UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

A New Monetary Policy Shock with Text Analysis

Adrian
C.R. Ochs
University of Cambridge

Abstract

Measures of monetary shocks commonly give rise to the puzzling result where a monetary tightening has an expansionary effect. A possible reason is that agents may believe that monetary shocks contain information regarding the central bank's assessment of the economic environment (Nakamura and Steinsson, 2018). Under this hypothesis, the estimated response to monetary policy shocks would contain two conflating effects: the actual effect of monetary policy and the reaction of private agents to the newly acquired information. This paper overcomes this problem by extracting a novel series of monetary shocks using text analysis methods on central bank documents. The resulting text-based variables contain the informational content from changes in the policy rate. Thus, they can be used to extract exogenous changes in monetary policy that are orthogonal to any central bank information. Using this information-free measure of monetary policy shocks reveals that a monetary tightening is not expansionary, even when estimated on more recent periods.

Reference Details

2148	Cambridge Working Papers in Economics
2021/21	Cambridge-INET Working Paper Series
Published	9 June 2021
Websites	www.econ.cam.ac.uk/cwpe
	www.inet.econ.cam.ac.uk/working-papers

A New Monetary Policy Shock with Text Analysis

Adrian C.R. Ochs*

First Version: December 12, 2019
This Version: June 9, 2021

Abstract

Measures of monetary shocks commonly give rise to the puzzling result where a monetary tightening has an expansionary effect. A possible reason is that agents may believe that monetary shocks contain information regarding the central bank's assessment of the economic environment (Nakamura and Steinsson, 2018). Under this hypothesis, the estimated response to monetary policy shocks would contain two conflating effects: the actual effect of monetary policy and the reaction of private agents to the newly acquired information. This paper overcomes this problem by extracting a novel series of monetary shocks using text analysis methods on central bank documents. The resulting text-based variables contain the informational content from changes in the policy rate. Thus, they can be used to extract exogenous changes in monetary policy that are orthogonal to any central bank information. Using this information-free measure of monetary policy shocks reveals that a monetary tightening is not expansionary, even when estimated on more recent periods.

*University of Cambridge, Department of Economics, Austin Robinson Building, Sidgwick Ave., Cambridge CB3 9DD, United Kingdom (email: acro2@cam.ac.uk). I received helpful comments from Irene Benito, Giancarlo Corsetti, Lukas Freund, Stephen Hansen, Konrad Kuhmann, Michael McMahon, Silvia Miranda-Agrippino, Christopher Rauh, Pontus Rendahl, Christina Romer, David Romer, Jón Steinsson, and Lutz Weinke. I also gratefully acknowledge generous support from the Cambridge Trust and Wolfson College.

When estimated on recent periods, many of the conventional monetary policy shock (MPS) measures obtained via SVAR analysis, narrative methods or, more recently, high-frequency approaches show puzzling results where a monetary tightening has expansionary effects (Barakchian and Crowe, 2013; Miranda-Agrippino and Ricco, 2017). Assuming that these methodologies successfully isolate exogenous movements in the policy rate, Nakamura and Steinsson (2018) argue that private agents could still face a signal extraction problem, which gives rise to the so-called *information channel*. For example, suppose private agents interpreted an exogenous increase in the policy rate to signal an improving economy. In that case, production could increase and unemployment decrease, leading to an expansionary effect after a policy tightening (Romer and Romer, 2000; Nakamura and Steinsson, 2018). Hence, the results of MPS, even if econometrically exogenous, depend on whether private agents believe that they carry information on the central bank's assessment of the economy (Ramey, 2018).

This paper derives a new monetary policy "text shock"[1] series that addresses the information channel problem by using natural language processing techniques on central bank documents. After every policy meeting, the federal open market committee (FOMC) publishes so-called minutes to explain their policy decisions to the public. Quantifying the information within these minutes captures the explanations of policy decisions. Therefore, it is a direct proxy for the signal extraction problem facing private agents, deciding whether a surprise move in the policy rate is due to superior central bank information or genuinely exogenous.

To extract macroeconomic information within the FOMC minutes, I use a directional sentiment approach. More precisely, I construct three so-called dictionaries containing keywords that describe output, inflation and unemployment, respectively. These keywords are then counted for every sentence. Similarly, I adopt a directional-term dictionary with words such as "increase" or "decrease". Combining the directional-term count with the keyword counts allows assessing how much was talked about the three economic variables in question and in which direction.

I then regress the central banks' policy rate on the textual information from the FOMC to obtain the monetary policy shock series. Interpreting the FOMC minutes as the private agents' information set on monetary policy decisions, anything that is not explained by the text measures is a surprise move in the policy rate, which can be interpreted as MPS to private agents.[2] This is similar to the seminal work by Romer and Romer (2004), who

[1]In a recent article that I was not previously aware of, Handlan (2021) calls her shock series "text shocks" as well. While both shocks are derived using text analysis, they are not the same.

[2]This is a necessary assumption. Assume the central bank does not have information beyond what

use Greenbook[3] forecasts to free the policy rate from anticipatory movements. However, these forecasts are not available to the public at the time of policy setting. Thus, even if regressing the target rate on central bank forecasts produced exogenous shocks in the econometric sense, these shocks would only coincide with private agent monetary policy shocks if the central bank and the private agents had the same information set (Ramey, 2018).

This paper focuses on resolving the information channel problem by controlling for the private agent information set. Nevertheless, for the shock series to be exogenous, it cannot be anticipated by the central bank. Suppose the central bank had information that private agents do not and acted on it. Then the estimated effect of monetary policy comprises the causal effect and the realisation of the economic developments the fed foresaw (Romer and Romer, 2004). To show that the central bank does not anticipate the new shock series, I regress the text shock series on the forecast variables contained with the FOMC Greenbook. None of the forecasts is significant, and hence, the new shock series is exogenous to the central bank's information set. This also implies that the FOMC reveals much of its information in the minutes and that the language processing tools applied can extract data from text.

I use the newly identified shock series as an instrument in a local projection instrumental variable approach (LP-IV) to produce impulse response functions (IRFs) (Jordà et al., 2015). The empirical IRFs of output and employment are back in line with standard monetary theory. That is, a monetary tightening leads to an economic contraction. Contrary to conventional shock measures, this holds even for recent time periods, excluding the volatile Volcker period pre-1983. I obtain comparatively large results where a monetary policy shock leading to a one percentage point increase in the federal funds rate decreases industrial production by approximately 4.5% at its trough. Similarly, unemployment increases by up to 0.8% per cent.

I also show that the more rigorous conduct of monetary policy is a possible explanation for why conventional shock measures used to obtain theory-consistent results when estimated on earlier time periods but not when confronted with more recent data. Contrary to other shock series that are only available from the 1990s, the new text shock series allows studying earlier periods. Analysing earlier periods reveals that monetary policy was less systematic in earlier periods and that differentiating between the central bank

is contained in the FOMC minutes, but private agents believed otherwise. Then the shock series would still be contaminated with what private agents infer from the movement in the policy rate even though there was no actual information revealed.

[3]The Greenbook contains macroeconomic forecasts published internally by FOMC staff members one week before every meeting.

and private agent information set was less important. Indeed, comparing the text shock to the Romer and Romer (2004) narrative shock shows that the IRFs of the new shock, estimated on 1983–2007, are similar to what Romer and Romer (2004) found for their time period, 1969–1996. Therefore, the information channel appears to have become more important in recent years resulting in misleading results of conventional shock measures.

Literature This paper contributes to several strands of literature. Firstly, methodologically, this paper is most similar to Hansen and McMahon (2016) and a recent working paper by Handlan (2021). Both articles analyse FOMC statements using topic modelling and neural networks, respectively, to analyse the effect of forward guidance on macroeconomic aggregates. Focusing on FOMC minutes instead, I am able to derive a longer shock series dating back to 1969 and study why previous shock measures led to theory consistent results when estimated for earlier but not when estimated for more recent periods.

Furthermore, although in recent years topic modelling (Blei et al., 2003) has become the almost default for analysing textual data in economics (Hansen and McMahon, 2016; Hansen et al., 2018; Dybowski and Adämmer, 2018; Mueller and Rauh, 2018), I show that the appropriate way to analyse texts depends very much on the task at hand. Using simple text analysis methods can be powerful when the economic information to look for in texts is already known and where the underlying text source is technical. The latter point means that technical documents are explicit about their language such that, for example, "inflation" always implies that the document is talking about inflation. This simplifies text analysis since taking account of textual context through more advanced techniques becomes less critical. This transparent approach can therefore be easily adapted for other empirical research beyond monetary policy shocks.

Second, to take account of the possible signal that policy changes entail, the derivation of MPS has shifted from controlling for the central bank's information set (e.g. Romer and Romer, 2004) to controlling for the information set of private agents (e.g. Gertler and Karadi, 2015). So far, all studies controlling for the private agent information set rely on high-frequency data of federal funds futures. However, Miranda-Agrippino and Ricco (2017) show that high-frequency identified shocks are anticipated by the central bank and, therefore, not exogenous. Two recent papers have sought to solve this problem. Jarociński and Karadi (2020) use sign restrictions to differentiate between an information and a "true" monetary policy shock. Wolf (2020) has recently cautioned that the minimal requirement where only the shock of interest can satisfy sign restrictions is not sufficient when a linear combination of other shocks might lead to the same results. Miranda-Agrippino and Ricco (2017) regress changes in federal funds futures (private sector information set) on Greenbook forecasts (central bank information set). However, as pointed out before, this only yields valid exogenous shocks when the private agents' information is the same as the central bank's information set. This paper can be seen as an alternative, testing whether different assumptions, no sign restrictions, and solely using a private-sector information set can lead to similar results.

Third, there is an ongoing debate in empirical macroeconomics on using local projection or SVAR techniques. Plagborg-Møller and Wolf (2021) have shown that both methods produce the same IRFs in population with an unrestricted lag structure. Using LP-IV methods, I obtain large real effects of monetary policy compared to two other recent SVAR based studies by Miranda-Agrippino and Ricco (2017), and Jarociński and Karadi (2020). Nevertheless, these larger results from an LP-IV approach become very similar when estimating the effects in an SVAR framework as in Miranda-Agrippino and Ricco (2017) and Jarociński and Karadi (2020). This is related to the fact that differences in empirical studies on the effect of monetary policy can be due to varying methodologies, a point which was made theoretically in Wolf (2020).

Outline Section 2 discusses the data sources with a particular focus on the unstructured data of FOMC minutes. The third section describes the derivation of the new monetary policy text shock series. This contains the methodology of obtaining the text-based measures. Fourth, impulse response studies show the effects of the new shock series. The fifth section presents several robustness checks, and the final section concludes.

2. Data

This section describes the data used in the empirical analysis to derive a new monetary policy shock measure. One of the novel parts of this paper is that it combines structured (quantitative) and unstructured (textual) data for 1969−2007. I first report the structured data sources and then focus in greater detail on the description of the FOMC minutes and the pre-processing steps conducted for text analysis.

The structured data for the main analysis is obtained from Romer and Romer (2004) and Coibion et al. (2017). The central bank's policy rate in the form of the intended funds rate series stems directly from Romer and Romer (2004) for 1969 − 1996 and was extended up until 2007 by Coibion et al. (2017). For comparability, I use the data on output, inflation and unemployment from the Romer and Romer (2004) data set. To check whether the new text shock series is forecastable, I use macroeconomic forecast variables on unemployment, output and inflation from the FOMC Greenbook.[4] Additionally, I draw on the FRED monthly data set of the Federal Reserve Bank St. Louis to obtain a large set of macroeconomic variables for a second forecastability check.[5] All FOMC related data (intended funds rate, Greenbook, and minutes) are at the frequency of meetings.

Leaving the target rate aside, the critical data source for this paper is the minutes of the FOMC meetings. The information from these texts acts as a proxy for the private agents' information set. FOMC minutes are roughly structured in four parts (Jegadeesh and Wu, 2017). The first part is administrative and contains information on who participated in the meeting and previous policy actions. The second part reviews the go-around session during FOMC meetings where central bankers present their prepared statements on the economic situation. The second part is augmented by information on the FOMC members' discussion of the current economic environment and their projections. The final section contains policy decisions, their rationalization and potentially, discussions on future policy actions.

One caveat of using FOMC minutes rather than FOMC statements is their publication lag. From 1967−1976, FOMC minutes (then called Record of Policy Actions) were released 90 days after each meeting. Starting in 1976, they were published a few days after the subsequent meeting, and since 1993, the FOMC minutes are published after three weeks. This time lag means that the source of information is not available to the public at the time of the decision making of the FOMC. Hence, using FOMC minutes is only a proxy to the signal extraction problem of private agents, albeit an improved one compared to the Greenbook data, which is published with a five-year lag. Furthermore, given that we are interested in the effect on slow-moving macro variables, FOMC minutes are still a good proxy for the signal extraction problem and the information available to private agents despite their time lag of three weeks.

FOMC statements, which are a short document on the policy decision published immediately after each meeting, are not used in the analysis because they were started only in 1994, making comparisons to earlier periods impossible. An extension of this paper could combine FOMC statements with a federal funds shadow rate (rather than the target rate) or high-frequency movements of federal funds futures to derive monetary policy shocks for more recent periods, including the zero lower bound period.

The FOMC meeting documents are directly scraped from the website of the federal reserve and then undergo light pre-processing (described in Appendix A for the dictio-

[4]These variables are also kindly provided by Romer and Romer as well as Coibion.
[5]Code to download this data set is kindly provided by McCracken and Ng (2015).

nary methods used in the following section.[6] This pre-processing results in 360 minute documents and a total of 35919 sentences for 1969 – 2007.

3. Extracting Text Data from FOMC Minutes

This section approximates the signal extraction problem of private agents using directional sentiment analysis, a form of dictionary methods. In particular, I construct three individual directional sentiment indices that track the monetary policy stance on output, inflation and unemployment, and one overall sentiment measure tracking overall economic sentiment.

3.1 Text Measure: Directional Sentiment

This subsection describes the construction of the individual directional sentiment indices. Following this, the directional sentiment measures are illustrated graphically to show that they track their quantitative counterparts well.

The construction of individual sentiment measures is similar to Apel and Grimaldi (2014) but extended in three ways.[7] In their simplest form, dictionary methods count the occurrence of keywords specified in a dictionary to measure the amount of discussion about a certain concept within a text. However, solely counting keywords such as "inflation" might not provide information on whether inflation increases or decreases. Suppose the FOMC talks more about inflation whenever it is both high or low. In this case, counting keywords would not reveal whether inflation increases or decreases. To obtain directional text measures, Apel and Grimaldi (2014) suggest combining keywords with terms that indicate directional sentiment. I adopt a directional term list from Hansen and McMahon (2016) shown in Table 1 that differentiates words into groups that indicate an increase or decrease.[8] Counting the combination of these directional terms with keywords provides a directional sentiment measure.

The first extension of the Apel and Grimaldi (2014) approach is that rather than developing only an overall measurement of the monetary policy stance, I differentiate between sentiment on inflation, employment and output by constructing keyword lists for each variable. The keyword lists shown in Table 1 are constructed following Apel and

[6]Phone conferences are not accounted for since no minutes are published for phone calls. Similar to Romer and Romer (2004), I also exclude the executive session of the 29$^{\text{th}}$ of March 1976 from my analysis.

[7]Appendix B provides example sentences from FOMC minutes to illustrate the algorithm.

[8]I thank S. Hansen for kindly providing the directional term list used in the paper Hansen and McMahon (2016).

Grimaldi (2014) and extended with commonly appearing words when the minutes talk about output, inflation and unemployment.

The differentiation of the keywords allows extracting information on inflation, output and unemployment, separately. Furthermore, it allows controlling for cases where an "increase in unemployment" and "an increase in employment" would both be classed as indicating an increase even though they should be classed as going in opposite directions. These cases can now be accounted for by interchanging the directional terms such that *increase* become *decrease* terms (and vice versa) whenever the keywords "unemployment" or "deflation" are counted.

Table 1: Dictionary Words and Directional Terms

Key-Words			Directional Terms	
Inflation	*Employment*	*Output*	*Increase*	*Decrease*
inflation	employment	output	accelerate	collapse
price	unemployment	gdp	boom	contract
cpi	labor	gnp	expand	cool
ppi	job	economic growth	fast	decelerate
deflation	hire	economic activity	gain	decrease
		production	high	fall
		productivity	improve	lose
		sale	increase	loss
			rise	low
			strong	moderate
			strength	slow
			strengthen	soften
				subdue
				weak
				weaken
				weakness

Note: This table presents the keywords used to class a sentence as being about the topic of inflation, output or employment. Counting the number of directional terms in every classed sentence and subtracting the *decrease* from the *increase* terms provides a directional measure for the sentence.

Since the minutes are economic documents, and the FOMC uses the word "inflation" when it talks about inflation, this simple approach is powerful. Unlike texts with more literary acclaim, this characteristic of the FOMC minutes allows for simple identification of whether a sentence is about inflation or GDP growth. Furthermore, knowing the variables we are looking for avoids the necessity for more advanced tools such as topic models. In fact, this sentiment approach identifying sentences with keywords is similar to a guided topic model where we know the topics in advance.

The second extension classes sentences as *increase* or *decrease* if one of the directional terms appears within the proximity of the keyword. This is different to Apel and Grimaldi (2014) who only search for direct combinations, e.g. ”increased GDP”. In particular, I test for combinations where a directional term appears with a keyword within a sentence.[9] Compared to direct keyword directional term combinations, this will capture more combinations and is closer to speech patterns in FOMC meetings. For example, in many transcripts, the keyword ”unemployment” is never used directly with any directional words. When allowing for a somewhat looser measurement, however, discussions on employment are captured again.

The final extension accounts for negations, which improve the fit of dictionary methods when compared with human judgement (Shapiro et al., 2017). Negations are sentences such as ”there was no increase in output”. A naive dictionary would falsely class this sentence as *increase*, not accounting for the ”no” as a negator of the sentence. To control for this, I use a list of the most common negators by Rinker (2019).

Equipped with the keywords and directional terms from Table 1 as well as negations, the algorithm determining the directional sentiment can be outlined below. Firstly, the algorithm goes through every sentence for each keyword and tags the sentence if the keyword appears. To ensure consistent matching, the keywords, the directional terms as well as the text was lemmatized. The lemmatization implies that keywords such as ”price” will also tag sentences that mention ”oil prices” or ”energy prices”.

The next step classes the tagged sentences as either *increase* or *decrease*. To do so, all directional words within a tagged sentence are counted. This yields a total count of terms indicating an *increase* and *decrease* for every sentence.

Thirdly, all negations within the tagged sentences are counted whenever there is an odd (even) number of negators, the directional sentiment of the sentence flips (stays the same).

All sentiment measures are calculated on the sentence level. For every keyword I count how many *decrease* and *increase* terms appear in the keyword sentence and then subtract the *decrease* counts from the *increase* counts. The result is then multiplied by plus or

[9]I also tested for more restricted combinations where directional terms appear four words before and two words after the keyword. Both yield similar results. Hence, I present the more straightforward method of counting sentence combinations.

minus one depending on whether the number of negators in the same sentence was even or odd. To aggregate this count to the meeting level, I sum all sentence-level sentiment measures and divide the result by the total number of words used in the minutes for every keyword. Dividing by the total number of words is a standardization to account for the varying length of the minutes. This provides a sentiment measure for every keyword in the three subgroups. To create the final index for output, inflation and employment, I take the average over all the keywords in every subgroup.

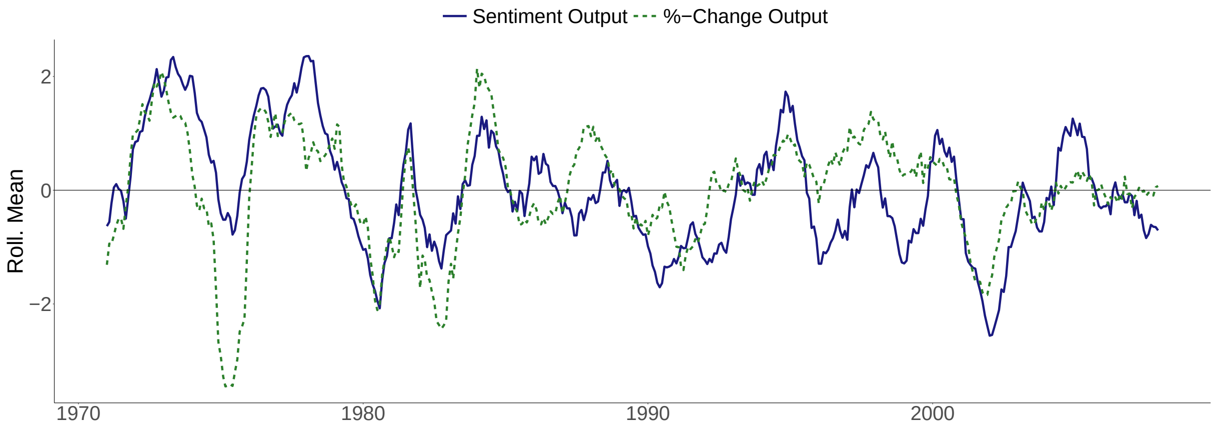
Figure 1 compares the yearly rolling mean of each directional sentiment measure to its quantitative counterpart from the FRED database measuring industrial production, consumer price index and the unemployment rate. This reveals that the dictionary methods capture information on these three macroeconomic variables well.[10] For the comparison, all variables are standardized.

Panel 1a juxtaposes the sentiment measure of output growth (solid-blue line) with industrial production growth (short-dashed green line). It can be observed that the text variable tracks the quantitative measure very closely, apart from the late-1980s, where the dynamics briefly go in opposite directions. Note that it was unnecessary to take the first difference of the directional sentiment measure to compare it to the growth in output. This is intuitive since the gross domestic product or industrial production is mostly discussed as percentage growth and rarely as its level. The directional text measure will pick this up directly and can be interpreted as measuring the percentage change in output rather than the overall level. Many economic studies using text analysis employ first-differences of their text measures in regression exercises. However, since the text variable already appears to measure changes in quantitative variables, the intuition of taking first differences would not be immediate, showing the importance of mapping the text to quantitative variables.

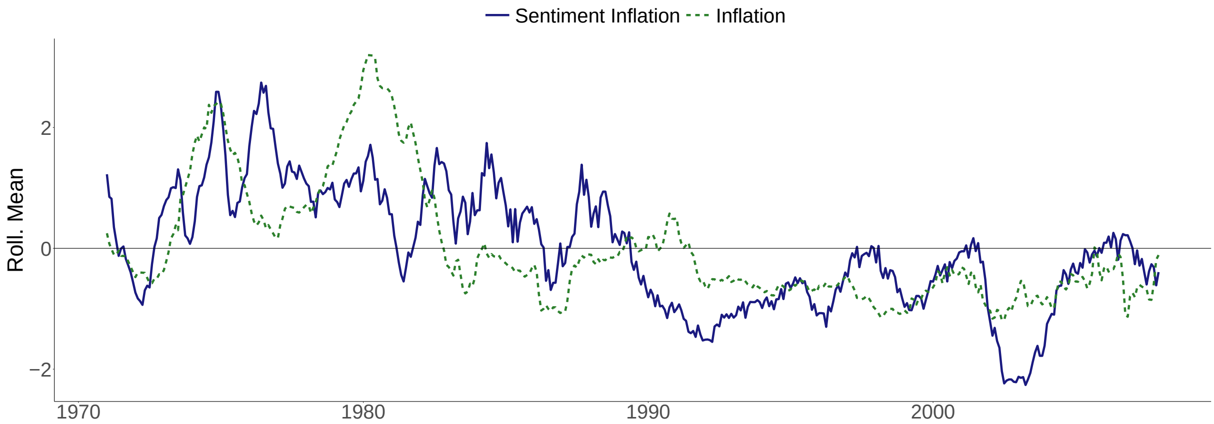
Similar observations can be made for the sentiment measure on inflation in Panel 1b, although the two measures do not seem to follow each other as closely as the sentiment on output and industrial production. Including more keywords in the inflation dictionary did not improve the measure substantially. The variable used from the FRED database is the monthly consumer price index. To obtain the inflation measure shown in Panel 1b the monthly percentage change is calculated and presented as a rolling mean. Again, inflation is mainly discussed as a percentage rather than the level of the consumer price index. Hence, the directional sentiment will measure inflation directly rather than the level of the consumer price index.

[10]The macroeconomic data stems from the FRED database, and the codes for output, inflation and unemployment were INDPRO, CPIAUCSL, UNRATE.

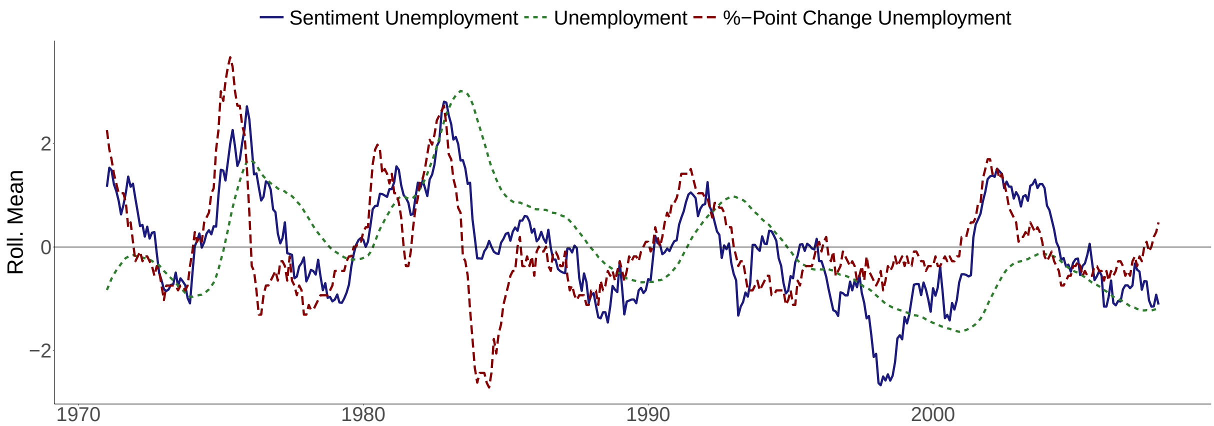
Figure 1: Directional Sentiment vs. Macroeconomic Counterpart



(a)



(b)



(c)

Note: This figure compares the directional sentiment measures for output, inflation and employment to their quantitative counterparts. The macroeconomic variables were obtained from the FRED monthly data set. For ease of comparison, the graph depicts a yearly rolling mean of the standardized variables.

Panel 1c reveals that the measure for unemployment sentiment[11] tracks the unemployment rate very well. This time, the graph includes the level of the unemployment rate (green short-dashed line) since it is the overall level of unemployment that FOMC minutes discuss rather than the monthly change in the unemployment rate (red long-dashed line). The sentiment measure does appear to track the dynamics of both quantitative measurements closely.

3.2 Text Measure: Overall Sentiment

This subsection derives the overall economic sentiment measure to complement the individual indices. Although it is important to leave as much variation within the monetary shock as possible, overall economic sentiment measures have strong predictive powers (Shapiro et al., 2017), and I include it as an additional control for economic forecasts.

To measure the stance on the overall economy, I use a dictionary approach with positive and negative terms. Differently from the above, I do not count keyword and directional term combinations. Instead, I count all the sentiment terms within the minutes. Since the FOMC minutes are solely about economics, this sentiment measure is unlikely to be measuring anything but the sentiment on economics. It can therefore be seen as an overall sentiment measure on the economy.

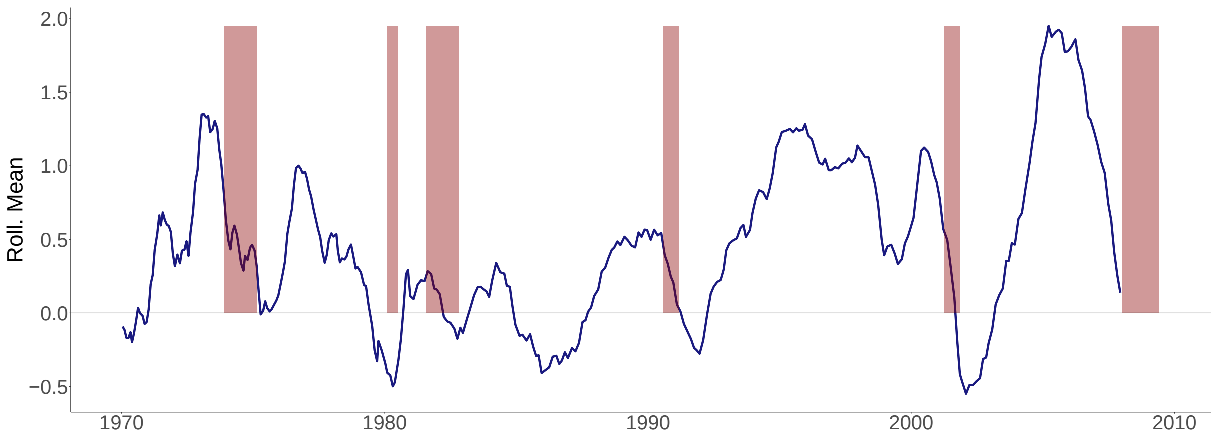
Following suggestions by Shapiro et al. (2017), I combine the sentiment dictionaries of Loughran and Mcdonald (2011) and Hu and Liu (2004) and account for negations. These dictionaries comprise approximately 4500 negative and 1800 positive terms.[12] In every sentence, the positive and negative words are counted and then subtracted from each other to obtain an overall sentence-level sentiment measure. Whenever the number of negations in a sentence is odd (even), the sign of the overall sentiment is flipped. To aggregate from the sentence to the meeting level, I sum all overall sentiment measures per sentence and divide them by the total number of words within the respective minutes.

Figure 2 shows that the overall sentiment consistently declines before and during recessions. The graph depicts the rolling mean of the overall economic sentiment for the FOMC minutes of every meeting since 1970. The shaded areas are NBER recessions. Overall sentiment starts to decline before and continues to decline further during all six NBER recessions. This would indicate that overall sentiment is a good predictor for recessions and that the text measure appears to extract valuable information from the

[11] Throughout the text, I have used a measure for employment. Here, the negative of the employment text variable measures unemployment sentiment for ease of comparison with the unemployment rate.

[12] The dictionaries are available on their website as well as in the replication code provided for this paper.

Figure 2: Overall Sentiment



Note: The overall sentiment of each FOMC minutes depicted as a rolling mean from 1970−2007. The red bars indicate NBER recession.

FOMC minutes.

4. Derivation of the Monetary Policy Text Shock

The previous section empirically approximated private agents' signal extraction problem by constructing variables based on the information contained within central bank minutes. These text-based variables are now used to purge the federal funds target rate from anticipatory movements and information effects. I then show that the resulting MPS series is not forecastable with private information of the central bank or factors of macroeconomic aggregates.

4.1 Specification

This subsection specifies the derivation of the new MPS series, regressing the change in the federal funds target rate on the text measures constructed in the previous section. The resulting residuals are MPS - movements in the target rate unexpected by private agents and not interpreted as signals about the economy.

Similar to Romer and Romer (2004), I estimate an empirical Taylor rule that takes account of output, inflation, and unemployment forecasts and use the residuals as MPS. It is not necessary to account for all the information private agents and central banks possess at the time of the committee meeting. Instead, to study the effects of monetary policy on inflation, output and unemployment, it suffices to control for the information and anticipation on these three variables (Romer and Romer, 2004). What is essential is that the shock removes the problem of reverse causality from the output, unemployment, and inflation forecasts (Cochrane, 2004). This is by construction since anything that remains in the residual is orthogonal to these three variables. Additionally, this allows for more variation in the residuals, which is helpful for identification and alleviates the problem that monetary shocks are nowadays rare and small (Ramey, 2018).

The difference to the Romer and Romer (2004) approach is that I do not use the central bank's but the private agents' information set to control for output, inflation and unemployment. This allows accounting for the information channel. Controlling with the central banks' information set might produce econometrically exogenous shocks. However, private agents who produce the data that we estimate the impulse response functions with might interpret these movements in the target rate as new information producing misleading impulse response functions.

$$\Delta ff_m = \alpha + \beta_1 ffb_m + \beta_2 y^s_m + \beta_3 \pi^s_m + \beta_4 emp^s_m + MPS_m \tag{1}$$

To derive the new shock series, equation (1) depicts the regression of the target rate, ff on three sentiment measures for output, inflation and employment (y^s, π^s, emp^s) as well as the old target rate, ffb to control for mean reversion (Romer and Romer, 2004). The remaining residual of equation (1), ϵ is the monetary policy shock on the frequency of FOMC meetings.

Table 2 shows the regression results from equation (1). Again, it should be noted that the goal was not to predict the central bank's Taylor rule perfectly. Hence, the low fit of roughly 37% is unsurprising. Moreover, if the text measures are good approximations for the information set of private agents, the low fit is a good sign leaving more variation in the error term. Additionally, repeating the same exercise but with Greenbook forecasts instead of text measures following Romer and Romer (2004) yields a similarly low fit of 44%, indicating that the text measures capture useful information similar to quantitative variables from the Greenbook (see Appendix D).

Analysing each explanatory variable, both sentiment on *output* and the *overall* sentiment measures are highly significant and positive. The positive sign would indicate that an increase in the directional sentiment of output correlates with a positive change in the intended federal funds rate. This is what we would expect the central bank to do to protect the economy from overheating. Similarly, when *overall* sentiment increases, this indicates a positive stance on the overall economy, and hence, the central bank would

Table 2: Identification of the Text Shock

	Δ Target Rate	
	Coefficients	*Standard Error*
Constant	-0.137^{***}	0.035
Initial level of Target Rate	0.004	0.006
Sentiment Measures:		
Output	0.047^{***}	0.012
Inflation	0.005	0.015
Employment	0.007	0.007
Overall	0.115^{***}	0.017

* $(p < 0.1)$, ** $(p < 0.05)$, *** $(p < 0.001)$
Adjusted $R^2 = 0.366$, N= 200

Note: This table shows the regression of the change in the intended Federal Funds rate on sentiment measures. The resulting residual is interpreted as a monetary policy shock.

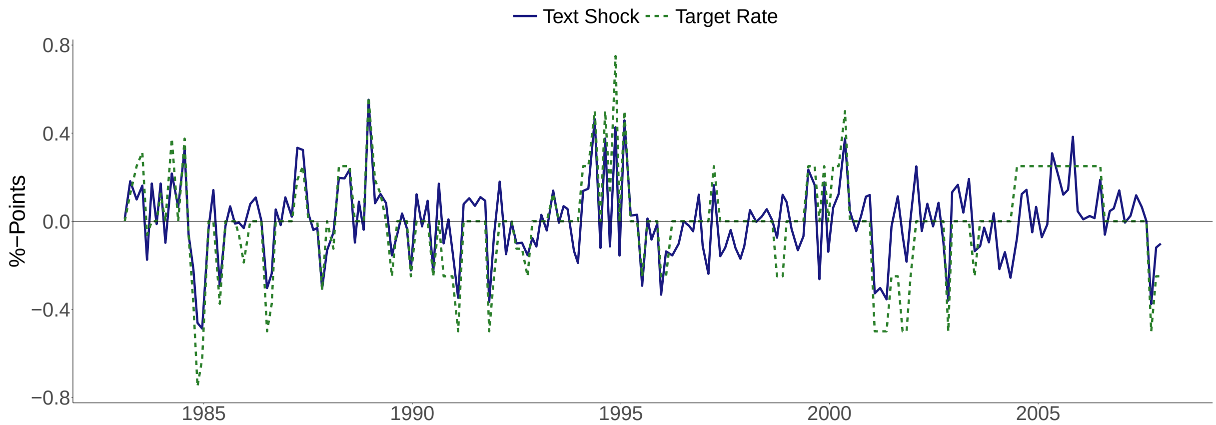
increase interest rates.

While both *inflation* and *employment* are statistically non-significant, their positive signs can be interpreted similarly to the above. Moreover, the insignificance might be due to some multicollinearity between the individual and the overall sentiment measure. Indeed, repeating the above regression without the *overall* sentiment and only with the directional sentiments does make the employment sentiment measure significant, hinting at some collinearity between the *overall* and the directional measures (see Appendix C). Furthermore, the results (including the later impulse response studies) without the *overall* measure are similar to including it. However, explanatory power drops roughly 10 percentage points. Hence, including *overall* sentiment acts as an exogeneity insurance controlling for information on the overall economy within FOMC minutes.

$$MPS_t = \sum_{i=1}^{12} \beta_i MPS_{t-i} + MPS_t^{ac}. \tag{2}$$

To obtain monthly shocks, I accumulate the series and set the shock to zero for months in which there were no meetings (Romer and Romer, 2004). Following Miranda-Agrippino and Ricco (2017), the monthly shock series is then regressed on twelve lags (i.e. one year) to control for autocorrelation. The new shock series, free of autocorrelation, is the residual, MPS^{ac} resulting from regression (2).

Figure 3: Comparing the Target Rate to the Text Shock Measure



Note: This figure compares the new shock measure (solid-blue line) to the change in the policy instrument (dashed-green line). The comparison is on meeting-level frequency.

Figure 3 shows the resulting shock series plotted against the change in the target federal funds rate. For illustrative reasons, Figure 3 shows the target and shock series on the meeting level frequency to avoid plotting zeros whenever there was no meeting. The shock series remains similar to the changes in the target rate, which is unsurprising given an explanatory power of the empirical Taylor rule of approximately 37%. Notably, the differences between the two series increase whenever there were comparatively large changes in the target rate. This shows that the model does well in predicting small changes around the policy instrument but that larger jumps will be interpreted as monetary policy shocks.

4.2 Is the shock forecastable?

Having derived the new MPS series, it remains to show that it is exogenous and, thus, a valid shock measure. Following Miranda-Agrippino and Ricco (2017), the exogeneity assumption can be analysed by testing whether it is predictable (1) by the Greenbook forecasts of the FOMC used in Romer and Romer (2004) and (2) by macroeconomic variables from the FRED database (McCracken and Ng, 2015).

The first forecastability check regresses the shock series on the Greenbook forecasts used in Romer and Romer (2004) and shows that there is no explanatory power. The Greenbook forecasts can be seen as the information set of the FOMC, and thus, a correlation between the shock and the forecasts might indicate that our shock series contains

Table 3: Forecast Test 1: Greenbook Forecasts

	Text Shock	
	Coefficients	*Standard Error*
Constant	−0.001	0.064
Forecasted Inflation		
Quarters Ahead:		
-1	0.016	0.017
0	0.022	0.020
1	−0.017	0.035
2	0.003	0.037
Change in Forecasted Inflation		
Quarters Ahead:		
-1	−0.011	0.025
0	−0.041	0.028
1	0.045	0.048
2	0.037	0.056
Forecasted Output Growth		
Quarters Ahead:		
-1	−0.005	0.009
0	0.018	0.015
1	0.015	0.021
2	−0.009	0.021
Change in Forecasted Output Growth		
Quarters Ahead:		
-1	0.004	0.016
0	0.030	0.019
1	−0.020	0.023
2	0.036	0.024
Forecasted Unemp. Rate (Current Quarter)	−0.018	0.013

* ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.001$)
Adjusted $R^2 = 0.058$, N= 200

Note: The new shock series is regressed on Greenbook forecasts used in Romer and Romer (2004). It is shown that the Greenbook variables have no explanatory power and thus, the shock series is also orthogonal to the information set the central bank possesses at the time of setting the new target rate.

anticipatory movements from the central bank. Table 3, containing the regression results, shows that none of the Greenbook forecasts is significant. Furthermore, the adjusted R-squared is low at approximately 6%. Hence, the new shock series is orthogonal to the information set of the central bank, which is necessary for a valid exogenous shock series (Romer and Romer, 2004).

Table 4: Forecast Test 2: Macroeconomic Factors

	Text Shock	
	Coefficients	*Standard Error*
Constant	0.607	0.379
Lagged Factors:		
F1	0.039	0.063
F2	0.058	0.035
F3	−0.085	0.060
F4	−0.117**	0.042
F5	−0.051	0.046
F6	−0.001	0.050
F7	−0.090	0.057
F8	0.065	0.044
Lagged Controls in Logs:		
Ind. Prod. (1)	0.025	0.047
Inflation (1)	0.015	0.046
Unemp. (1)	−0.008	0.064
Ind. Prod. (2)	−0.028	0.047
Inflation (2)	−0.014	0.046
Unemp. (2)	−0.020	0.063

* ($p < 0.05$), ** ($p < 0.01$)
Adjusted $R^2 = 0.059$, N= 297

Note: This table shows the regression of the new shock series on factors from the FRED monthly macroeconomic variable database from 1983 – 2007.

The second forecastability check regresses the shock series on macroeconomic factors summarising a large set of monthly variables from the FRED database. A completely exogenous monetary policy shock should be unpredictable by any information, and regressing the new shock series on the information contained within the FRED database is a strong check for this (Miranda-Agrippino and Ricco, 2017). Nonetheless, it is to note that the shock constructed here has to be solely exogenous to the variables in question. I follow McCracken and Ng (2015) in summarising the variability of the whole FRED database using principal component analysis. The analysis suggests summarising the 123 monthly macroeconomic variables available for $1983 - 2007$ into eight factors. Additionally, I add controls that will later be used in the impulse response exercise.

Regressing the text shock on the eight components from the PCA in Table 4 reveals that factor four is significant and hence, correlates with the new shock series. This factor is related to different Treasuries (see Appendix F, which shows the six most important macroeconomic variables for each of the eight factors). Nevertheless, for two reasons, the problem of a significant factor might not be too severe. Firstly, the shock was constructed to be orthogonal only to the key variables - output, inflation and unemployment - leaving as much variation from economic variables in the series as possible. Hence, it is unsurprising that some factors have explanatory power since we did not control for every possible macroeconomic variable. Therefore, the significance of some factors is not problematic as long as this does not entail anticipatory movements concerning the three key variables. Moreover, as shown before, the shock series cannot be anticipated by the Greenbook forecasts. Secondly, the overall model only explains 5.9% and thus, has still low explanatory power.

5. Impulse Response Functions of the New Shock Series

This section shows that the new shock series produces IRFs in line with economic theory for the post-1983 period. I first specify the local projection regressions used to estimate IRFs. Following this, I present the IRFs focusing on the post-1983-period where conventional MPS measures produce counter theoretical results.

LP-IV is used to estimate IRFs for two reasons. Firstly, Ramey (2016) used local projections to reveal that the Romer and Romer (2004) shock yields opposite results when estimated on the 1983 - 2007 period as opposed to the original 1969 - 1996 period. Using a similar methodology facilitates a direct comparison focusing on the new shock series rather than using a different estimation technique. Secondly, the reason for extending the Ramey (2016) methodology with an IV part is that using the new shock series as an instrument rather than the actual shock series makes the estimation robust to measurement error (Stock and Watson, 2012; Plagborg-Møller and Wolf, 2021). Additionally, Plagborg-Møller and Wolf (2021) show that using a valid instrument allows disentangling the fundamental shock from the potential linear combination of other shocks. Hence, LP-IV estimates the correct (relative) IRFs even if the true monetary policy shock is not invertible - that is, even when the true shock cannot be recovered by lags and leads of observable macroeconomic data (Stock and Watson, 2018; Plagborg-Møller and Wolf, 2021).

I specify local projections following Ramey (2016) but this time extending it by using the new shock series as an instrument rather than directly in the projection. Similar to the usual IV-estimation, the instrument has to satisfy two restrictions. Borrowing from Stock and Watson (2018) and letting z_t denote the instrument and ϵ_j the shock of interest and ϵ_{-j} all other shocks, the two restrictions are:

1. $Cov(z_t, \epsilon_{j,t}) \neq 0$ and $Cov(z_t, \epsilon_{j,s}) = 0$, $\forall s \neq t$.
2. $Cov(z_t, \epsilon_{-j,s}) = 0$, $\forall s$.

Condition one states that the instrument has to be relevant, i.e. correlated with the true shock we are after. However, the instrument has to be unrelated to any leads or lags of the shock. The second condition can be interpreted as the exclusion restriction stating that the instrument is not related to any other shocks and neither their lags nor their leads.

The estimation can then be written in two steps for convenience[13]

$$ffr_t = \alpha + \beta_1 shock_t + \beta_2 Controls + v_t \tag{3}$$

$$y_{t+h} = \alpha_h + \theta_h \hat{ffr}_t + \gamma_h Controls + \epsilon_{t+h} \tag{4}$$

where Equation (3) is the first stage regressing the federal funds rate on the instrument and the other exogenous control variables. The fitted values from this equation are then used in the second stage, Equation (4). In Equation (4), y is the variable of interest such as the log of industrial production, log of the GDP deflator, the unemployment rate or the federal funds rate. Controls include two lags of all endogenous variables. Then, θ_h is the estimate of the impulse response of the dependent variable at horizon h to a shock at time t. The IRF is then just the point estimates at each horizon, h.

[13]An alternative computation to see why the LP-IV computes relative IRFs is to write the first stage as in Equation (3) but in the second stage estimate the reduced form as $y_{t+h} = \alpha_h + \theta_h shock_t + \gamma_h Controls + \epsilon_{t+h}$. The impulse response coefficient θ_h^{IV} would then be $\theta_h^{IV} = \frac{\theta_h}{\beta_1}$ (Plagborg-Møller and Wolf, 2021; Stock and Watson, 2018).

5.1 Results

Having specified the local projection, we can now analyse the resulting IRFs for the post-1983 period. Figure 4 shows the IRFs to the new text shock series in Panel 4a and to the original Romer and Romer shock re-estimated on $1983 - 2007$ in Panel 4b. Both impacts of the shock series are scaled to increase the federal fund rate by one percentage point. The grey areas are 95% confidence intervals, and the light blue areas are 68% confidence intervals.

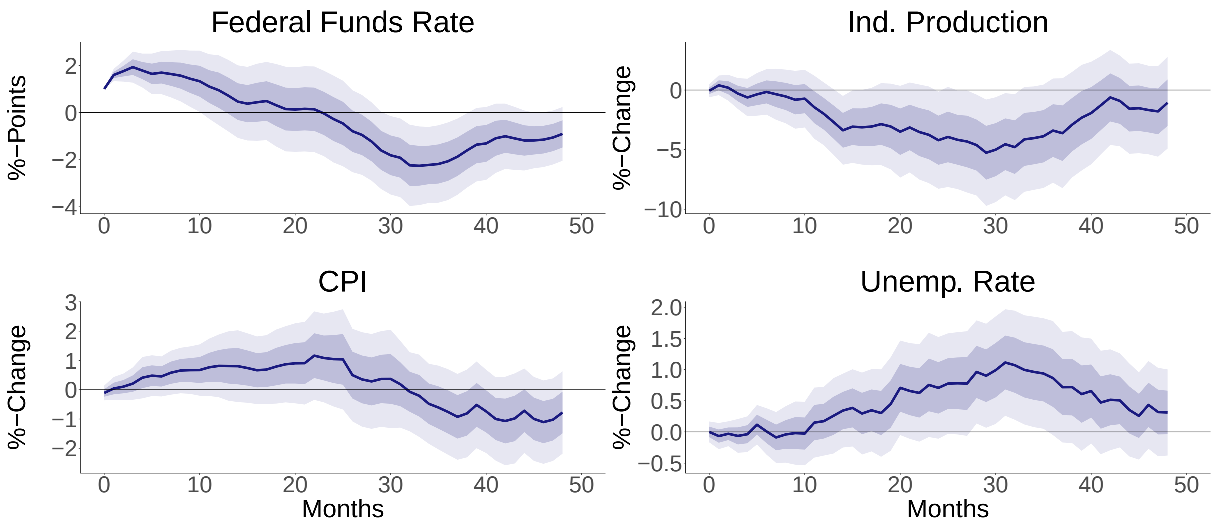
Starting with the Romer and Romer shock in Panel 4b, it can be observed that in response to a contractionary monetary policy shock, industrial production increases significantly and starts returning to its original level after around 42 months. This is a troubling result since we expect contractionary monetary policy actions to have contractionary effects on the economy (Romer and Romer, 2004; Ramey, 2016). Similarly, in response to a contractionary monetary policy shock, the unemployment rate decreases. These puzzling results have been found by others such as Ramey (2016), and Barakchian and Crowe (2013) for the Romer and Romer shocks, and VAR identified shocks such as in Christiano et al. (1998). More recent methods using high-frequency approaches (Gertler and Karadi, 2015) have also been shown to lead to counter theoretical results (Miranda-Agrippino and Ricco, 2017).

Panel 4a shows that contrary to the Romer and Romer shock, the new text shock series produces IRFs that are back in line with economic theory. Panel 4a, top-right shows that in response to a one percentage point increase in the federal funds rate, industrial production declines and unemployment increases. In other words, a contractionary monetary policy shock is again contractionary.

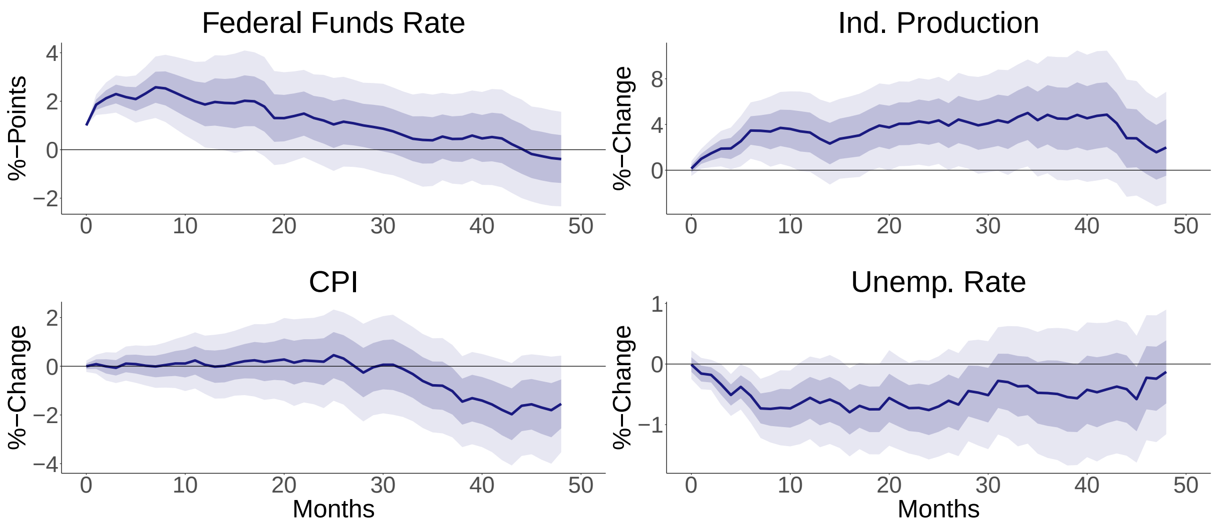
Indeed, the results appear similar to what Romer and Romer (2004) found for the 1969 - 1996 period (see Figure 6b) and are at the larger end to that what is found in other studies (Miranda-Agrippino and Ricco, 2017; Jarociński and Karadi, 2020). For example, the trough for the log of industrial production is at approximately 5%. As a comparison, Miranda-Agrippino and Ricco (2017) and Jarociński and Karadi (2020) find lower effects around $1.6 - 2\%$ while for their original time-period (1969 - 1996) Romer and Romer (2004) found a maximum trough of 4.3%. Similarly, the peak of unemployment at 0.8% is larger than the approximately 0.2% that Miranda-Agrippino and Ricco (2017) find but in line with previous findings by Romer and Romer (2004).

The robustness section converts this LP-IV exercise into a hybrid SVAR thereby showing, that the estimates decrease close to the results found by Miranda-Agrippino and Ricco (2017) and Jarociński and Karadi (2020). As will be discussed, this indicates that the

Figure 4: Impulse Response Functions for 1983 - 2007



(a) Text Shock



(b) Romer and Romer Shock

Note: This figure shows IRFs to the new text shock series (Panel 4a) and the original Romer and Romer shock (Panel 4b). The light grey areas are 95% and the light blue areas are 68% confidence intervals. Standard errors are calculated using 2SLS-Newey-West standard errors. Both shocks are identified on the 1983 − 2007 period.

larger magnitude is mainly due to the estimation method used.

Analysing the effect of monetary policy on inflation[14] in the bottom-left panel of Figure 4a reveals a price puzzle that does not appear when using the Romer and Romer shock for the same period (Figure 4b, bottom-left panel). Several studies suggest including commodity prices to overcome the problem of price puzzles (Bernanke and Ilian, 1998; Bernanke et al., 2005). Unfortunately, this does not alleviate the problem, and the price puzzle remains.

Interestingly, using the original Romer and Romer shock in the here presented LP-IV approach also reveals a price puzzle for the 1969 - 1996 period that is similar in magnitude to Figure 4a (see Figure 6). However, in their SVAR analysis, Romer and Romer (2004) do not find a price puzzle, which could suggest that the price puzzle found here might be due to the LP-IV methodology used. Indeed, repeating the exercise in a hybrid SVAR study (rather than LP-IV) does improve the prize puzzle suggesting that the differences are due to methodology (see section 6.2).

Another possible explanation for the price puzzle might be that the text measure for inflation is not capturing all the information on the development of prices. Table 2 in the previous section showed that the directional sentiment on inflation is insignificant, which could suggest that the text measure does not adequately capture all the information on prices. Extending the dictionary on inflation with other price-related words does not solve the price puzzle. However, given that Romer and Romer (2004), using Greenbook forecasts on inflation similarly find a price puzzle might indicate that more price-related information is needed to overcome the price puzzle.

Another difference can be observed in the dynamics of the federal funds rate in Figure 4. In Panel 4a, the federal funds rate becomes negative after around 25 months and then drops to roughly offset the initial increase. A similar, though slightly lower in magnitude, finding is presented by Miranda-Agrippino and Ricco (2017) and Jarociński and Karadi (2020). In response to their shock measures, the one-year treasury bond rate increases on impact and becomes negative after 12 horizons for Miranda-Agrippino and Ricco (2017) and after 20 horizons for Jarociński and Karadi (2020). The trough is reached shortly after that and remains at roughly negative 20% and 50%, respectively of their impact measure of the one-year treasury bond rate.

To conclude on the central question of this paper, Figure 4 shows that controlling for the information set of private agents rather than the information set of the central bank

[14]For reasons of comparability, I have used the CPI here. This could be problematic because, until the 1980s, interest rates entered directly into the index. Using personal consumption expenditure (implicit deflator) instead does not change the results (see Appendix E).

results in monetary policy shocks that produce IRFs that are back in line with what theory would predict. This supports the hypothesis that the information channel is responsible for the fact that monetary shock series that do not fully take account of the information channel lead to counter-theory results.

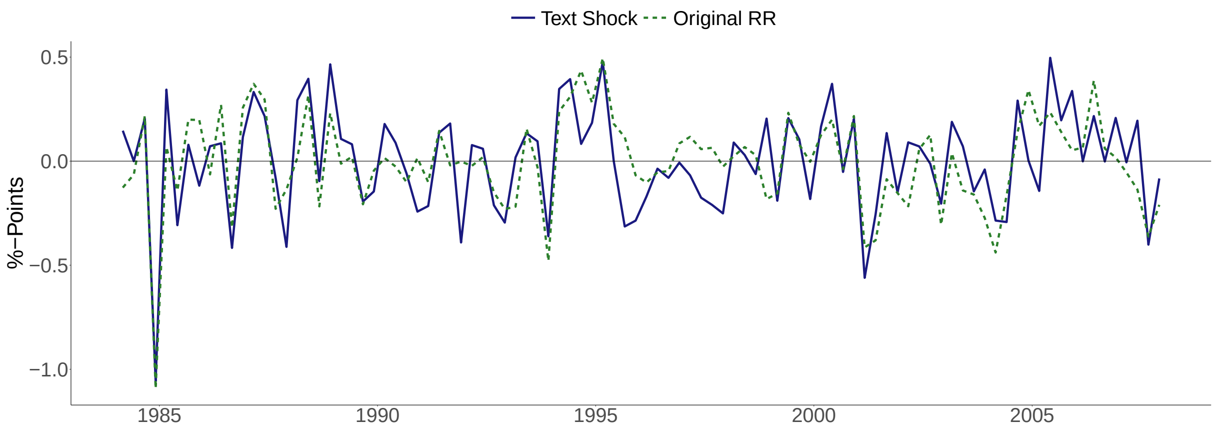
Nevertheless, this also raises the question of why the information channel did not lead to problems in earlier periods. A possible reason for this might be that approximately until after the Volcker period, shocks were still large and monetary policy setting was not as systematic as it is today, so that private agents did not interpret surprising changes in the target rate as a signal extraction problem (Ramey, 2016, 2018). To test this, Appendix D estimates the Romer and Romer (2004) shocks for the original time period as well as for 1983 - 2007, excluding the Volcker period. Repeating the Romer and Romer (2004) exercise for different periods reveals that the Greenbook forecasts only explain 22% of the change in the federal funds target rate for $1969 - 1996$ as compared to 44% in the post-1983 period. The doubling in the explanatory power of the Greenbook forecasts is evidence that monetary policy has become more systematic in recent years. Furthermore, since the model explains more of the movements in the target rate, less volatility remains in the error term, and MPS become smaller.

The hypothesis that pre-1983 monetary policy shocks were more similar for the central bank and private agents than in the later period is further supported by comparing the Romer and Romer and the new text shock series in Figure 5 (for illustration purposes, I aggregated the shocks to quarterly frequency). Panel 5a depicts the two shock series for the post-Volcker time-period of $1983 - 2007$ and Panel 5b for the original Romer and Romer period of $1969 - 1996$. It appears that the earlier shock series are more closely related. This is further evidence that it is more important to control for the private agent information set for more recent periods. These observations are confirmed when calculating the correlation coefficients. For the original Romer and Romer period of 1969 - 1996, the correlation coefficient between the Romer and Romer shock and the new text shock series is 0.828. This correlation is lower for the more recent period and estimated to be around 0.715. Hence, it has become more important to take the information channel into account in recent years since central bank shocks differ from private agent shocks.

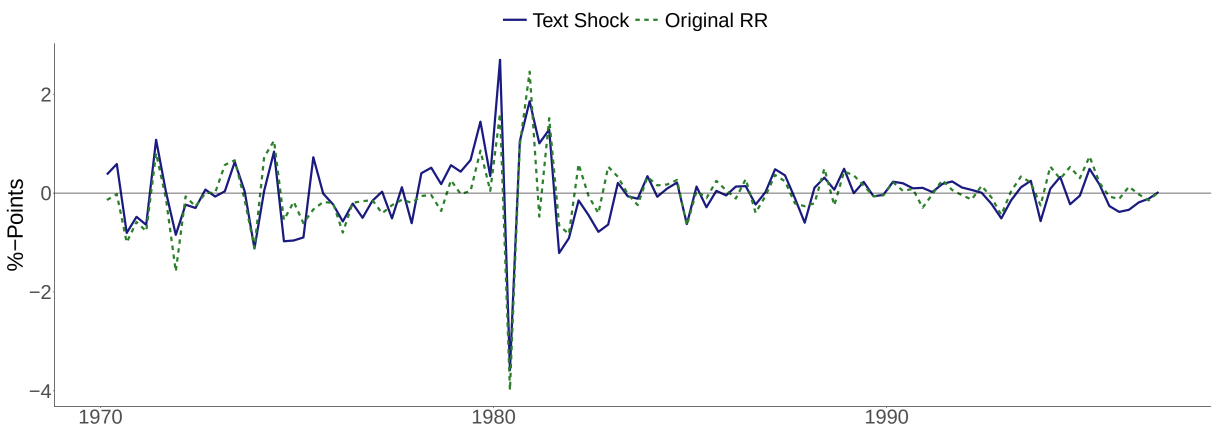
6. Robustness Checks

This section executes several robustness checks showing that the new shock series is robust to changes in the estimation period, estimation technique, and the methodology

Figure 5: Comparing the original Romer and Romer to the Text Shock



(a)



(b)

Note: Panel 5a, 1983 − 2007 (post-Volcker period): Shock comparison of the Romer and Romer shock and the new text shock series. The correlation coefficient is 0.715. Panel 5b, 1969 − 1996 (original Romer and Romer period): Shock comparison of the original Romer and Romer shock to the new text shock series. The correlation coefficient is 0.828.

of quantifying FOMC minutes.

6.1 IRFs for the original Romer and Romer Period - 1969-1996

Repeating the identification exercise above between 1969 and, 1996 show that the IRFs are robust to changes in periods. Different to other recent monetary policy shocks (Miranda-Agrippino and Ricco, 2017; Gertler and Karadi, 2015; Jarociński and Karadi, 2020) the identification method presented here allows to derive shock series for earlier periods. This is because most current methods rely on high-frequency data, which is only available from the 1990s, whereas the FOMC minutes are available in slightly changing form since 1967 (then Minutes of Action).

Figure 6 shows the IRFs of the new text shock series in Panel 6a for 1969 - 1996 and compares them to the original Romer and Romer shock in Panel 6b. Two observations can be made from this comparison. Firstly, the IRFs in both panels look very similar. The similarity indicates that the text shock identification is robust to changes in the time period. Similar to what Romer and Romer (2004) find for $1969 - 1996$, industrial production (top-right graph) as well as unemployment (bottom-left graph) both show an initial counter theoretical increase and decrease, respectively. After approximately five to ten horizons, industrial production becomes negative and unemployment positive, as we would expect. The magnitudes of the effect of a shock that increases the federal funds rate by one percentage point appear to be similar.

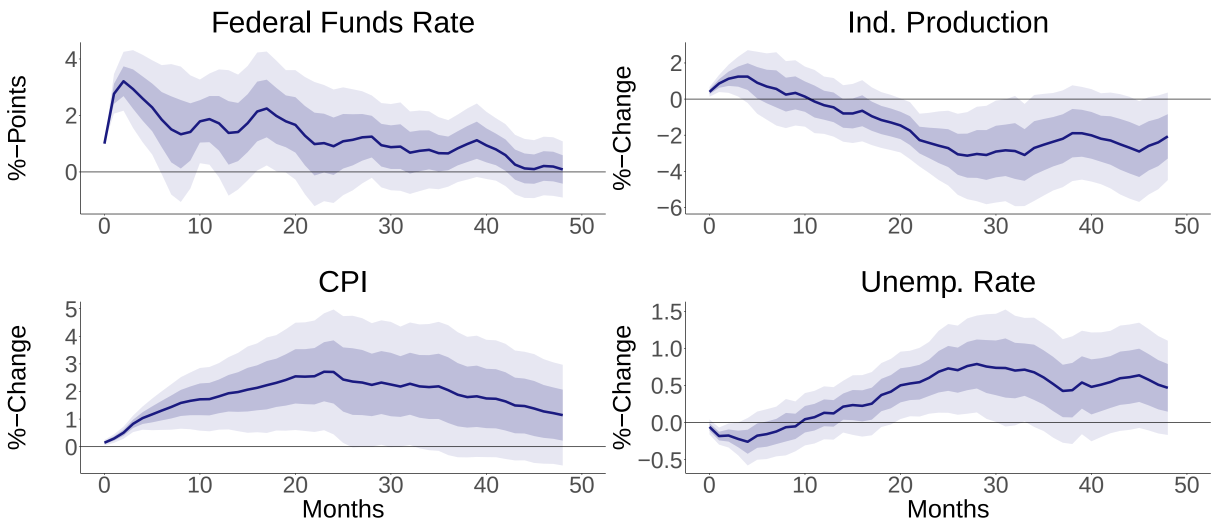
Secondly, the bottom-left graphs in figure 6 show a pronounced price puzzle for both IRFs. Adding commodity prices as controls does improve the price puzzle for the Romer and Romer IRFs somewhat but has almost no effect on the text shock IRFs. Similar to the discussion for the 1983 - 2007 period, a possible reason for the price puzzle with the text shock might be that the directional sentiment measure is the only text measure that is not significant (the directional sentiment on unemployment is significant at the 10% level). Hence, one might need to add other information on prices to overcome this puzzle.

6.2 Estimation Methodology: Hybrid SVARs

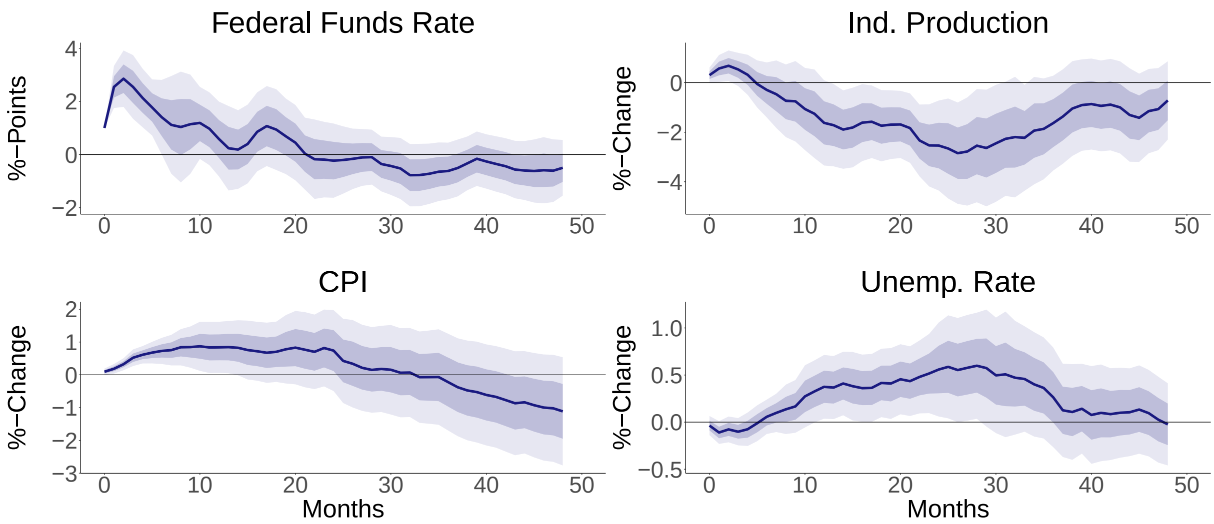
Another question that often arises in empirical studies is whether to use a VAR or LP estimation technique. I compare the LP-IV technique to the VAR literature by using a hybrid VAR and endogenously including the external shock series.

Plagborg-Møller and Wolf (2021) show that the VAR equivalent of an LP-IV estimation is to order the instrument first in a VAR. The resulting VAR model can therefore be

Figure 6: Impulse Responses 1969 – 1996



(a) Text Shock



(b) Romer and Romer Shock

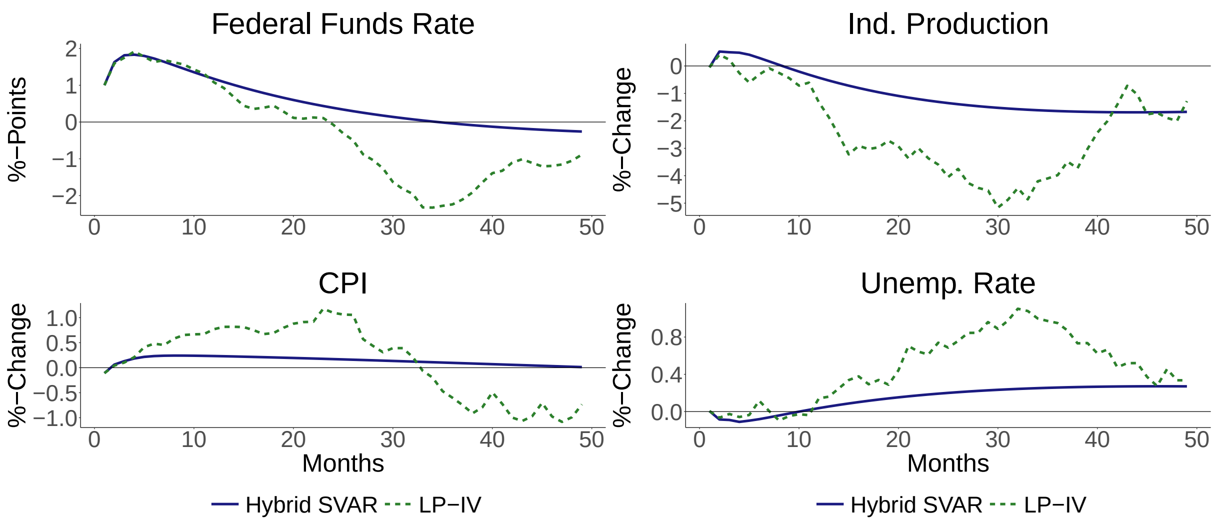
Note: This figure shows the IRFs to the original Romer and Romer shock (Panel 6b) and the new text shock series (Panel 6a). The light grey area is the 95% and the light blue area the 68% confidence intervals. Standard errors are calculates using 2SLS-Newey-West standard errors. Both shocks are identified on the 1969 – 1996 period.

written as:

$$\vec{y}_t = \sum_{j=1}^{2} \mathbf{F}_j \vec{y}_{t-j} + \vec{u}_t \tag{5}$$

where $\vec{y}$ contains the text shock, industrial production, inflation, unemployment, and federal funds rate in that order, and u_t are reduced form residuals. The lag length is two, equal to the LP-IV specification.

Figure 7: Impulse Responses using Two Different Methodologies



Note: This figure shows the IRFs to the new shock series using two different methodologies. The solid blue line depicts the IRFs identified using a hybrid VAR ordering the shock instrument first. The dashed-green line is the IRFs estimated using LP-IV. Both IRFs are estimated for the time period of 1983 – 2007.

Figure 7 shows that the IRFs resulting from the hybrid VAR are consistently smaller than the LP-IV IRFs after few horizons. The similarity for short horizons would be expected according to Plagborg-Møller and Wolf (2021) who show that LPs and VARs produce the same IRFs up to the maximum lag number used. After a few horizons, the difference between the two IRFs becomes greater. The results from the hybrid VAR appear very similar in magnitude to the findings in Miranda-Agrippino and Ricco (2017), and Jarociński and Karadi (2020) who both use a (Proxy) SVAR approach for their estimation.

The reason for this difference in magnitude might be due to a possible persistence of the shock measure. Alloza et al. (2019) show that due to the forward iteration in LPs, shock persistence biases the IRF point estimates upwards even when the shock was corrected for auto-correlation (i.e. regressing the shock on its lags). This is because by iterating forward, the correlation between future shock measures and the shock measure at t affects the point estimates, and this possible correlation has not been controlled for by regressing the shock measure at t on its lags. However, using the Box-Pierce test for serial correlation as Alloza et al. (2019) suggest, the null hypothesis of no serial correlation cannot be rejected (p-value = 0.177). Similarly, the results remain unchanged when including leads of the instrument in the estimation of the IRFs with the LP-IV approach to account for possible shock persistence. Hence, there does not seem to be persistence in the shock measure, and the difference appears to be due to estimation techniques.

6.3 Topic Model on FOMC Minutes

To test whether the methodology of converting textual data to quantitative variables affects the derivation of the new shock series, I apply a correlated topic model to find latent topics within FOMC minutes and repeat the identification exercise using these topics instead of the directional sentiment measures.

One of the drawbacks of using a dictionary approach is that the researcher has to determine the words for the dictionary. Overcoming this subjective input, the correlated topic model (Blei et al., 2003; Blei and Lafferty, 2007), a so-called unsupervised machine learning algorithm, has been developed, finding the latent topics within a text without requiring the researcher to label the data set. I delegate the interested reader for a detailed description of the methodology and processing steps to Appendix H and here only sketch a brief outline to provide sufficient understanding.

A correlated topic model provides the researcher with a probabilistic topic distribution over all FOMC meeting minutes. The only parameter that has to be provided by the researcher is the number of topics the algorithm should look for in each document. Then, for each of the FOMC minutes, the algorithm provides a probability of a specific topic. For example, given that we specified a topic model with $K = 10$ topics, the model's output would be of the following form. Topic 1 appears in FOMC meeting A with 20% probability, topic 2 with 75% and the remaining eight topics appear with low probabilities. We observe these probabilities for each FOMC meeting and obtain a time series of topic distributions that can control for the information contained within FOMC minutes.

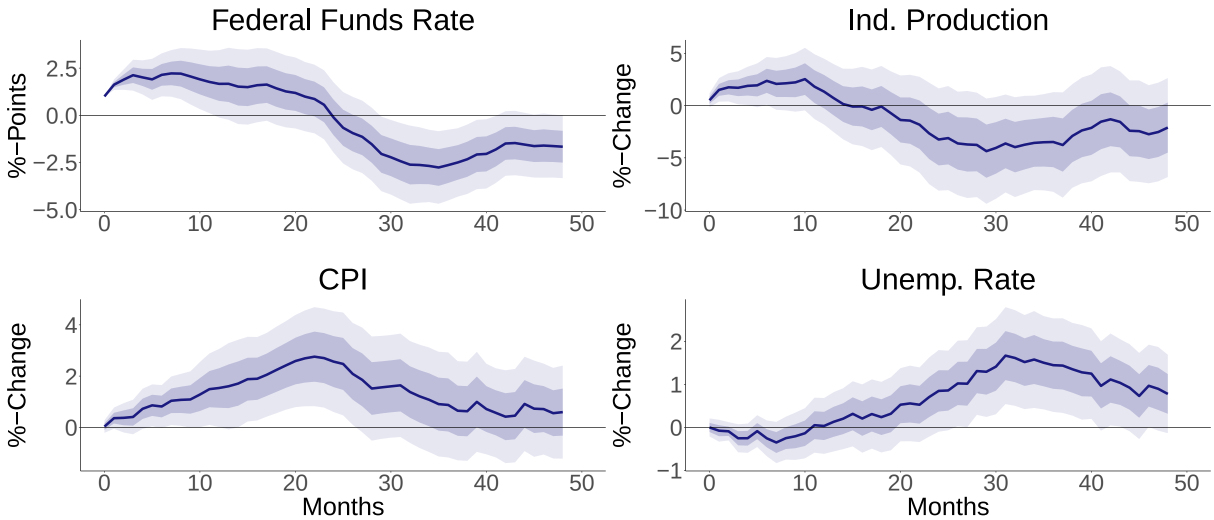
The algorithm does not label the topics themselves. However, each topic contains a probability distribution over the whole vocabulary of the text corpus (i.e. all FOMC minutes). The probability attached to each word within a topic allows the researcher to give intuitive labels to each topic. For example, if the word "inflation" or "energy prices" have high probabilities within topic A, then the researcher could label this topic "inflation". Those labels have no impact on the empirical analysis but are provided for convenience and tentative interpretability.

Loosely speaking, a topic receives a high probability within a document when its high probability terms appear often. This is very similar to the sentiment approach, where I count how often a word within a dictionary appears in a text. The difference is that the topics or word groups were found by a data-driven algorithm and not by the researcher. On their own, the topics do not carry any directional information but rather measure how much the FOMC talked about a particular topic. If, however, the topic of employment increases whenever employment is high or low, we would not be able to disentangle the direction of the topics. For this reason, I follow an approach by Hansen and McMahon (2016) who created directional topics by counting directional words like "increase" or "decrease" within each topic and obtain a directional word count for each topic.

Since the algorithm does not provide the topic labels, selecting the topics to be included in the regression exercise has to be done with another approach. Again, I follow Hansen and McMahon (2016) and conduct a bootstrap LASSO analysis for the model selection. The bootstrap entails repeating LASSO selection N times and keeping only those variables that have been selected more than a certain threshold. Here, I keep all variables that are chosen more often than 65% of the time. However, moving this threshold up or down does not alter the results. After the selection process, I then regress the target rate change on the selected topics and obtain the remaining residual as the monetary policy shock on the meeting level. This shock series is cumulated to the monthly level and regressed on twelve of its lags to control for autocorrelation. This shock series is then used in the same LP-IV specification as shown previously to estimate IRFs.

The resulting IRFs are presented in Figure 8 and show that the identification procedure is robust to the methodology used to quantify textual data. Two main observations can result from Figure 8. Firstly, the dynamics remain similar to using directional sentiment measures instead of topics from a topic model. The similarity confirms that the identification procedure is robust to the methodology used to transform textual data into quantitative variables. Secondly, the effects of the topic shock series appear to be similar in magnitude overall. Note that using dictionary methods above allows controlling for the "topics" to be searched for in the text and included in the regression. The topic model approach does not provide such a clear distinction. Hence, a selection approach for the topics was needed. However, the selection was based on which topics best explain the target rate. Thus, less variation was left in the error term than when one only controls for output, inflation, and employment, as was done above. Although more information was purged from the target rate in this process, both the dynamics and magnitudes are similar to the more straightforward text analysis method.

Figure 8: Impulse Responses to Topic Shock



Note: This figure shows the IRFs to the new shock series using topics from a correlated topic model as controls for the private agent information set. The light grey areas are 95% and the light blue areas are 68% confidence intervals. Standard errors are calculates using 2SLS-Newey-West standard errors. Both shocks are identified on the 1983 − 2007 period.

7. Conclusion

This paper derived monetary policy shock series free from an information effect that allows estimating IRFs consistent with monetary theory for recent periods. It was argued that unanticipated changes in the policy rate might contain information for private agents on the central banks' economic projections. Not controlling for this so-called information channel can lead to unexpected results in IRFs.

To overcome the information channel problem, this paper used natural language processing techniques to obtain text measures of the informational content of FOMC minutes. FOMC minutes contain public information on discussions and policy decisions during FOMC meetings. As such, the extraction of information from the FOMC minutes can be interpreted as a proxy for the signal extraction problem and the private agents' information set.

A directional sentiment approach allows extracting the information contained within FOMC minutes and turning the unstructured data into quantifiable variables. More precisely, three sentiment dictionaries were constructed containing keywords to describe output, inflation and unemployment. These keyword counts were then combined with directional terms such as ”increase” or ”decrease” to assess the direction of the sentiment. Furthermore, to augment this analysis, an overall sentiment measure was constructed for each meeting minutes capturing the general sentiment on the economy.

To construct the new MPS series, changes in the target rate were regressed on the directional and overall sentiment. Using these text measures as controls for the private agent information set, anything not explained by these variables could be interpreted as surprise movements in the target rate. Hence, the remaining residual of this regression was interpreted as the new MPS.

The new MPS series is then used in an LP-IV approach to estimate IRFs. Contrary to more conventional shock measures (Romer and Romer, 2004) and high-frequency approaches (Gertler and Karadi, 2015), the estimated IRFs produce results that are back in line with what theory would predict. A contractionary policy shock leads to a contraction in the economy. Namely, industrial production decreases, and unemployment goes up. However, the price puzzle remains.

Another advantage of the new shock identification strategy is that it allows for the analysis of earlier periods. It was shown that monetary policy was less systematic for earlier periods, and thus, private agents interpreted surprise movements in the target rate not as signals but as true monetary policy shocks. Hence, the information channel is less critical for earlier time periods.

References

Alloza, Mario, Jesus Gonzalo, and Carlos Sanz, "Dynamic Effects of Persistent Shocks," *SSRN Electronic Journal*, 2019.

Apel, Mikael and Marianna Blix Grimaldi, "How Informative Are Central Bank Minutes?," *Review of Economics*, January 2014, *65* (1).

Barakchian, S. Mahdi and Christopher Crowe, "Monetary policy matters: Evidence from new shocks data," *Journal of Monetary Economics*, 2013, *60* (8), 950–966.

Bernanke, Ben S and Mihov Ilian, "Measuring Monetary Policy," *Quarterly Journal of Economics*, 1998, *113* (3), 35.

_ , Jean Boivin, and Piotr Eliasz, "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *QUARTERLY JOURNAL OF ECONOMICS*, 2005, p. 36.

Blei, David M. and John D. Lafferty, "A correlated topic model of Science," *The Annals of Applied Statistics*, 2007, *1* (1), 17–35.

Blei, David M, Andrew Y. Ng, and Michael I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 2003, *3*, 993–1022.

Boukus, Ellyn and Joshua V. Rosenberg, "The Information Content of FOMC Minutes," *SSRN Electronic Journal*, 2006.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans, "Monetary Policy Shocks: What Have We Learned and to What End?," Technical Report w6400, National Bureau of Economic Research, Cambridge, MA 1998.

Cochrane, John, "Comments on "A new measure of monetary shocks: Derivation and implications"," *NBER EFG meeting*, 2004.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia, "Innocent Bystanders? Monetary policy and inequality," *Journal of Monetary Economics*, 2017, *88*, 70–89.

Dybowski, T.P. and P. Adämmer, "The economic effects of U.S. presidential tax communication: Evidence from a correlated topic model," *European Journal of Political Economy*, 2018, *55*, 511–525.

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

Handlan, Amy, “Text Shocks and Monetary Surprises:,” *Working Paper*, 2021, pp. 1–65.

Hansen, Stephen and Michael McMahon, “First Impressions Matter: Signalling as a Source of Policy Dynamics,” *The Review of Economic Studies*, 2016, *83* (4), 1645–1672.

— , — , and Andrea Prat, “Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach*,” *The Quarterly Journal of Economics*, 2018, *133* (2), 801–870. Read influence measures.

Hu, Minqing and Bing Liu, “Mining and Summarizing Customer Reviews,” in “SIGKDD” KDD04, New York: Association for Computing Machinery, 2004, pp. 168–177.

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

Jegadeesh, Narasimhan and Di (Andrew) Wu, “Deciphering Fedspeak: The Information Content of FOMC Meetings,” *SSRN Electronic Journal*, 2017.

Loughran, Tim and Bill Mcdonald, “When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks,” *The Journal of Finance*, 2011, *66* (1), 35–65.

McCracken, Michael W and Serena Ng, “FRED-MD: A Monthly Database for Macroeconomic Research,” 2015, p. 31.

Miranda-Agrippino, Silvia and Giovanni Ricco, “The Transmission of Monetary Policy Shocks,” *CEPR Discussion Paper*, 2017, (13396).

Mueller, Hannes and Christopher Rauh, “Reading Between the Lines: Prediction of Political Violence Using Newspaper Text,” *American Political Science Review*, 2018, *112* (2), 358–375.

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

Plagborg-Møller, Mikkel and Christian K Wolf, “Local Projections and VARs Estimate the Same Impulse Responses,” *Econometrica*, 2021, p. 36.

Ramey, V. A., “Chapter 2 - Macroeconomic Shocks and Their Propagation,” in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Amsterdam: Elsevier, 2016, pp. 71–162.

Ramey, Valerie A, “Discussion of: The Transmission of Monetary Policy Shocks (by S. Miranda-Agrippino and G. Ricco),” ASSA Meetings 2018.

Rinker, Tyler, "sentimentr: Calculate Text Polarity Sentiment," 2019.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley, "Package for Structural Topic Models," *Journal of Statistical Software*, 2019, *91* (2).

_ , _ , and Edoardo M. Airoldi, "A Model of Text for Experimentation in the Social Sciences," *Journal of the American Statistical Association*, 2016, *111* (515), 988–1003.

Romer, Christina D and David H Romer, "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 2000, *90* (3), 29.

_ and _ , "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, 2004, *94* (4), 30.

Shapiro, Adam H., Moritz Sudhof, Kanjoya, Daniel Wilson, and Federal Reserve Bank of San Francisco, "Measuring News Sentiment," *Federal Reserve Bank of San Francisco, Working Paper Series*, January 2017, pp. 01–22.

Stock, James H. and Mark W. Watson, "Disentangling the Channels of the 2007 - 2009 Recession," *Brookings Papers on Economic Activity*, 2012.

_ and _ , "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 2018, *128* (610), 917–948.

Wolf, Christian K., "SVAR (Mis)Identification and the Real Effects of Monetary Policy Shocks," *American Economic Journal: Macroeconomics*, October 2020, *12* (4), 1–32.

Òscar Jordà, Moritz Schularick, and Alan M. Taylor, "Betting the House," *Journal of International Economics*, 2015, *96*, S2–S18.

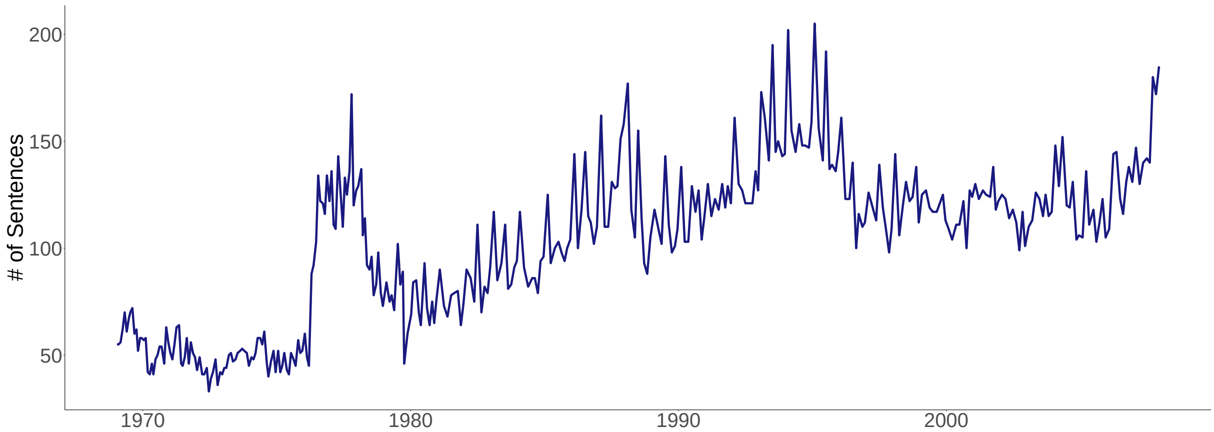
A. Appendix: Pre-Processing of FOMC Minutes for Dictionary Methods

Firstly, I transform all pdfs into computer-readable text and read in each meeting at the sentence level. Secondly, the first part on administrative details is deleted for the analysis following Jegadeesh and Wu (2017), and Boukus and Rosenberg (2006). Thirdly, I correct for some bad read-ins due to the transformation of older pdfs to text files (for example, ”uncertain” to ”uncertain”). Fourthly, I account for contractions transforming ”don’t” to ”do not”. This simplifies the sentiment analysis later on when allowing for negations. Fifthly, the text is lemmatized. Lemmatization is the reduction of words to their word stems. For example, ”going” would be reduced to ”go” or ”meetings” to ”meeting”. This process makes it unnecessary to look for both ”inflation” and ”inflationary”. Finally, I transform some key nouns to their acronyms whenever they were not already in this form to ensure a consistent count of these terms. This entails transforming ”gross domestic product” to ”GDP”, ”gross national product” to ”gnp”, ”producer price index” to ”PPI” and ”consumer price index” to ”cpi”.[15]

Figure 9 shows how the length (measured in the number of sentences) of the FOMC minutes increased over time. Particularly at the beginning of the sample, FOMC minutes were much shorter, containing only around 50 sentences compared to roughly 150 on average for later periods. Since the primary analysis of this paper focuses on the time after the volatile Volcker years from 1983 − 2007, the increasing length of the FOMC minutes appears less problematic. Figure 9 shows that the average length of the more recent minutes was reached from the mid-1980s. Nevertheless, the varying length of the minutes will still necessitate a normalization of the word counts in the dictionary methods used in the next section.

[15]Preparing the text data for the topic model in the robustness section entails further pre-processing steps outlined in appendix H.

Figure 9: Number of Sentences per FOMC Minutes



Note: This figure depicts the development of the length (measured in the number of sentences) of FOMC minutes over time from 1969 – 2007.

B. Appendix: Sentiment Analysis Examples

Below I show several examples of the algorithm's workings to derive the directional sentiment measures for inflation, unemployment and output. This is mainly for illustrative purposes showing the advantages and disadvantages of the classification procedure.

Inflation: An example (not pre-processed) for a sentence classed to be about inflation in the February 6. 1991 FOMC minutes:

> ”With oil **prices** lower and some added slack expected in resource utilization, the staff projected a slowing in the pace of increases in prices and labor costs in coming quarters.”

The word ”prices”, marked in bold, tags the sentence to be about the topic of inflation. The algorithm then searches for directional terms, which are underlined. Here, ”lower” and ”slowing” are counted as *decrease* words and ”increase” as *increase* such that the sentence gets classed as *decrease* with a count of minus one.

Output: An example (not pre-processed) for a sentence classed to be about output in the September 21. 1976 FOMC minutes:

> ”One member questioned whether the strike under way in the automobile industry might *not* have a significantly adverse effect on expansion in aggregate **output**, at least over the near term–although others stated that in the past the bulk of output losses resulting from major strikes had generally tended to be made up within a short period.”

The algorithm marks the sentence to be about output due to the word ”output” appearing in the sentence. Note that the underlined term ”losses” increases the directional term count of decrease words by one. ”Expansion” is not included in the directional term list, only its verb forms, ”expand, expanded, expanding, expands”. This follows Hansen and McMahon (2016). Furthermore, the emphasized negator, ”not”, flips the sign of the sentence such that the sentence is classed as increase with a count of one. This shows some shortcomings of this counting approach since a reader would likely class this sentence as negative/decreasing since it is cautioned against an adverse effect.

Unemployment: An example (not pre-processed) for a sentence classed to be about unemployment in the September 21. 1976 FOMC minutes:

> Service industries continued to post large gains in **employment** in June; however, hiring at retail establishments was markedly slower than earlier in the year.

The sentence above is tagged to be about ”employment”. The directional terms ”gains” and ”slower” counteract each other such that the sentence gets a count of zero.

C. Appendix: Shock Derivation without Overall Sentiment

Table 5: Identification of the Text Shock without Overall Sentiment

	Δ Intended Federal Funds Rate	
	Coefficients	*Standard Error*
Constant	−0.041	0.035
Initial level of Intended Funds Rate	−0.010	0.006
Sentiment Measures:		
Output	0.073***	0.013
Inflation	0.014	0.016
Employment	0.018*	0.007

* ($p < 0.5$), ** ($p < 0.01$), *** ($p < 0.001$)
Adjusted $R^2 = 0.237$, N= 200

Note: This table shows the regression results of a change in the intended Federal Funds rate on sentiment measures but without the overall sentiment. The resulting residual is interpreted as a monetary policy shock.

D. Appendix: Original Romer and Romer Shock Identification

Table 6: Original Romer and Romer Identification

	Δ Target Rate	
	Coefficients	*Standard Error*
Constant	0.171	0.141
Initial level of Target Rate	-0.021^{*}	0.012
Forecasted Inflation		
Quarters Ahead:		
-1	0.021	0.024
0	-0.044	0.029
1	0.010	0.044
2	0.052	0.047
Change in Forecasted Inflation		
Quarters Ahead:		
-1	0.057	0.045
0	0.003	0.048
1	0.031	0.074
2	-0.062	0.081
Forecasted Output Growth		
Quarters Ahead:		
-1	0.007	0.010
0	0.003	0.019
1	0.010	0.032
2	0.022	0.032
Change in Forecasted Output Growth		
Quarters Ahead:		
-1	0.050^{*}	0.030
0	0.152^{***}	0.030
1	0.021	0.046
2	0.021	0.051
Forecasted Unemp. Rate (Current Quarter)	-0.048^{**}	0.021

* $(p < 0.1)$, ** $(p < 0.05)$, *** $(p < 0.001)$
Adjusted $R^2 = 0.229$, N= 263

Note: This table replicates the original Romer and Romer shock identification (1969 – 1996). The intended funds rate is regressed on Greenbook forecasts. The resulting residual is interpreted as a monetary policy shock.

Table 7: Original Romer and Romer Identification

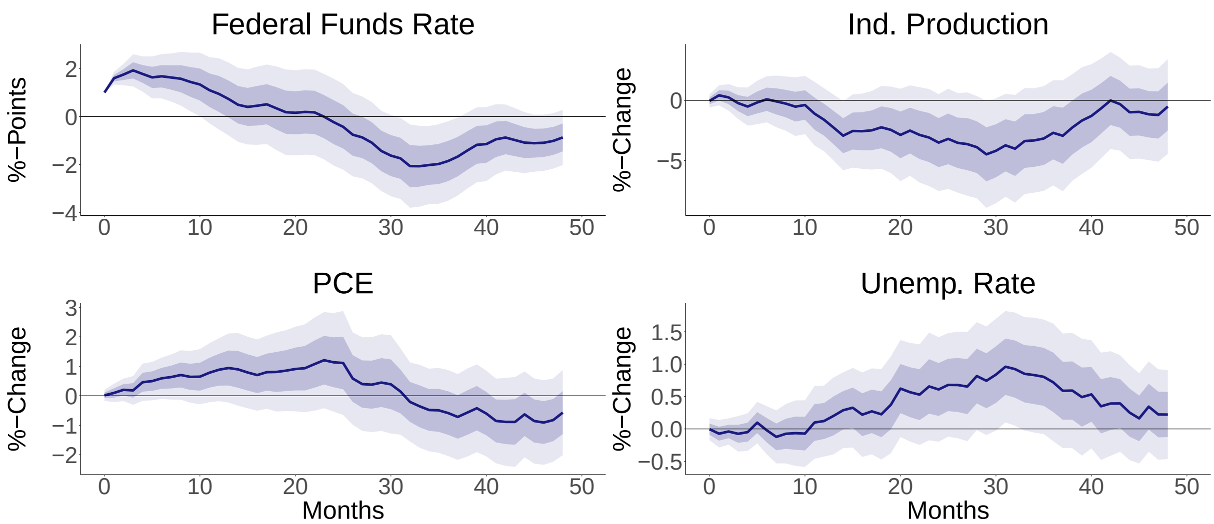
	Δ Target Rate	
	Coefficients	*Standard Error*
Constant	0.071	0.064
Initial level of Intended Funds Rate	-0.057^{***}	0.010
Forecasted Inflation		
Quarters Ahead:		
-1	0.025	0.017
0	0.050^{**}	0.020
1	0.027	0.035
2	0.042	0.038
Change in Forecasted Inflation		
Quarters Ahead:		
-1	0.004	0.025
0	-0.059^{**}	0.028
1	0.033	0.047
2	0.006	0.055
Forecasted Output Growth		
Quarters Ahead:		
-1	0.005	0.009
0	0.055^{***}	0.015
1	0.042^{**}	0.021
2	-0.033	0.021
Change in Forecasted Output Growth		
Quarters Ahead:		
-1	0.011	0.016
0	0.045^{**}	0.019
1	0.008	0.023
2	0.063^{***}	0.024
Forecasted Unemp. Rate (Current Quarter)	-0.054^{***}	0.013

* ($p < 0.1$), ** ($p < 0.05$), *** ($p < 0.001$)
Adjusted $R^2 = 0.442$, N= 200

Note: This table replicates the original Romer and Romer shock identification (1983 – 2007). The intended funds rate is regressed on Greenbook forecasts. The resulting residual is interpreted as a monetary policy shock.

E. Appendix: IRFs using Personal Consumption Expenditure (PCE) 1983 - 2007

Figure 10: Impulse Responses to Text Shock



Note: This figure shows the IRFs to the new text shock series using PCE instead of CPI. The light grey areas are 95% and the light blue areas are 68% confidence intervals. Standard errors are calculates using 2SLS-Newey-West standard errors. Both shocks are identified on the 1983 – 2007 period.

F. Appendix: Description of Factors

Table 8: Factor Description from PCA

Factor	FRED Code	Loading	Factor	FRED Code	Loading
F1	USGOOD	0.661	F5	DSERRG3M086SBEA	0.262
	IPMANSICS	0.628		RETAILx	0.233
	INDPRO	0.610		CMRMTSPLx	0.228
	CUMFNS	0.570		ISRATIOx	0.224
	DMANEMP	0.565		M2SL	0.218
	MANEMP	0.556		TOTRESNS	0.214
F2	CUSR0000SA0L2	0.8620241	F6	AAAFFM	0.288
	CUSR0000SA0L5	0.857		BAAFFM	0.279
	CPIAUCSL	0.852		T10YFFM	0.189
	CUSR0000SAC	0.826		M2SL	0.155
	DNDGRG3M086SBEA	0.812		M1SL	0.145
	CPIULFSL	0.808		EXSZUSx	0.140
F3	PERMITW	0.693	F7	T5YFFM	0.256
	PERMIT	0.664		T10YFFM	0.240
	HOUST	0.648		T1YFFM	0.208
	HOUSTW	0.639		AAAFFM	0.161
	PERMITS	0.507		IPCONGD	0.159
	HOUSTS	0.489		BAAFFM	0.153
F4	TB6SMFFM	0.403	F8	S&P 500	0.537
	GS5	0.394		S&P: indust	0.535
	GS10	0.377		S&P div yield	0.506
	TB3SMFFM	0.373		S&P PE ratio	0.447
	GS1	0.343		UMCSENTx	0.202
	AAAFFM	0.306		W875RX1	0.115

Note: This table shows the six most important variables for each of the eight factors from the principal component analysis of 123 macroeconomic monthly variables from the FRED database.

G. Appendix: Text Shock Identification 1969 - 1996

Table 9: Identification Text Shock 1969 – 1996

	Δ Target Rate	
	Coefficients	*Standard Error*
Constant	-0.200^{**}	0.078
Initial level of Intended Funds Rate	0.0001	0.009
Sentiment Measures:		
Output	0.097^{***}	0.023
Inflation	0.016	0.023
Employment	0.026^{*}	0.013
Overall	0.160^{***}	0.037

* $(p < 0.1)$, ** $(p < 0.05)$, *** $(p < 0.001)$
Adjusted $R^2 = 0.195$, N= 263

Note: This table shows the regression of the change in the intended Federal Funds rate on sentiment measures. The resulting residual is interpreted as a monetary policy shock.

H. Appendix: Correlated Topic Model

This section describes the correlated topic model used for the robustness check. A brief description of the algorithm is provided. Then the data pre-processing is laid out in detail. Following this, the number of topics is determined in a data-driven way. Given this parameter (the number of topics), the topic model is estimated, and the topics are shown. These topics are then transformed into directional topics and finally used in the shock identification procedure.

The strategy of a correlated topic model is to assume a generative model and then estimate the most likely parameter values from the data (Roberts et al., 2019). To describe this assumed data generating process, I will first fix some notation and terminology. The only observables when working with textual data are words within documents. All documents put together make up a corpus. Following convention, let $w_{d,n}$ denote the nth word in document d. The correlated topic model relies on the bag-of-words assumption, meaning that only the word count per document is important, not their position in the text. Hence, we can represent the corpus in a document-term matrix that contains the number of times a unique term is used in every document. Furthermore, a topic is a probability distribution β over all unique terms. A topic contains all unique terms of the entire corpus, and each term has a probability attached to it though most words' probabilities will be close to zero. The terms with the highest probability within a topic are usually used to label topics and cluster the text. Even though a topic contains all unique terms, it is assumed that a word within the text stems from a particular topic. This is often referred to as a topic assignment $z_{d,n}$, where n is the nth word and d the dth document (Blei and Lafferty, 2007). Finally, the key variables for the empirical exercise are the topic proportions, θ_d. Topic proportions are probability distributions over all topics for a document, d. Thus, a topic proportion, θ_d, depicts the probability with which a document is about a certain topic. Put differently, θ_d encapsulates the probability with which the words in a given document, d are drawn from the topics, K. Note that following Blei and Lafferty (2007), I use a parametrization such that $\eta = log(\theta_i/\theta_K)$.

Having laid out the terminology, the generative process assumed to be underlying a document will be explained following Blei and Lafferty (2007). Given a number of topics, K, which is set by the researcher:

1. Draw $\beta \sim Dir(\gamma)$.
2. Draw $\eta_d|[\mu, \Sigma] \sim \mathcal{N}(\mu, \Sigma)$.
3. Then for $n \in 1, ..., N_d$:

(a) Draw a topic assignment $z_{d,n}|\eta_d$ from a Mult$(f(\eta_d))$.

(b) Given the topic assignment and the topics, draw word $w_{d,n}|z_{d,n}, \beta_{1:K}$ from a Mult$(\beta_{z_{d,n}})$.

The generative model for a CTM varies from an LDA topic model in that it draws the topic proportions from a multivariate Gaussian rather than a Dirichlet distribution. This change allows the consideration of correlation between topics, which is captured in the covariance of the Gaussian, Σ. The cost of this advancement is that the multivariate is no longer conjugate to the multinomial, and hence, more complex sampling algorithms have to be used for inference (Blei and Lafferty, 2007).

This paper uses a spectral initialization and a partially collapsed variational EM algorithm by Roberts et al. (2016) to infer the parameters of the model from the data.[16] The inference of the model relies on a trade-off that is imposed through the Dirichlet and multinomial Gaussian priors. The trade-off is that a document is populated by as few topics as possible and similarly, that a topic attaches high probability to a few words and low probability to the rest.

To make the modelling assumption that only a few topics populate a document more plausible, I follow Hansen et al. (2018) in their suggestion to estimate the topic model on the paragraph rather than the document level. This is argued to yield better interpretable topics for long documents since paragraphs are more likely to contain fewer topics than a whole transcript. However, estimating the model on the paragraph level creates the difficulty of aggregating the probabilities to the document level after the estimation. Hansen et al. (2018) propose to estimate the topic model on the paragraph level and then use this topic model to predict the topic proportions for the transcripts on the document level as if it was new data. The approach used in this paper calculates a weighted average for the document level. Both approaches yield similar results.

The following list best describes the pre-processing steps undertaken before estimating a topic model:

1. Read in minutes at paragraph level.
2. Set M1, M2 and M3 to “moneyone”, “moneytwo”, “moneythree”. This avoids loosing these terms in later pre-processing steps.
3. Identify contractions. Change e.g. ”don’t” to ”do not” or ”we’re” to ”we are”.
4. Eliminate all non-alphabetical characters and set all characters to lower-case.
5. Exclude paragraphs with less than 20 words.

[16]This algorithm is calculated with the stm-package in R, developed by Roberts et al. (2019)

6. Remove common stop words. The stop-word list was kindly provided by Andrew Wu (Jegadeesh and Wu, 2017).
7. Find Collocations. That is find words such as "Open Market Committee" or "Labor market" and connect them as one. Find tri- and bigrams. That is find words such as "Open Market Committee" or "Labor market" and connect them as one.
8. Write out all acronyms to count them consistently and not drop any later on. For example, "ppi" to "producer price index".
9. Lemmatize the words.
10. Remove words with less than three characters.

The critical parameter to be specified by the researcher is the number of topics. Several measures have been suggested in the literature to choose the number of topics in a data-driven manner. Nevertheless, there is a trade-off between the interpretability of the topics and the fit of the model (Blei and Lafferty, 2007; Hansen et al., 2018). Measures such as held-out-likelihood and semantic coherence versus exclusivity would suggest a number of topics between 25-30. However, the highest interpretability is achieved with a number of topics between 10 - 15 topics, which is also found by Jegadeesh and Wu (2017). For tractability, I choose ten topics. However, the same results follow through even with higher numbers of topics around 40.

Estimating the topic model results in a probability distribution over each FOMC meeting overall topics and a probability distribution over all words within a topic. Table 10 shows the five most important words for each topic. Given these top-words, the topics are labelled for ease of illustration.

To create directional topics, I follow Hansen and McMahon (2016) and count directional words within each paragraph, sum all positive and all negative direction words, take the difference of the two and divide by the total number of directional words used in the paragraph. This index is then multiplied with the topic distribution of each paragraph, where the topic distribution acts as a weight on the index. Finally, following Hansen and McMahon (2016) I also construct first differences of these directional topics.

As described in the main part, the topic labels are just for illustrative purposes. The topics to be used in the shock identification have to be chosen via model selection. I conduct a bootstrap LASSO analysis with $N = 300$ and choose all variables that are selected more than 80% of the time. This threshold is chosen since there appears to be a cut from the last variable chosen at 81% to the next variable at 73%. However, extending the threshold to all variables that are chosen more often than, for example, 65% does not

Table 10: Topic Labels

Label	Highest Probability Terms
1. Economy	business, economic, consumer, economy, market, demand, effect, continue, financial, pressure
2. Inflation	inflation, energy, labor, consumer, cost, core, food, producer, index, compensation
3. Money Stock	monetary, aggregate, economic, money, policy, broad, interest, behavior, market, monetary_aggregate
4. Debt	debt, dollar, market, nonfinancial, domestic, currency, nonfinancial_debt, domestic_nonfinancial, aggregate, monetary
5. Inventory	inventory, sale, vehicle, motor, motor_vehicle, manufacture, consumer, production, output, retail
6. reserve	restraint, pressure, directive, degree, market, acceptable, monetary, reserve_restraint, position
7. Construction	construction, house, equipment, employment, business, home, sale, capital, gain, start
8. Policy	policy, inflation, risk, economic, ease, economy, pressure, market, view, tighten
9. Trade	unite, state, export, economic, unite_state, foreign, trade, import, economy, deficit
10. Bond Yield	market, interest, yield, fund, long, short, balance, bond, economic, policy

Note: Topics from topic model estimated for 1983 − 2007. This table shows the ten most important terms for each topic. The topics are labelled subjectively.

change the results and does not increase the adjusted-R^2 when regressing the change in the intended funds rate on the chosen topic measures. The resulting regression results of the change in the target rate on the chosen topics are depicted in table 11.

From table 11 it can be seen that the model fit is slightly higher at approximately 40% as compared to 37% when using sentiment measures. This is unsurprising since we have chosen the topics that best describe the change in the target rate and not solely focused on controlling for output, inflation, and employment. Again, the remaining residual of regressing the change in the target rate on the topic measures is interpreted as the monetary policy shock. The resulting IRFs are discussed in the main section on robustness checks.

Table 11: Identification Topic Shock

	Δ Target Rate	
	Coefficients	*Standard Error*
Constant	0.000	0.055
Initial level of Intended Funds Rate	-0.327^{***}	0.062
Topic Measures:		
2. Inflation	0.097^{***}	0.023
5.Inventory	0.316^{***}	0.077
6.Reserve	0.194^{***}	0.072
8.Policy	0.332^{***}	0.074
10.Bond yield	0.123^{**}	0.061
Δ 4.Dollar	-0.055	0.056
Δ 5.Inventory	-0.249^{***}	0.069
Δ 6.Reserve	-0.139^{**}	0.069
Δ 8.Policy	-0.140^{**}	0.070

* $(p < 0.1)$, ** $(p < 0.05)$, *** $(p < 0.001)$
Adjusted $R^2 = 0.397$, N= 200

Note: This table shows the regression of the change in the intended Federal Funds rate on directional topic measures and first differences thereof. The resulting residual is interpreted as a monetary policy shock.