Bargaining over Words?

Text Analysis of a Model of Monetary Policy by a Committee

Hie Joo Ahn[1] Taeyoung Doh[2] Thomas Cook[2]

[1]Federal Reserve Board
[2]Federal Reserve Bank of Kansas City

Central Banks and Information, ASSA 2026
January 2026

The Focus of This Paper: Microfoundation of FOMC's Reactions

Interactions over words about the statement language among the FOMC members FG

1. How is voting members' sentiment aggregated to the FOMC statement?
   - Uncover the aggregation regimes by linking the topic-sentiment in each FOMC statement with the individual topic-sentiment in the corresponding meeting transcript

2. Explore novel sources of monetary policy shocks
   - Surprises of topic-sentiment in FOMC statements relative to the market expectation
   - Changes in the voting regimes

... with minimal human readings of text or intervention

Main Findings

1. Heterogeneity in individual members' preferences across topics in the statements
2. The chair-centric regime is prevalent (60%) on average.
   - deviating from the median voter model and varying over time → Intense bargaining
   - Policy guidance (87%) > Real activity (44%)
   - Yellen (82%) > Greenspan (57%) ≈ Bernanke (54%)
3. Surprises in topic-sentiment and changes in voting regimes → monetary policy shocks.
   - Unexpected positivity on policy guidance, ⇓ (orthog.) monetary policy shock
   - Chair-centric regime about policy action → monetary tightening, and ...
   - Departure from the equal-weight regime about policy action → monetary loosening.

Literature Review

- **Text analysis of monetary policy research**: Aruoba and Drechsel (2024), Baerg (2020), Chappell Jr et al. (2005), Cieslak and McMahon (2023), Doh et al. (2025), Handlan (2022), Hansen et al. (2018), Schonhardt-Bailey (2013), Shapiron and Wilson (2022), Banerjee et al. (2025), Ahn et al. (2025) and growing!
- **Modeling the committee decision process in monetary policy**: Hansen et al. (2014), Riboni and Ruge-Murcia (2010, 2014, 2020).
- **Topic/LLM**: Blei et al. (2003), Griffiths and Steyvers (2004)/Hartmann et al. (2023).
- **Monetary Policy Shocks**: Bauer and Swanson (2023)

.

Plan

1. Data
2. Topic Classification
3. Sentiment Assessment
4. Sentiment Aggregation : Voting Regime Identification
5. Implications for Monetary Policy
6. Conclusion

Data

FOMC Statements and Meeting Transcripts

Sample period

- Statements: March 2004 – September 2024
- Transcripts: March 2004 – December 2017
  - Greenspan (March 04-Jan 06), Bernanke (Jan 06- Jan 14), Yellen (Jan 14 - Dec 17)

Data cleaning

- Remove stop words, common contractions (e.g., Mrs, MR), and procedural words (e.g., intermeeting, committee) WORD CLOUD
- Transcripts: Focusing on the voting members' remarks

Empirical Methodology

Overview of Empirical Strategy

Step 1: Identify topics from the statements, and assign a topic score for each "sentence"

- Topic model : Latent Dirichlet Allocation (LDA) identifies the cluster of words based on their co-frequency patterns

Step 2: Apply the topic classification (assign a score) for each "remark" of the transcripts

Step 3: Evaluate sentiment of each sentence and each remark

- Sentiment model : SIEBERT (pretrained from various sources)

Step 4: Estimate the sentiment aggregation model (the voting regime) for each topic

- Links committee sentiment in the statements to individual sentiment in the transcripts

1. Topic Model

Latent Dirichlet Allocation

Cluster Words based on the Term Co-Frequency Feature

Sentence (W)
: total #D

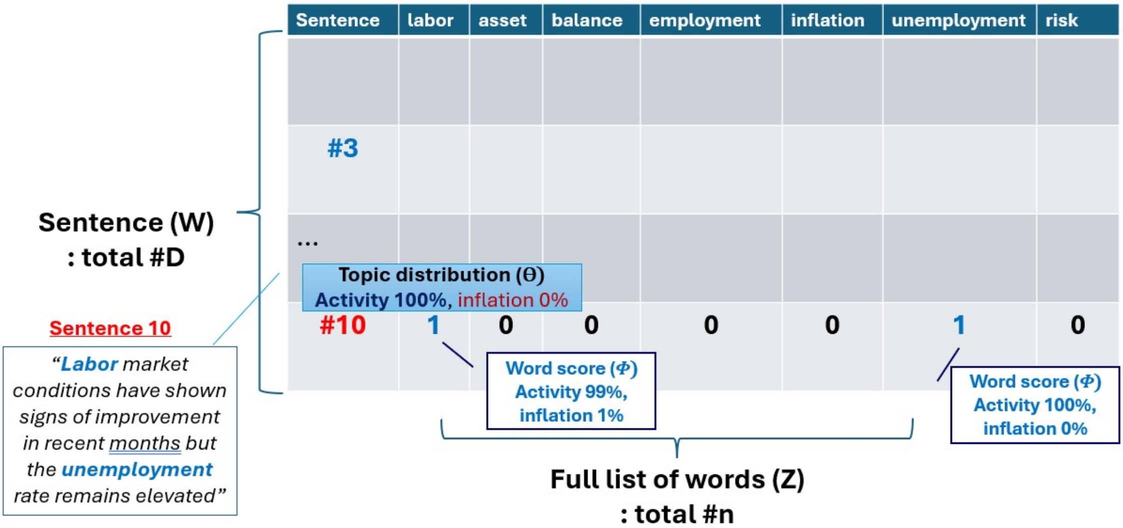
Sentence 10

"***Labor*** *market conditions have shown signs of improvement in recent months but the* ***unemployment*** *rate remains elevated*"

Sentence	labor	asset	balance	employment	inflation	unemployment	risk
#3							
...							
#10	1	0	0	0	0	1	0

Full list of words (Z)
: total #n

Uncovers Topic Distribution for a Sentence & Topic Score for a Word



Topic Distribution for a Sentence (Θ) & Topic Score for a Word (Φ)

Sentence 3

*“The Committee judges that the risks to achieving **employment** and **inflation** goals are roughly in balance.”*

Sentence (W)
: total #D

Sentence 10

*“**Labor** market conditions have shown signs of improvement in recent months but the **unemployment** rate remains elevated”*

Sentence	labor	asset	balance	employment	inflation	unemployment	risk
#3	0	0	1	1	1	0	1
...							
#10	1	0	0	0	0	1	0

Topic distribution (Θ)
Activity 50%, inflation 50%

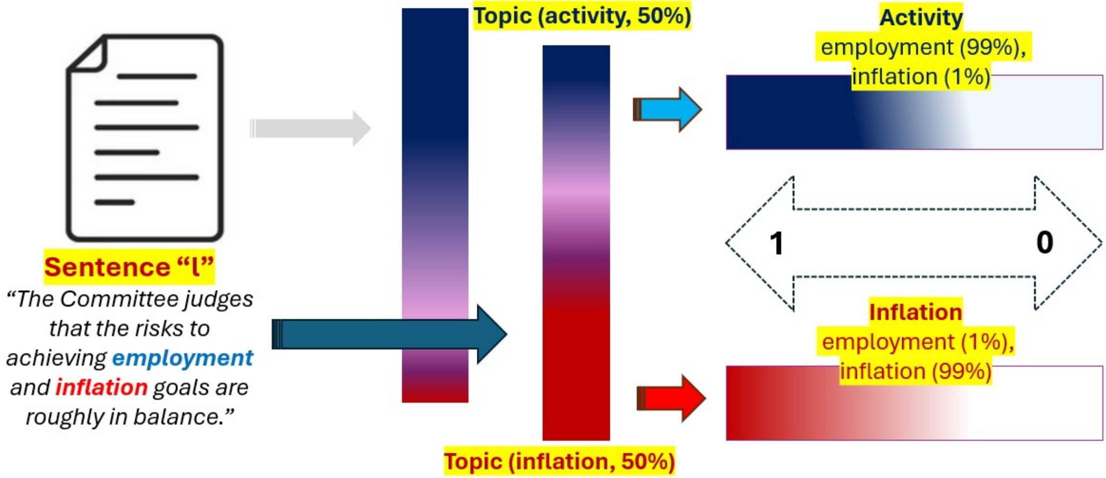
Topic distribution (Θ)
Activity 100%, inflation 0%

Word score (Φ)
Activity 99%,
inflation 1%

Word score (Φ)
Activity 0%,
inflation 100%

Full list of words (Z)
: total #n

Sentence	Topic Distribution of Sentence (l)	(k)-Topic score for each word in (l)
$\boldsymbol{W} = (W_1, W_2, .. \boldsymbol{W_l}, \dots W_D)$	$\boldsymbol{\theta}^{\boldsymbol{l}} = (\theta_{1,l}, \dots \boldsymbol{\theta}_{k,l}, \dots, \theta_{K,l})$	$\boldsymbol{\Phi}^{\boldsymbol{k}} = (\Phi_{k,1}, \dots, \Phi_{k,N})$



Opening up the Black Box: Structure of the LDA Model

With text data (W), estimate the parameters of word distribution for a topic (Φ) and those of the topic distribution of a sentence (Θ) in a Bayesian way (Wallach et al. (2009))

Prior $\times$ **Likelihood** $\Rightarrow$ **Posterior**

- **Dirichlet priors** for Φ and Θ with hyperparameters of α and β respectively
- **Likelihood**: $p(W|\Phi, \Theta)$
- **Posterior estimates**: $p(\Phi, \Theta|W, \alpha, \beta)$: **Collapsed Gibbs sampling** to make the posterior kernels of Θ and Φ tractable Detail

Estimation Results

Topic Classification

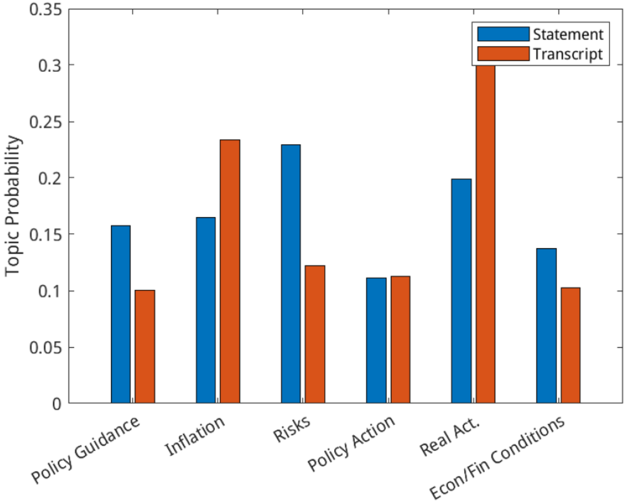
Total Six Topics: Topic Coherence and Sufficient Coverage

Topic	Example Sentence
1. Policy Guidance	"In determining how long to maintain this target range, the Committee will assess progress—both realized and expected—toward its objectives of maximum employment and 2 percent inflation."
2. Inflation	"The rise in energy prices, however, has not notably fed through to core consumer prices."
3. Outlook Risks	"The economic outlook is uncertain, and the Committee is attentive to the risks to both sides of its dual mandate."
4. Policy Actions	"The Committee is maintaining its existing policies of reinvesting principal payments from its holdings of agency debt and agency mortgage-backed securities in agency mortgage-backed securities and of rolling over maturing Treasury securities at auction."
5. Economic Activity	"Firms have brought inventory stocks into better alignment with sales."
6. Economic and Financial Conditions	"This assessment will take into account a wide range of information, including measures of labor market conditions, indicators of inflation pressures and inflation expectations, and readings on financial and international developments."

Notes: We select the example sentence for each topic based on the highest probability for the topic across the

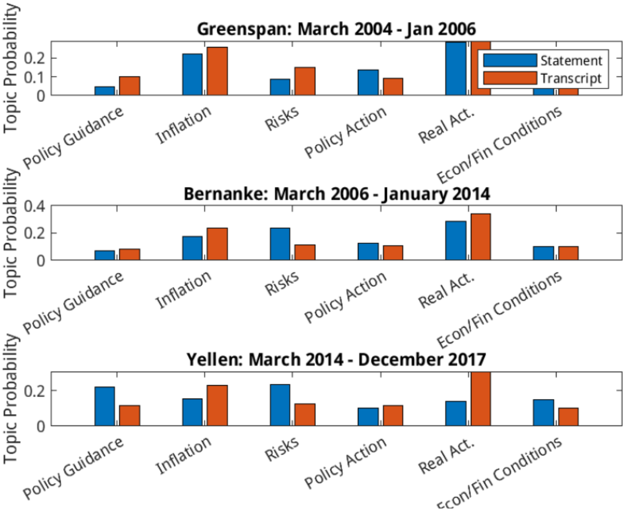
Topic Distribution (Sample Average)

More inflation & activity in the transcripts, more outlook risks in the statements



Topic Prevalence Changes Over Time.

FOMC statements became more forward-looking → $\Uparrow$ importance of forward guidance



- **Outlook risks** :
  $\Downarrow$ Greenspan,
  $\Uparrow$ Bernanke & Yellen
  (ZLB)

- **Inflation & Activity** :
  $\Uparrow$ Greenspan,
  $\Downarrow$ Bernanke & Yellen

Lift-off

2. Sentiment Model

SiEBERT

Sentiment Evaluation with SiEBERT

For each sentence, we quantify the sentiment by using SiEBERT (Hartmann et al. (2023)).

- Analyzes Sentiment in English with RoBERTa-large.
- A pre-trained large language model that is a fine-tuned version of RoBERTa
- Binary sentiment classification (i.e. "positive" vs "negative") on English text.
- Sentiment score ranging from -1 (perfectly negative) to 1 (perfectly positive). SiEBERT

Sentiment by Topic

Policy Action **-0.99**

The Committee will continue reducing↓ its holdings− of Treasury securities− and agency MBS−, as described in its previously announced plans *(June 15, 2022* **FOMC statement)**

Policy Guidance **-0.41** ***Inflation*** **-0.21**, **frequency weight** (0.17)

I do not support raising the federal funds rate↓↓ today, I think our decision today is a close one and a judgment call but I don't think achievement of our symmetric 2 percent inflation↓ target will↓ be well served by once again raising the funds rate↓ while inflation↓ substantially underruns 2 percent and inflation↓ expectations↓ are too low to be consistent with our symmetric objective ..... *(Evans, December 2017* **FOMC transcript)**

3. Sentiment Aggregation

Sentiment Evaluation : FOMC Statements

The k-th topic-sentiment score on the l-th sentence

$$s_{l,k} = \underbrace{\theta_{l,k}}_{\text{topic share}} \underbrace{s_l}_{\text{sentiment}}$$

The statement-level sentiment

$$s_k^{FOMC} = \sum_{p=1}^{P} \frac{\theta_{p,k} s_p}{P}$$

- p: the index of a sentence in the respective FOMC statement
- P: total number of sentences in the statement

Apply the Same Methodology to the Meeting Transcripts

Sentiment on topic k of member i

$$s_k^{Member,i} = \sum_p \hat{\mathbf{freq}}^{\mathbf{i}}_{\mathbf{k,p}} \underbrace{\theta^i_{p,k}}_{\text{topic share}} \underbrace{s^i_p}_{sentiment}$$

$\hat{\mathbf{freq}}^{\mathbf{i}}_{\mathbf{k,p}}$: The relative frequency of topic key words among total words in each remark (p)

Downplay Less Relevant Text

A remark by Chair Yellen:

.... I first need a motion from a Board member to increase the interest rates on required and excess reserve balances to 1 percent effective December 14 2017 ....

~~.... I guess many people consider 10:15 too early for lunch, Laughter. Boxed sandwiches and salads are available in the anteroom for you to consume whenever you regard as appropriate. And there will be for anyone who wants to stay around a TV in the Special Library if you want to watch the press conference which begins at 2:30.~~

- Frequency : 0.15
- Sentiment score : 0.99

Aggregation of Statement Topic Scores at the Transcript Level

FOMC's sentiment on topic k

$$s_k^{FOMC} = \sum_i w_{i,k} s_k^{Member,i}$$

$$\begin{aligned}
&\textbf{Equal-weight} \quad w_{i,k} &&= \frac{1}{N} && (1)\\
&\textbf{Median-voter} \quad w_{i,k} &&= \begin{cases} 1, & \text{if } s_{i,k}^{\text{Member}} = Q_{0.5}\left(s_k^{\text{Member}}\right), \\ 0, & \text{otherwise.} \end{cases} && (2)\\
&\textbf{Chair-centric} \quad w_{i,k} &&= \frac{1 - \dfrac{(s_k^{\text{Member},i} - s_k^{\text{Chair}})^2}{\max_l (s_k^{\text{Member},l} - s_k^{\text{Chair}})^2}}{N - \dfrac{\sum_j (s_k^{\text{Member},j} - s_k^{\text{Chair}})^2}{\max_l (s_k^{\text{Member},l} - s_k^{\text{Chair}})^2}}. && (3)
\end{aligned}$$

Aggregation of Statement Topic Scores at the Transcript Level

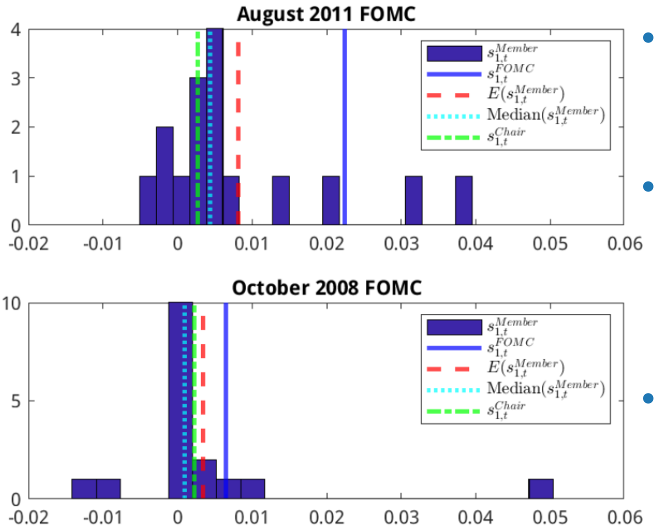
Consider three models for $w_{i,k}$ & estimate the weights by minimizing the loss function

$$\begin{aligned} & \mathrm{Loss}\big(s_k^{\mathrm{FOMC}},\, s_{1,k}^{\mathrm{Member}}, \ldots, s_{M,k}^{\mathrm{Member}}\big) \\ = \quad & \sum_{r=1}^{3} P_r \left(s_k^{\mathrm{FOMC}} - \sum_{i=1}^{M} w_{i,k}^{(r)}\, s_{i,k}^{\mathrm{Member}} \right)^2, \\ & \sum_{r=1}^{3} P_r = 1. \end{aligned}$$

Estimation Results

Sentiment Aggregation

Sentiment Distribution and Aggregation



- Chair's stance (green), similar in both meetings.
- Statement score (blue) ≈ Mean (red) > Median (light blue) or Chair's (green) → Equal weight
- More outliers in 2011: Evans (positive) pushed the FOMC's position to his preferred one.

Chair-centric Regime Prevails, Followed by the Equal-weight Regime

Table: The Preferred Aggregation Regime Probability by Chairs

Period	**Equal Weight**	**Median**	**Chair Centric**
Greenspan (16 Statements)	0.25	0.18	0.57
Bernanke (64 Statements)	0.21	0.25	0.54
Yellen (31 Statements)	0.14	0.04	0.82
Sum (111 Statements)	0.20	0.18	0.62

- Importance of Chair's agenda setting power (Riboni and Ruge-Murcia (2020))

Regime Aggregation Varies by Topics

Chair-centric regime, preferred less about topics on real economic activity and policy

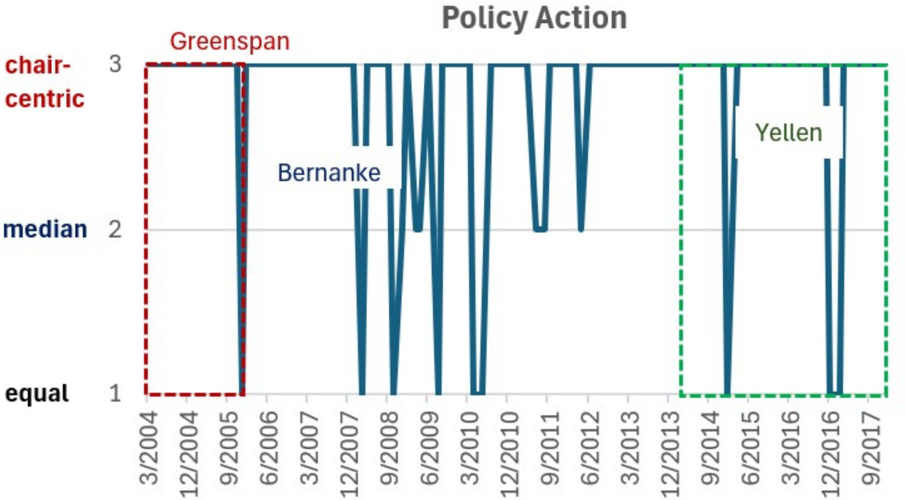
Table: Aggregation Regime Probability by Topics

Topic	Equal Weight	Median	Chair Centric
Policy Guidance	0.27	0.28	0.45
Inflation Conditions	0.08	0.05	0.87
Outlook Risks	0.26	0.21	0.53
Policy Action	0.12	0.14	0.74
Real Economic Activity	0.33	0.23	0.44
Economic and Financial Conditions	0.12	0.16	0.72

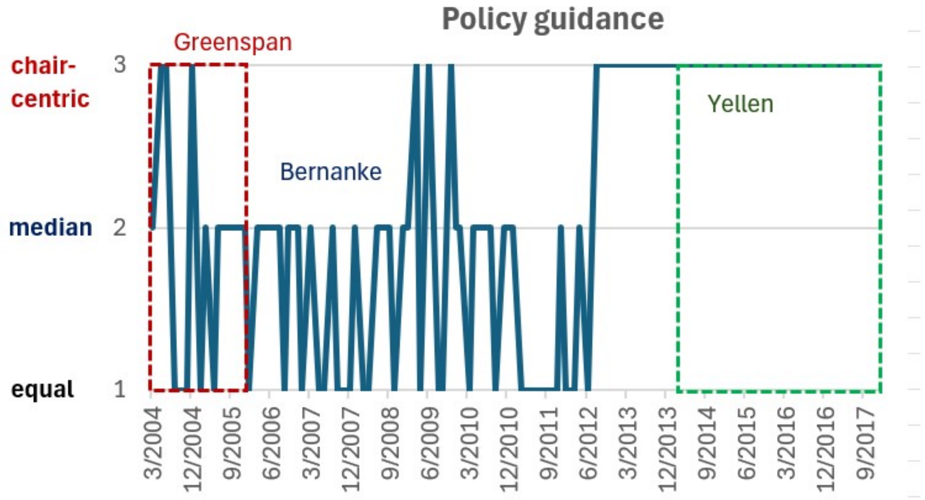
- Likely reflecting Reserve Bank presidents' discussions on district economic conditions

Figures

Voting Regime about Policy Action : Chair-centric



Policy Guidance : Less chair-centric before Yellen



Source of Monetary Policy Shocks

1. Surprises in FOMC statements
2. The Voting Regime and its Evolution

Assessing Statement Surprises by Topic

Identify a surprise component of an FOMC statement by

- Applying our topic classification/sentiment model to the market's expected statements
  - Proprietary data on "proxy" statements written by a consulting firm and circulated among paid subscribers including major investment banks a few days before FOMC meetings (provided by a former governor Larry Meyer)
- Comparing the topic-sentiment scores between the expected and realized statements

Meyer's Draft on Oct. 27, 2008 for the Oct. 28-29 Meeting

FOMC BRIEFING

STATEMENT DRAFT

Below is a draft of the FOMC statement that we see as likely. We did not track the changes from the previous statement since they are so considerable:

The Federal Open Market Committee decided today to lower its target for the federal funds rate 50 basis points to 1 percent.

Incoming economic data suggest that the pace of economic activity has slowed markedly in recent months. Moreover, the intensification of financial market turmoil is likely to exert additional restraint on spending, partly by further reducing the ability of households and businesses to obtain credit. Over time, the substantial easing of monetary policy, combined with ongoing measures to foster market liquidity, should help to promote moderate economic growth.

Inflation has been high, but the Committee believes that the decline in energy and other commodity prices and the weaker prospects for economic activity should lead to a moderation of inflation. Still, the inflation outlook remains uncertain.

The upside risks to inflation have diminished, while the downside risks to growth have increased. The Committee will monitor economic and financial developments carefully and will act as needed to promote sustainable economic growth and price stability.

Voting for the FOMC monetary policy action were: Ben S. Bernanke, Chairman; Timothy F. Geithner, Vice Chairman; Elizabeth A. Duke; Richard W. Fisher; Donald L. Kohn; Randall S. Kroszner; Sandra Pianalto; Charles I. Plosser; Gary H. Stern; and Kevin M. Warsh.

In taking this action, the Board approved the requests submitted by the Boards of Directors of the Federal Reserve Banks of Boston, New York, Cleveland, Atlanta, Chicago, Minneapolis, Dallas, and San Francisco.

Assessing Statement Surprises by Topic

- $s_k^{Surprise}$: The statement surprise component

$$s_k^{Surprise} = s_k^{FOMC} - s_k^{Meyer}$$

- y_t: a financial market response during the 30-minute window around an FOMC announcement

$$y_t = \alpha + \sum_{k=1}^{K} \beta_k s_{k,t}^{Surprise} + \epsilon_t$$

- Bauer-Swanson (2022) : **Surprises** = Fed's response to news + Orthogonalized shock

Surprises in Policy Guidance Sentiment → Monetary Policy Shocks

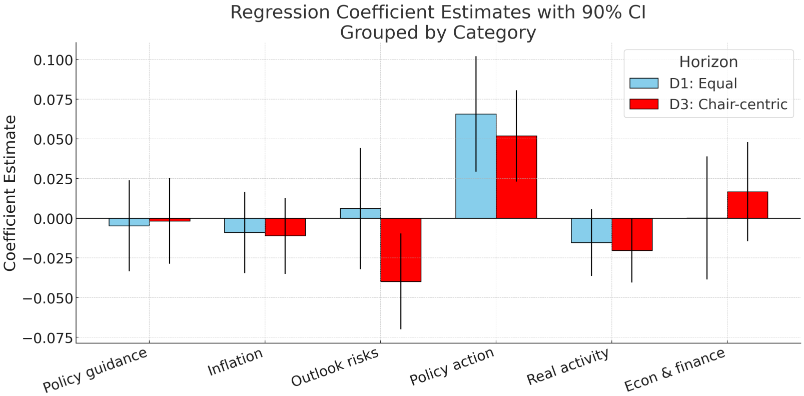
Table: High-frequency Bond Market Responses to Surprises in Topic Sentiment Scores

Coefficient	Unorthogonalized	Orthogonalized	News
β_1 (policy guidance)	-0.081	**-0.187***	**0.106***
β_2 (inflation)	0.003	-0.054	**0.057***
β_3 (outlook risks)	0.041	-0.016	**0.057****
β_4 (policy action)	0.028	0.008	0.020
β_5 (real activity)	-0.007	-0.028	0.021
β_6 (econ & financial conditions)	0.016	-0.050	**0.066****
R^2	0.021	0.092	0.228

- Unexpected positive sentiment on “policy guidance" → more accommdative policy
- Close correlations with Fed’s responses to news

Chair-centric: Outlook Risks Expansionary, Policy Action Tightening

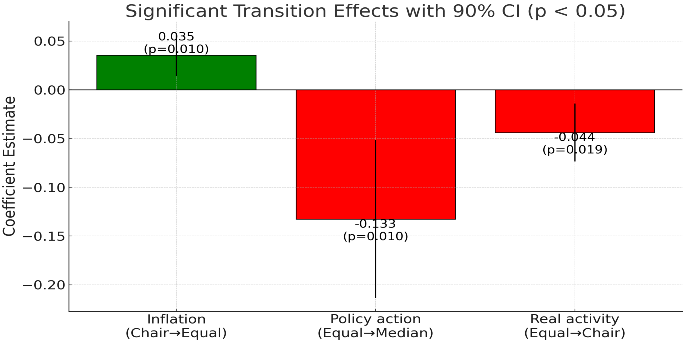
$$\underbrace{y_t^{orth}}_{\text{orthogonalized}} = \alpha + \sum_{k=1}^{K} D_k^1 R_{k,t}^{Equal} + \sum_{k=1}^{K} D_k^3 R_{k,t}^{Chair} + \epsilon_t$$



FedResp

Changes in Voting Regimes on Policy Action, an MP shock

$$\underbrace{y_t^{orth}}_{\text{orthogonalized}} = \alpha + \sum_{k=1}^{K} \sum_{p=e,m,c} \sum_{q=e,m,c} D_k^{p.q} R_{k,t}^{p.q} + \epsilon_t$$



Switching out of the equal-weight regime is expansionary for policy action and real activity.

Conclusion

Conclusion

- We model FOMC decision making as bargaining over statement language.
- By combining topic and sentiment models, we identify topic sentiment scores for both the committee and individual members.
- Aggregation regimes deviate from the media voter model frequently and vary over time, reflecting the internal dynamics of the FOMC.
- Changes in topic sentiment and voting regimes can be underlying sources of monetary policy shocks.

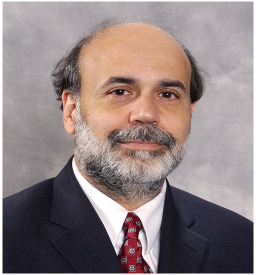
To do list: Suggestions and Ideas Welcome!

- A more tight connection between a bargaining model and empirical analysis (e.g, implicit and explicit dissents).
- High-frequency financial market response beyond the bond market to the surprise component in the statement.
- Digging into the determinants of shifts in the preferred aggregation regime.
- A bargaining model under construction ...

Appendix

Changes in Forward Guidance

August 09, 2011



Just to get a sense of the Committee's preferences ...,

if we were to go back to **alternative B** as written
with **the forward-leaning language** of the last paragraph,
without the "financial conditions",

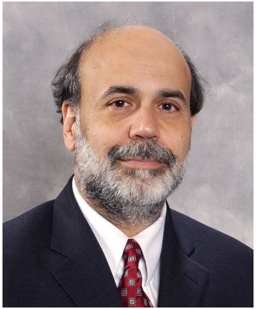
how many dissents would I have?

No reference to mid-2013?

Alternative B

… The Committee continues to anticipate that economic conditions … are likely to warrant exceptionally low levels for the federal funds rate for an extended period. The Committee will complete its purchases of $600 billion of longer-term Treasury securities by the end of this month and will maintain its existing policy of reinvesting principal payments from its securities holdings. …

RETURN

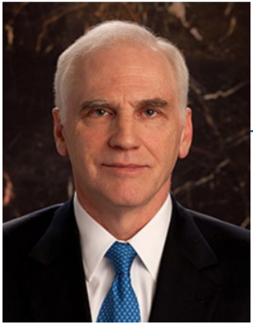


That's right!

Alternative B as written,

except we take out the "financial conditions" language and

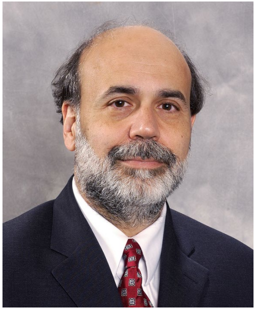
we use the more forward-leaning paragraph 4.



So no action at all?

Alternative B

… The Committee continues to anticipate that economic conditions … are likely to warrant exceptionally low levels for the federal funds rate for an extended period. ~~The Committee will complete its purchases of $600 billion of longer-term Treasury securities by the end of this month and will maintain its existing policy of reinvesting principal payments from its securities holdings~~. …



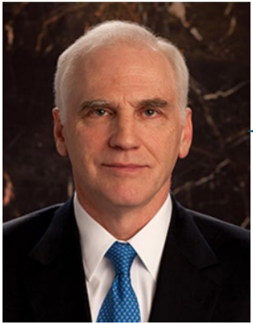
That's right!

Alternative B as written,

except we take out the "financial conditions" language and

we use the more forward-leaning paragraph 4.

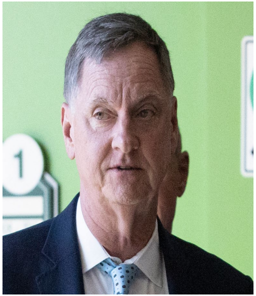
No action at all



So no action at all?

Alternative B

… The Committee continues to anticipate that economic conditions … are likely to warrant exceptionally low levels for the federal funds rate for an extended period. ~~The Committee will complete its purchases of $600 billion of longer-term Treasury securities by the end of this month and will maintain its existing policy of reinvesting principal payments from its securities holdings~~. …



I could not.

I mean, *it just seems clear that we would simply be postponing that very discussion until September, when we would have the very same outcome*.

And so I could not support that.

You might have other negative votes

Statement

The Committee continues to anticipate that economic conditions… are likely to warrant exceptionally low levels for the federal funds rate **at least through mid-2013**. The Committee also will maintain its existing policy of reinvesting principal payments from its securities holdings.

Word Cloud from the FOMC Statement Data

Panel A: Unfiltered

Panel B: Filtered

RETURN

Posterior Estimation: Challenges and Solutions

Posterior

$$p(\Phi, \Theta \mid W, \alpha, \beta) \propto p(W \mid \Theta, \Phi)\, p(\Theta \mid \alpha)\, p(\Phi \mid \beta)$$

Posterior Estimation: Challenges and Solutions

Posterior

$$p(\Phi, \Theta \mid W, \alpha, \beta) \propto p(W \mid \Theta, \Phi)\, p(\Theta \mid \alpha)\, \underbrace{p(\Phi \mid \beta)}_{\substack{\text{Prior} \\ \text{for } \Phi \\ \text{(word)}}}$$

Posterior Estimation: Challenges and Solutions

Posterior

$$p(\Phi, \Theta \mid W, \alpha, \beta) \propto p(W \mid \Theta, \Phi) \; \underbrace{p(\Theta \mid \alpha)}_{\substack{\text{Prior} \\ \text{for } \Theta \\ \text{(topic)}}} \; \underbrace{p(\Phi \mid \beta)}_{\substack{\text{Prior} \\ \text{for } \Phi \\ \text{(word)}}}$$

Posterior Estimation: Challenges and Solutions

Posterior

$$p(\Phi, \Theta \mid W, \alpha, \beta) \propto \underbrace{p(W \mid \Theta, \Phi)}_{\text{Likelihood of } \Theta \text{ \& } \Phi} \; \underbrace{p(\Theta \mid \alpha)}_{\text{Prior for } \Theta \text{ (topic)}} \; \underbrace{p(\Phi \mid \beta)}_{\text{Prior for } \Phi \text{ (word)}}$$

Posterior Estimation: Challenges and Solutions

Posterior

$$p(\Phi, \Theta \mid W, \alpha, \beta) \propto \underbrace{p(W \mid \Theta, \Phi)}_{\substack{\text{Likelihood of} \\ \Theta \,\&\, \Phi}} \; \underbrace{p(\Theta \mid \alpha)}_{\substack{\text{Prior} \\ \text{for } \Theta \\ \text{(topic)}}} \; \underbrace{p(\Phi \mid \beta)}_{\substack{\text{Prior} \\ \text{for } \Phi \\ \text{(word)}}}$$

Challenges and Solution

- Integrating the posterior kernel wrt Θ and Φ is intractable because the posterior kernel depends on the summation of the product of Θ and Φ (Blei et al., 2003). Summation

Posterior Estimation: Challenges and Solutions

Posterior

$$p(\Phi, \Theta \mid W, \alpha, \beta) \propto \underbrace{p(W \mid \Theta, \Phi)}_{\text{Likelihood of } \Theta \,\&\, \Phi} \; \underbrace{p(\Theta \mid \alpha)}_{\text{Prior for } \Theta \text{ (topic)}} \; \underbrace{p(\Phi \mid \beta)}_{\text{Prior for } \Phi \text{ (word)}}$$

Challenges and Solution

- **Collapsed Gibbs sampling**: obtain posterior mean estimates of Θ and Φ from posterior draws of Z via Collapsed Gibbs sampling where $p(z_{i,l}|Z_{-(i,l)}\alpha, \beta)$ can be factorized into $p(z_{i,l}|Z_{-(i,l)}, \alpha)p(z_{i,l}|Z_{-(i,l)}, \beta)$.
- Since $z_{i,l}$ follows a multinomial distribution given θ^l, we obtain the posterior of θ^l using Dirichlet–multinomial conjugacy. RETURN

Collapsed Gibbs Sampling: Griffiths and Steyvers (2004)

$$p(\theta, \Phi \mid \alpha, \beta, W) = \frac{p(\theta \mid \alpha)\, p(\Phi \mid \beta)\, p(W \mid \theta, \Phi)}{\int p(\theta \mid \alpha)\, p(\Phi \mid \beta)\, p(W \mid \theta, \Phi)\, d\theta\, d\Phi}$$

But marginalizing w.r.t. θ and Φ is intractable.

$$p(W \mid \theta, \Phi) = \prod_{l=1}^{D} \sum_{j=1}^{K} \prod_{i=1}^{n} (\theta_j^l \Phi_{j,i})^{W_{i,l}}$$

We circumvent this by calculating:

$$p(\theta, \Phi \mid z, W) = \prod_{l=1}^{D} \prod_{i=1}^{n} \left[\Big(\prod_{j=1}^{K} (\theta_j^l)^{\mathbf{1}(z_{i,l}=k)} \Big) \Phi_{z_{i,l},i} \right]^{W_{i,l}}.$$

Why Does This Work?

- By conditioning on z, the likelihood is transformed from a mixture of multinomial distributions to a simple multinomial distribution.
- The dependence of the posterior distribution of z on α and β can be factorized.
- $p(\theta, \Phi \mid z, W)$ follows a Dirichlet distribution due to the conjugacy of a Dirichlet with a multinomial distribution!
- The posterior draws of z given W can be obtained using Gibbs sampling: draw z_j iteratively conditional on W and $z_{-j} = z \setminus z_j$.
- Start from a random draw of z given α and iteratively update each element of z based on Gibbs sampling.

Details of the Algorithm

Due to conjugacy of a Dirichlet prior $\left((\theta_j^l)^{\alpha-1}, \Phi_{z_{i,l},i}^{\beta-1}\right)$ with a multinomial distribution $\left((\theta_j^l)^{\mathbf{1}(z_{i,l}=k)}, \Phi_{z_{i,l},i}^{W_{i,l}}\right)$, the posterior distribution of $z_{i,l}$ is:

$$p(z_{i,l} = k \mid z_{-(i,l)}, W, \alpha, \beta) \propto p(z_{i,l} = k \mid z_{-(i,l)}, W, \alpha)\, p(z_{i,l} = k \mid z_{-(i,l)}, W, \beta)$$

$$= \frac{n_{(k,.,l)}^{-(i,l)} + \alpha}{\sum_{j=1}^{K} n_{(j,.,l)}^{-(i,l)} + K\alpha} \cdot \frac{n_{(k,w(i,l),.)}^{-(i,l)} + \beta}{n_{(k,.,.)}^{-(i,l)} + n\beta}$$

- $n_{(k,.,l)}$: # of words in document l assigned to topic k,
- $n_{(k,w(i,l),.)}$: # of times word $w(i,l)$ is assigned to topic k,
- $n_{(k,..,l)}^{-(i,l)} = n_{(k,..,l)} - \mathbf{1}(z_{i,l} = k)$. $\qquad n_{(k,v,..)}^{-(i,l)} = n_{(k,v,..)} - \mathbf{1}(z_{i,l} = k)$.

Details of the Algorithm RETURN

1. Sample z iteratively until $\log p(W \mid z, \beta)$ stabilizes.
2. Once converged, compute posterior means:

$$E(\theta_k^l \mid W, \alpha) = \frac{n_{(k,.,l)} + \alpha}{\sum_{j=1}^{K} n_{(j,.,l)} + K\alpha},$$

$$E(\Phi_{k,i} \mid W, \beta) = \frac{n_{(k,i,.)} + \beta}{\sum_{v=1}^{n} n_{(k,v,.)} + n\beta}.$$

Inside SiEBERT

1. Transforms an input text to a vector of tokens

Inside SiEBERT

1. Transforms an input text to a vector of tokens
2. Apply a deep neural network architectures (prediction models) to obtain the numeric representation of the entire input text, which is called the "embedding representation."

Inside SiEBERT

1. Transforms an input text to a vector of tokens
2. Apply a deep neural network architectures (prediction models) to obtain the numeric representation of the entire input text, which is called the "embedding representation."
3. The model trains (estimates) parameters that generate the embedding representation of a given text to maximize the accuracy of sentiment classification, and then

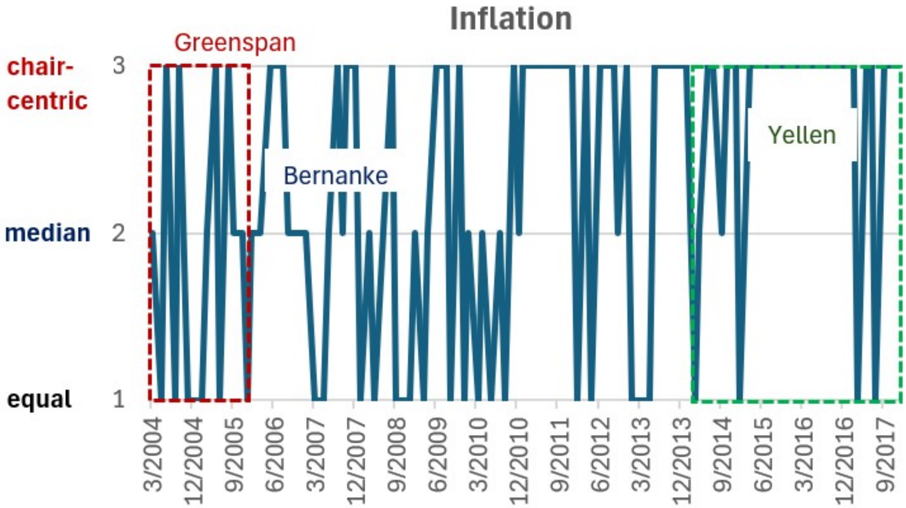
Inside SiEBERT

1. Transforms an input text to a vector of tokens
2. Apply a deep neural network architectures (prediction models) to obtain the numeric representation of the entire input text, which is called the "embedding representation."
3. The model trains (estimates) parameters that generate the embedding representation of a given text to maximize the accuracy of sentiment classification, and then
4. Optimizes the embedding representation by fine-tuning additional neural network layers on top of the output from RoBERTa with 15 different pre-labeled datasets.

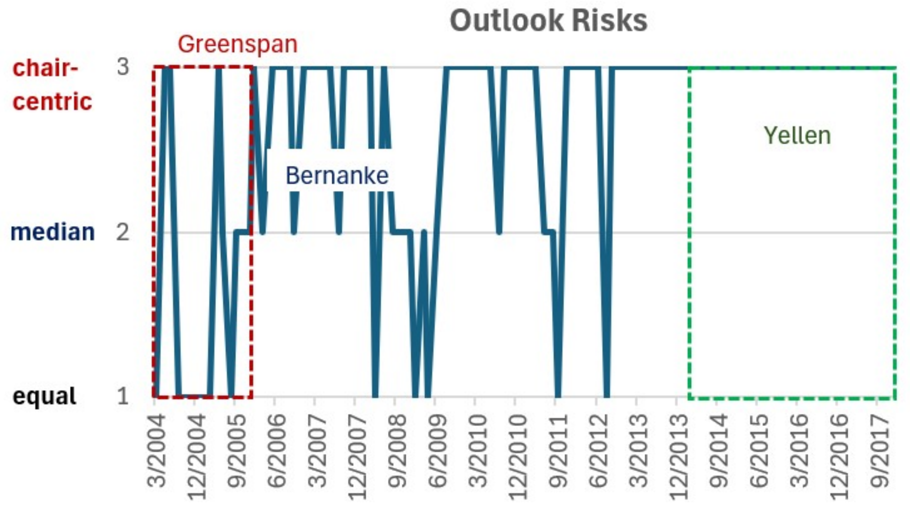
Inside SiEBERT

1. Transforms an input text to a vector of tokens
2. Apply a deep neural network architectures (prediction models) to obtain the numeric representation of the entire input text, which is called the "embedding representation."
3. The model trains (estimates) parameters that generate the embedding representation of a given text to maximize the accuracy of sentiment classification, and then
4. Optimizes the embedding representation by fine-tuning additional neural network layers on top of the output from RoBERTa with 15 different pre-labeled datasets.
5. Sentiment score ranging from -1 (perfectly negative) to 1 (perfectly positive). RETURN

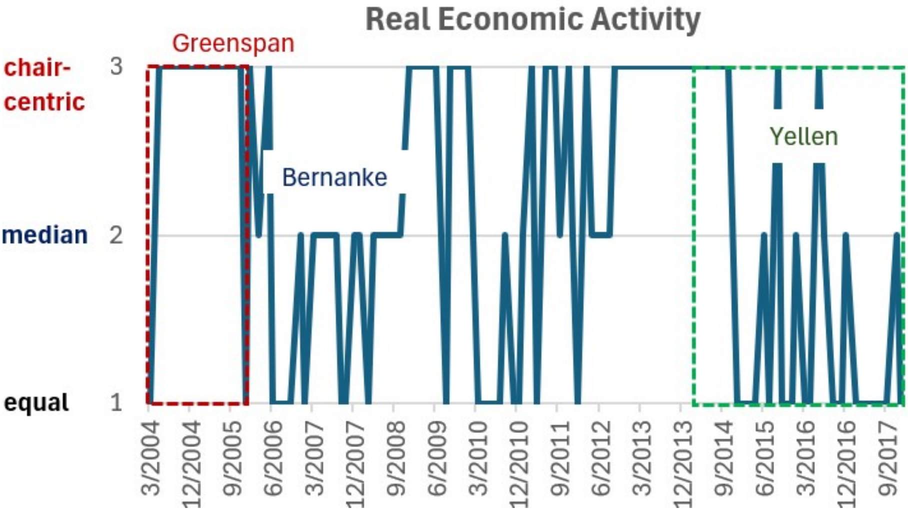
Inflation RETURN



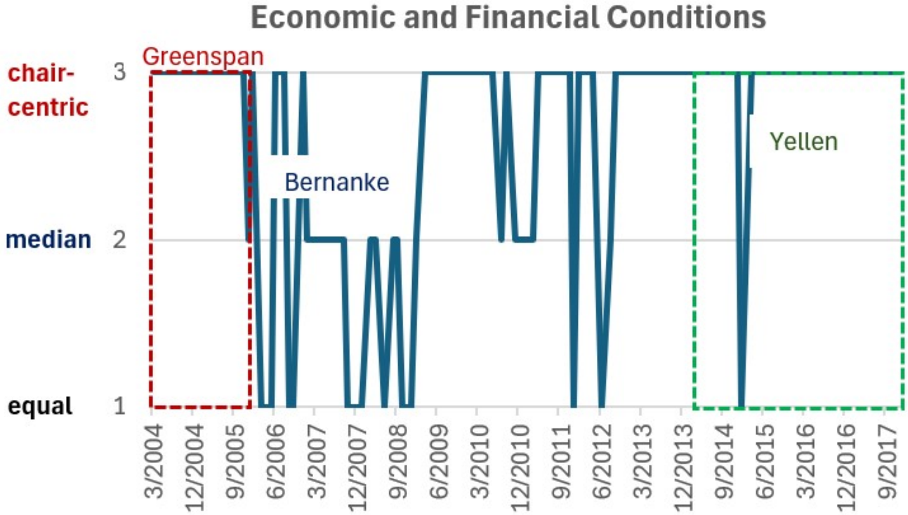
Outlook Risks



Real Economic Activity



Economic and Financial Conditions



Effects on Fed's response to News RETURN

$$\underbrace{y_t^{orth}}_{\text{orthogonalized}} = \alpha + \sum_{k=1}^{K} D_k^1 R_{k,t}^{Equal} + \sum_{k=1}^{K} D_k^3 R_{k,t}^{Chair} + \epsilon_t$$