Algorithmic Examination of Monetary Policy Deliberations: an Analysis of the Informational Content of FOMC Meetings

J. Daniel Aromí
IIEP UBA-Conicet, FCE UCA

Daniel Heymann
IIEP UBA-Conicet

Abstract:

We apply natural language processing techniques to infer sentiment expressed in FOMC meetings. The sample period covers the Great Recession and its aftermath (2003-2012). We infer meetings' tone implementing large language models (BERT, NLI, ChatGPT) and traditional dictionary methods (Loughran & MacDonald 2011, Aromí 2020). Suggesting policymakers are advantageously informed, we find that tone in FOMC meetings anticipates media sentiment, consumers' confidence, and financial market dynamics. Furthermore, meetings' tone also anticipates growth forecast errors from Fed staff and private sector analysts. The findings are robust to changes in text processing methodologies and show a persistent anticipatory ability that extends over multiple quarters. We observe that, despite some discrepancies and evidence of underreaction, the tone of FOMC is closely replicated in meetings' minutes. Our analysis shows that, despite its availability, analysts fail to incorporate the information on policymakers' deliberations in an adequate manner.

Keywords: forecasts, monetary policy, transparency, underreaction

1. Introduction

A diverse collection of theoretical and empirical studies in macroeconomics shows an growing interest in the beliefs characterizing different economic actors (Malmendier & Nagel 2021, Angeletos et al. 2023, Farhi & Werning 2019, Greenwood & Shleifer 2014). At the same time, advances in natural language processing (NLP) techniques expand our ability to infer subjective features in macroeconomic contexts (Heymann & Sanguinetti 1998). These tools can be used to characterize economic assessments by key economic actors and describe the way these assessments are communicated. In this way, there are new opportunities to shed light on the rationale behind macroeconomic behavior and advance our understanding of the resulting macroeconomic dynamics.

In this work we use rich set of NLP techniques to estimate US monetary policymakers' assessments of the economy as reflected by the transcripts of FOMC meetings. The sample period covers the buildup of the housing bubble, the Great Recession, and its aftermath (2003-2012). In this way, we contribute a characterization of beliefs surrounding one of the most relevant events in macroeconomic history.

After inferring the tone of the deliberations, we use this estimate in a set of forecasting exercises that document the extent to which policymakers are advantageously informed. Complementarily, with a focus on transparency of communication, we analyze how this information is transmitted publicly through meetings' minutes. At the same time, we use this evaluation to characterize how economic analysts respond to the information released by policymakers.

We implement an agnostic NLP methodology that combines traditional techniques with novel large language models. That is, acknowledging that the performance of alternative methods is an unknown function of aspects such as model expressiveness and domain expertise, we characterize policymakers' deliberations using 5 different strategies. We consider two traditional dictionary methods that are adapted to the economic context (Loughran & MacDonald 2011, Aromí 2020). In addition, we consider 3 techniques based on large language models (LLMs). First, we use a finetuned model (BERT, Devlin et al. 2018) for sentiment classification. Also, we estimate sentiment though an LLM that was fine-tuned to perform natural language inference (NLI). Under this approach, an LLM is trained to identify entailment between pair of sentences, and, in this way, it can perform tasks such as text classification. Lastly, we use OpenAI's API to classify sentiment through ChatGPT.

Our findings indicating that the tone of policymakers' deliberations anticipates a rich set of indicators that shape macroeconomic and financial dynamics. The tone of deliberations of participants in the meetings contains information about subsequent media and consumer sentiment. Expected volatility and stock market returns are also anticipated by meetings' tone. Additionally, there is a positive association between tone and subsequent mean revisions in growth forecasts by Fed staff and private forecasters. From these findings, we conclude that policymakers seem to have informational advantage (either through access to information that private agents do not possess or a comparative advantage in interpreting data). Also, we find that a substantive amount of information on meetings' tone is transmitted through written minutes of the exchange of views. While we find some evidence of lags in incorporating information, the anticipatory ability of tone inferred from minutes is similar to that documented in the case of transcripts. On the other hand, we find that private professional forecasters fail to incorporate this information in their projections in a timely manner. That is, we document a new instance of inattention linked to information provided by policymakers.

This article is related to a growing literature that uses text analysis techniques in the context of monetary policy. An important part of these contributions has focused on the immediate impact of statements released at the time a policy is announced (Gürkaynak et al. 2005, Ehrmann et al. 2020, Cieslak & Schrimpf 2019). Similarly, Apel et al. (2022) uses a dictionary approach to estimate the policy instance transmitted by minutes (hawkish vs. dovish) and evaluate its information content in terms of the future policy decision. With a different focus, Shapiro & Wilson (2022) use FOMC meetings transcripts to approximate policymakers' preferences regarding inflation. Cieslak & Vissing-Jorgensen (2021) characterize the "Fed put" inspecting mentions of the stock market in FOMC transcripts. Sharpe et al. (2023) use a dictionary approach to characterize the information content of text in Greenbooks/Tealbooks. These are documents produced by Fed staff in preparation of FOMC meetings. Beyond text analysis, Gorodnichenko et al. (2023) show that the tone transmitted by the voice of the Chairperson during FOMC press conference contains information regarding subsequent economic and financial outcomes.

This article contributes to this literature through several contributions. First, we provide novel evidence on policymakers' foresight with a focus on the Great Recessions, the biggest economic perturbation since the Great Depression. Second, we provide a careful novel evaluation of monetary policymakers' communication. Complementarily, we establish that public information released by monetary authorities is not adequately incorporated by professional forecasters. At the same time, this article contributes informative evidence regarding the performance of alternative NLP techniques for alternative classes of monetary policy documents.

This article is organized as follows. In the next section, we discuss the data and methodology. In section 3 we present the results on foresight. We analyze transparency in section 4. Section 5 presents concluding remarks.

2. Data and methodology

In this study, we process a corpus of documents related to regular FOMC meetings spanning the years 2003 to 2012. Each year, 8 regular meetings are held, typically taking place during the first and last months of each quarter. In these meetings, participants discuss economic and financial conditions, monetary policy stance, and risks related to long-term goals. The participants in these discussions comprise the 12 FOMC members, along with Fed staff and Presidents of Federal Reserve Banks.

The first corpus used in this study correspond to meetings' transcripts. The 80 transcripts, which cover the sample period, are large documents with an average of 2713 sentences and 50105 words. There is a time lag of 5 years before this data becomes publicly available.

In addition to the transcripts, we also consider the minutes of these meetings for the same time span (2003-2012). The average size of these documents amounts to 210 sentences, totaling around 5726 words. Since 2005, these minutes become publicly accessible 3 weeks after each meeting. In the previous period, minutes became available at the time of the next regular FOMC meeting.

In this study, we employ natural language techniques to capture the underlying tone of FOMC meetings using transcripts and minutes. It is crucial to note that the efficacy of NLP in tasks like opinion classification depends on various factors such as domain expertise, model expressiveness, and corpus features.

Given the uncertainty surrounding the performance of alternative techniques of data analysis, we consider five distinct methods. First, we adopt two conventional dictionary-based approaches. Under this approach, tone quantification relies on the frequency of specific keywords. The dictionary methods implemented estimate the intensity of manifestation of negative tone and uncertainty. The list of 2345 negative words is a subjectively curated list tailored to the financial domain and was proposed in Loughran & MacDonald (2011). The list of 500 uncertainty-related words was identified using similarity measures from word embeddings trained on WSJ content, as proposed by Aromi (2017, 2020).

Additionally, we construct three indicators derived from the outputs of different large language models (LLMs). Compared to dictionary methods, LLMs have the benefit of being able to interpret words in their context and, in this way, grasp the meaning of discourse in a more comprehensive manner. The indices from LLMs are constructed, in each case, as follows:

i. BERT

One of the employed methodologies involves finetuning Google's BERT (Bidirectional Encoder Representations from Transformers). This is a deep neural network that implements a network architecture known as transformer. In our exercises, we use “bert-base-cased”, a pretrained specification of the model with 88 million parameters. For the finetuning process, 2000 sentences from the transcripts' corpus were manually labeled for sentiment using labels between 1 and 10. These sentences were randomly selected to ensure a representative sample. The finetuning procedure is implemented through the Huggingface "Transformer" library, which offers a comprehensive framework for training and adapting transformer-based models. (Reference: Devlin et al., 2018; Huggingface documentation: https://huggingface.co/docs/transformers/training). Given a document and the corresponding BERT classified set of sentences, the value of the index is equal to the average sentiment value for sentences with at least 10 words.

ii. NLI:

The concept behind natural language inference (NLI) involves training a language model to determine whether a given sentence (premise) implies or contradicts another sentence (hypothesis). A distinctive attribute of this approach is its "zero-shot learner" capability. This means that the model can perform tasks it hasn't been explicitly trained for, such as topic classification and sentiment analysis.

In this implementation, we employ Facebook's BART, which shares the architecture with BERT but comes with optimized training parameters. The finetuning data corresponds to the Multi-Genre Natural Language Inference dataset (MultiNLI, Williams et al. 2017). This dataset consists of 433k sentence pairs, each labeled for entailment or contradiction. This carefully curated dataset is expected to provide an adequate performance in a diverse set of contexts.

We use this model to classify the sentiment of sentences found in FOMC transcripts and minutes. We use the sentences to be classified as premise. The hypothesis is given by the template “This is an example of [BLANK] sentiment” where the blank space is filled by one of three labels: “negative”, “positive” and “neutral”. Given these inputs, the model responds with an output in the form of probabilities assigned to each label. Our indicator of sentence sentiment is equal to the difference between the probability assigned “positive” and “negative”. As in the previous case, for each document, the value of the index is equal to the average value for document sentences with at least 10 words.

Table 1: Sample sentence sentiment classification from NLI method (probabilities)

Sentence	Positive	Negative	Neutral
First, on financial markets and financial market functioning, I think as Bill Dudley suggested that they are much improved from our last meeting.	0.974	0.006	0.018
So all of this suggests to me that the outlook at the moment is not very promising, and a significant policy response on our part is appropriate.	0.014	0.966	0.020
I don’t know if the numbers will be revised up, but I would like to go cautiously, recognizing that they have been revised up in the past.	0.314	0.303	0.383

iii. ChatGPT

ChatGPT is a generative class of model with 175 billion parameters that has been finetuned to follow instructions. We access GPT-3.5-turbo using OpenAI’s API and the accompanying Python library (https://platform.openai.com/docs/api-reference/introduction).

We construct a prompt template to get sentiment labels for paragraphs of FOMC’s transcripts and minutes. The template we use is the following:

"The following text between quotes discusses economic policy and macroeconomic conditions: [[FOMC TEXT]]. Given the previous text, provide a rough estimate of its economic sentiment in the form of a number between 1 and 10 (where 1 indicates very negative and 10 very positive). Do not reply with text. That is, just provide a number (between 1 and 10) indicating an estimate of the economic sentiment”.

As in the previous cases, the tone of the document is given by the average sentiment assigned to the fractions of the document submitted through the API.

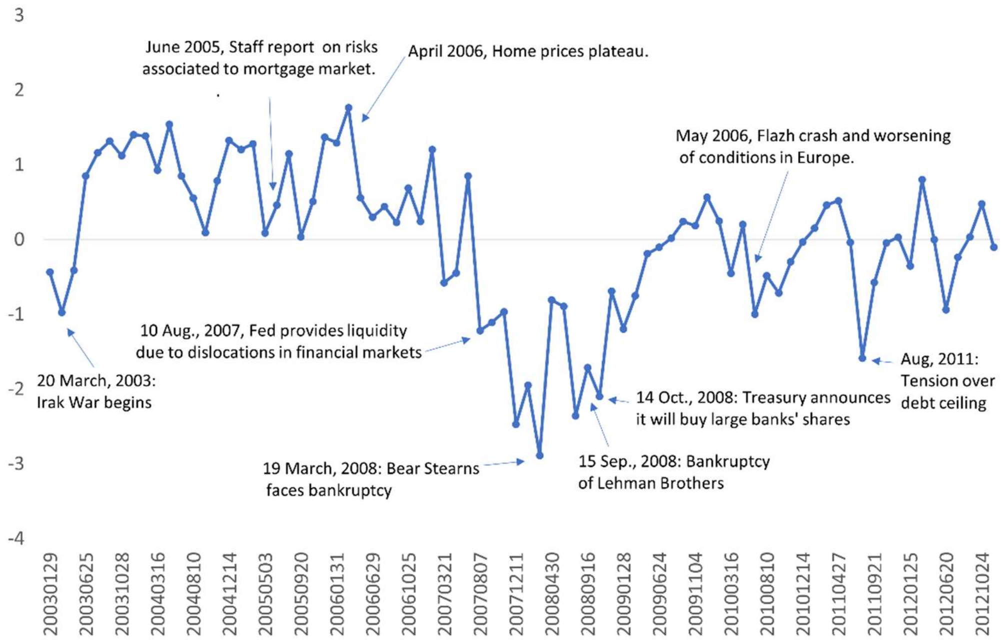
In this manner, we implement 5 methods to estimate the tone FOMC meetings. We integrate both traditional and contemporary NLP methodologies. In the evaluations presented below, the indices associated of each approach are combined after computing z-scores of each metric. The combination is implemented simply by adding the corresponding z-scores. In this way, we obtain three indices: “Tone” (the index that combines the 5 methods), “Tone_LLM” (the index that results from combining the output of 3 LLM-based estimations) and “Tone_Dict” (the combination of 2 dictionary methods).

It's important to consider that, given the objective of the present study, not all content within the transcripts may be equally relevant. This could result in introducing noise or inadequate weighting of different sections of the document. To address this concern in a simple manner, considering that the current study is centered around a recession triggered by a financial crisis, we build an alternative set of indicators focusing on sentiment of text discussing financial issues.

This alternative set of indices is formulated based solely on sentences classified as "finance-related". To identify the relevant sentences, in a first stage, they are categorized by topic through the application of NLI (Natural Language Inference). Out of 7 labels, the classification includes three financial labels: "finance," "credit market," and "asset market". A sentence is deemed "finance-related" if the sum of the probability assigned to the three financial topics surpasses the threshold of 0.75. Approximately 10% of sentences fall within this category, both in the case of transcripts and minutes.

Figure 1 shows the baseline indicator that combines the information resulting from 5 different NLP methods applied to the full transcript content. With the exemption of the first three meetings that took place around the beginning of the Irak war, in relative terms, the metric of sentiment is high in the years that precede the financial crisis. It is worth noting that in June 2005 the FOMC meeting included 5 presentations by Fed staff in which they discussed the implications of developments in the real-estate and mortgage markets. In line with the human assessment we carried out for this transcript, the positive value of the standardized index suggests that, at the time of this meeting, policymakers were not seriously concerned about these developments. Approximately 3 years after, in March 2008, the indicator of tone reaches a minimum value of -3 standard deviations in the context of the failure of Bear Stearns, a large broker-dealer. In the subsequent meetings the index increases. This is an interesting feature since by many indicators (e.g. expected volatility, stock indices) the worst part of the crisis took place by the end of 2008 and the beginning of 2009. Finally, it is worth observing that in the last years of the sample period, average tone is clearly lower than that observed in the early part of the sample.

Figure 1: Index of FOMC meeting tone



Notes: Standardized index resulting from the sum of z-scores of the metrics of FOMC meetings' transcripts under the five proposed NLP methodologies (2 dictionary methods and 3 LLMs).

2.1 Evaluation of information content

The empirical model employed in this study assesses the information content using a parsimonious approach. It can be expressed as follows:

$$Y_{m+h} = \alpha + \beta_{lag} Y_{m+h} + \beta Tone_m + u_{m+h} \quad (1)$$

Where m is the meeting index, the h-meeting-ahead outcome of interest is Y_{m+h} and $Tone_m$ is a metric of tone. The outcomes of interest considered in this study are: asset returns, expected volatility, media and consumer sentiment, and revisions in growth forecasts (Fed staff and private sector). The estimated parameter of interest is β, that is, the estimate of the change in the expected future outcome as a function of the present metric of FOMC meeting tone. As described above we consider multiple specifications of the metric of tone as a function of the NLP methodology, the FOMC type of document or the selection of text inside each document.

3. Analysis of foresight

Our study of foresight reveals that the tone within FOMC deliberations holds predictive value for multiple indicators of economic and financial conditions. For this diverse set of relevant indicators, a statistically and economically significant anticipatory ability is documented. Also, it is worth noting that the estimated slopes increase with forecast horizon. This suggests that the forward-looking information is not concentrated in the immediate future. On other words, our estimations show that FOMC meetings' tone anticipates economic news that arrive over the relatively long-time window.

First, we consider the anticipation of sentiment metrics transmitted by other economic actors. We estimate the model reported in equation 1 for alternative specifications of the tone index. Table 2 shows the estimated slope for the tone metric, $\hat{\beta}$, corresponding to the alternative model specifications. We consider consumer sentiment from the University of Michigan's household survey. The indicator is used in real time, that is, the value of the index corresponding to a given meeting is the last available observation at the time of the corresponding meeting. Also, we analyze media sentiment processing economic headlines from The Wall Street Journal. The index was computed using NLI following the same procedure as in the case of FOMC documents. The index associated to each meeting is equal to the average value during the most recent 28 days.

For these metrics of sentiment, a statistically and economically significant anticipatory ability is verified for different specifications of the FOMC tone metric. For example, an increase of one standard deviation in the index is followed by an increment of more than 0.4 standard deviations in expected consumer sentiment. In the case of WSJ sentiment, the expected increment is, for most specifications, above 0.3 standard deviations. It is worth noting that the results are quite similar under alternative specifications of the tone index.

Table 3 shows the estimated coefficients for two metrics of financial conditions. First, we consider a popular metric of expected stock market volatility: CBOE's Volatility Index (VIX). The estimates show that FOMC's tone anticipates the evolution of expected volatility. The sign of the estimated coefficient is inline with expectations. An increment in tone is followed by a drop in expected volatility. In addition, we verify that the different metrics of tone are able to anticipate stock returns as expressed by the Standard & Poor's 500 index. The estimated coefficients indicate that an increment of one standard deviation in tone is followed, eight-meeting-ahead, by an increment of approximately 6% in mean cumulative returns.

Table 2: Sentiment metric forecasts

A. Consumer sentiment (U. of Michigan Survey)

Transcripts: All

	h=1	h=2	h=4	h=8
Tone	0.244 ***	0.2355 ***	0.351 ***	0.4823 ***
	(0.05)	(0.08)	(0.11)	(0.14)
Tone_LLM	0.2655 ***	0.2347 ***	0.3037 **	0.3996 **
	(0.06)	(0.09)	(0.13)	(0.18)
Tone_Dict	0.1799 ***	0.198 ***	0.3472 ***	0.4944 ***
	(0.05)	(0.07)	(0.10)	(0.09)

Transcripts: Finance

	h=1	h=2	h=4	h=8
Tone	0.2385 ***	0.2566 ***	0.3537 ***	0.4845 ***
	(0.05)	(0.06)	(0.10)	(0.10)
Tone_LLM	0.2385 ***	0.2738 ***	0.3406 ***	0.4675 ***
	(0.05)	(0.07)	(0.12)	(0.14)
Tone_Dict	0.1912 ***	0.1864 ***	0.2978 ***	0.406 ***
	(0.04)	(0.06)	(0.10)	(0.11)

B. Media sentiment (The Wall Street Journal)

Transcripts: All

	h=1	h=2	h=4	h=8
Tone	0.262 ***	0.235 *	0.274 **	0.314 *
	(0.09)	(0.13)	(0.12)	(0.17)
Tone_LLM	0.242 ***	0.202 *	0.219 *	0.211
	(0.08)	(0.12)	(0.12)	(0.17)
Tone_Dict	0.243 ***	0.243 *	0.305 **	0.409 ***
	(0.09)	(0.13)	(0.14)	(0.14)

Trancripts: Finance

	h=1	h=2	h=4	h=8
Tone	0.246 ***	0.193	0.287 **	0.310 *
	(0.09)	(0.13)	(0.13)	(0.17)
Tone_LLM	0.219 **	0.140	0.264 **	0.221
	(0.09)	(0.13)	(0.12)	(0.18)
Tone_Dict	0.205 **	0.185	0.241 *	0.353 ***
	(0.09)	(0.12)	(0.14)	(0.13)

Notes: The table reports standardized estimated coefficients for the corresponding tone metric ($\hat{\beta}$). WSJ sentiment is computed for economic headlines using NLI positivity and negativity scores. Heteroskedasticity-autocorrelation robust standard errors reported in parenthesis.

Table 3: Financial market forecasts

A. Expected Volatility (VIX)

Transcripts: All

	h=1	h=2	h=4	h=8
Tone	0.9681	1.9312	3.2437	6.4912
	(0.71)	(1.42)	(0.71)	(0.71)
Tone_LLM	0.843	1.507	2.3103	4.2498
	(0.74)	(1.49)	(0.71)	(0.71)
Tone_Dict	1.0406	2.3419 *	4.272 **	9.1593 ***
	(0.66)	(1.27)	(0.71)	(0.71)

Transcripts: Finance

	h=1	h=2	h=4	h=8
Tone	1.2271 *	2.3394 *	4.2161 **	7.8264 **
	(0.71)	(1.34)	(2.11)	(3.67)
Tone_LLM	1.5243 **	2.6649 *	4.484 **	7.3162 *
	(0.75)	(1.38)	(2.19)	(3.84)
Tone_Dict	0.7085	1.6338	3.6177 *	8.0122 **
	(0.63)	(1.13)	(1.91)	(3.70)

B. Stock Market Returns (S&P 500)

Transcripts: All

	h=1	h=2	h=4	h=8
Tone	-2.125 *	3.931 *	-3.878 **	-7.5038 ***
	(1.15)	(2.37)	(1.57)	(1.39)
Tone_LLM	-2.257 *	-4.011	-3.708 **	-6.4343 ***
	(1.23)	(2.49)	(1.59)	(1.56)
Tone_Dict	-1.613	-3.183 *	-3.435 **	-7.5175 ***
	(0.89)	(1.92)	(1.56)	(1.27)

Transcripts: Finance

	h=1	h=2	h=4	h=8
Tone	-2.311 **	-3.830 *	-3.507 **	-6.6228 ***
	(1.10)	(2.03)	(1.51)	(1.62)
Tone_LLM	-2.787 **	-4.191 **	-3.457 **	-5.3714 **
	(1.27)	(2.07)	(1.50)	(2.14)
Tone_Dict	-1.298 *	-2.309 *	-2.587 **	-6.179 ***
	(0.66)	(1.20)	(1.25)	(1.60)

Notes: The table reports the estimated slope for the corresponding standardized tone metric ($\hat{\beta}$). Heteroskedasticity-autocorrelation robust standard errors reported in parenthesis.

In the last set of evaluations of foresight, we consider the ability to anticipate GDP growth forecasts revisions. We evaluate forecasts generated by the staff of the Board of Governors. These forecasts are distributed to FOMC members one week before each meeting and, hence, are revised, approximately, every 6.5 weeks. We analyze revisions in four-quarter-ahead cumulative growth forecasts. We estimate forecast models in which tone in meeting m, $Tone_m$, is used to anticipate the forecast revision observed in the next meeting, Rev_{m+1}. We also estimate a longer horizon forecast model in which the sum of two subsequent forecast revisions is anticipated.

In an extension of this analysis, we evaluate private sector forecasts. These projections correspond to the Survey of Professional Forecasts (SPF) and are collected by the Philadelphia Federal Reserve. Projections are released with a quarterly frequency around the middle of each quarter. Again, we analyze quarterly revisions in four-quarter-ahead cumulative growth forecasts. In these forecast exercises, the metric of tone corresponding to a quarter is given by that computed using the transcript of the first meeting of the corresponding quarter.

The estimated forecast models, reported in table 4, show a robust association between tone and subsequent revisions in growth forecasts. This association is verified both in the case of Fed staff forecasts and private forecasts. While the estimated coefficients are larger in the case of private forecasts, it must be noted that revisions to Fed staff forecasts occur twice more frequently.

As in the previous analysis, we observe that when the forecast horizon is extended, the size of the estimated coefficient increases. This indicates that the documented anticipatory ability is not concentrated in the immediate future. The information content of FOMC deliberations is seen to be spread over multiple subsequent periods.

In essence, our findings show the predictive nature of tone in FOMC deliberations. The documented regularities suggest FOMC members and Fed staff are advantageously informed. Complementarily, the reported exercises also result in methodological insights dealing with the performance of different NLP methods. When analyzing transcripts, we note consistent results between the application of large language models (LLMs) and dictionary-based methods. In principle, dictionary methods are at a disadvantage because they are not able to assess the meaning of a word in its context. Nevertheless, we conjecture that due to the large size of the analyzed documents, this disadvantage does not result in a perceptible difference in the ability to approximate meetings' sentiment.

Table 4: Anticipation of growth forecast revisions

A. Greenbook/Tealbook

Next meeting revision (Rev_{m+1})

	Transcripts: All			Transcripts: Finance		
	Tone	**Tone_LLM**	**Tone_dict**	**Tone**	**Tone_LLM**	**Tone_dict**
$\hat{\beta}$	0.109 *	0.110 *	0.095 *	0.134 **	0.151 **	0.090 **
	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.05)

Sum of revisions in next two meetings ($Rev_{m+1} + Rev_{m+2}$)

	Transcripts: All			Transcripts: Finance		
	Tone	**Tone_LLM**	**Tone_dict**	**Tone**	**Tone_LLM**	**Tone_dict**
$\hat{\beta}$	0.188 *	0.177	0.182 *	0.222 **	0.247 **	0.148 *
	(0.10)	(0.11)	(0.09)	(0.10)	(0.11)	(0.08)

B. Survey of Professional Forecasters

Next quarter revision (Rev_{q+1})

	Transcripts: All			Transcripts: Finance		
	Tone	**Tone_LLM**	**Tone_dict**	**Tone**	**Tone_LLM**	**Tone_dict**
$\hat{\beta}$	0.338 ***	0.310 **	0.340 ***	0.368 ***	0.363 ***	0.288 **
	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)	(0.13)

Sum of revisions in next two quarters ($Rev_{q+1} + Rev_{q+2}$)

	Transcripts: All			Transcripts: Finance		
	Tone	**Tone_LLM**	**Tone_dict**	**Tone**	**Tone_LLM**	**Tone_dict**
$\hat{\beta}$	0.548 **	0.478 **	0.598 ***	0.603 ***	0.580 ***	0.488 **
	(0.20)	(0.23)	(0.20)	(0.20)	(0.23)	(0.20)

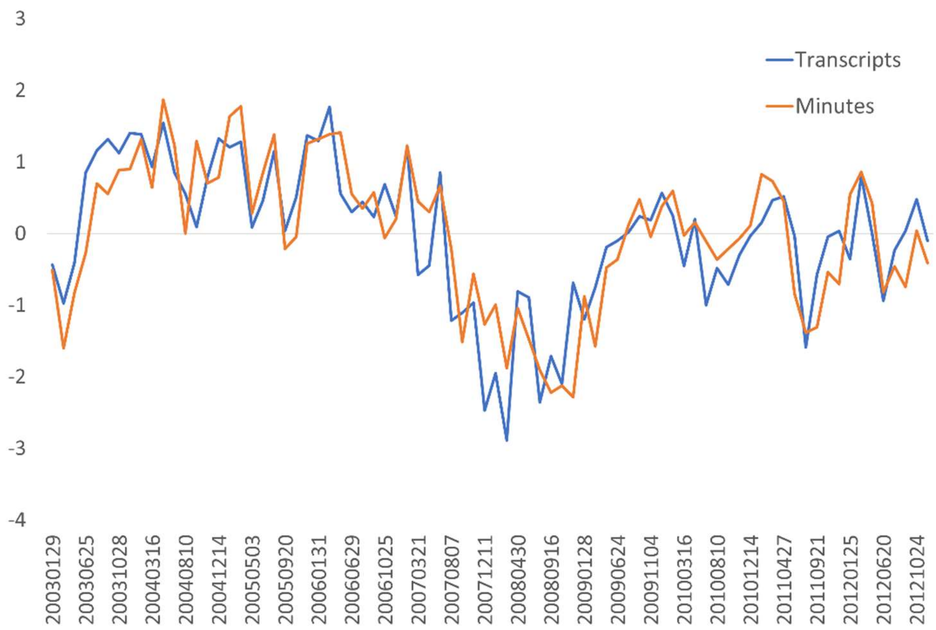
Notes: The table reports the estimated slope for the corresponding standardized tone metric ($\hat{\beta}$). Heteroskedasticity-autocorrelation robust standard errors reported in parenthesis.

4. Evaluation of transparency

In this section, we delve into the transmission of FOMC meeting tone information through publicly available minutes and its subsequent incorporation into economic projections by economic analysts. For that purpose, we compute indices of tone using text in minutes. Next, we carry out descriptive analysis and formal forecasting exercises to identify the information content of these indices versus those computed from transcripts' content. Our analysis of anticipatory ability is implemented for the case of private sector forecast revisions. By the real-time nature of the implemented forecast exercise, the analysis will simultaneously provide evidence on the extent to which economic analysts incorporate this information in an adequate manner.

Figure 2 shows that, despite some discrepancies, there exists a very close association between the two metrics of tone. We verify a high Pearson correlation coefficient of 0.84. This simple metric of association suggests that the transmission of meetings' tone by minutes is satisfactory. This strong association is indicative of a transparent policy. In the exercises presented below, we extend the analysis through some specific evaluations that will contribute to a more informative description of the association between minutes' and transcripts' tone.

Figure 2: Indices of FOMC meetings' tone – Trancripts vs. Minutes



Notes: Standardized indices resulting from the sum of z-scores of the metrics of FOMC meetings' transcripts or minutes under the five proposed NLP methodologies (2 dictionary methods and 3 LLMs).

Inspecting the discrepancies between the standardized indices shown in figure 2, an interesting pattern is observed around the years of the financial crisis. We observe that in the early stage of the financial crisis transcripts' tone is typically lower than minutes' tone. In contrast, during the late stages of the crisis, the index computed using minutes is lower that the index computed using transcripts. This pattern is compatible with instances in which meetings tone is reflection in the minutes with a delay.

Motivated by this observation, we implement a simple exercise that contributes to a more comprehensive characterization of the way in which minutes reflect transcripts' tone. More specifically, we conjecture the presence of information rigidities or underreaction. Under this conjecture, recent changes in meetings' tone will be only partially reflected in minutes' content. As a result, in this scenario, we would observe a positive association between recent changes in meetings tone ($\Delta Tone\_Transcr_m$) and the difference between transcripts' and minutes' tone ($Tone\_Transcr_m - Tone\_Min_m$). We evaluate this hypothesis through the following model:

$$Tone\_Transcr_m - Tone\_Min_m = \alpha + \beta \Delta Tone\_Transcr_m + u_m \qquad (2)$$

Table 6 shows the estimated coefficients. It can be verified that increments in tone are, on average, associated to a larger difference between the tone of transcripts vs minutes. This suggests that there is underreaction in the tone reported in the minutes. The second row of estimated coefficients suggests that this inefficiency is only observed for the most recent change, that is, the change between the current meeting tone and the immediately preceding meeting. In contrast to what we find in the analysis of foresight, the reported pattern is only observed for the shortest horizon.

Table 6: Changes in transcripts' tone and discrepancies

	All Text			Finance Related		
	Tone	Tone_LLM	Tone_Dict	Tone	Tone_LLM	Tone_Dict
$Tone_m - Tone_{m-1}$	0.3001 ***	0.2660 ***	0.376 ***	0.2966 ***	0.191 **	0.524 ***
	(0.05)	(0.04)	(0.08)	(0.11)	0.08	(0.14)
$Tone_m - Tone_{m-2}$	0.2408 ***	0.245 ***	0.292 ***	0.2941 ***	0.144	0.631 ***
	(0.06)	(0.04)	(0.09)	(0.11)	(0.10)	(0.18)

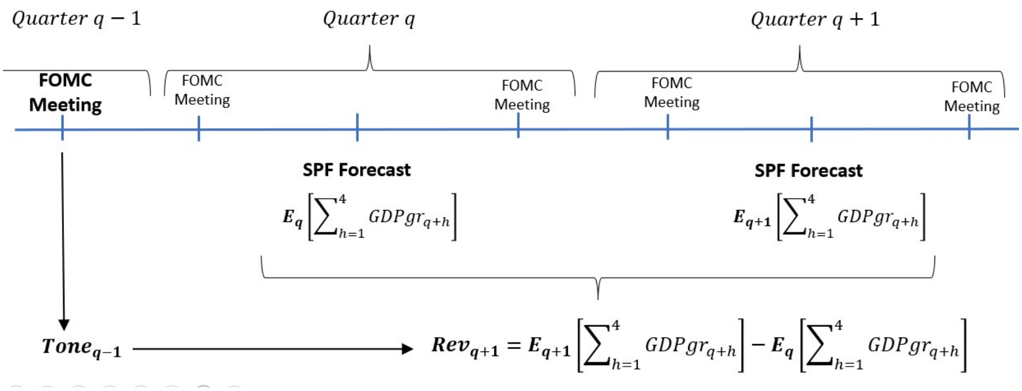
Notes: The table reports the estimated slope for the corresponding change in the standardized tone metric ($\hat{\beta}$). Heteroskedasticity-autocorrelation robust standard errors reported in parenthesis.

Another way in which the communication of meetings tone can be characterized involves comparing the anticipatory ability of minutes' vs. transcripts' indices. For this analysis we will consider the case of private sector forecast revisions. In contrast to financial or sentiment metrics, this is a metric with a straightforward interpretation.

It must be noted that this analysis is implemented in a form that will characterize the way in which analysts incorporate the information in FOMCs' minutes. That is, since we evaluate the ability to anticipate revisions in SPF growth forecasts in pseudo real-time. As a result, any anticipatory ability from minutes' content will indicate a scenario in which analysts fail to incorporate available information in an adequate manner.

Figure 3 shows a timeline that describes the real-time forecast exercise we implement. We use the tone of the minute corresponding to the second meeting of quarter $q-1$ to anticipate the difference between the forecasts submitted in quarters $q+1$ and q. If minutes' tone were incorporated in an adequate manner, there would be no ability to anticipate changes in expectations submitted after minutes information becomes available.

Figure 3: Timeline of FOMC meetings and SPF forecasts



This exercise involves processing the shorted texts of FOMC's minutes. We found that, in this context, Language Model Models (LLMs) yielded superior results compared to traditional dictionary-based methods. This is another contribution of the current article, we verify that LLMs are particularly valuable in the case of indices computed from short pieces of texts. In the analysis reported below, we focus on the case of indices resulting from LLMs.

Table 6 shows the results of the real-time forecasting exercise. As shown in the previous section, we verify that transcripts provide valuable information regarding forecast revisions. In addition, we find that minutes also provide information useful to anticipate revisions in GDP growth forecasts. In both cases, more positive tone anticipates upward revisions in GDP growth. In both cases, transcripts and minutes, this pattern is more precisely estimated in the case of indices computed using finance-related discussions. Suggesting relevant information is transmitted in an adequate manner, the estimated coefficient in the case of minutes' content is similar to that observed for indices computed using transcripts' content. Also, it is interesting to observe that the estimated coefficient increases with forecast horizon. Hence, in line with what we observed when we analyzed foresight, the information in minutes is not only relevant to anticipate revisions in the near future, but it is also relevant to project revisions observed 2 and 3 quarters ahead.

Table 6: Anticipating SPF forecast revisions in real time (transcripts vs. minutes)

	All text		Finance Related	
	Transcripts	Minutes	Transcripts	Minutes
Rev_{q+1}	0.244 **	0.165	0.225 **	0.211 **
	(0.12)	(0.12)	(0.10)	(0.09)
$\sum_{h=1}^{2} Rev_{q+h}$	0.441 *	0.296	0.472 **	0.368 *
	(0.23)	(0.22)	(0.21)	(0.20)
$\sum_{h=1}^{3} Rev_{q+h}$	0.611 *	0.383	0.700 **	0.560 *
	(0.34)	(0.29)	(0.31)	(0.31)

Notes: The table reports the estimated slope for the corresponding standardized tone metric ($\hat{\beta}$). Heteroskedasticity-autocorrelation robust standard errors reported in parenthesis.

5. Concluding remarks

The study's key findings shed light on the anticipatory ability of the FOMC's meeting tone, suggesting that policymakers are advantageously informed. Additionally, we find that, despite some information rigidities, tone gets conveyed through the minutes of the meetings; however, analysts currently fail to adequately incorporate this valuable information into their assessments. We also report insights regarding the performance of alternative NLP methods. While dictionary-based methods prove effective for larger texts, their relative performance diminishes when applied to shorter text pieces such as minutes.

Looking ahead, several promising avenues for research emerge. First, going beyond the presented metric of tone analysis, we see value in deriving and evaluating more specific indicators of transcripts' content. One such extension could involve identifying the information content of frequency of topics discussed in the meetings. Also, we could inspect the information content of tone conditioning on the time-orientation of the statements. We anticipate that the utilization of Language Models (LLMs) holds a significant role in implementing this type of extensions. The possibility of attributing statements to their authors, be it the Chair, members, or staff, presents another promising path for future investigation.

Looking further ahead, we consider that there is value in exploring how these tools can contribute to our understanding how misperceptions can drive the business cycle (Heymann & Sanguinetti 1998) and the appropriateness of different interpretations of equilibrium concepts (Heymann & Pascuini 2021). In a similar spirit, this algorithmic approach can advance the understanding of narratives as drivers of macroeconomic dynamics as proposed in Shiller (2017). For example, policymakers' discourse could be mapped into economic models or chains of thought. In this way, we could provide new insights regarding the rationale behind behavior and advance the interpretation of the resulting macroeconomic dynamics.

References:

Apel, M., Blix Grimaldi, M., & Hull, I. (2022). How much information do monetary policy committees disclose? Evidence from the FOMC's minutes and transcripts. *Journal of Money, Credit and Banking*, *54*(5), 1459-1490.

Aromi, J. D. (2020). Linking words in economic discourse: implications for macroeconomic forecasts. *International Journal of Forecasting*, *36*(4), 1517-1530.

Cieslak, A., & Schrimpf, A. (2019). Non-monetary news in central bank communication. *Journal of International Economics*, *118*, 293-315.

Cieslak, A., & Vissing-Jorgensen, A. (2021). The economics of the Fed put. *The Review of Financial Studies*, *34*(9), 4045-4089.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Ehrmann, M., & Talmi, J. (2020). Starting from a blank page? Semantic similarity in central bank communication and market volatility. *Journal of Monetary Economics*, *111*, 48-62.

Gorodnichenko, Y., Pham, T., & Talavera, O. (2023). The voice of monetary policy. American Economic Review, 113(2), 548-584.

Gürkaynak, R. S., Sack, B. P., & Swanson, E. T. (2004). Do actions speak louder than words? The response of asset prices to monetary policy actions and statements. *The Response of Asset Prices to Monetary Policy Actions and Statements (November 2004)*.

Heymann, D., & Pascuini, P. (2021). On the (In)consistency of RE modeling, Industrial and Corporate Change, Volume 30, Issue 2, Pages 347–356.

Heymann, D., & Sanguinetti, P. (1998). Business cycles from misperceived trends. *ECONOMIC NOTES-SIENA*, 205-232.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of finance, 66(1), 35-65.

Shapiro, A. H., & Wilson, D. J. (2022). Taking the fed at its word: A new approach to estimating central bank objectives using text analysis. *The Review of Economic Studies*, *89*(5), 2768-2805.

Sharpe, S. A., Sinha, N. R., & Hollrah, C. A. (2023). The power of narrative sentiment in economic forecasts. *International Journal of Forecasting*, *39*(3), 1097-1121.

Shiller, R. J. (2017). Narrative economics. *American economic review*, *107*(4), 967-1004.

Williams, A., Nangia, N., & Bowman, S. R. (2017). A broad-coverage challenge corpus for sentence understanding through inference. *arXiv preprint arXiv:1704.05426*.