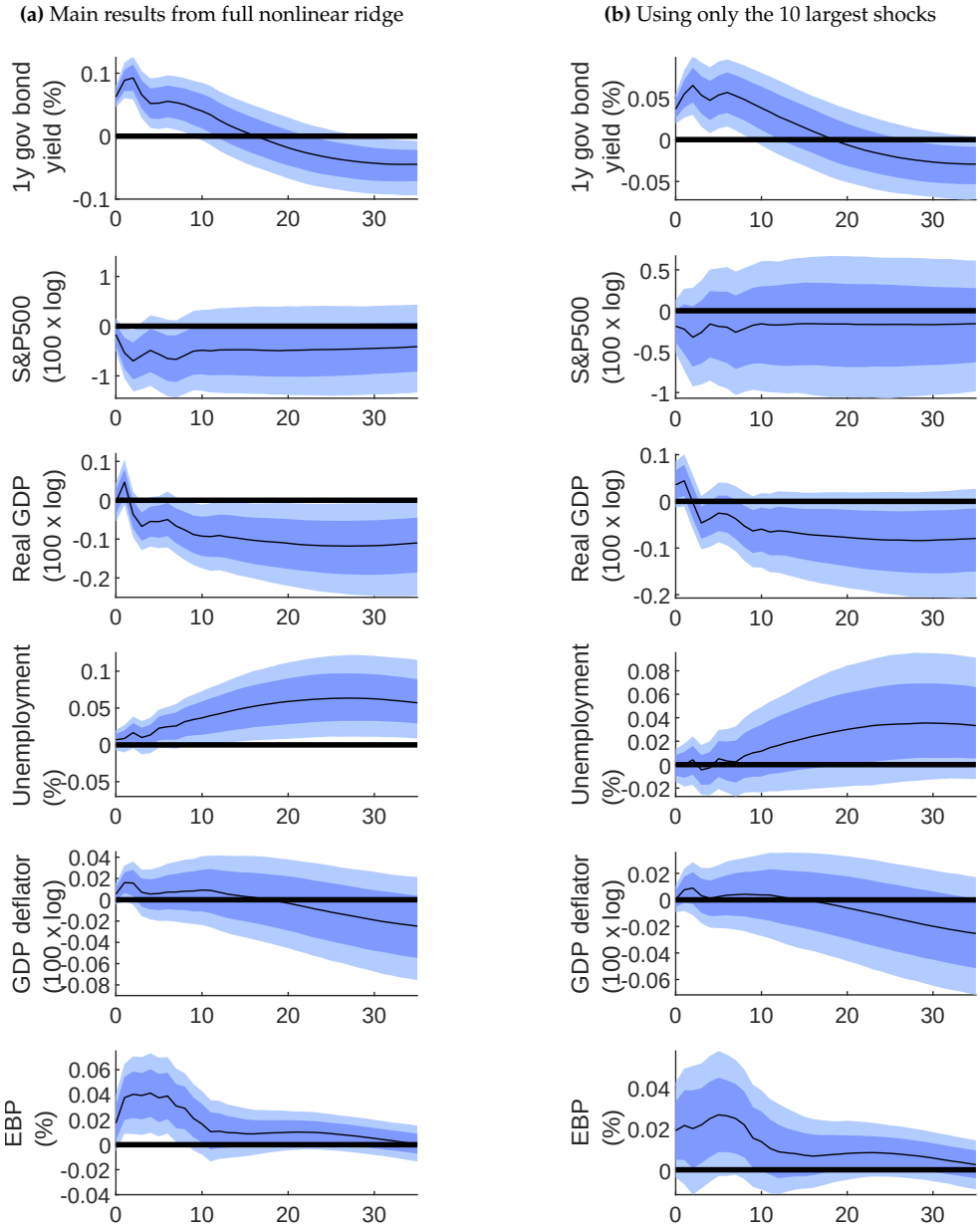
Notes. The two panels correspond to those in Figure 6, but impose the additional sign restrictions suggested by Jarocinski and Karadi (2020) to separate monetary policy shocks from central bank information shocks. Specifically, the IRFs shown here are for monetary policy shocks which are assumed to create a negative covariance between interest rates and stock prices.

Figure H.3: IRFS ESTIMATED FROM INTERMEDIATE SHOCK VERSIONS



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

Figure H.4: IRFS TO ALL SHOCKS AND THE 10 LARGEST SHOCKS IN COMPARISON



Notes. Panel (a) repeats our main IRFs (Figure 6, Panel (a)). Panel (b) applies the same BVAR specification but only using the 10 largest observations in absolute value for the time series of the monetary policy shocks, setting the shock for all other meetings to zero. 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.

Table H.1: Variance contribution of monetary policy shocks in the BVAR

Panel (a): Using shocks from full ridge model

Variable	6m	12m	24m	36m
1y gov bond	0.11	0.08	0.06	0.07
S&P500	0.03	0.03	0.03	0.04
Real GDP	0.03	0.04	0.06	0.07
Unemployment	0.02	0.04	0.06	0.08
GDP deflator	0.02	0.02	0.02	0.03
EBP	0.05	0.06	0.07	0.06

Panel (b): Using shocks from Romer-Romer OLS

Variable	6m	12m	24m	36m
1y gov bond	0.06	0.06	0.04	0.04
S&P500	0.01	0.02	0.03	0.03
Real GDP	0.02	0.02	0.02	0.02
Unemployment	0.01	0.01	0.02	0.02
GDP deflator	0.01	0.01	0.02	0.02
EBP	0.01	0.02	0.02	0.02

Notes. Share of variance explained by monetary policy shocks, for the six variables included in the BVAR at four different horizons. Panel (a) shows the results for the shocks estimated based on our main specification and Panel (b) for our replication of the Romer-Romer shocks. Calculations are based on the BVAR variance decomposition following Jarocinski and Karadi (2020).

Figure H.5: IRFS TO DIFFERENT MONETARY POLICY SHOCKS USING LOCAL PROJECTIONS

(a) Using shocks from full ridge model

(b) Using shocks from Romer-Romer OLS

Notes. IRFs analogous to Figure 6 in the main text, but based on a frequentist local projections approach (Jordà, 2005) rather than a BVAR. Panel (a) uses our proposed measure of monetary policy shocks, estimated using the full nonlinear ridge model on the extended set of numerical forecasts and our sentiment indicators from FOMC documents. Panel (b) shows the analogous IRFs when a simpler empirical specification is used to estimate the shocks, which includes only the original set of numerical forecasts in a Romer-Romer OLS regression. The solid line represents the median, and the 5th and 95th percentiles are captures by the bands. The sample used to estimate the IRFs is 1984:02-2008:10.

I Extracting shocks from recent FOMC meetings

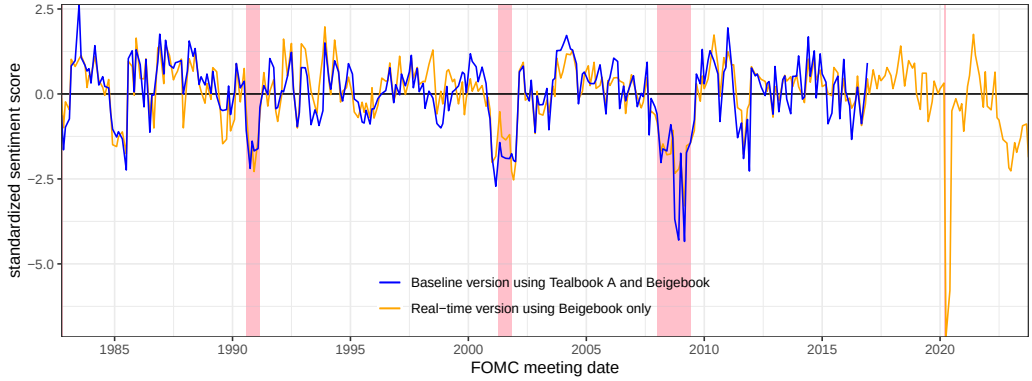
As an extension, we demonstrate how our method can be used to extract monetary policy shocks from the FOMC's more recent decisions. While the Tealbooks are available only with a five-year delay, the Beigebooks are available prior to every FOMC meeting. These summarize regional economic conditions for each individual Federal Reserve district. We already use the Beigebooks alongside the Tealbooks over our main sample period 1982-2008. The idea behind this section is to show that constructing our sentiment indicators only from the Beigebook text provides at least a limited proxy for the FOMC's information set.

We verify how well this proxy works: while in our main analysis we use both the Tealbook A and the Beigebook, we find that using only the Beigebook over our main 1982 to 2008 sample gives us strongly correlated sentiment indicators, as illustrated for "economic activity" in Figure I.1. Running our main ridge regression with these sentiments, we find that the deviance ratio from using only Beigebook sentiments to estimate (3) is 0.68, compared to 0.94 with information from Tealbooks and Beigebooks combined. The resulting shocks have a correlation of 0.92 with each other. We further confirm that the BVAR IRFs we study in the previous section look qualitatively similar for the shocks constructed using only the Beigebook. It is important to emphasize that leveraging the Beigebooks is not possible in the original Romer and Romer (2004) approach, as the Beigebooks do not contain any numerical forecasts. This is a further advantage of our NLP approach.

As a "proof of concept", we run the Beigebook-only ridge, with 4 lags and squared terms, over the period December 2015 to October 2023. This sample starts after the 2008 to 2015 zero lower bound period, and it includes all interest rate increases that the Fed undertook during the inflationary period in 2022-2023.[4] This is not feasible in our baseline because Tealbooks are not yet available over this sample. Towards the end of this period, our procedure measures sharp changes in the sentiment indicators around various economic concepts in the Beigebooks. For example, the sentiment around "inflation" drops massively in late 2021, with a reduction of more than 6 standard deviations (in terms of its 1982-2023 variability). A main contributors to this pattern is a sharp increase in the use of the negatively

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

Figure I.1: SENTIMENT SURROUNDING ECONOMIC ACTIVITY: BASELINE VS. BEIGEBOOK-ONLY



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

connotated word "concern" from the Loughran and McDonald (2011) in proximity to inflation. Other concepts around which the sentiment deteriorates strongly into negative territory in the runup to the first tightening decisions are "recession", "fuel", and "China".

We find that the fit from estimating the Beigebook-only version of (3) over the 2015 to 2023 sample is 98%, suggesting only a small role for monetary policy shocks. Recall that this is the case despite the fact that we can only include the Beigebook sentiments, without using Tealbook sentiments and numerical forecasts which add significant predictive power in the 1982 to 2008 period. While the total increase in the FFR target between March 2022 and October 2023 amounted to 525 bp, the estimated shock component cumulates to around 21 bp over this period. In other words, our method implies that the tightening starting in 2022 entailed only mild contractionary monetary policy shocks.

To conclude, we think researchers should use our baseline measure whenever they can, even if it means dropping a number of observations at the end of their sample due to the availability of the Tealbooks. In situations where this will be very costly, the Beigebook-only version provides a viable alternative.