Understanding the Federal Reserve

What is the price of a word?

Participants:

Liam Chentoufi, Student

Dr. Mary Kaltenberg, Faculty Mentor

Dyson College of Arts and Sciences

Provost's 2023-2024 Student-Faculty Undergraduate Research and Creative Inquiry Award

Methodology:

Sentiment analysis, a machine learning method, is applied to predict Federal Open Market Committee interest rate decisions. Textual data from policymaker's public speeches is used to glean implied monetary policy stance of committee members.

Purpose:

By analyzing FOMC actions, we can better understand its "reaction function", namely, participants' projections and the actions they feel are necessary to meet policy goals. From the perspective of the Fed, transparency and communication is necessary for the effective transmission of monetary policy.

Outcomes:

Preliminary model shows 63.8% accuracy on unseen data in predicting interest rate decision.

Simple regression appears to suggest that sentiment may have a larger effect on decision-making during downturns, and less of an influence before hiking campaigns.

Understanding the Federal Reserve

What is the price of a word?

This project's *Dual Mandate*

- Scrape speech data over the 2006-2023 timeframe; utilize a machine learning model to categorize speeches across time based on their expressed sentiment. (Hawkish/Dovish/…Neither?)

- Attempt at answering: What is the relationship between speech attitudes and the real economy?

Background, Theoretical Assumptions

- The Federal Reserve (Fed) controls key interest rates that, through arbitrage, affect other market rates and the broader economy.

- It is assumed that Fed officials leave hints in their remarks, regarding their outlook and future path of their policy rate.

So, why not do this manually?

- “Fedspeak” is inherently vague as officials seek to maintain credibility. Commitment absent of action degrades trust; word choice is critical.

- Machine learning can allow us to analyze longer time frames using greater amounts of data, yielding more actionable results than singular speeches.

"Since I've become a central banker, I've learned to mumble with great incoherence. If I seem unduly clear to you, you must have misunderstood what I said."[1]

Greenspan, 1987

1. Chair Greenspan speaking to a Senate Committee, 1987. *Oxford Reference*

"I continue to strongly believe that monetary policy is most effective when the FOMC is forthcoming in addressing economic and financial developments such as those I have discussed in these remarks, and when we speak clearly about how such developments may affect the outlook and the expected path of policy."[2]

Yellen, 2016

2. Chair Yellen speaking at the Economic Club of New York, 2016. *Federal Reserve Board of Governors*

In the Literature

Chan, Ken. "Interpreting the Fedspeak: Text Analysis on FOMC Statements - BBVA Research." ***BBVA Research*****, 27 Sept. 2016, www.bbvaresearch.com/wp-content/uploads/2016/09/160927_US_Fedspeak.pdf.**

(Fed sentiment impact on market uncertainty)

Zhang, Ken Zhan, and Qingyi Huang. "Our Patented Fed Sentiment Indicator." ***Morgan Stanley*****, Morgan Stanley, 25 Apr. 2023, www.morganstanley.com/articles/mnlpfeds-sentiment-index-federal-reserve.**

(Fed sentiment impact on interest rates, yield curve, U.S. exchange rate)

In the Literature (cont.)

Adams, Travis Adams, et al. "More than Words: Twitter Chatter and Financial Market Sentiment." ***The Federal Reserve Bank of New York*****, The Federal Reserve Bank of New York, 2023, www.federalreserve.gov/econres/feds/files/2023034pap.pdf.**

(Media sentiment helps forecast changes to stance of U.S. monetary policy.)

Shah, Agam, et al. "Trillion Dollar Words: A New Financial Dataset, Task & Market Analysis." ***ACL Anthology*****, ACL Anthology 2023, 13 May 2023, aclanthology.org/2023.acl-long.368.pdf.**

(RoBERTa categorization based on statement sentiment)

Figure 1: US Equity Price Movement and the Dec 16, 2015 FOMC Statement[2]

Source: Schnidman, Evan. "The Prattle Machine Learning Algorithm How and Why - Hubspot." *Prattle White Paper*, Prattle - Hubspot, 7 Dec. 2016, cdn2.hubspot.net/hubfs/1784239/Prattle_Content/The%20Prattle%20Machine%20Learning%20Algorithm%20How%20and%20Why.pdf.

Hawk/Dove in the Context of "Fedspeak"

Hawkish: "*We still have a ways to go. Until then, I support continued rate increases and ongoing reductions in the Fed's balance sheet to restrain aggregate demand.*"[3]

Governor Kugler, 2024

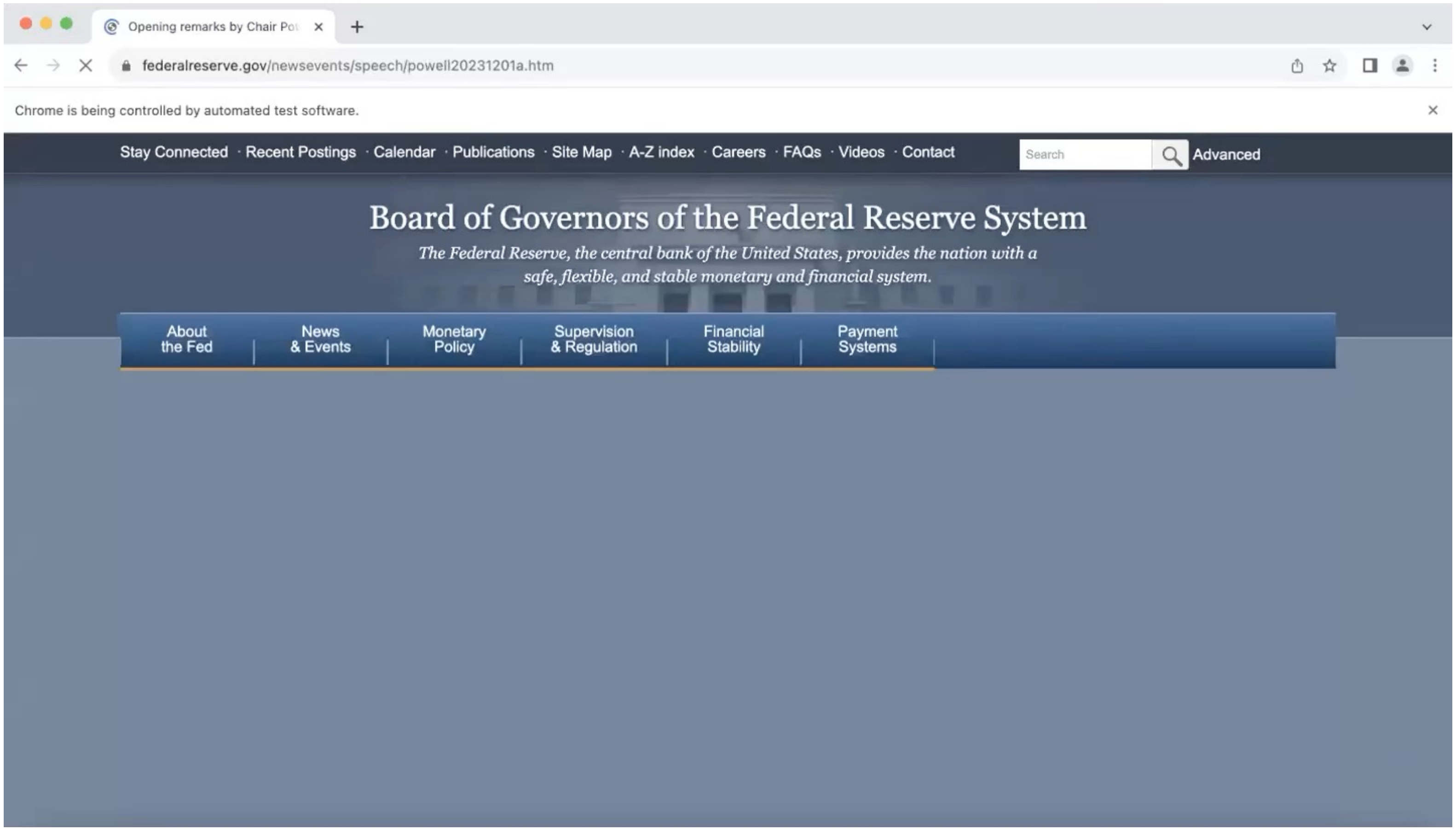
Dovish: "*In its most recent announcement, the FOMC indicated that risks remain weighted toward economic weakness.*"[4]

Vice Chair Ferguson, 2001

3. Governor Kugler speaking at the Weidenbaum Center on the Economy, Government, and Public Policy, 2024. *Federal Reserve Board of Governors*
4. Vice Chair Ferguson speaking at the Bay Area Council 2001 Outlook Conference, 2001. *Federal Reserve Board of Governors*

Text Collection and Labeling

Text Collection



Sentence Labeling!

Manual Labeling

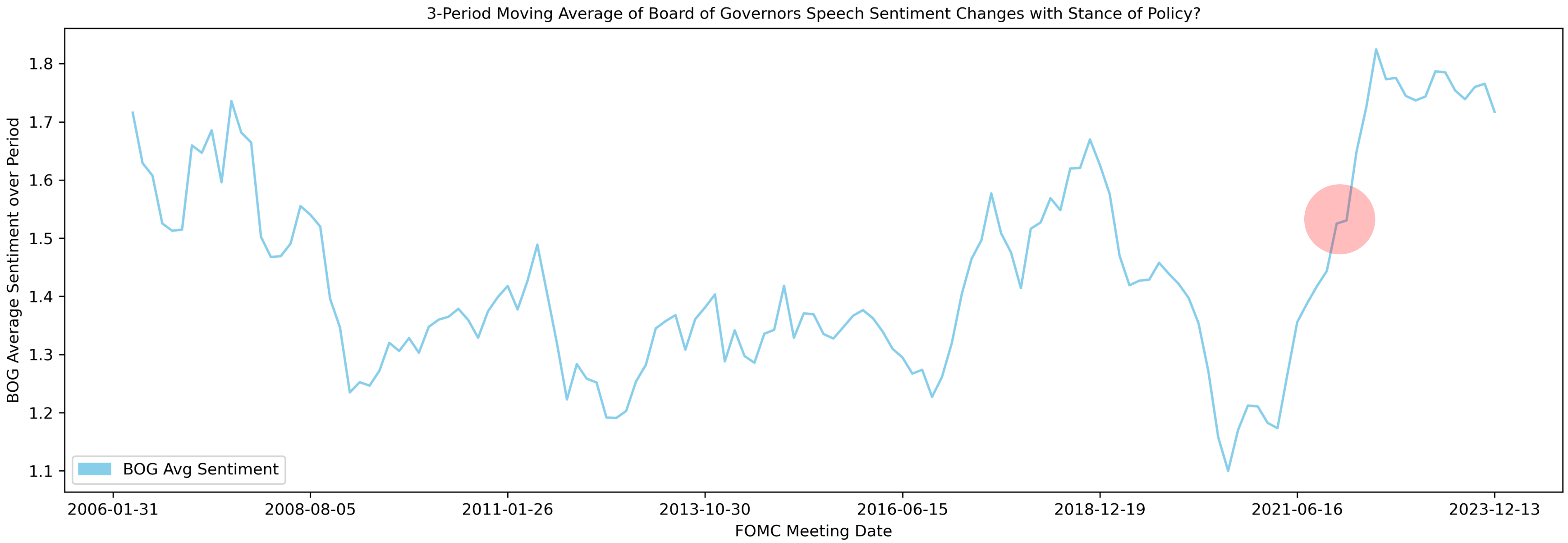
	sentence	my_label
0	Thank you to Columbia University and the organ...	n
1	It is a fitting time to share some reflections...	n
2	This was the first major bank failure since th...	n
3	Experienced and well-respected staff from arou...	n
4	But the review also found that Federal Reserve...	n
...	...	...
796	Like an exchange, a central counterparty impos...	n
797	That said, many OTC contracts are already elig...	n
798	Assessing recent financial market performance ...	n
799	I would like to highlight two themes that I be...	n
800	Product complexity and a lack of transparency ...	n

Shah et al. RoBERTa Classifier

	sentence	label	period
0	Thank you to Columbia University and the organ...	2	2024-03-20
1	It is a fitting time to share some reflections...	2	2024-03-20
2	This was the first major bank failure since th...	2	2024-03-20
3	Experienced and well-respected staff from arou...	2	2024-03-20
4	But the review also found that Federal Reserve...	2	2024-03-20
...	...	...	...
50070	The difficulty is partly that improvements alo...	2	2006-01-31
50071	Another part of the difficulty is determining ...	2	2006-01-31
50072	However, the exact dates are difficult to pin ...	2	2006-01-31
50073	With the growing use of mobile, battery-powere...	2	2006-01-31
50074	Because technology feeds into various macroeco...	2	2006-01-31

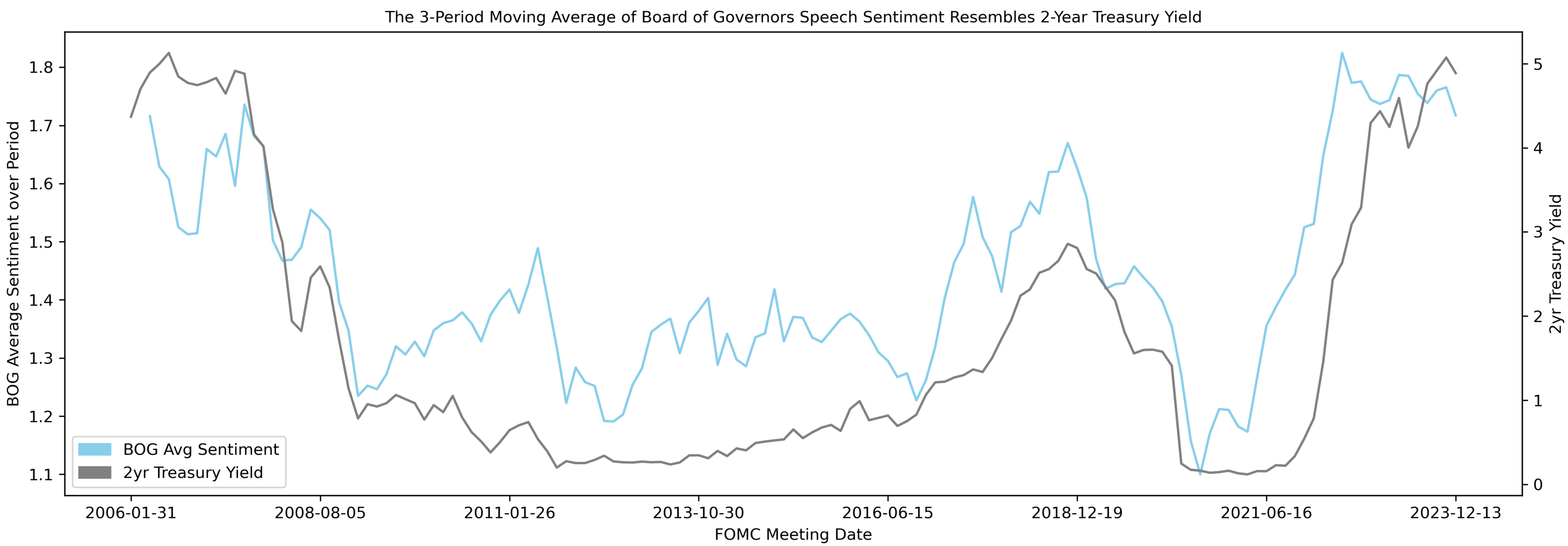
- Due to the nuance that comes with natural language, large datasets are a cornerstone of language modeling. More examples during training helps the model to generalize to unseen data.

Sentiment appears to proceed policy action.



Source: Board of Governors of the Federal Reserve System, Author's Calculations.

…And markets. But primarily during hiking campaigns.



Source: Board of Governors of the Federal Reserve System, Author's Calculations.

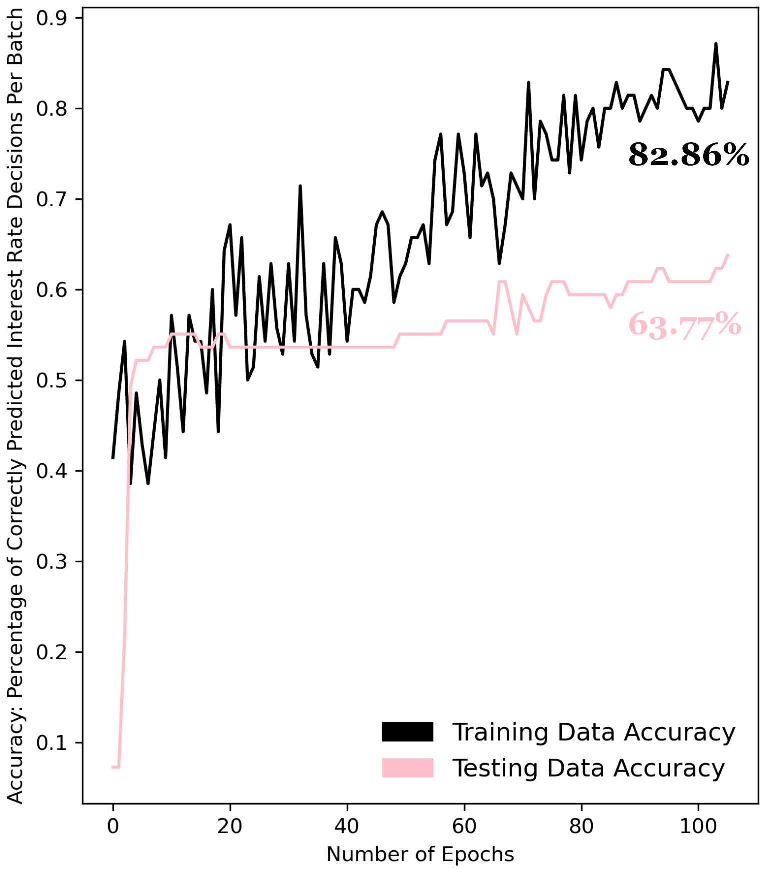
Economic variables and sentiment are grouped by meeting period.

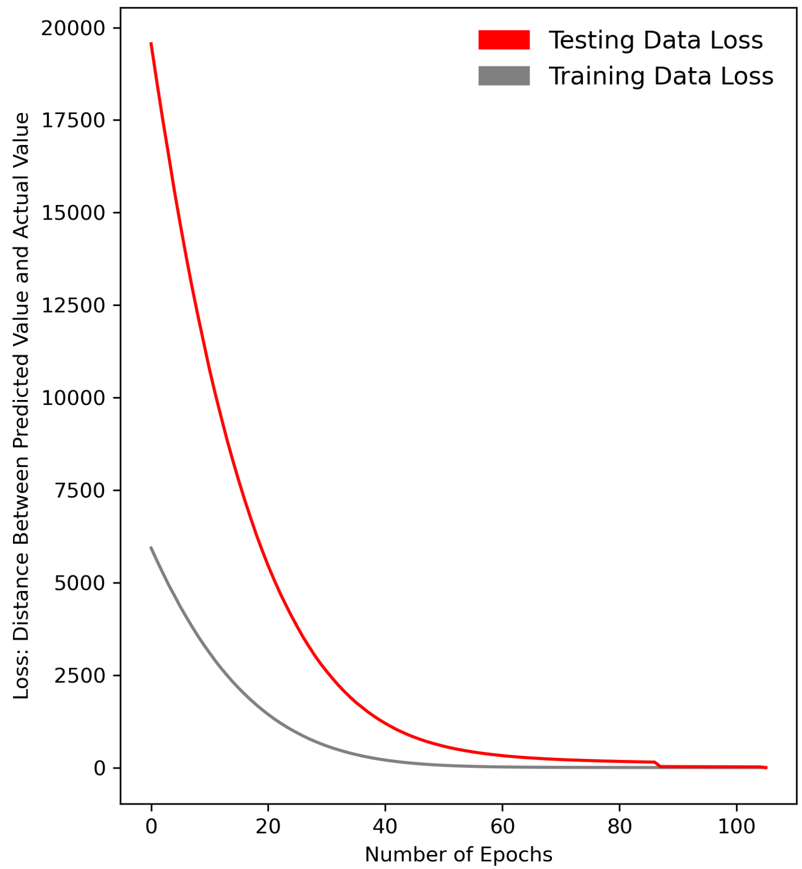
	period	avglabel_excludingneutral	PSAVERT	MORTGAGE30US	HOUST	PERMIT	UNRATE	U6RATE	CIVPART	LNS11300060	...	T10YIE	DFII5	DFII10
0	2006-01-31	1.916667	3.00	6.210	2273.0	2212.0	4.70	8.40	66.00	82.80	...	2.35	1.995	2.05
1	2006-03-28	1.619403	3.25	6.270	2044.0	2129.5	4.75	8.30	66.15	82.95	...	2.53	2.010	2.09
2	2006-05-10	1.611111	3.00	6.510	1881.5	1951.5	4.65	8.15	66.10	82.80	...	2.60	2.255	2.41
3	2006-06-29	1.656250	2.90	6.620	1802.0	1867.0	4.60	8.40	66.20	82.80	...	2.61	2.340	2.46
4	2006-08-08	1.555556	2.40	6.760	1693.5	1742.5	4.70	8.45	66.15	82.95	...	2.59	2.450	2.49
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
136	2023-06-14	1.696078	4.80	6.480	1418.0	1441.0	3.60	6.90	62.60	83.50	...	2.20	1.590	1.46
137	2023-07-26	1.805970	4.40	6.745	1451.0	1443.0	3.50	6.70	62.60	83.40	...	2.24	1.870	1.56
138	2023-09-20	1.777778	4.20	7.105	1330.5	1506.0	3.80	7.05	62.80	83.50	...	2.34	2.120	1.90
139	2023-11-01	1.712500	4.00	7.530	1376.0	1498.0	3.80	7.20	62.70	83.30	...	2.37	2.420	2.40
140	2023-12-13	1.660870	3.95	7.365	1539.0	1480.0	3.70	7.05	62.65	83.25	...	2.26	2.210	2.16

141 rows × 46 columns

Sentiment and economic conditions by period guide our predictions.

Model begins overfitting on training data after 106 epochs. Training is ceased at this point.





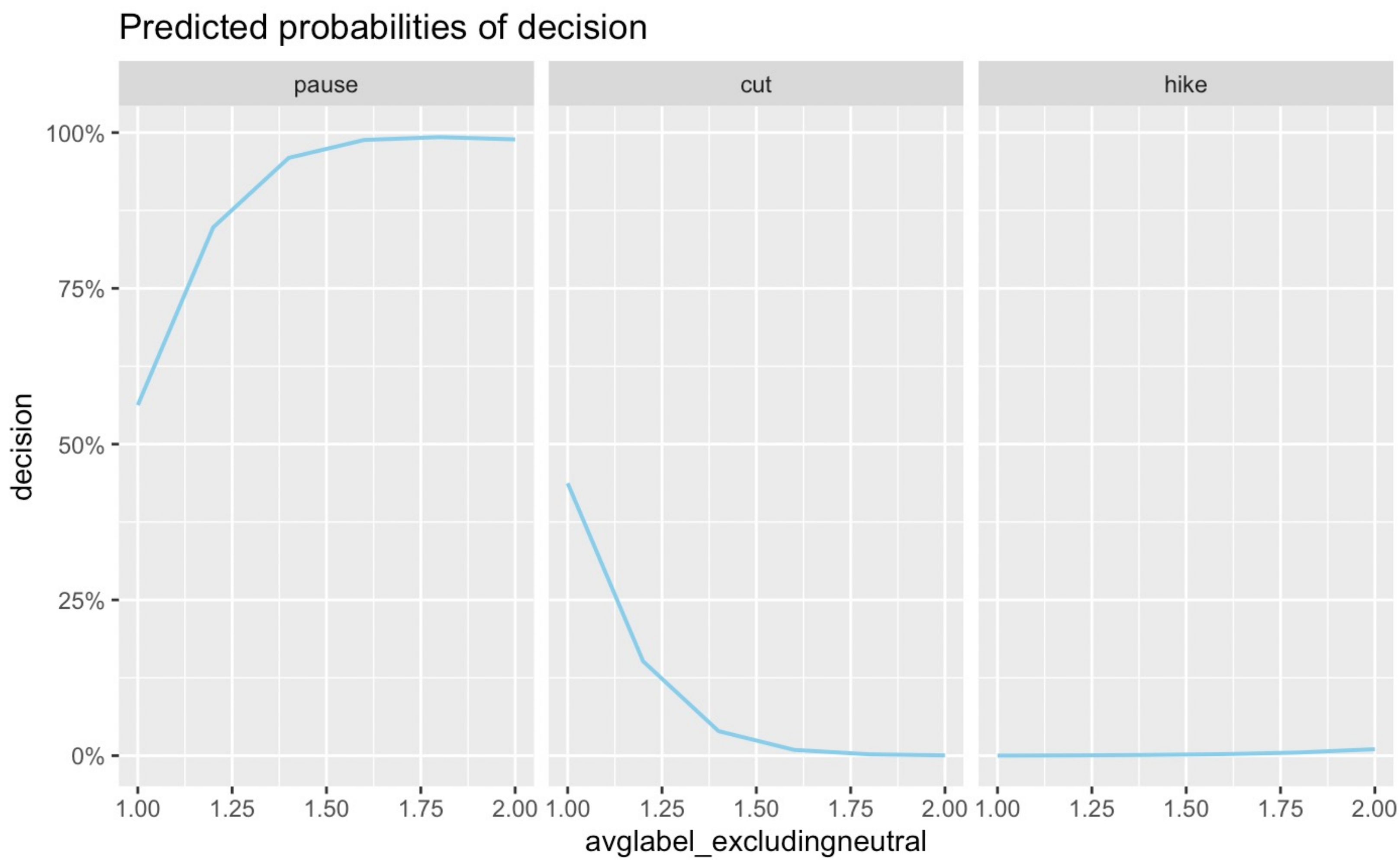
To estimate the impact of the speech sentiment data alone we can turn to a regression.

Variable of Interest

Group <chr>	Term <chr>	Contrast <chr>	Estimate <chr>	Std. Error <chr>	z <chr>	Pr(>|z|) <chr>	2.5 % <chr>	97.5 % <chr>
cut	avglabel_excludingneutral	mean(dY/dX)	-0.459	0.215	-2.139	0.0324	-0.880	-0.0385
hike	avglabel_excludingneutral	mean(dY/dX)	0.229	0.189	1.209	0.2268	-0.142	0.5996
pause	avglabel_excludingneutral	mean(dY/dX)	0.231	0.271	0.851	0.3945	-0.300	0.7619

- This is the partial output of a multinomial logistic regression which estimates the effect of the meeting period sentiment and control variables on the FOMC's interest rate decision.
- Control variables were selected based on the variables assumed to be most heavily considered in an FOMC meeting decision. (Incl. federal funds rate, inflation rate, inflation expectations for next five years, and unemployment rate.)

Looking a bit deeper, we find...



Concluding thoughts.

- Both our preliminary machine learning model and traditional regression suggest a meeting period's published speeches grant insight into committee member's attitudes.
- Attitudes, not only appear to impact interest rate decisions, but may also have a larger influence on committee member actions during times of economic stress when compared with actions in response to an overheating economy.
- One may argue that the latter case has a higher penalty for acting too early, which leads to reliance on data. The former causes undue hardship if not addresses expeditiously, leading to an increased chance of acting present opinions.

Looking ahead…

- Model accuracy can be enhanced by either, increasing data quality and quantity, or increasing model complexity.
- Once a sufficient amount of data is manually labeled, tailored results may follow. More freedom is granted when the user has entire control of input data.
- MTurk and other approaches will hopefully be utilized to both increase quantity of data and reduce risk of biased labels, due to subjectivity involved in deeming a statement “hawkish” over “dovish”.

Thank you!

Works Cited

Source of speeches (2006-2024):

https://www.federalreserve.gov/newsevents/speeches.htm

Quoted Officials:

Governor Kugler:

https://www.federalreserve.gov/newsevents/speech/kugler20240403a.htm

Chair Yellen:

https://www.federalreserve.gov/newsevents/speech/yellen20160329a.htm

Vice Chair Ferguson:

https://www.federalreserve.gov/boarddocs/speeches/2001/20010112/default.htm

Chair Greenspan :

https://www.oxfordreference.com/display/10.1093/acref/9780191843730.001.0001/q-oro-ed5-00005054#:~:text=Alan%20Greenspan%201926%E2%80%93&text=Since%20I've%20become%20a,have%20misunderstood%20what%20I%20said.&text=How%20do%20we%20know%20when%20irrational%2

Related Literature:

Chan, Ken. “Interpreting the Fedspeak: Text Analysis on FOMC Statements - BBVA Research.” *BBVA Research*, 27 Sept. 2016, www.bbvaresearch.com/wp-content/uploads/2016/09/160927_US_Fedspeak.pdf.

Zhang, Ken Zhan, and Qingyi Huang. “Our Patented Fed Sentiment Indicator.” *Morgan Stanley*, Morgan Stanley, 25 Apr. 2023, www.morganstanley.com/articles/mnlpfeds-sentiment-index-federal-reserve.

Adams, Travis Adams, et al. “More than Words: Twitter Chatter and Financial Market Sentiment.” *The Federal Reserve Bank of New York*, The Federal Reserve Bank of New York, 2023, www.federalreserve.gov/econres/feds/files/2023034pap.pdf.

Shah, Agam, et al. “Trillion Dollar Words: A New Financial Dataset, Task & Market Analysis.” *ACL Anthology*, ACL Anthology 2023, 13 May 2023, aclanthology.org/2023.acl-long.368.pdf.

Additional Info: Sentence Labeling

Manual Labeling (Lower accuracy; as of now)

- Web-scraped speeches are split by every second sentence.
- Context stays consistent: larger chunks have higher chance of introducing more than one topic.
- Also, length of input text must fall within model architecture limits.
- Two-sentence chunks of speech text are human labeled based on the implied sentiment.

Shah et al. RoBERTa Classifier (Higher accuracy)

- The Shah et al. classifier is a pre-trained model used to bucket text into hawk/dove/neutral categories.
- Similar to manual approach, model is trained on human annotated Federal Reserve text.
- Classifier is used to label personally scraped speech text.
- Pretrained model unlocks ability to classify, more quickly, a larger number of sentences than manual approach. More data will enhance model accuracy.

Model Specific Information, Reference During Q&A

```
In [15]: model=tf.keras.Sequential([
                         tf.keras.Input(shape=X_reshaped.shape[1:]),

                         tf.keras.layers.LSTM(5, activation='relu', return_sequences=True, kernel_initializer=HeNormal(), ker
                         tf.keras.layers.Dropout(rate=0.6),
                         tf.keras.layers.BatchNormalization(),

                         tf.keras.layers.LSTM(5, activation='relu', kernel_initializer=HeNormal(), kernel_regularizer=regular
                         tf.keras.layers.Dropout(rate=0.6),
                         tf.keras.layers.BatchNormalization(),

                         tf.keras.layers.Dense(4, activation='relu', kernel_initializer=HeNormal(), kernel_regularizer=regula

                         tf.keras.layers.Dense(3, activation='softmax')])

         model.summary()
```

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 1, 5)              1000

 dropout (Dropout)           (None, 1, 5)              0

 batch_normalization (BatchN  (None, 1, 5)             20
 ormalization)

 lstm_1 (LSTM)               (None, 5)                 220

 dropout_1 (Dropout)         (None, 5)                 0

 batch_normalization_1 (Batc  (None, 5)                20
 hNormalization)

 dense (Dense)               (None, 4)                 24

 dense_1 (Dense)             (None, 3)                 15

=================================================================
Total params: 1,299
Trainable params: 1,279
Non-trainable params: 20
```

Example of using Shah et al. Classifier, Reference During Q&A

```
classifier = pipeline('text-classification',
                      model=model, tokenizer=tokenizer,
                      config=config, framework="pt")

results = classifier('These are the unfortunate costs of reducing inflation. But a failure to restore price stability would mean far greater pain.',
                      batch_size=128)

results[0]['label']
```

```
'LABEL_1'
```