Quantifying Economic Narratives:

The Transmission of FOMC Communication to Household Expectations via the Media*

By Eleftherios Bethmage and Johannes Zahner†

Households know little about central banking, yet research shows that they adjust their expectations around monetary policy announcements. We address this puzzle by showing that Federal Reserve inflation narratives – causal explanations for rising inflation articulated in Federal Open Market Committee (FOMC) communications – are transmitted to households via the media, thereby influencing how they form expectations. Using hand-annotated FOMC press conference transcripts, we evaluate large language models (LLM) in their capability to identify inflation narratives. We then link these narratives to contemporaneous New York Times coverage and household expectation data to trace how FOMC communication shapes household expectations. By employing an event-study IV approach, we exploit shifts in FOMC narrative emphasis to instrument high-frequency variation in media coverage. Each additional FOMC inflation narrative raises household inflation expectations by about half a percentage point through the media channel. The effect is concentrated among financially literate university graduates.

JEL: E58, C55, Z13, G50
Keywords: Central Bank Communication, Household Expectations, Media Transmission, Narratives, Natural Language Processing, LLM

* We are grateful for helpful comments from Michael Haliassos, Peter Andre, Refet Gürkaynak, Olivier Coibion, Claes Bäckman, Yigitcan Karabulut, Cristian Badarinza, Nils Wehrhoefer, and the FRGHF participants. We thank Mahesh Sathiamoorthy and Bespoke Labs for granting us academic access to the MiniCheck Python API.

† Affiliations: Bethmage, Zahner: Goethe University Frankfurt. Zahner: Marburg Centre for Institutional Economics. Declarations of interest: none (applies to all authors).

Corresponding author: Johannes Zahner, Goethe University Frankfurt, Chair of Macroeconomics and Finance, Theodor-W.-Adorno-Platz 3, 60323 Frankfurt am Main, Germany. E-Mail: johannes.zahner@gmail.com.

1. Introduction

During the Federal Open Market Committee (FOMC) press conference on 15 December 2021, Federal Reserve Chair Jerome Powell attributed the post-pandemic surge in the inflation rate to supply bottlenecks caused by the the COVID-19 recovery as follows:

> *"Supply and demand imbalances related to the pandemic and [to] the reopening of the economy have continued to elevated levels of inflation. In particular, bottlenecks and supply constraints are limiting how quickly production can respond to higher demand in the near term."*
> – FOMC Press Conference Transcript; December 15, 2021 — 14:00[1]

Three hours later, the New York Times (NYT) published an article echoing the same causal story – pandemic-related supply disruptions causing inflation – without referring to the FOMC or Powell by name:

> *"Inflation [...] is now soaring [...] The main trigger has been the jagged reawakening of economies that were largely shut down during the pandemic lockdowns during parts of 2020 and 2021. The surge in activity has caused supply-chain problems, hampered further by labor shortages [...]."*
> – New York Times; December 15, 2021 — 17:55[2]

Household expectations responded sharply. Within a day, the average inflation expectation in the New York Fed's Survey of Consumer Expectations (SCE) rose from 5.3% to 5.9%, while the variance among respondents fell from 19% to 13%. This paper shows that the pattern is not anecdotal but systematic. FOMC inflation narratives – causal explanations regarding inflation developments – are systematically transmitted through the media and shape household expectations in predictable fashion.

This transmission may help reconcile two conflicting empirical patterns in the literature on central bank communication. First, despite substantial efforts by policymakers to engage with households directly (Blinder et al., 2024), most Americans know little about central banking: many cannot name the Chair of the Federal Reserve or describe basic policy instruments (Kumar et al., 2015; Binder, 2017*a*). Second, despite this limited knowledge, survey experiments demonstrate the ability of central bank communication to affect household inflation expectations (Coibion, Gorodnichenko and Weber, 2022). Similarly, event studies find that households adjust their expectations around salient monetary policy events in the U.S. (Lewis, Makridis and Mertens, 2019; Binder, Campbell and Ryngaert, 2024), the euro area (Stiefel and Vivès, 2022; Bethmage, 2024), as well as in Australia (Claus and Nguyen, 2020).

[1] `www.federalreserve.gov/mediacenter/files/FOMCpresconf20211215.pdf` (accessed: July 1, 2025)
[2] `www.nytimes.com/2021/12/15/business/uk-inflation.html` (accessed: July 1, 2025)

One natural explanation for this apparent paradox lies in the role of the media in mediating this transmission. Traditional news outlets remain the predominant source of monetary policy information (Hayo and Neuenkirch, 2018; Conrad, Enders and Glas, 2022) and prior work has documented that news coverage of monetary policy announcements shapes household expectations through both its volume and tone (Berger, Ehrmann and Fratzscher, 2011; Lamla and Lein, 2014; Binder, 2017*b*). However, to the best of our knowledge no prior research has explicitly traced communication originating at the central bank through the media to household expectations. Our analysis provides the first evidence of this transmission channel for central bank–driven inflation narratives.

We focus on narratives given their pivotal role in framing households' perceptions of economic fluctuations and, consequently, expectations and economic decisions (Shiller, 2017). Recent work by Collier and Tuckett (2021) highlights the role of narratives within the careful communication of central bankers, whereas Andre et al. (2024) demonstrate their influence on household inflation expectations. We extend this line of research by investigating the transmission channel: whether the media acts as an intermediary through which central bank narratives are disseminated to households.

We make three contributions to the literature. First, we assemble two novel text corpora: FOMC press conference transcripts and the universe of NYT articles from 2019 to 2024. The press conferences are hand-annotated with Andre et al. (2024)'s inflation narratives, providing a gold-standard benchmark for narrative detection. We document strong prevalence of supply-sided narratives in the FOMCs communication regarding the post-pandemic hike in the inflation rate, corroborating Fraccaroli, Arel-Bundock and Blyth (2025)'s result while differentiating across distinct sources of supply pressure, such as Supply Chain Issues, Pandemic Impact and the Energy Crisis.

Second, we address the measurement challenge of identifying narratives in large text corpora. Existing NLP methods, such as dictionaries or topic models capture mostly topics but not causal explanations, i.e. narratives. However, *"a topic is not a narrative"* (Borup et al., 2023, p. 3). We assess existing methods along three criteria for narrative extraction in econometric text analysis: methods must produce reliable and valid indicators that identify causal direction, while being scalable for broader application. Existing approaches satisfy only a subset of these criteria each. To address this shortcoming, we adapt a state-of-the-art grounded factuality model (*MiniCheck*) from the computational linguistics literature. *MiniCheck* satifies all three criteria. In a horse-race prediction exercise comparing performance across methods, *MiniCheck* consistently outperforms the existing approaches.

Third, applying *MiniCheck* to the NYT corpus in an event-study framework, we show that FOMC narratives are transmitted through the media to households. Exploiting the sharp timing of pre-scheduled press conferences, we show that certain inflation narratives receive systematically more NYT coverage following a mention by the FOMC. On average, an inflation narrative discussed in a press conference leads to approximately one-quarter of an additional NYT article referencing the same narrative. Using FOMC narratives as an instrument for media amplification, we find that each additional narrative-related article raises household inflation expectations by about 1.5 percentage points and significantly reduces within-respondent dispersion. These effects are concentrated among financially literate respondents with a university degree.

Our paper connects three strands of the literature. First, we extend work on quantifying central bank communication using Large Language Models (LLMs) (Hansen, McMahon and Prat, 2017; Ferrara et al., 2022; Baumgärtner and Zahner, 2025; Bertsch et al., 2025; Geiger et al., 2025). Second, we build on research documenting inflation (narratives) in policymaker communication and the media (Carroll, 2003; Dräger, 2015; Ash, Gauthier and Widmer, 2024; Müller et al., 2022; Weinig and Fritsche, 2024; Fraccaroli, Arel-Bundock and Blyth, 2025). Finally, we contribute to the growing literature on the economic effects of narrative framing (Shiller, 2017; Reinhart and Rogoff, 2009; Borup et al., 2023), and in particular to recent work linking inflation narratives to household expectations (Andre et al., 2024).

The remainder of the paper is organized as follows. Section 2 defines inflation narratives and their role in policy communication. Section 3 establishes our two novel text corpora used the evaluation and the empirical exercise. Section 4 discusses currently used methodologies of text quantification and introduces *MiniCheck*. Section 5 evaluates the presented methods systematically. Section 6 presents the economic application, studying the diffusion of FOMC narratives into media and household expectations before Section 7 concludes.

2. Defining Economic Narratives

We begin by distinguishing topics, stories, and narratives. A *topic* refers to an economic concept, such as inflation, interest rates, or the Financial Crisis, that may be expressed through different terms (e.g., 'rising prices,' 'CPI,' 'inflationary pressure'). A *story* connects two or more topics in chronological order. For instance, 'inflation rose and then interest rates rose.' A *narrative* adds causality to the story, it explains *why* the sequence occurred. Following Roos and Reccius (2024):

> *"A narrative is a special kind of story [... that] has a deeper meaning for the speaker and the listener [...] told with the intention to understand the world and to interpret some data, event or action."*

Narratives thus provide a causal interpretation of why events occur rather than providing a mere chronology – a distinction that matters empirically. To illustrate this, consider the following statements about the two topics *inflation* and *labor markets*:

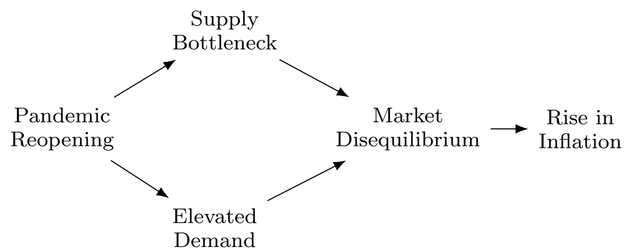
> **Statement 1:** In today's policy brief, the Governor will discuss recent trends in *inflation* and the extraordinarily *tight labor market.*
>
> **Statement 2:** The recently *tight labor market* raised wages and, in turn, pushed up consumer price *inflation.*

Statement 1 merely lists two topics without specifying a relationship between them. By contrast, Statement 2 establishes a causal link. Because knowledge of causal relationships enables individuals to make predictions and thus form expectations (Shiller, 2017), detecting narratives requires identifying causal direction – that is, distinguishing between mere co-occurrence of topics, as in Statement 1, and causal accounts, as in Statement 2.
We focus on *backward-looking* narratives that offer an explanation for the *rise* in the inflation rate. This choice has two advantages. It restricts attention to realized events and provides directionality, thus rendering an ex post separation of increases and decreases in the inflation rate narrative unnecessary. We adopt the taxonomy of household inflation narratives developed by Andre et al. (2024, Table 1), who elicit and systematize causal accounts for inflation drivers from a large-scale survey experiment. The authors find a broad set of twelve candidate explanations for the recent inflation surge, including fiscal policy (government spending, debt levels, taxes), monetary policy, demand-side accounts (pent-up demand, demand shifts), supply-side accounts (supply chains, labor shortages, energy crisis, pandemic disruptions, war in Ukraine), and expectation-based mechanisms (inflation expectations, price-gouging, government mismanagement, base effects). In addition, two residual categories capture supply and demand narratives not covered by the narrative types listed above.
To conceptualize this, consider once more the paragraph from the December 2021 FOMC press conference:

> *"Supply and demand imbalances related to the pandemic and [to] the reopening of the economy have continued to contribute to elevated levels of inflation. In particular, bottlenecks and supply constraints are limiting how quickly production can respond to higher demand in the near term."*
>
> — FOMC Press Conference Transcript, December 15, 2021

This is a bona fide narrative: it specifies an outcome (an elevated level of inflation) and articulates a causal chain (pandemic reopening leading to supply constraints and higher demand). Eliaz and Spiegler (2020) and Andre et al. (2024) formalize such causal structures using directed acyclic graphs (DAGs), as illustrated in Figure 1, where nodes represent variables and arrows denote perceived causal links. What defines a narrative is not the set of nodes alone, but the direction of the arrows. Accordingly, the DAG highlights two key requirements for valid narrative inference: (i) identifying the relevant topics (nodes) and (ii) establishing causal direction (arrows).

Figure 1 : DAG linking pandemic reopening to inflation



The current literature has focused on unsupervised (machine-learning) methods to identify narratives. For example, dictionaries link the frequency of negative words in business reports to business cycle fluctuations (Flynn and Sastry, 2022), while so-called topic models have been applied to measure central bank communication (Hansen, McMahon and Prat, 2017; Hansen, McMahon and Tong, 2019) and inflation-related media coverage (Müller et al., 2022; Borup et al., 2023; Weinig and Fritsche, 2024). The transmission of central bank communication to households has also been studied using topic models: Ter Ellen, Larsen and Thorsrud (2022) show that narrative surprises in central bank statements affect subsequent media coverage, and Weinig and Fritsche (2024) find that supply-side narratives in the media Granger-cause household expectations. However, while such unsupervised methods are well suited for identifying topics (the nodes) they cannot detect the causal direction (the arrows).

3. Text Corpora

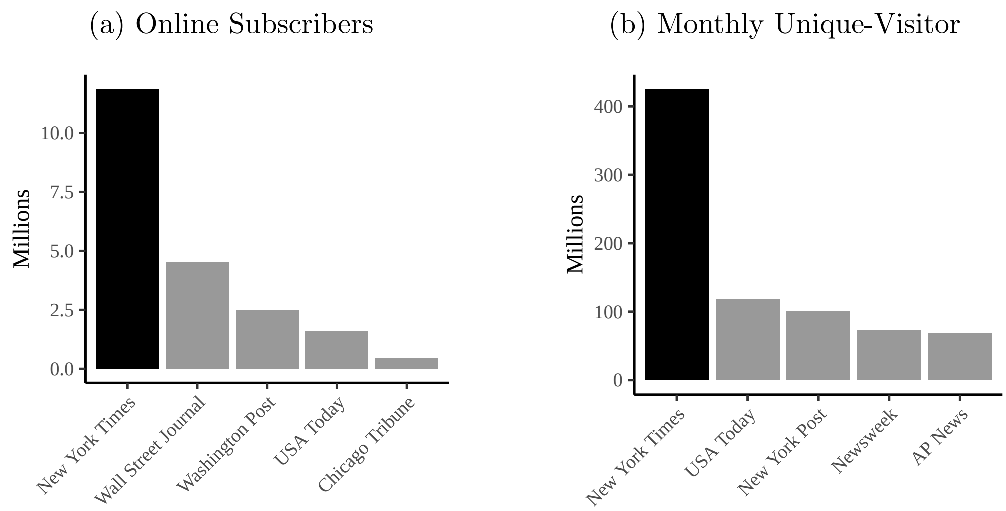
Next, we introduce the two text corpora that underpin our analysis. First, we compile the transcripts of all FOMC press conferences held between January 2019 and December 2023. Among the Fed's various communication channels, these press conferences receive the most media attention and serve as the primary way the Committee explains and justifies its monetary policy decisions to the public (Binder, 2017*b*).[3] Their sharp predetermined timing enables a clean event-study design, in line with the literature on high-frequency identification around monetary policy announcements (Gürkaynak, Sack and Swanson, 2005; Nakamura and Steinsson, 2018; Bauer and Swanson, 2023).

Second, we collect the universe of NYT articles over the same horizon to measure changes in the information set available to households around each press conference. We focus on traditional print media – and in particular on the NYT – rather than social media (e.g., Ehrmann and Wabitsch, 2022; Angelico et al., 2022; Born et al., 2024; Masciandaro, Peia and Romelli, 2024) or TV broadcasts

[3] Press conferences are held immediately after the release of the FOMC's policy decision, which itself reports the outcome but provides little reasoning.

(e.g., Binder, Frank and Ryngaert, 2025; Çekin and Polattimur, 2025) for three reasons. First, traditional news outlets remain households' primary source of information about monetary policy (Hayo and Neuenkirch, 2018; Conrad, Enders and Glas, 2022), and the main channel through which they are exposed to inflation narratives (Andre et al., 2024), Second, any performance benchmark derived from the hand-labeled textual dataset of FOMC press conferences is most valid when applied to similarly structured written texts. Accordingly, focusing on written news articles ensures methodological consistency and direct comparability for the narrative-detection algorithms. Finally, the NYT is the largest subscription newspaper in the United States. Its paid readership and online traffic exceeds that of the next four outlets combined (Figure 2). Accordingly, the NYT offers the most representative coverage of traditional print information and best approximates the households' media information set. Moreover, as shown by Herbert, Istrefi and Sagna (2024), cross-news-outlet variation declines systematically around FOMC press conferences, further strengthening the NYT's validity as a measure of media coverage during the event window.

Figure 2 : Top Five Newspapers by Online Subscribers and Traffic in 2025



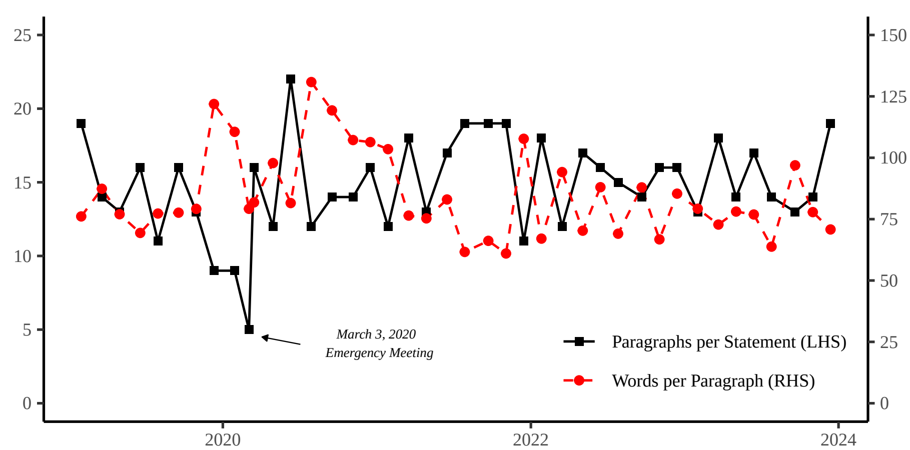
Notes: Data on Online Subscriptions are sourced from each newspaper's most recent earnings call or financial disclosure. For example, the New York Times Company's Q2 2025 report is available at nytco-assets.nytimes.com/2025/08/Q2-2025-Earnings-Release.pdf. Figures on Monthly Unique Visitors are obtained from Press Gazette.

FOMC Corpus

We commence our analysis in January 2019, when Federal Reserve Chairman Jerome Powell initiated that press conferences would take place following every FOMC meeting. This marked a departure from the Bernanke-era practice, where press conferences were limited to meetings accompanied by the release of the Summary of Economic Projections (De Pooter, 2021). Our dataset includes all regularly scheduled press conferences up to December 2023, along with the two unscheduled meetings held in March 2020 at the onset of the COVID-19 pandemic. Altogether, the FOMC corpus covers 41 events.

We segment each press conference transcript into paragraphs based on the Fed's official formatting, which follows a consistent structure progressing sequentially through distinct economic topics. Inflation narratives are therefore concentrated within dedicated sections, making the paragraph a natural unit of analysis. This level of segmentation preserves causal direction and cross-sentence reasoning while providing within-document variation. Paragraph-level segmentation expands the dataset from 41 press conferences to 723 paragraphs, thereby increasing statistical power and precision. Importantly, we find no systematic changes in paragraph length over time (Figure 3), alleviating concerns that shifts in communication style might confound our analysis.

Figure 3 : FOMC Statements (2019–2024)



NYT Corpus

To capture the diffusion of inflation narratives into public discourse, we construct a corpus of NYT articles published around FOMC press conferences, retrieved via the official NYT Article Search API.

We filter for articles containing monetary policy and inflation-relevant terms such as 'Federal Reserve', 'FOMC', 'inflation', and 'deflation'. To ensure comprehensive coverage, we also include linguistic variants (e.g., 'inflationary') and references to recent chairman ('Powell', 'Yellen', 'Bernanke'). In addition, we include all articles authored by the NYT's Federal Reserve correspondents, regardless of keyword presence. Supplementary analysis confirms that inflation narratives are virtually absent outside this sample.
Our final NYT corpus comprises 15,605 articles, 491,263 paragraphs, and roughly 19 million words. Table 1 compares the two matched corpora. While the average document length is similar, the NYT corpus contains nearly 700 times more paragraphs, underscoring the need for automated methods. Paragraphs also differ in style. Technical FOMC paragraphs average 81 words, compared with only 42 in the NYT, reflecting the mix of reporting, analysis, and commentary in journalism. Despite these differences, the paragraph remains the natural unit of analysis, as causal arguments span multiple sentences but rarely extend across paragraphs.

Table 1: Summary Statistics

	Number of:				
	Documents	Paragraphs	Words	Words/Document	Words/Paragraph
FOMC	41	723	58.600	1.429	81
NYT	15.605	491.263	19.316.062	1.242	42

4. Methodologies

The quantification of economic narratives can be pursued through two paradigms: *unsupervised* and *supervised* methods. Their goals differ fundamentally. Unsupervised approaches, such as word-list and topic models, are suited for exploratory tasks, where the aim is to discover patterns in a corpus without prior assumptions. By contrast, supervised methods target verification where researchers specify candidate labels ex ante and then test whether these labels can be predicted from the text. This latter paradigm aligns closely with our objective of detecting the predefined inflation narratives of Andre et al. (2024) in central bank communication and subsequent media coverage. Following manual classification as our baseline gold standard, we organize the methodological survey around this supervised-unsupervised distinction.

4.1. Manual Classification

Manual classification—actual reading of statements, briefings, or news articles and hand-coding them by expert judgment—remains the gold standard. Romer and Romer (2023), building on their earlier "narrative approach" to monetary (Romer and Romer, 2004) and fiscal policy shocks (Romer and Romer, 2010), emphasize that while classification can, in principle, be delegated to research assistants or algorithms, feasibility depends on the sophistication of the task. They distinguish between relatively simple exercises, such as coding the tone of Fed discussions, and more complex interpretative work that requires identifying causal relationships and subtle policy intentions. Narrative detection falls into the latter category.

We therefore begin with a manually coded benchmark. We independently coded the complete FOMC corpus, examining all 723 paragraphs for the presence of the twelve inflation narratives defined by Andre et al. (2024). The definition of a narrative was agreed upon ex ante, and paragraphs were randomized to prevent spillover bias. Each paragraph was classified on a binary scale for each narrative, reflecting the non-exclusive nature of causal explanations for inflation. The coding interface can be seen in Figure B1 in the Appendix. Coding disagreements were rare and were resolved jointly.[4] This manual classification required roughly 13 hours, represents a substantial time investment, which underscores both the cognitive complexity of narrative identification and the scalability constraints that motivate automated approaches.

The final dataset constitutes our ground truth: a human-validated mapping of inflation narratives across FOMC communication. This manual benchmark provides the foundation for evaluating automated approaches in the next section. Table 2 summarizes the results. Out of the twelve inflation narratives identified by Andre et al. (2024), only a subset appears with notable frequency in FOMC communications. For our analysis we restrict attention to narratives occurring in more than 1% of all paragraphs, effectively the six most prevalent narratives, since methods cannot be reliably evaluated on only two observations.

Similar to Fraccaroli, Arel-Bundock and Blyth (2025, Figure 3), Table 2 shows that most FOMC inflation narratives emphasize supply-side shocks and the labor market. Specifically, most narratives center on the COVID-19 pandemic, supply chain disruptions, labor shortages, and the energy crisis. By contrast, some narratives are entirely absent. Price-gouging, for instance, does not appear at all – consistent with Andre et al. (2024)'s finding that households, but not experts, advance this explanation. Similarly, narratives centered on government spending are absent, despite the period of large-scale pandemic relief programs. This suggests that households acquire such views from sources outside central bank communication.

Figure 4 documents the temporal evolution of narratives in FOMC communication. Pandemic-related explanations dominate early in the sample, while energy and war-related narratives only gain prevalence after the Russian invasion of Ukraine in 2022, reflecting contemporaneous geopolitical and commodity-price shocks. Consistent with a pattern of stable prices – apart from a brief uptick in mid-2019

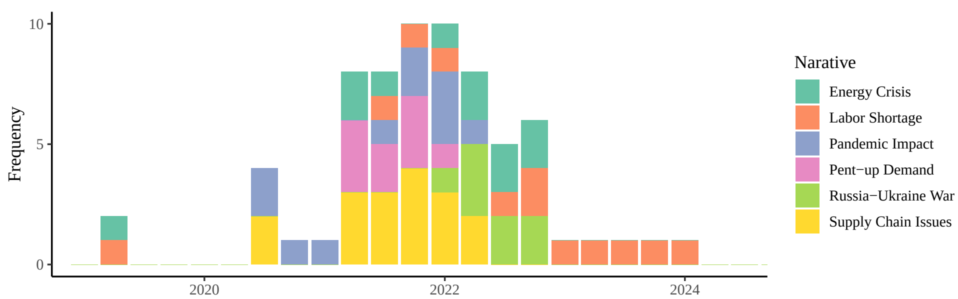
[4]Disagreements between coders were documented and discussed collaboratively. Most misalignments arose from edge cases where the language did not clearly distinguish between backward- and forward-looking instances.

Table 2: FOMC Press-Conference Inflation Narrative Count (2019-2023) retrieved by manual classification

	Narrative	Frequency	Relative Frequency in %
1	Supply Chain Issues	17	2.4
2	Labor Shortage	12	1.7
3	Pandemic Impact	11	1.5
4	Energy Crisis	11	1.5
5	Pent-up Demand	9	1.2
6	Russia-Ukraine War	8	1.1
7	Demand Shift	2	0.3
8	Loose Monetary Policy	2	0.3
9	Supply Shocks (MISC)	2	0.3
10	Government Spending	0	-
11	Inflation Expectations	0	-
12	Price-Gouging	0	-

– and our focus on narratives that explain *rises* in the inflation rate, inflation narratives are nearly absent both before the pandemic and again in 2023, when inflation started to decline back to target levels.

Figure 4 : FOMC Press-Conference Inflation Narratives 2019-2024



Notes: Quarterly narrative frequencies retrieved from manual classification as outlined in Section 4.4.1. The low–frequency categories 'Demand Shift', 'Loose Monetary Policy', and 'Supply Shocks (Misc.)' have been omitted. Data cover manually coded FOMC statements from January 2019 to December 2024.

4.2. Unsupervised Methods

Dictionaries

The simplest automated approach is word counts, so-called dictionary-methods. Applications to narratives include Shiller (2017), who counts mentions of the bigram *'Laffer Curve'*, and Flynn and Sastry (2024) using a positive-negative-word list by Loughran and McDonald (2011). In the context of central bank communication, however, such dictionaries have been shown to be dominated by noise (Hayo and Zahner, 2023), and no dictionary exists that meaningfully distinguish across different inflation narratives. We therefore exclude dictionary approaches from our main evaluation but show in Section 6 that our results are robust to the inclusion of inflation-related article counts.

Topic Modeling

The current literature on narrative quantification predominantly employs so-called topic models to extract multidimensional representations – typically ranging from 10 to 50 latent dimensions, or 'topics' – from text corpora (e.g., Hansen, McMahon and Prat, 2017; Hansen, McMahon and Tong, 2019; Müller et al., 2022; Herbert, Istrefi and Sagna, 2024; Weinig and Fritsche, 2024). The most widely used approach is Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003), which assumes that each document is a probabilistic mixture of topics and each topic a probabilistic mixture of words.[5]

In practice, researchers infer a topic's latent meaning from its word–topic distribution. For example, Figure 5 (Panel a) displays the word distribution of topic 28 from the LDA model trained on the FOMC corpus. The prominence of inflation-related terms indicates that this topic centers on inflation. However, it is important to note that this interpretation is inherently subjective, as it relies solely on the frequency and co-occurrence of words within the topic rather than on any underlying linguistic structure or causal relationships. Although the term inflation must be featured heavily in this topic, the context in which it is discussed can only be inferred since the model, by construction, cannot differentiate between Type 1 and Type 2 like statements.

Given this constraint, each document is represented by a vector of topic probabilities, where each element corresponds to the share of a given topic in that document. For example, the following vectors correspond to two paragraphs from the December 15, 2021 FOMC Press Conference referenced earlier:

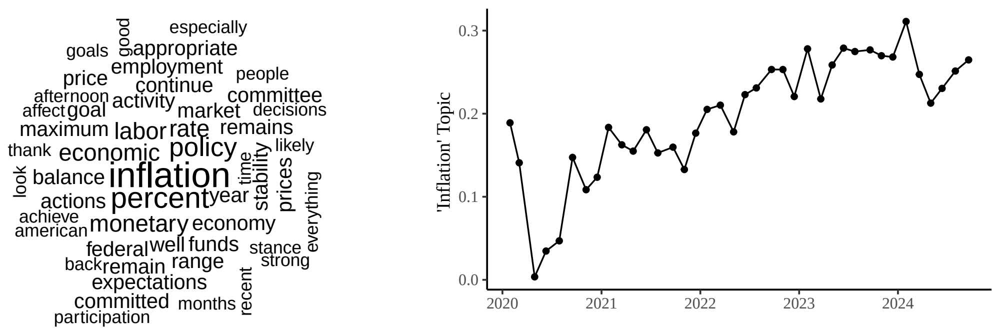
$$\mathrm{v}_{2021/12/15\ \textit{Paragraph}\ 4} = [0.012, 0.214, 0.005, ...]$$
$$\mathrm{v}_{2021/12/15\ \textit{Paragraph}\ 5} = [0.006, 0.013, 0.006, ...]$$

This topic–document distribution can be used to construct a time-series index of topic prevalence, whose temporal evolution (Figure 5, Panel b) may serve as input into a regression that estimates whether a press conference features any of the narratives defined by Andre et al. (2024). This yields a data structure aligned with our manual labels, which serve as the benchmark for evaluating the automated approach.

[5] Our discussion focuses on LDA, but the same considerations apply to more advanced topic-modeling techniques.

Alternatively, one may remain agnostic about the interpretation of individual topics and include all topic dimensions in a regression model jointly – typically with regularization – to 'let the data to speak', as in Larsen, Thorsrud and Zhulanova (2021). We adopt this latter approach for LDA in Section 5.5.2.

Figure 5 : LDA derived FOMC Inflation Topic



(a) Word Cloud (Topic 28)

(b) Topic Intensity Over Time

Notes: The LDA model is estimated on the FOMC corpus. Text was pre-processed with removal of numbers, punctuation, symbols and English stop-words. Tokens with term-frequency < 5 or document-frequency < 2 were discarded. We set $\alpha = 0.1$. The Inflation topic is selected by keyword–based matching. The time-series traces the mean share of Topic 28.

The distributional assumptions on both word and topic distributions imposed by LDA are strong. Much of the applied literature adopts a standard parameterization by Griffiths and Steyvers (2004) and Kintsch (2006), which result in nearly uniform distribution of topic shares across documents. We illustrate the effect of this prior choice in the Appendix B.B2. However, the notion that every narrative is equally represented across texts is inconsistent with our FOMC evidence, where narrative prevalence varies markedly and no narrative appears in more than 2.5% of all paragraphs.

RELATIO: SEMANTIC ROLE LABELING

A recent methodological advancement designed specifically for narrative quantification is *relatio* (Ash, Gauthier and Widmer, 2024), an unsupervised method that extracts narrative structures from text using semantic role labeling (SRL). *relatio* proceeds in multiple stages. First, SRL decomposes sentences into grammatical components: agents (entities performing actions), verbs (actions), and patients (entities affected by actions). The resulting sentences take the following form:

$$\text{agent} \xrightarrow{\text{action}} \text{patient}$$

Next, *relatio* applies systematic dimension reduction to the raw SRL output. Named-entity recognition consolidates references to people, organizations, and locations, while word-embedding–based clustering groups semantically similar topics. Finally, narrative salience is measured by counting the frequency of each agent-action-patient triplet, producing a quantitative measure of prominence within the corpus.

For our application to inflation narratives, *relatio* offers two advantages. First, it preserves causal direction, and second, it operates without supervision, allowing for the discovery of narrative structures not specified ex ante. We apply *relatio*, as implemented by the authors, to our FOMC corpus.[6] The resulting narratives from the best-performing model (see next section) are presented in Table B3 in the Appendix. Using these narratives as inputs, we apply the same agnostic regression framework described above—following the setup of Larsen, Thorsrud and Zhulanova (2021)—to predict the paragraph-level occurrence of the six manually identified narratives from Andre et al. (2024).

4.3. Supervised Methods

Compared to the previous approaches, supervised methods require labeled data to train classifiers. While in principle one could fine-tune a model on our hand-coded FOMC corpus (e.g., Baumgärtner and Zahner, 2025), we rely on pre-trained models. Specifically, we evaluate the performance of generative LLMs, which represent the most widely adopted supervised approach in recent economic applications. Generative LLMs require a prompt – a the textual instruction – that may strongly influence results. To limit discretionary influence, we adopt a single generic prompt structure applied uniformly across all narratives. In line with our narrative-definition, each prompt is phrased in the past tense (*have contributed*) and explicitly refers to a rise in the inflation rate:

```
[Narrative] have contributed to the rise in the
current inflation rate.
```

The exact formulations for all `[Narrative]` prompts are provided in Table 3.

ChatGPT

The most widely recognized generative LLM is *ChatGPT*. Large language models of this type have been rapidly adopted in economics (Korinek, 2024; Jha et al., 2024; Chatterji et al., 2025) and even applied to the analysis of central bank communication (Alonso-Robisco and Carbó, 2023; Hansen and Kazinnik, 2024).

[6] Specifically, we experimented with a range of cluster-counts $K = [2, 3, 6, 9, 14, 25, 50, 75]$ for the phrase embeddings, and for each K we evaluated both silhouette score and elbow method (inertia) to assess clustering quality and compactness. We then inspected the resulting clusters qualitatively, and selected $K = 6$ as the best trade-off between coherence, interpretability, and dimensional coverage in our sample.

Table 3: LLM Prompts

Narrative	LLM Prompt
Supply Chain Issues	Global supply chain disruptions/bottlenecks have contributed to the rise in the current inflation rate.
Labor Shortage	Worker shortages have contributed to the rise in the current inflation rate.
Pandemic Impact	The COVID-19 pandemic has contributed to the rise in the current inflation rate.
Energy Crisis	The energy crisis has contributed to the rise in the current inflation rate.
Pent-up Demand	Pent-up consumer demand has contributed to the rise in the current inflation rate.
Russia-Ukraine War	The Russian invasion of Ukraine has contributed to the rise in the current inflation rate.

Because the transformer architecture preserves word order and relational structure, *ChatGPT* is in principle capable of capturing causal direction in text.
Our procedure for *ChatGPT* was as follows. We organized the input into a spreadsheet, arranging paragraphs by rows and assigning one narrative to each column. The model was then prompted with the following instruction:

```
Think like an economist. For each paragraph in the spreadsheet,
evaluate whether the following statements are supported by the
text:
- Supply Chain Issues: Global supply chain [...]
- Labor Shortage: [...]
Record your answers as binary indicators (1 = narrative present,
0 = narrative absent) in the corresponding columns of the
spreadsheet.
```

The narrative–prompt pairings used to query ChatGPT are again presented in Table 3. The resulting binary spreadsheet serves as the direct input for our evaluation against our manually labeled data.[7]

MiniCheck

MiniCheck (Tang, Laban and Durrett, 2024) is a fact-checking model designed to reduce hallucinations in LLM outputs – statements that appear plausible but lack evidential grounding. These hallucinations often take the form of overgeneralization or inaccurate combinations of information not actually present in the underlying

[7]We used OpenAI's ChatGPT-4 Turbo, accessed via the Web API in July 2025.

source material. *MiniCheck* addresses this challenge by verifying claims against a set of grounding documents. If a claim can be supported by the evidence, it is marked as factual; if not, it is flagged as unsupported.

What distinguishes *MiniCheck* from other generative LLM's is its training regime. Beyond conventional entailment data, it was trained on claim–document pairs specifically designed to teach the model to verify whether a claim – down to its atomic facts – is supported by evidence dispersed across a document. This specialized training enables *MiniCheck* to match state-of-the-art levels accuracy while remaining computationally lightweight and scalable (Tang, Laban and Durrett, 2024).[8]

We invert *MiniCheck*'s intended workflow for narrative detection. Specifically, we assert that a particular inflation narrative is present in a document and use *MiniCheck* to verify this claim based on the textual evidence. Thus, we operationalize the presence of inflation narratives as testable causal propositions, that can be fact-checked. The causal framing of the 'fact' – that is, the narrative itself – naturally distinguishes a Statement-1-type text from a Statement-2-type text.

In practice, *MiniCheck* evaluates each paragraph-narrative pair and produces a probability score $\in [0, 1]$ indicating the degree to which that paragraph supports the narrative. We threshold these scores at varying cutoffs to generate binary classifications.

5. Evaluation

Quantifying economic concepts, such as inflation narratives, means translating a theoretical concept in a numerical representation. We compare five approaches to do so: manual classification, unsupervised machine learning models (LDA and *relatio*), and supervised machine learning models (*ChatGPT* and *MiniCheck*). Each method is applied to the FOMC press-conference corpus on the paragraph-level, following the procedures outlined in Section 4 and the respective literature. The resulting predictions are presented in Table 4. We experiment with alternative parameter settings to test robustness, some of which we report in Section 5.5.2. The full parameterization and replication codes are provided in the Online Appendix. Because the mapping from a narrative to its numeric representation is neither directly observable nor unique, we evaluate all methods both qualitatively and quantitatively – by predictive power against the manually coded benchmark – to identify the best performing approach.

5.1. Qualitative Evaluation

The candidate methods are assessed against three criteria that we view as central to narrative quantification in large-scale corpora:

[8] We use the Bespoke-MiniCheck-7B model. The model can be run both locally or remotely via the BespokeLabs API.

Table 4: Prediction Frequencies of Narrative Classification Methods (in %)

	Manual Classification	LDA	*relatio*	*ChatGPT*	*MiniCheck*
Supply Chain Issues	2.4	0.4	0.8	0.7	2.5
Labor Shortage	1.7	0.0	0.8	0.0	2.1
Energy Crisis	1.5	0.0	0.6	2.1	1.4
Pandemic Impact	1.5	0.0	0.7	10.8	2.5
Pent-up Demand	1.2	0.4	0.7	0.4	1.0
Russia-Ukraine War	1.1	0.0	0.0	0.0	1.2

Note: Unit of observation: paragraph × narrative. LDA: 50 topic model with lasso l1-penalized logistic regression model. *relatio*: 50 dimension model with lasso l1-penalized logistic regression model. ChatGPT: binary classification. MiniCheck: threshold scores at 50%.

1) *Reliability*: Methods should produce indicators that are stable across resamples, annotators, and model initializations.

2) *Validity*: Methods should produce indicators that capture the intended narrative construct.

3) *Scalability*: Methods should remain computationally feasible when applied to large corpora.

We evaluate all methods against these criteria using our manually coded FOMC corpus as the benchmark. The next section provides a synopsis of the results.

Reliability

Reliability refers to the consistency of a measurement across contexts, time periods, and resamples. In our framework, it captures the extent to which repeated application of a method, under reasonable perturbations of inputs or parameters, yields consistent narrative indicators.

Manual classification performs well: the Pearson correlation of the binary inflation narrative classifications between the two annotators was 0.99, indicating near-perfect agreement conditional on both annotators being domain experts. Recent literature suggests though, that this conditionality might be binding (Romer and Romer, 2023; Naef, 2024).

The unsupervised methods perform poorly. For LDA, re-estimation or minor specification changes (e.g., increasing the number of topics from 25 to 75) result in substantially different outcomes. For example, the "inflation" topic's time series in Figure 5 correlates at only 0.67 when re-estimating the model using a larger topic number (see Appendix B.B2).

Supervised methods perform substantially better. While *ChatGPT* introduces stochasticity through its sampling mechanism, we observed only limited variability across repeated runs and prompt modifications. *MiniCheck*, by contrast, is deterministic: given the same inputs, it produces identical outputs across runs.

Also, when testing the paraphrase sets, Cronbach's Alpha, a standard measure of internal consistency based on average pairwise correlations for each narrative classification consistently exceeds 0.9, considered 'excellent' (Kline, 2013), for all narrative clusters (see Table B4). These results suggest that MiniCheck is not only deterministic but also robust to variation in formulations.

Validity

Validity concerns the degree to which a measure captures the intended construct. While proving validity outright is rarely feasible, it can be falsified by documenting systematic misclassifications, most readily through false positive (Type I errors) and false negatives (Type II errors). Under the assumption that our hand-classification provides the most valid measurement of the inflation narratives, the other methods can be evaluated against this ground truth. We systemize this evaluation in Section 5.5.2.

Topic models raise substantive validity concerns. Standard LDA hyperparameters impose nearly uniform distributions of topic probabilities across documents, in contrast to our results where most paragraphs pertain none or a single narrative. Consequently, almost no narrative reaches a meaningful probability threshold (Table 4). When focusing on the intensive margin (Table 4), these diffuse priors inflate both Type I and Type II errors. For instance, we find a spurious 'pandemic' topic predating the COVID-19 outbreak, driven by generic macroeconomic terms such as *recovery, stability,* and *employment* (Figure B3).

relatio faces similar validity challenges as it is also based on dimensionality reduction with frequency-based weighting. Consequently, the model requires a large and diverse corpus to generate stable embeddings. In our smaller corpus, infrequent yet important narratives are systematically down-weighted. The most prominent narrative we find (`We affect I`, Table B3) lacks semantic coherence, merely echoing the standardized welcome paragraph of each press-conference. The only inflation-related narrative we found (`Inflation affect Inflation`) is tautological. After extensive manual adjustment, the model yields one additional inflation-related link (`Monetary Policy affects Inflation`), yet both the causal direction – whether policy actions increased or moderated inflation – and the low frequency of this pattern in our manual labeled dataset (Table 2) call its validity for our corpus into question.

ChatGPT exhibits a the opposite issue for certain narratives (e.g., Pandemic, Table 4), where it treats any keyword mention as conclusive evidence, producing Type I errors. Beginning in September 2020, for instance, the Chairman's standard preface *"Since the beginning of the pandemic, we have taken forceful actions to provide some relief and stability, to ensure that the recovery will be as strong as possible [...]"* is repeatedly misclassified as a pandemic-driven inflation narrative. Across our sample, *ChatGPT* produced 78 such pandemic-driven classifications, compared with 11 in the hand-coded set. By contrast, *MiniCheck* exhibits similar levels of classifications as the gold standard (Table 4). Disagreements between both are rare (see Section 5.5.2) and occur primarily in passages where the backward-looking orientation is implicit.

Scalability

Scalability captures the feasibility of applying a classification method to thousands of documents at low marginal cost. While large-scale manual annotation could produce highly reliable and valid labels, it is prohibitively costly for widespread application. For illustration, coding the 41 FOMC press conferences required roughly 13 hours per annotator. Extending this effort to the 15,605 NYT articles in our sample would be practically infeasible.

In our case, *relatio* imposed the heaviest computational burden: end-to-end runs on our (comparatively small) FOMC corpus took on average 7-8 minutes on a standard workstation, driven mainly by the embedding construction and entity recognition. By contrast, LDA fits the same corpus within seconds, and both *MiniCheck* and *ChatGPT* allow for near-instantaneous online inference. As these differences are first order, we would recommend LDA, *MiniCheck*, or *ChatGPT* for larger corpora. Another important though with respect to scalability concerns data cleaning. For small, structured corpora such as the FOMC corpus, it is possible to manually refine and standardize the text—for instance, by removing repetitive and non-substantive introductions or closing remarks—to improve performance in methods such as *relatio*. However, this approach is infeasible for large and heterogeneous datasets like the NYT corpus, where transformer-based models such as *ChatGPT* or *MiniCheck* provide an inherent advantage.

Taking Stock

Table 5 provides a summary of the strengths and limitations of each method. *MiniCheck* stands out by meeting all three criteria, making it a particularly promising tool for empirical work in narrative economics.

Table 5: Overview of Narrative Quantification Methods

	Reliable	Valid	Scalability
Manual Classification	X	X	
Topic Modeling (e.g. LDA)			X
Semantic Role Labeling (e.g. *relatio*)	X		
ChatGPT	X		X
MiniCheck	X	X	X

Notes: The table evaluates methodologies based on their alignment with the criteria essential for robust narrative quantification. X's indicate that the method satisfactorily meets the criterion.

5.2. Quantitative Evaluation

We next conduct a quantitative evaluation in form of a prediction task. In a horse-race design, each method is tested against the human-coded benchmark at the paragraph–narrative level. For each of the six narratives, the evaluation sample comprises 723 paragraphs (Table 1 and Table 2).

For the unsupervised methods, we employ both LDA and *relatio* with $k \in 10, 25, 50$. The resulting dimensions are mapped to narrative labels via narrative-specific l_1-penalized logistic regressions, following Larsen, Thorsrud and Zhulanova (2021) to eliminate irrelevant dimensions.[9] Accordingly, evaluation of the unsupervised methods is conducted *in-sample*.[10] For the supervised methods, we employ *ChatGPT*, which produces binary classifications, and *MiniCheck*, which generates probability scores that we threshold at 50% and 90% (Section 4.4.3). As these models do not require pre-training on our corpus, their performance is assessed entirely *out-of-sample* using the prompts outlined in Section 4.4.3

The results are summarized in Table 6, averaged across the six narratives. We evaluate performance using standard confusion–matrix statistics in a binary-label framework. Given the rarity of positive cases (17/605 or fewer), overall accuracy is uninformative, so we focus on $Precision = TP/(TP + FP)$, $Sensitivity = TP/(TP + FN)$, and the harmonic mean $F_1 = 2 \cdot Precision \cdot Sensitivity/(Precision + Sensitivity)$. Precision penalizes false positives, sensitivity penalizes false negatives, and the F_1 score balances both.

Our main findings are as follows. First, consistent with the qualitative assessment, LDA and *relatio* models systematically under-predict narratives, thus producing high precision (few false positives) but very low sensitivity (many false netagives). Second, relative the unsupervised approaches, *ChatGPT* performs better. Though both precision and sensitivity remain modest – consistent with our earlier observation of over-classifying the pandemic narrative – the resulting F_1 score is substentially higher. Finally, *MiniCheck* with a 50% threshold provides the best balanced performance, combining high precision with high sensitivity to achieve the highest F_1. Raising the threshold to 90% reduces F_1 as expected, but provides a more conservative classification when avoiding false positives is paramount.

[9] Regression results are robust to alternative specifications; see the Online Appendix.

[10] We tested a 80/20 train–test split, but the small number of positive cases left several narratives with too few test observations to yield meaningful metrics.

Table 6: Quantitative Evaluation I: FOMC Narratives

	Accuracy	Precision	Sensitivity	F-Score
Unsupervised Machine Learning				
LDA (10 Topics)	0.98	0.00	0.00	0.00
LDA (25 Topics)	0.99	1.00	0.09	0.16
LDA (50 Topics)	0.99	1.00	0.09	0.16
relatio (10 narratives)	0.98	0.00	0.00	0.00
relatio (25 narratives)	0.98	1.00	0.03	0.06
relatio (50 narratives)	0.99	0.80	0.06	0.11
relatio (10 narratives; small cluster)	0.99	0.80	0.06	0.11
relatio (25 narratives; small cluster)	0.99	0.80	0.06	0.11
relatio (50 narratives; small cluster)	0.99	0.65	0.25	0.36
Supervised Machine Learning				
ChatGPT	0.97	0.17	0.25	0.20
MiniCheck (50% Threshold)	0.99	0.77	0.87	<u>0.81</u>
MiniCheck (90% Threshold)	0.99	0.87	0.38	0.53

Note: Accuracy, Precision, Sensitivity, and F_1 are averaged across six narratives. Unit of observation: paragraph × narrative. LDA(k): l_1-penalized logistic regression on document–topic proportions from an LDA(k) Gibbs sampler, with λ chosen by five-fold cross-validation. ChatGPT: binary classification. MiniCheck: threshold scores at 50% or 90%.

One concern may be that *MiniCheck* does not provide information beyond what is already captured by the other models. To test this, we jointly include the full 50-dimension LDA model, the 50 narrative *relation* model, the *ChatGPT* prediction, and the *MiniCheck* scores in narrative-specific Lasso regressions. The selected variables are then re-estimated in a simple OLS regression to obtain standard errors, reported in Table 7. For brevity, we only report coefficients for the two supervised methods, and only report the number of included LDA and *relatio* coefficients in the last two rows. The full table is the Appendix (Table B1). With the exception of the Labor Supply Narrative, *MiniCheck* is consistently selected and enters with large, highly significant coefficients, indicating that it provides information beyond the other methods. Once *MiniCheck* is included, *ChatGPT*-based predictions become practically irrelevant: when selected at all, the coefficients are small in magnitude. Finally, the chosen LDA Topics and *relatio* narratives (Table 7, last two rows) are non-exlusive, with some topics contributing to multiple narratives. While not problematic *per se*, this underscores the lack of one-to-one mapping between LDA topics and narratives, limiting its interpretability for narrative detection.

Table 7: Quantitative Evaluation II: Lasso Regression

	Dependent variable: Narrative					
	Supply Chain Issues	Labor Shortage	Pandemic Impact	Energy Crisis	Pent-up Demand	Russia-Ukraine War
	(1)	(2)	(3)	(4)	(5)	(6)
MiniCheck	0.49***		0.50***	0.83***	0.40***	0.98***
	(0.05)		(0.04)	(0.04)	(0.09)	(0.02)
ChatGPT			0.04***	0.14***		
			(0.01)	(0.02)		
Incl. LDA Topics	2	4	7	1	3	2
Incl. *relatio* Narratives	2	3	2	-	2	-
Observations	723	723	723	723	723	723
R^2	0.65	0.15	0.42	0.66	0.53	0.88
Adjusted R^2	0.64	0.14	0.42	0.66	0.53	0.88

Note: Standard errors in parentheses: *p<0.1; **p<0.05; ***p<0.01. Explanatory variables are selected by a Lasso logistic model with 5-fold cross-validation. Selected predictors were then re-estimated in OLS for interpretability. LDA: 50 topic model; *relatio*: 50 narrative model with small clusters; ChatGPT and MiniCheck: binary indicators.

6. From FOMC Narratives to household expectations

We next examine whether inflation narratives articulated in Federal Open Market Committee (FOMC) press conferences diffuse into broader public discourse and whether variation induced by the FOMC in media coverage affects household expectations. We focus on the six backward-looking inflation narratives – *Pent-up Demand*, *Supply Chains*, *Labor Shortages*, *Energy*, *Pandemic*, and *Russia–Ukraine* – that consistently appeared in our manual coding of FOMC transcripts from 2019 to 2024 (see Table 2).

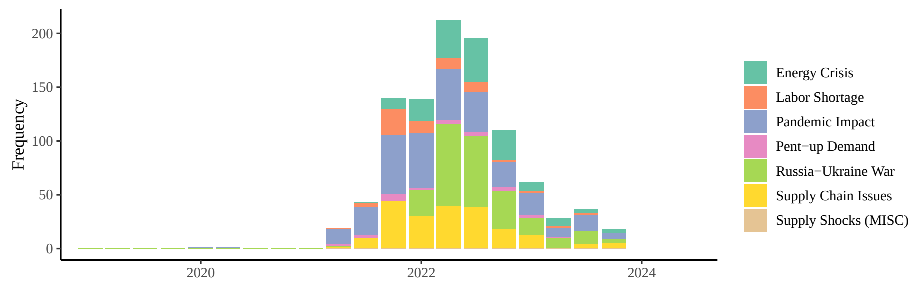
To measure the transmission of central bank communication to the media, we combine the two corpora introduced in Section 3: (i) FOMC press conference transcripts and (ii) the universe of NYT articles between 2019 and 2024. For the FOMC transcripts, we construct a meeting-level narrative intensity $FOMC_k \in [0,1]$, defined as the maximum of the paragraph-level probabilities assigned to narrative k within each meeting by our best-performing classifier, *MiniCheck*. On the media side, we label a NYT article as containing narrative k if at least one of its paragraphs attains a probability above 0.90.[11] Of the 15,606 articles containing inflation-related terms published during our sample period, only 612 (4 percent) feature at least one of our six inflation narratives.[12] This sparsity suggests that word-counts may provide a noisy proxy for the presence of an economic narrative.

[11]The 0.90 cutoff is deliberately conservative to limit false positives; given the relatively small number of narrative mentions, even a few spurious classifications could distort the estimates.

[12]To assess potential false negatives, we manually inspected a random sample of more than 300 articles classified as containing no narrative. None contained an inflation narrative. Rather articles referred to inflation in a factual or descriptive manner, without providing explanations or accounts of its underlying causes.

Figure 6 shows quarterly counts of NYT articles by narrative, which broadly track the FOMC narrative series in Figure 4 – a pattern we formalize below. Summary statistics for the daily time-series variables, as well as all variables used in this chapter are reported in Appendix A.A1.

Figure 6 : Quarterly NYT Articles by Narrative. Notes: Count of NYT articles containing each inflation narrative identified by MiniCheck.



6.1. Hypotheses and Empirical Strategy

Three hypotheses structure our empirical evaluation of how FOMC inflation narratives propagate through the media to households. First, the FOMC and the NYT both respond to the same underlying set of macroeconomic shocks. In this case, inflation narratives in FOMC press conferences should coincide with coverage of the same narrative in the NYT:

> **H1:** FOMC press conferences inflation narrative intensity co-moves with contemporaneous NYT coverage of the same narrative.

Second, if the FOMC causally affects media framing, we should observe an increase in narrative-specific NYT coverage following press conferences that emphasize that narrative:

> **H2:** Higher FOMC inflation narrative intensity increases post-meeting media coverage of the same narrative.

Finally, if households form expectations based on FOMC-driven narratives transmitted via media coverage, those narratives that are amplified by the media should affect household beliefs:

> **H3:** NYT inflation narrative intensity, driven by FOMC narrative intensity, shifts post-meeting household inflation expectations.

6.2. H1: Common-shocks – FOMC and NYT co-movement

Consistent with the notion that both the Fed and the NYT editorial board are likely respond to common macroeconomic shocks, we first test whether FOMC narrative intensity co-moves with pre-meeting NYT coverage. Both corpora contain precise intraday timestamps, allowing us to aggregate narrative NYT intensity into 24-hour windows around each press-conference (14:00 ET). Figure 7 illustrates the set-up. This narrow window offers two advantages. First, it improves identification by minimizing confounding news and random variation in unrelated reporting. Second NYT coverage in this short window around the press-conference becomes more similar to that of other outlets (Herbert, Istrefi and Sagna, 2024) making the NYT more representative and the results correspondingly more generalizable.

Figure 7 : NYT Treatment Windows around the FOMC Press Conference

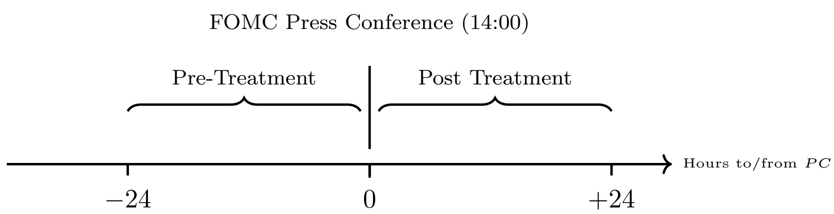


Figure 8 plots, for each narrative, the number of NYT articles in the 24 hours preceding the press-conference (NYT_{t-1}) against the corresponding FOMC press conference narrative intensity ($FOMC_t$). The scatter plot shows a positive link, suggesting that narratives receiving greater emphasis in FOMC press conferences tend to coincide with a higher number of *pre*-meeting NYT articles featuring the same narrative.
To systematically test the relationship, we estimate the following regression:

$$FOMC_t = \alpha + \beta NYT_{t-1} + \epsilon_t \tag{1}$$

where NYT_{t-1} denotes the number of NYT articles containing a specific inflation narrative in the 24 hours preceding the FOMC meeting, and $FOMC_t$ measures the corresponding narrative intensity during the respective press conference as defined above.
The results are shown in Table 8. The estimated coefficients $\hat{\beta}_k$ are positive and, for most specifications, statistically significant. Estimating Equation (1) jointly across narratives, one additional NYT is associated with an increase in FOMC narrative intensity of one-third, or about one standard deviation. Estimating Equation (1) separately by narrative shows that this effect is driven by four narratives (*Supply Chains*, *Energy*, *Pandemic*, *Russia–Ukraine*), with effect sizes reaching up to 0.55.

Figure 8 : Pre-meeting NYT coverage and FOMC narrative intensity

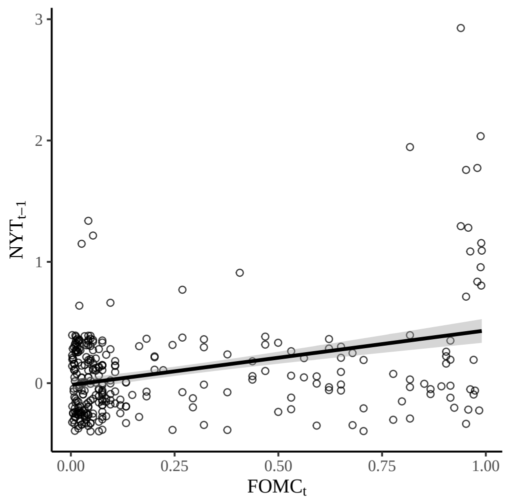


Table 8: FOMC Narrative Emphasis and New York Times Coverage

	Dependent variable: $FOMC_t$						
	All Narratives	Supply Chain	Labor Shortage	Pandemic Impact	Energy Crisis	Pent-up Demand	Russia-Ukraine War
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
NYT_{t-1}	0.33^{***}	0.38^{**}	0.13	0.27^{**}	0.26^{***}		0.54^{***}
	(0.05)	(0.15)	(0.27)	(0.11)	(0.09)		(0.11)
Constant	0.18^{***}	0.28^{***}	0.28^{***}	0.32^{***}	0.19^{***}	0.16^{***}	0.11^{**}
	(0.05)	(0.06)	(0.04)	(0.06)	(0.05)	(0.04)	(0.05)
Observations	246	41	41	41	41	41	41
R^2	0.19	0.13	0.01	0.13	0.18	0.00	0.39

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01. Regressions based on Equation (1); Narrative indicators are constructed by thresholding the article-level probabilities at 0.9; New York Times Article Count is based on 24h before FOMC press conference.

The two remaining narratives show little co-movement, reflecting their low base rates: *Pent-up Demand* does not appear in the NYT in any *pre*-meeting window, while *Labor Shortages* appears only three times.

6.3. H2: FOMC-to-media transmission

We next examine whether the FOMC, through its press conferences, shapes subsequent NYT coverage of specific inflation narratives. To isolate directional transmission, we compare pre- and post-meeting coverage within the same meeting using a fixed-effects specification:

$$NYT_{t,m} = \alpha + \beta FOMC_m + \gamma(POST_t \times FOMC_m) + \mu_m + \varepsilon_{m,t}, \tag{2}$$

where NYT denotes the number of NYT articles categorized for a narrative, $FOMC_m$ captures the corresponding intensity in meeting m, $POST_t$ is a binary variable equals one after the press conference and zero otherwise, and μ_m denotes a meeting fixed to absorb shocks common to the event window of meeting m. In robustness tests, we control for the total number of inflation-related articles and for monetary policy surprises from Swanson (2021); results remain unchanged. The parameter of interest, γ, measures the average within-meeting change in NYT coverage from the pre- to the post-meeting window associated with FOMC narrative intensity.

Table 9 presents the results. The joint specification yields a positive and significant interaction coefficient, $\hat{\gamma}_k$. The magnitudes is economically meaningful: A press conference that does mention one of the six narrative increases NYT coverage of that narrative by roughly one-fourth of an additional article in the following 24 hours – around half of a standard deviation in NYT coverage (Table A1).

Columns (2)–(7) of Table 9 show that the effect is largely homogeneous across narratives. The estimates are positive for five of the six narratives and statistically significant for three: *Labor Shortages*, *Pent-up Demand*, and *Russia–Ukraine*. Effects for *Energy* and *Pandemic* are smaller and imprecise.

Table 9: Event Study: FOMC Narrative Emphasis on New York Times Coverage

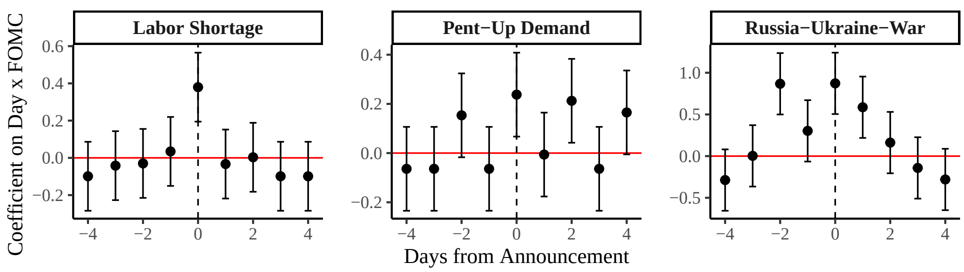
	Dependent variable: New York Times Article Count$_t$						
	All Narratives	Supply Chain	Labor Shortage	Pandemic Impact	Energy Crisis	Pent-up Demand	Russia-Ukraine War
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FOMC×Post$_t$	0.28***	0.36	0.41**	0.08	−0.05	0.30**	0.72***
	(0.09)	(0.21)	(0.18)	(0.22)	(0.25)	(0.14)	(0.21)
PC FE	X	X	X	X	X	X	X
Observations	492	82	82	82	82	82	82
R^2	0.37	0.67	0.60	0.67	0.60	0.54	0.77
Adjusted R^2	0.30	0.31	0.18	0.32	0.16	0.04	0.53

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01. Regressions based on Equation (2). Narrative indicators are constructed by thresholding the article-level probabilities at 0.9; New York Times Article Count is based on 24h around FOMC press conference. All specifications include meeting fixed effects and control for the total number of inflation-related articles.

We substantiate the baseline results with a two-way fixed-effects difference-in-differences specification around each press conference, focusing on the three significant narratives. Figure 9 plots the estimated interaction coefficients for a 9 day window around the press-conference with 95% confidence intervals. The horizontal axis denotes days relative to the press conference, and the vertical axis shows the change in expected NYT article counts associated with $FOMC_m$. The dashed vertical line marks the press-conference day.

Three results emerge. First, we observe a sharp and statistically significant increase in coverage in the 24 hours following the press conference across all three narratives, with magnitudes slightly larger than the baseline fixed-effects estimates in Table 9. Second, pre-press-conference point estimates are insignificant, consistent with parallel pre-trend assumption for all narratives except for one observation of the *Russia–Ukraine* narrative. Third, effects persist modestly to $t = 2$ for *Pent-up Demand* and *Russia–Ukraine*, indicating short-lived but measurable persistence of the effect. Taken together, these results reinforce the interpretation of Table 9 that FOMC narrative emphasis causes a near-instantaneous increase in media coverage.

Figure 9 : Difference-in-Difference estimates of FOMC narratives on NYT coverage



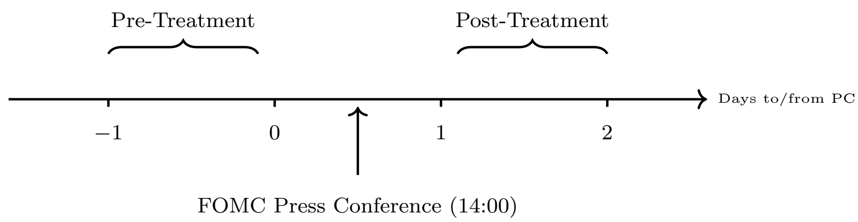
Notes: Estimates from two-way fixed effects DiD regressions with a ±10*-day window. The horizontal axis indicates days relative to the FOMC press conference (day 0), and the vertical axis reports the estimated coefficients for each relative day. Vertical bars represent 95% confidence intervals. The figure displays estimates for days –4 through +4; coefficients for omitted periods are reported in the Online Appendix.*

6.4. H3: $FOMC \rightarrow NYT \rightarrow \pi^e$

We next test whether media amplification translates into changes in household inflation expectations. For household expectations, we harmonize the text-based indicators with the New York Fed's Survey of Consumer Expectations (SCE). As survey participation does not vary systematically around FOMC announcements with respect to observable household characteristics (Binder, Campbell and Ryngaert, 2024), we exploit daily variation in the SCE's inflation expectations. However since the SCE is available only at a daily frequency, we cannot distinguish pre- and post-announcement responses within the same day (Figure 10). The press-conference day may thus represent a mixed-treatment observation. We therefore report results both including and excluding the press conference day.

To estimate the causal effect of NYT on household expectations, we implement a two-stage least squares (2SLS) specification in which the FOMC narrative intensity serves as an instrument for narrative-specific, post-meeting NYT coverage. As shown in the previous subsection, the instrument is relevant: stronger FOMC narrative emphasis increases post-meeting NYT coverage of the same narrative.

Figure 10 : SCE Treatment Windows around the FOMC Press Conference



The instruments' exogeneity follows from households' limited knowledge of central banking (Kumar et al., 2015; Binder, 2017*a*), which implies little direct exposure to FOMC communication.

Table 10 reports the estimates for mean household inflation expectations. First, Column (1) tests the endogeneity concern, showing that a simple OLS relationship between NYT coverage and π^e is negative, suggesting that higher media attention tends to correlate with periods of weaker inflation expectation. Second, Column (2)-(4) show that this relationship reverses once instrumented with FOMC narrative intensity. The 2SLS coefficient is positive and statistically significant. Excluding the press conference day (Columns (2) vs. (3)) improves both precision and the magnitude, suggesting that the press conference day does a mixed-treatment period for households. The effect size is economically meaningful: one additional NYT article referencing an inflation narrative is associated with an increase in mean household inflation expectation of 1.67 percentage points, or about one standard deviation over the sample period. Together with the first-stage coefficient (third row Table 10), this implies that one additional FOMC narrative during the press conference raises expected inflation by about $0.27 \times 1.67 = 0.45$ percentage points. Finally, restricting the sample to the three significant narratives identified in the previous subsection (Column (4)) increases the estimated effect to roughly 2 percentage points. However, the smaller sample size lowers the first-stage F-statistic below the conventional threshold of 10, rendering this specification less reliable. We therefore regard the full-sample specification, which includes all six narratives, as our preferred baseline.

We next examine whether inflation narratives in the media, instrumented by FOMC narratives, affect household inflation expectations beyond their short-term mean expectations. Specifically, using the same 2SLS specification, we test for effects on the within-respondent dispersion and the term structure of inflation expectations. The findings, reported in Table 11, show that within-respondent variance declines statistically significantly by about 10 percentage points – corresponding to two standard deviations over the sample period. This result suggests that exposure to news articles covering inflation narratives affects not only the level of inflation expectations but also reduces their heterogeneity. This finding is consistent with evidence for expectation convergence following

Table 10: 2SLS Event Study: Household 12 month inflation expectations

	Dependent variable: π_t^e			
	OLS		*2SLS*	
	(1)	(2)	(3)	(4)
NYT	-0.13^{**}	1.00^{*}	1.67^{**}	2.02^{*}
	(0.07)	(0.53)	(0.71)	(1.14)
First-stage $FOMC \times Post$:		0.30^{***}	0.27^{***}	0.31^{**}
First-stage F-statistic:		14.06	10.17	5.86
Wu-Hausman (p-Value):		0.008	0.000	0.001
Event FE	X	X	X	X
PC Day excluded			X	X
Selective Narratives				X
Observations	636	636	420	210

Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01. Regressions are estimated by two-stage least squares (2SLS), with the first-stage instrument reported in the second row. All specifications include meeting fixed effects, the lagged dependent variable, and a control for the total number of inflation-related articles. Sample is truncated to Inflation Expectations at ± 20 percentage points.

salient news shocks (Coibion, Gorodnichenko and Weber, 2022). Turning to the term structure of inflation expectations, Columns (4)–(6) show that the effect is concentrated in the short run. The effect size diminishes at 24 months and disappears entirely at 60 months, suggesting that inflation narratives primarily shift short-term expectations with little influence on long-term expectations, again consistent with evidence presented by Coibion, Gorodnichenko and Weber (2022). Taken together, these results suggest that exposure to inflation narratives anchors household inflation expectations at a higher level and with higher certainty. This pattern is consistent with expectation formation where information shocks alter perceived inflation, and households update their expectations adaptively.

Table 11: 2SLS Event Study: Variance and Long-term household inflation expectations

	$Var(\pi_t^e)$			π_t^e		
				12m	24m	60m
	(1)	(2)	(3)	(4)	(5)	(6)
NYT	-8.05^{***}	-9.99^{***}	-10.23^{*}	1.67^{**}	0.70^{*}	-0.57
	(3.09)	(3.70)	(5.27)	(0.70)	(0.42)	(0.67)
First-stage $FOMC \times Post$:	0.30^{***}	0.27^{***}	0.31^{**}	0.27^{***}	0.28^{**}	0.38^{*}
First-stage F-statistic:	14.06	10.17	5.86	10.88	13.57	3.96
Wu-Hausman (p-Value):	0.008	0.000	0.001	0.000	0.044	0.281
Event FE	X	X	X	X	X	X
PC Day excluded		X	X	X	X	X
Observations	636	420	210	420	420	180

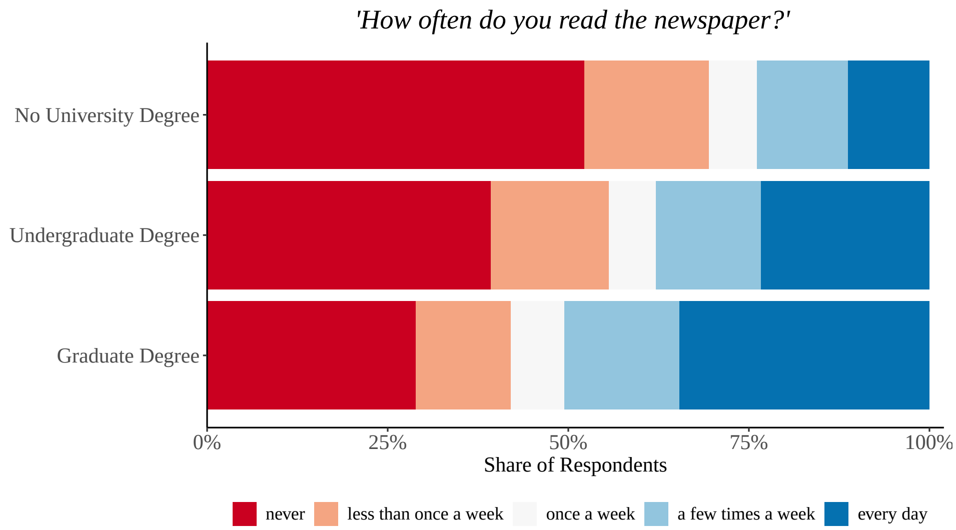
Note: Standard errors in parentheses; *p<0.1; **p<0.05; ***p<0.01. Regressions are estimated by two-stage least squares (2SLS), with the first-stage instrument reported in the second row. All specifications include meeting fixed effects, the lagged dependent variable, and a control for the total number of inflation-related articles, and control for the total number of inflation-related articles. Sample is truncated to Inflation Expectations at ± 20 percentage points.

6.5. H3: Household Heterogeneity

If FOMC narratives reach households through the media, the strength of the treatment effect should depend on information access and cognitive ability at the household level. As shown in Figure 11, respondents with university degrees are substantially more likely to read a newspaper "every day" or "a few times a week," and less likely reading one "never." This pattern implies that exposure – and thus expected treatment intensity – systematically increases with the level of education. To test this prediction, we estimate the 2SLS specification at the household level, partitioning the sample along the following household characteristics: education, financial literacy and financial responsibility (Gomes, Haliassos and Ramadorai, 2021; Arrondel et al., 2022; Coibion, Gorodnichenko and Weber, 2022; D'Acunto et al., 2024; Doh, Lee and Park, 2025).

The household-level 2SLS estimates in Table 12 confirm that the average effect NYT narratives on household inflation expectations are positive and statistically significant. Column (1) yields a slightly different coefficient to the aggregated regression (1.57 versus 1.67), reflecting the non-uniform sampling across days. Partitioning the sample by education, Column (2)-(4) shows that the treatment effect is concentrated among university graduates. The response is strongest for respondents with a doctoral degree, whose estimated coefficient is nearly twice as large as the sample average. Finally, we examine heterogeneity by financial literacy and financial responsibility. Columns (5)-(6) show that the effect is confined to households with high financial literacy, while Columns (7)-(9) indicates a similar pattern among respondents solely responsible for household financial decisions.

Figure 11 : Heterogeneity in Newspaper reading



Notes: Data are from the General Social Survey (GSS), conducted by NORC at the University of Chicago. Only household responses from 2019–2024 are used. Household graduate categories are clustered to align with the brackets used in the Survey of Consumer Expectations (SCE).

Both traits imply greater engagement with financial information and potential greater capacity to process economic news.
Taken together, these results point to a single mechanism. FOMC narratives reach households through news media, but are constrained to translate into expectations by exposure and cognitive ability on the household side. Specifically, households that more likely to read newspapers, possess high financial literacy, and actively make financial decisions update their expectations; others show little systematic response.

To summarize, our findings suggest that the communication channel between central banks and households is active but imperfect. On average, each inflation-related narrative during an FOMC press conference generates roughly one-quarter of an extra article in the NYT with that specific narrative on the following day. Each additional article increases household inflation expectations by, on average, 1.5 percentage points and reduces the expected variance of future inflation rates by ten percentage points. However, the effects are concentrated among well educated and financially literate households – a group whose expectations are already more informed and better anchored.

Table 12: Micro Level IV Event Study: 12 month ahead household inflation expectations

	Dependent variable: π^e_t								
		University Education			**Fin. Literacy**		**Fin. Responsibility**		
		No Degree	Bachelor/ Master	PhD/ more	Low	High	None	Shared	Sole
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
NYT	1.57*	−1.50	2.29**	2.95**	−2.59	2.65**	0.77	1.27	2.09**
	(0.90)	(1.75)	(1.02)	(1.51)	(1.91)	(1.05)	(1.01)	(1.40)	(0.87)
Event FE	X	X	X	X	X	X	X	X	X
PC Day excl.	X	X	X	X	X	X	X	X	X
Observations	12,984	4,938	7,212	798	3,270	9,714	318	5,346	6,162
R^2	0.06	0.10	0.05	0.24	0.06	0.04	0.76	0.10	0.06

Note: Two-way clustered standard errors (by observation date and respondent) are reported in parentheses; *p<0.1; **p<0.05; ***p<0.01. Regressions are estimated by two-stage least squares (2SLS). All specifications include meeting fixed effects, the lagged dependent variable, and a control for the total number of inflation-related articles, and control for the total number of inflation-related articles. Sample is truncated to Inflation Expectations at ±20 percentage points.

7. Conclusion

This paper addresses a striking puzzle in monetary economics: households know little about central banking, yet their expectations move systematically around monetary policy announcements. We show that the media acts as a critical transmission mechanism, channeling Federal Reserve inflation narratives—causal explanations for inflation dynamics—from policymakers to the public.

We make four contributions. First, we compile two novel corpora for 2019–2024: inflation narrative hand-annotated FOMC press-conference transcripts and the universe of New York Times (NYT) articles referencing inflation. We document that both policymakers and journalists during this period predominantly advanced supply-side inflation narratives.

Second, we adapt a state-of-the-art supervised Large Language Model (LLM) from the computational linguistics fact-checking literature. Specifically, we modify *MiniCheck* to verify whether a given narrative is supported by textual evidence, thereby enabling the empirical identification of narratives within large text corpora. Using the hand-annotated FOMC transcripts as a benchmark, we evaluate *MiniCheck* against existing approaches and show that it offers superior performance in detecting narratives.

Third, applying *MiniCheck* to the NYT corpus and exploiting the sharp timing of FOMC press conferences, we show that inflation narratives of the FOMC are systematically amplified in media coverage. Specifically, each additional FOMC inflation narrative yields on average one-quarter of an extra inflation narrative article the following day.

Finally, using this exogenous variation in media reporting, we show that one additional NYT inflation narrative article raises household inflation expectations by 1.5 percentage points, and reduces expected inflation variation by 10 percentage points. The effect is concentrated among well-educated, and financially literate households.

Taken together, our findings demonstrate that central banks have the ability to reach households using communication, but only indirectly via media transmission. Exposure to news media and cognitive ability to interpret it are binding constraints to this channel. Narrative framing is therefore a selective yet powerful instrument by which central banks can shape public expectations.

References

Alonso-Robisco, Andres, and José Manuel Carbó. 2023. "Analysis of CBDC narrative by central banks using large language models." *Finance Research Letters*, 58: 104643.

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart. 2024. "Narratives about the Macroeconomy."

Angelico, Cristina, Juri Marcucci, Marcello Miccoli, and Filippo Quarta. 2022. "Can we measure inflation expectations using Twitter?" *Journal of Econometrics*, 228(2): 259–277.

Arrondel, Luc, Hector Calvo-Pardo, Chryssi Giannitsarou, and Michael Haliassos. 2022. "Informative social interactions." *Journal of Economic Behavior & Organization*, 203: 246–263.

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2024. "Relatio: Text Semantics Capture Political and Economic Narratives." *Political Analysis*, 32(1): 115–132.

Bauer, Michael D, and Eric T Swanson. 2023. "An alternative explanation for the "fed information effect"." *American Economic Review*, 113(3): 664–700.

Baumgärtner, Martin, and Johannes Zahner. 2025. "Whatever it takes to understand a central banker-Embedding their words using neural networks." *Journal of International Economics*, 104101.

Berger, Helge, Michael Ehrmann, and Marcel Fratzscher. 2011. "Monetary policy in the media." *Journal of Money, Credit and Banking*, 43(4): 689–709.

Bertsch, Christoph, Isaiah Hull, Robin L Lumsdaine, and Xin Zhang. 2025. "Central bank mandates and monetary policy stances: Through the lens of federal reserve speeches." *Journal of Econometrics*, 105948.

Bethmage, Eleftherios. 2024. "Puppies or Policy: Measuring Household (In) Attention to Central Bank Communication." *Available at SSRN 5076585.*

Binder, Carola. 2017*a*. "Fed speak on main street: Central bank communication and household expectations." *Journal of Macroeconomics.*

Binder, Carola. 2017*b*. "Federal reserve communication and the media." *Journal of Media Economics*, 30(4): 191–214.

Binder, Carola Conces, Jeffrey R Campbell, and Jane M Ryngaert. 2024. "Consumer inflation expectations: Daily dynamics." *Journal of Monetary Economics*, 145: 103613.

Binder, Carola, Pascal Frank, and Jane M Ryngaert. 2025. "The Causal Effect of News on Inflation Expectations." National Bureau of Economic Research.

Blei, DM, AY Ng, and MI Jordan. 2003. "Latent dirichlet allocation." *Journal of machine Learning research.*

Blinder, Alan S, Michael Ehrmann, Jakob De Haan, and David-Jan Jansen. 2024. "Central bank communication with the general public: Promise or false hope?" *Journal of Economic Literature*, 62(2): 425–457.

Born, Benjamin, Hrishbh Dalal, Nora Lamersdorf, Jana-Lynn Schuster, and Sascha Steffen. 2024. "Inflation Expectations in the Social Media Age."

Borup, Daniel, Jorge Wolfgang Hansen, Benjamin Dybro Liengaard, and Erik Christian Montes Schütte. 2023. "Quantifying investor narratives and their role during COVID-19." *Journal of Applied Econometrics*, 38(4): 512–532.

Carroll, Christopher D. 2003. "Macroeconomic expectations of households and professional forecasters." *the Quarterly Journal of economics*, 118(1): 269–298.

Çekin, Semih Emre, and Hamza Polattimur. 2025. "Televised inflation: Measuring TV news coverage and its effect on household expectations." ZEW Discussion Papers.

Chatterji, Aaron, Thomas Cunningham, David J Deming, Zoe Hitzig, Christopher Ong, Carl Yan Shan, and Kevin Wadman. 2025. "How people use chatgpt." National Bureau of Economic Research.

Claus, Edda, and Viet Hoang Nguyen. 2020. "Monetary policy shocks from the consumer perspective." *Journal of Monetary Economics*, 114: 159–173.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2022. "Monetary policy communications and their effects on household inflation expectations." *Journal of Political Economy*, 130(6): 1537–1584.

Collier, Paul, and David Tuckett. 2021. "Narratives as a coordinating device for reversing regional disequilibrium." *Oxford Review of Economic Policy*, 37(1): 97–112.

Conrad, Christian, Zeno Enders, and Alexander Glas. 2022. "The role of information and experience for households' inflation expectations." *European Economic Review*, 143: 104015.

De Pooter, Michiel. 2021. "Questions and Answers: The Information Content of the Post-FOMC Meeting Press Conference." *FEDS Notes*, 2021.0(2997).

Doh, Taeyoung, Ji Hyung Lee, and Woong Yong Park. 2025. "Heterogeneity in Household Inflation Expectations and Monetary Policy." *Journal of Financial Econometrics*, 23(1): nbae034.

Dräger, Lena. 2015. "Inflation perceptions and expectations in Sweden–Are media reports the missing link?" *Oxford Bulletin of Economics and Statistics*, 77(5): 681–700.

D'Acunto, Francesco, Evangelos Charalambakis, Dimitris Georgarakos, Geoff Kenny, Justus Meyer, and Michael Weber. 2024. "Household inflation expectations: An overview of recent insights for monetary policy."

Ehrmann, Michael, and Alena Wabitsch. 2022. "Central bank communication with non-experts–A road to nowhere?" *Journal of Monetary Economics*, 127: 69–85.

Eliaz, Kfir, and Ran Spiegler. 2020. "A Model of Competing Narratives." 110(12): 3786–3816.

Ferrara, Federico M, Donato Masciandaro, Manuela Moschella, and Davide Romelli. 2022. "Political voice on monetary policy: Evidence from the parliamentary hearings of the European Central Bank." *European Journal of Political Economy*, 74: 102143.

Flynn, Joel P., and Karthik A. Sastry. 2024. "The Macroeconomics of Narratives." *Social Science Research Network*.

Flynn, Joel P, and Karthik Sastry. 2022. *The macroeconomics of narratives.* SSRN.

Fraccaroli, Nicolò, Vincent Arel-Bundock, and Mark Blyth. 2025. "What do central bankers talk about when they talk about inflation? The rise and fall of inflation narratives." *New political economy*, 1–16.

Geiger, Felix, Dimitrios Kanelis, Philipp Lieberknecht, and Diana Sola. 2025. "Monetary-Intelligent Language Agent (MILA)." Technical Paper.

Gomes, Francisco, Michael Haliassos, and Tarun Ramadorai. 2021. "Household finance." *Journal of Economic Literature*, 59(3): 919–1000.

Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding scientific topics." *Proceedings of the National Academy of Sciences*, 101(suppl_1): 5228–5235.

Gürkaynak, Refet S, Brian Sack, and Eric Swanson. 2005. "The sensitivity of long-term interest rates to economic news: Evidence and implications for macroeconomic models." *American economic review*, 95(1): 425–436.

Hansen, Anne Lundgaard, and Sophia Kazinnik. 2024. "Can chatgpt decipher fedspeak?" *Available at SSRN 4399406.*

Hansen, Stephen, Michael McMahon, and Andrea Prat. 2017. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." *The Quarterly Journal of Economics*, 133(2): 801–870. Publisher: Oxford University Press (OUP).

Hansen, Stephen, Michael McMahon, and Matthew Tong. 2019. "The long-run information effect of central bank communication." *Journal of Monetary Economics*, 108: 185–202. Publisher: Elsevier BV.

Hayo, Bernd, and Edith Neuenkirch. 2018. "The influence of media use on layperson monetary policy knowledge in Germany." *Scottish Journal of Political Economy*, 65(1): 1–26.

Hayo, Bernd, and Johannes Zahner. 2023. "What is that noise? Analysing sentiment-based variation in central bank communication." *Economics Letters*, 222: 110962.

Herbert, Sylvérie, Klodiana Istrefi, and Béatrice Sagna. 2024. "Coordinating the Message: Media Coverage of Fed News and Market Reactions."

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang. 2024. "ChatGPT and corporate policies." National Bureau of Economic Research.

Kintsch, W. 2006. "Latent semantic analysis: A road to meaning." *Laurence Erlbaum.*

Kline, Paul. 2013. *Handbook of psychological testing.* Routledge.

Korinek, Anton. 2024. "Generative AI for economic research: LLMs learn to collaborate and reason." National Bureau of Economic Research.

Kumar, Saten, Hassan Afrouzi, Olivier Coibion, and Yuriy Gorodnichenko. 2015. "Inflation targeting does not anchor inflation expectations: Evidence from firms in New Zealand." *Brookings papers on economic activity*, 2015(2): 151–225.

Lamla, Michael J, and Sarah M Lein. 2014. "The role of media for consumers' inflation expectation formation." *Journal of Economic Behavior & Organization*, 106: 62–77.

Larsen, Vegard H., Leif Anders Thorsrud, and Julia Zhulanova. 2021. "News-driven inflation expectations and information rigidities." *Journal of Monetary Economics.*

Lewis, Daniel J, Christos Makridis, and Karel Mertens. 2019. "Do monetary policy announcements shift household expectations?"

Loughran, Tim, and Bill McDonald. 2011. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of finance*, 66(1): 35–65.

Masciandaro, Donato, Oana Peia, and Davide Romelli. 2024. "Central bank communication and social media: From silence to Twitter." *Journal of Economic Surveys*, 38(2): 365–388.

Müller, Henrik, Tobias Schmidt, Jonas Rieger, Lena Maria Hufnagel, and Nico Hornig. 2022. "A German Inflation Narrative." *DoCMA Working Paper;9.* Publisher: TU Dortmund.

Naef, Alain. 2024. "Blowing against the Wind? a narrative approach to central Bank foreign exchange intervention." *Journal of international money and finance.*

Nakamura, Emi, and Jón Steinsson. 2018. "High-frequency identification of monetary non-neutrality: the information effect." *The Quarterly Journal of Economics*, 133(3): 1283–1330.

Reinhart, Carmen M, and Kenneth S Rogoff. 2009. *This time is different: Eight centuries of financial folly.* princeton university press.

Romer, Christina D, and David H Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *American economic review*, 94(4): 1055–1084.

Romer, Christina D, and David H Romer. 2010. "The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks." *American economic review*, 100(3): 763–801.

Romer, Christina D, and David H Romer. 2023. "Presidential address: Does monetary policy matter? The narrative approach after 35 years." *American Economic Review*, 113(6): 1395–1423.

Roos, Michael, and Matthias Reccius. 2024. "Narratives in economics." *Journal of Economic Surveys*, 38(2): 303–341. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12576.

Shiller, Robert J. 2017. "Narrative Economics." *The American Economic Review.*

Stiefel, Michael, and Rémi Vivès. 2022. "'Whatever it takes' to change belief: evidence from Twitter." *Review of World Economics*, 158(3): 715–747.

Swanson, Eric T. 2021. "Measuring the effects of federal reserve forward guidance and asset purchases on financial markets." *Journal of Monetary Economics*, 118: 32–53.

Tang, Liyan, Philippe Laban, and Greg Durrett. 2024. "MiniCheck: Efficient Fact-Checking of LLMs on Grounding Documents."

Ter Ellen, Saskia, Vegard H Larsen, and Leif Anders Thorsrud. 2022. "Narrative monetary policy surprises and the media." *Journal of Money, Credit and Banking*, 54(5): 1525–1549.

Weinig, Max, and Ulrich Fritsche. 2024. "Going viral: Inflation narratives and the macroeconomy."

Appendix

A. Summary and Descriptive Statistics

A1. Summary Statistics

Table A1: Narrative Descriptive Summary

	N	Mean	St. Dev.	Min	Max
NYT Articles					
Pent-up Demand	1,825	0.14	0.44	0.00	4.00
Supply Chain Issues	1,825	1.28	1.49	0.00	11.00
Labor Shortage	1,825	0.16	0.47	0.00	4.00
Energy Crisis	1,825	0.37	0.90	0.00	9.00
Pandemic Impact	1,825	1.24	1.40	0.00	10.00
Russia-Ukraine War	1,825	0.40	0.99	0.00	7.00
FOMC Transcripts					
Pent-up Demand	41	0.16	0.24	0.00	0.96
Supply Chain Issues	41	0.33	0.37	0.01	0.97
Labor Shortage	41	0.28	0.26	0.01	0.71
Energy Crisis	41	0.23	0.36	0.00	0.98
Pandemic Impact	41	0.37	0.39	0.01	0.99
Russia-Ukraine War	41	0.19	0.36	0.00	0.99

Table A2: H1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
NYT	246	0.110	0.394	0	3
FOMC	246	0.262	0.340	0.003	0.990

Table A3: H2: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
NYT	492	0.175	0.500	0	3
FOMC	492	0.262	0.339	0.003	0.990
Post	492	0.500	0.501	0	1
# Inflation Articles	492	31.244	8.142	24	67

Table A4: H3: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
InfExp12	690	4.206	1.486	0.600	8.324
InfExp12Var	690	11.321	4.519	0.333	25.196
InfExp24	690	3.102	1.075	0.346	5.955
InfExp60	270	1.770	1.236	−1.654	4.110
FOMC	738	0.262	0.339	0.003	0.990
NYT	684	0.190	0.501	0	3
# Inflation Articles	684	8.246	6.428	1	27
POST	738	0.333	0.472	0	1

Table A5: H3: Summary Statistics (Mico Data)

Statistic	N	Mean	St. Dev.	Min	Max
InfExp12	20,208	4.129	4.421	−18.000	19.000
FOMC	20,208	0.257	0.340	0.003	0.990
NYT	18,930	0.187	0.494	0	3
# Inflation Articles	18,930	8.452	6.427	1	27
POST	20,208	0.294	0.455	0	1

A2. Intra-Narrative Correlation

Table A6: FOMC PC Narrative Correlation

	Pent-up Demand	Supply Chain Issues	Labor Shortage	Energy Crisis	Pandemic Impact	Russia-Ukraine War
Pent-up Demand	1	0.80	0.29	0.34	0.67	0.04
Supply Chain Issues	0.80	1	0.47	0.43	0.77	0.30
Labor Shortage	0.29	0.47	1	0.56	0.23	0.56
Energy Crisis	0.34	0.43	0.56	1	0.16	0.86
Pandemic Impact	0.67	0.77	0.23	0.16	1	0.01
Russia-Ukraine War	0.04	0.30	0.56	0.86	0.01	1

Table A7: NYT Narrative Correlation

	Pent-up Demand	Supply Chain Issues	Labor Shortage	Energy Crisis	Pandemic Impact	Russia-Ukraine War
Pent-up Demand	1	0.33	0.19	0.12	0.33	0.13
Supply Chain Issues	0.33	1	0.27	0.30	0.57	0.38
Labor Shortage	0.19	0.27	1	0.13	0.27	0.13
Energy Crisis	0.12	0.30	0.13	1	0.25	0.52
Pandemic Impact	0.33	0.57	0.27	0.25	1	0.34
Russia-Ukraine War	0.13	0.38	0.13	0.52	0.34	1

B. Additional Evaluation Results

B1. Manual Classification Interface

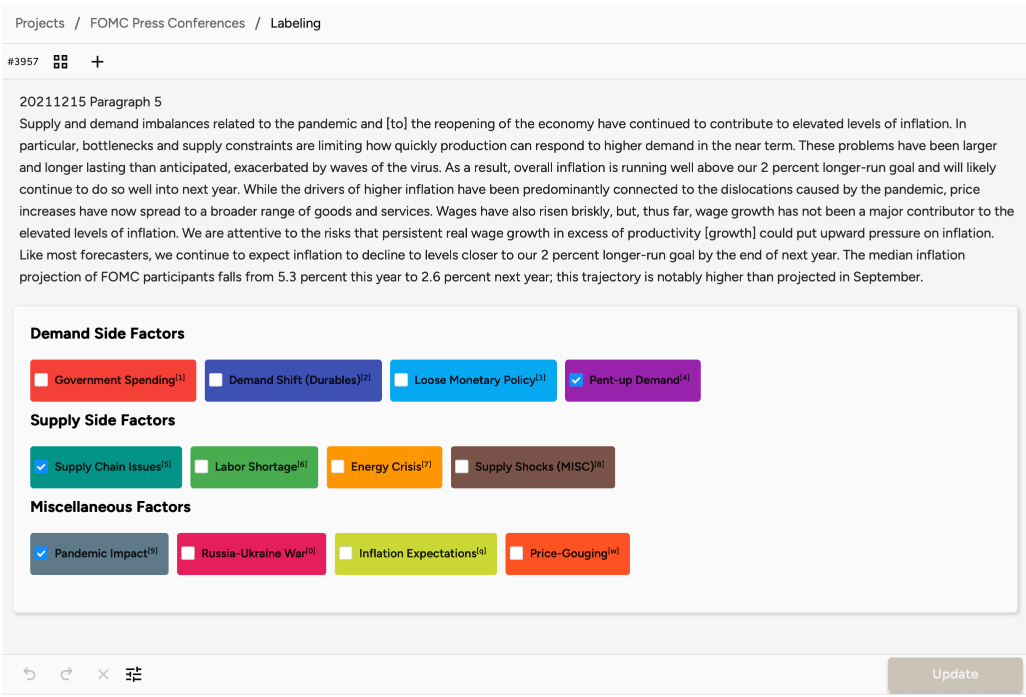


Figure B1 : User Interface for Manual Narrative Coding. The figure illustrates the annotation tool used to hand-code FOMC press conference paragraphs for inflation narratives. It provides a structured layout for paragraph-level text display, binary narrative tagging across twelve categories, and coder agreement tracking.

Table B1: Quantitative Evaluation II: Lasso Regression – Full Regression Table

	Dependent variable: Narrative					
	Supply Chain Issues	Labor Shortage	Pandemic Impact	Energy Crisis	Pent-up Demand	Russia-Ukraine War
	(1)	(2)	(3)	(4)	(5)	(6)
MiniCheck	0.49***		0.50***	0.83***	0.40***	0.98***
	(0.05)		(0.04)	(0.04)	(0.09)	(0.02)
ChatGPT			0.04***	0.14***		
			(0.01)	(0.02)		
LDA Topic 9		0.87***				
		(0.23)				
LDA Topic 18		1.04***				
		(0.16)				
LDA Topic 4			0.09			
			(0.17)			
LDA Topic 19			0.27**			
			(0.12)			
LDA Topic 20			0.04			
			(0.08)			
LDA Topic 31			0.05			
			(0.10)			
LDA Topic 32			0.17**			
			(0.08)			
LDA Topic 15					0.23	
					(0.27)	
LDA Topic 38	1.84***	0.58***	0.63***		1.17***	
	(0.18)	(0.22)	(0.20)		(0.16)	
LDA Topic 17						0.41***
						(0.16)
LDA Topic 43	2.05***	0.32**	0.17	1.68***	1.87***	0.14***
	(0.15)	(0.13)	(0.14)	(0.09)	(0.10)	(0.04)
relatio narrative 1	0.05***		0.07***			
	(0.01)		(0.02)			
relatio narrative 35	0.09***					
	(0.03)					
relatio narrative 15			−0.02		0.02	
			(0.03)		(0.02)	
relatio narrative 42					0.11***	
					(0.04)	
relatio narrative 19		0.07**				
		(0.03)				
relatio narrative 21		0.33***				
		(0.04)				
relatio narrative 22		0.06*				
		(0.04)				
Constant	−0.06***	−0.05***	−0.03***	−0.03***	−0.05***	−0.01***
	(0.01)	(0.01)	(0.01)	(0.003)	(0.01)	(0.003)
Observations	723	723	723	723	723	723
R^2	0.63	0.25	0.37	0.64	0.51	0.88
Adjusted R^2	0.63	0.24	0.36	0.64	0.51	0.88

Note: Standard errors in parentheses: *p<0.1; **p<0.05; ***p<0.01. Explanatory variables are selected by a Lasso logistic model with 5-fold cross-validation. Selected predictors were then re-estimated in OLS for interpretability. LDA: 50 topic model; *relatio*: 50 narrative model with small clusters; ChatGPT and MiniCheck: binary indicators.

B2. Extended Qualitative Evaluation: Topic Models

LDA reliability: Inflation Topic

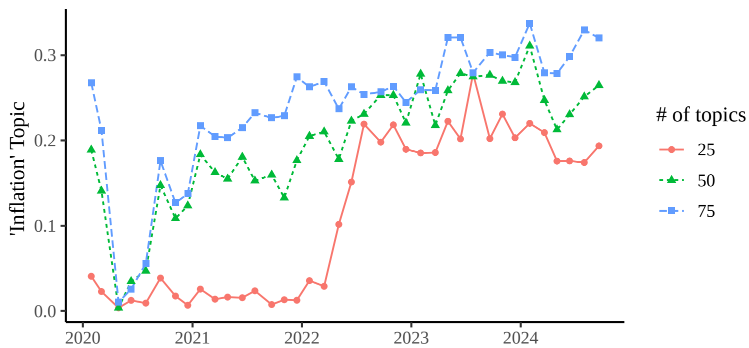
To illustrate the sensitivity of the Latent Dirichlet Allocation (LDA) model with respect to topic size, we re-estimate the LDA described in the paper for $k \in 10, 25, 50, 75$, selecting the inflation-related topic in each model according to the procedure outlined previously. Appendix B.B2 reports the pairwise correlations of the resulting "inflation" topic intensity time series across models. The corresponding time series are visualized in Appendix Appendix B.B2. The table shows that even when two LDAs are trained on an identical corpus, increasing K from 25 to 75 reduces the correlation of the identified inflation topic to 0.67. In terms of explained variance, this means that $R^2 = 0.67^2 \approx 0.45$, or less than half of the variation in the $k = 75$ Inflation series is explained by its $k = 25$ counterpart.

	10	**25**	**50**	**75**
10	1.000			
25	0.965	1.000		
50	0.820	0.829	1.000	
75	0.711	0.672	0.932	1.000

Table B2: Pairwise correlations of Inflation–topic prevalence across LDA models with different K

Notes: The LDA models are estimated on the FOMC press-conference corpus outlined in **??** *and constant* $\alpha = 0.1$*. The Inflation topic is selected by keyword–based matching following Weinig and Fritsche (2024), Müller et al. (2022), and Hansen, McMahon and Prat (2017). Entries are Pearson correlations of the document–level topic–prevalence vectors (*γ*) obtained from each LDA.*

The time series plot provides a visual representation of the inflation topic's evolution, indicating periods where inflation was more prominently discussed: Across topics, there is substantial divergence across models, with the 75-topic and 25-topic models differing by as much as 20 percentage points at certain points in time, which raises questions on their reliability.



LDA validity: Uniform Topic Distribution

We illustrate the issue by sampling 1,000 documents from a symmetric Dirichlet distribution using parameter values ($\alpha = 50/K$ with $K = 50$) suggested by Hansen, McMahon and Prat (2017) following Griffiths and Steyvers (2004) and Kintsch (2006). Figure B2 depicts the resulting histogram of topic shares for each document. Despite extensive sampling, no topic achieves a proportion higher than 12%. Although such uniformity alone is not necessarily problematic, the absence of clearly dominant topics contradicts intuitive expectations about the composition of documents.[13] In practical terms, this raises a fundamental challenge: the need to establish rigorous criteria for assigning documents to specific narratives—a step frequently neglected in the existing literature, where the focus tends to remain on intensive-margin variation in regression analyses. However, more troubling, is the inverse implication: most documents reflect each topic to a non-negligible degree, with even the lowest topic proportions remaining within one standard deviation of the mean.

[13] Alternative approaches, such as employing a more extreme prior (e.g., $\alpha = 1/K$ as in Müller et al. 2022), may alleviate this problem somewhat. However, simulations generally continue to produce similarly diffuse topic distributions, as observed in Figure B2.

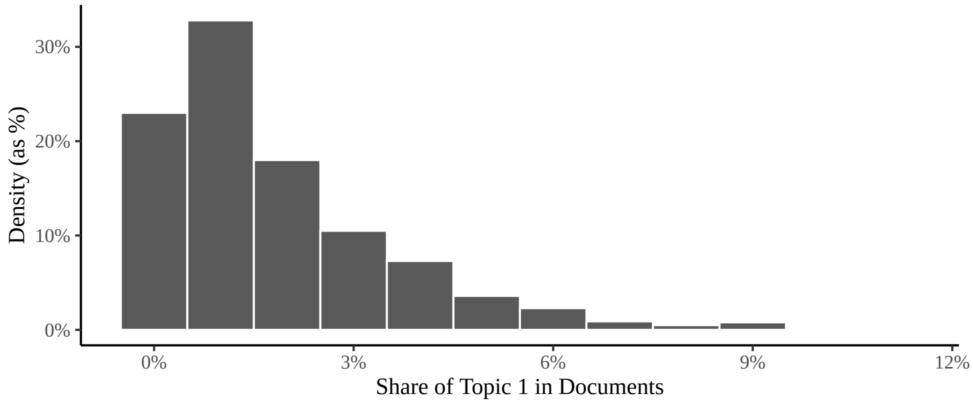
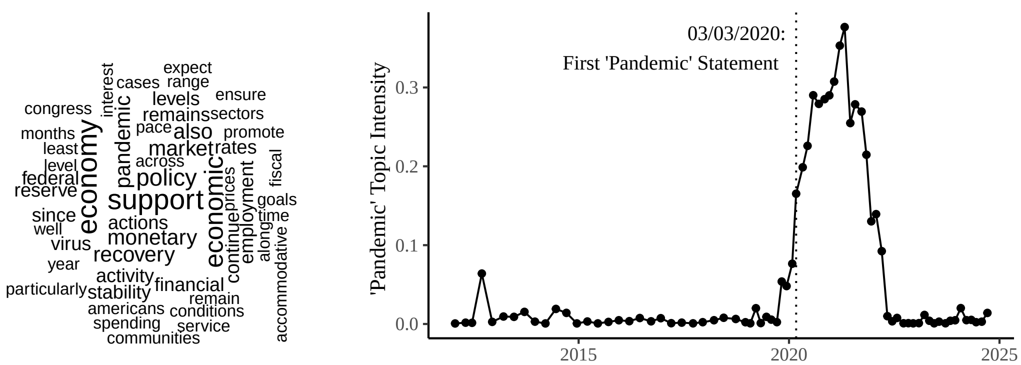


Figure B2 : Histogram of Topic Proportions Drawn from a Dirichlet Prior ($\alpha = 1$)

Notes: This figure presents the distribution of the share that Topic 1 occupies across 1,000 documents, sampled from a symmetric Dirichlet prior with $K = 50$ and $\alpha = 50/K$. The x-axis reports the proportion of Topic 1 within each of the sampled document. The y-axis reports the percentage of documents falling into each bin.

LDA validity: Pandemic Topic

We retrain the 75 LDA model on a dataset spanning from 2014 to 2025. The main objective is to track how discussions around the pandemic and COVID-19 evolved over time. Topic 37 is identified as the "Pandemic" topic, given its high association with terms like "pandemic" and "covid".



(a) Pandemic Topic Intensity Over Time

(b) Word Cloud of Pandemic Topic (Topic 37)

Figure B3 : LDA Analysis of Pandemic-Related Discourse

However, the corresponding time-series pattern (Panel b) indicates substantial topic intensity already before March 2020. Hence, the LDA model labels a significant presence of the Pandemic Topic even *prior* to the COVID-19 pandemic. This is likely due to terms such as "recovery," "stability," "employment," and "fiscal" (as seen in the word cloud), which suggest that pandemic-related vocabulary was often embedded in broader economic discussions concerning downturns and policy responses. This finding raises serious concerns about the topic's specificity, suggesting that it may not be exclusively capturing pandemic-related discourse, thereby challenging its validity as a distinct thematic category.

B3. relatio: Example Narratives

	Narrative	Example Sentence	Freq.
1.	WE AFFECT I	We understand that our actions affect communities, families, and businesses across the country.	77
2.	WE GUIDE INFLATION	The Federal Reserve's role is guided by our mandate from Congress to promote maximum employment and stable prices [...]	39
3.	I HAVE I	The virus is having a profound effect on people across the United States and around the world.	38
4.	WE DO I	We are committed to clearly explaining what we are doing and why we're doing it, both regarding the path of rates and also regarding management of the balance sheet.	36
	...		
25.	INFLATION AFFECT INFLATION	As the stance of monetary policy tightens further, it likely will become appropriate to slow the pace of increases while we assess how our cumulative policy adjustments are affecting the economy and inflation.	11

Table B3: Predominant Narratives identified by the *relatio* Model with $K = 6$

Notes: We ran the relatio application on $L = 50$ named entities, experimenting with cluster counts $K \in \{2, 3, 6, 9, 14, 25, 50, 75\}$ for the phrase embeddings. For each K, we evaluated clustering quality using the silhouette score and inertia (elbow method), followed by a qualitative inspection of cluster coherence. The configuration $K = 6$ provided the best balance between coherence, interpretability, and dimensional coverage.

B4. Manual classification: correlation between annotators

Table B4: Cronbach's Alpha values across narrative paraphrase sets.

	Cronbach's Alpha
Supply Chain Issues	0.958
Labor Shortage	0.983
Pandemic Impact	0.985
Energy Crisis	0.945
Pent-up Demand	0.958
Russia-Ukraine War	0.931