Monetary Policy Narratives and the Transmission of Monetary Policy*

Alexa Kaminski
University of Zurich

Alistair Macaulay
University of Surrey

Wenting Song
UC Davis

January 14, 2026

Abstract

This paper studies how the transmission of monetary policy varies with monetary policy narratives. Using an AI-based data classification algorithm guided by macroeconomic theory, we construct directed graphs of the causal mechanisms described in FOMC transcripts, which capture the narratives used to justify interest rate decisions. Even after purging these narratives of predictable components from contemporaneous macroeconomic conditions, we find substantial variation in narratives over time. Clustering the residual graphs yields three recurring types: an inflation narrative, a finance narrative, and a textbook narrative. Narrative-conditioned local projections reveal that the transmission of monetary policy is strongly narrative dependent, no narrative cluster exhibits the canonical joint decline in inflation and output, and the price puzzle is narrative specific. These results suggest that standard shock measures average over heterogeneous policy episodes and that narrative measurement provides a practical way to operationalize this heterogeneity.

JEL: C45, C55, E52, E58

Keywords: Monetary Transmission, Natural Language Processing, Narratives

*Kaminski (alexa.kaminski@econ.uzh.ch): University of Zurich. Macaulay (a.macaulay@surrey.ac.uk): University of Surrey. Song (wtgsong@ucdavis.edu): University of California, Davis.

1. Introduction

When studying the transmission of monetary policy, it is critical to understand *why* interest rates are being changed. Early work recognized the importance of separating exogenous monetary policy shocks from endogenous responses to the macroeconomic environment (Friedman and Schwartz, 1963; Romer and Romer, 1989; Kuttner, 2001). More recently, economists have sought to further divide identified shocks into those that communicate new macroeconomic information to financial markets (i.e., "information effects") and those that do not.[1] This distinction is important because monetary policy changes that convey information about the state of the economy may have very different transmission mechanisms from those that do not.[2] Estimates that fail to distinguish between these types of shocks conflate their effects and therefore identify neither transmission mechanism cleanly.

However, this is just one way of splitting up monetary policy shocks. In principle, there may be many different reasons for an unanticipated change in interest rates, and these may each come with distinct effects on the aggregate economy. In this paper, we take a step toward uncovering this broader heterogeneity, by analyzing the causal narratives invoked by central bankers to justify their decisions. We find substantial variation in the effects of monetary policy shocks depending on the narrative behind them.

Specifically, we combine recent advances in artificial intelligence (AI) tools with theory-based restrictions to extract the graph of causal relationships embodied in the discussions recorded in the transcripts of FOMC meetings. Following Eliaz and Spiegler (2020), we refer to these graphs as the policymaker *narratives*. Although this is the terminology used in recent literature, the idea of studying the causal reasoning of policymakers has a longer history. Romer and Romer (1989) read the same transcripts to manually identify episodes in which interest rates were raised because policymakers judged a recession was necessary to bring down excessive inflation. Indeed, they refer to this as the "narrative approach". Similarly, Blinder (1995) emphasizes that models involving causal reasoning, even informal

[1] The pioneering work by Nakamura and Steinsson (2018) documents the information effect of Fed announcements. Further work by Cieslak and Schrimpf (2019), Jarociński and Karadi (2020), Miranda-Agrippino and Ricco (2021), and Bauer and Swanson (2023), among others, decomposes the information contained in monetary shocks.

[2] Melosi (2017) and de Groot and Haas (2023), for instance, provide evidence on the heterogeneous transmission mechanisms.

ones, are at the heart of monetary policymaking.[3]

We build on and extend these insights. Our approach applies the narrative perspective systematically on rich text data in FOMC transcripts and lets the data dictate the dimensions along which monetary policy shocks should be divided. In doing so, we find that FOMC deliberations are organized around a small number of recurring narrative types, namely, an *inflation narrative*, a *finance narrative*, and a *textbook narrative*. Importantly, these narratives are not only descriptive. Conditioning on them in local projections reveals substantial heterogeneity in output and inflation responses, as well as a narrative-specific concentration of the price puzzle. This suggests that standard shock measures conflate heterogeneous policy episodes with systematically different reasoning used by policymakers and different associated macroeconomic effects.

Narrative measurement. Our measurement starts with the observation that macroeconomic models can be represented as graphs, where causal mechanisms link the relevant agents, variables, shocks, and forces.[4] To make use of this idea, we use a canonical, medium-scale dynamic stochastic general equilibrium (DSGE) model developed by the Federal Reserve Bank of New York (Del Negro et al., 2013) to construct a list of entities that plausibly capture all the major actors and forces in any policymaker's subjective model or narrative. We use this theory-driven entity list to prompt an LLM to scan each FOMC transcript and capture directed links between entities in sentences. We then construct graphs based on those identified links to capture the narratives contained in each transcript. Because of the guidance from theory in the prompt, related terms, such as "interest rates," "federal funds rate," and "policy rate" are extracted as a single entity, "monetary policy," rather than as separate nodes, yielding graphs that are interpretable through the lens of economic models.

This use of economic theory to pre-specify the entity list is the novel feature of this approach relative to existing algorithms to extract graphs from text (e.g., Ash et al., 2024). This constraint is particularly helpful for causal graphs in monetary policy contexts, where the relationships described in the text are often complicated and multi-dimensional, where policymakers may express the same economic concept in varied language, and where there

[3]He writes specifically that "*Some kind of a model — however informal — is necessary to do policy, for otherwise how can you even begin to estimate the effects of changes in policy instruments?*"

[4]See Auclert et al. (2021) for a recent application of this idea for solving heterogeneous-agent models.

is a relevant body of theoretical work that can be employed to generate the entity list.

We apply our method to FOMC transcripts from 1967–2011, converting each month's text into a weighted, directed graph of narrative links. Because our focus is on narratives that justify surprises in monetary policy, we purge each graph of links that reflect contemporaneous macroeconomic conditions, analogous to the way Romer and Romer (2004) residualize interest-rate changes with respect to current data. Specifically, we regress each narrative link's weight on internal Federal Reserve forecasts of current and next-quarter macroeconomic variables, and retain the residual edge weights. The residualized graphs therefore capture aspects of central-bank narratives that cannot be predicted by current inflation and output.

To validate our approach, we show that the residualized graphs for the six contractionary episodes classified by Romer and Romer (1989) align closely with the narratives those authors manually coded: in those months, central bankers became more concerned about the effects of monetary policy on inflation than contemporaneous macro aggregates would predict.

We then cluster the residualized graphs by computing pairwise cosine distances between narrative graphs and applying spectral clustering to the resulting distance matrix. This groups months with similar narrative structure. The gap statistic selects the optimal number of clusters, yielding three clusters that capture the dominant narrative types in the transcripts. This data-driven approach identifies the natural categories of monetary-policy narratives without imposing pre-defined labels.

Applying this method, we find that FOMC deliberations repeatedly organize around three recurring narrative types. The first is an inflation narrative, which frames monetary policy as a reaction to inflation and emphasizes underlying causes of inflation. The second is a finance narrative, which highlights financial intermediation, credit conditions and balance-sheets. The third is a textbook narrative, which centers around the canonical inflation-output tradeoff of monetary policy. Importantly, these narratives are identified purely based on the text data and do not rely on information about macroeconomic outcomes.

Heterogeneous transmission of monetary narratives Having partitioned monetary policy narratives into interpretable clusters, we study whether different rationales for changing interest rates are associated with different transmission effects. Using local projections

of macroeconomic variables on monetary policy shocks (Romer and Romer, 2004), we find substantial narrative-dependent differences in the transmission to both inflation and real activity. These differences are not limited to peak magnitudes but also involve timing, persistence, and short-run dynamics.

In particular, shocks associated with the inflation narrative generate faster and stronger disinflation but relatively weak real contraction, while shocks within the finance narrative cluster show the largest and most persistent decline in real activity but more muted inflation responses. Shocks accompanied by the textbook narrative lie between these extremes, yet display an initial rise in industrial production and a price puzzle — both inconsistent with standard models. No single narrative cluster produces the canonical joint decline in inflation and output following a contractionary monetary shock predicted by standard theories. Moreover, classic empirical puzzles such as the price puzzle are narrative specific, concentrated only in the finance and textbook narratives.

The heterogeneous transmission stands in contrast with the monetary transmission under full information rational expectations (FIRE), where a given monetary policy shock yields identical impulse responses regardless of the reasons behind the policy change. Our findings therefore implicitly highlight the importance of expectations in the transmission of monetary policy, and the differential effects that narratives have on those expectations.

Finally, our results also suggest that assessing monetary policy using average impulse responses can conceal substantial heterogeneity across policy episodes that differ in their underlying reasoning. As a result, both the canonical joint decline in output and inflation in response to contractionary monetary shocks and the classic price puzzle may partly reflect aggregation across episode types. More broadly, narrative measurement provides a practical way to operationalize heterogeneity in policy reasoning, and thereby refine empirical assessments of monetary transmission.

Related Literature. This paper relates to three main strands of literature. First, on the methodological front, we contribute to the growing literature using textual data to address questions in economics and finance (surveyed by Gentzkow et al., 2019; Ash and Hansen, 2023). Specifically, we develop a method that uses recent advances in large language models (LLMs) to extract the relationships between economic entities described in texts, in a way

that makes such causal chains amenable to empirical analysis. This is of particular relevance to the literature on economic narratives (Shiller, 2017), which can be formalized as graphs like those we measure in text data (Eliaz and Spiegler, 2020).

There are two main existing approaches to quantifying and testing narratives, both using text data. Andre et al. (2022) hand-classify the answers to open-ended survey questions on inflation. This careful manual approach has proven valuable in survey contexts (Haaland et al., 2024), where the length and the number of documents allow for manual classification. In contrast, Ash et al. (2024) develops an algorithm to automatically translate texts into relationship graphs. A key strength of their inductive method — a data-driven approach that discovers entities and their connections directly from the text without prior domain constraints — is its versability across social science disciplines.[5] We complement their broad discovery with a deductive approach, anchoring the graph construction around a set of entities grounded in a canonical structural macroeconomic model. By using a LLM to capture these theoretically defined nodes, we build on the scalability of the Ash et al. (2024) approach while ensuring the output is interpretable through theories of monetary policy.

Second, our application relates to the large literature analyzing texts from central banks. Previous studies have used this rich source of data to infer policymaker preferences (Chappell et al., 1997; Malmendier et al., 2021; Shapiro and Wilson, 2022), analyze the policymaking process (Hansen et al., 2018), to study the effects of central bank communication (Hansen and McMahon, 2016), and to improve the measurement of monetary policy shocks (Aruoba and Drechsel, 2025). We contribute to this literature by examining changes in central bank *narratives* using FOMC transcripts. The rich, high-dimensional nature of text data is important for this measurement: a single interest-rate change could be consistent with a wide variety of internal narratives, with different implications for the efficacy of monetary policy and the accuracy of predictions.

Finally, in pursuing this analysis we also contribute to the broader literature on the transmission of monetary policy and the content of monetary policy shocks (see reviews in

[5] Goetzmann et al. (2022) and Macaulay and Song (2022) also provide alternative ways to measure narratives, but these are specific to particular contexts and do not easily translate to monetary policy analysis. Similarly, the grammatical rule-based approach of Arold et al. (2025) is valuable when the text of interest consists of tightly defined standardized phrases, such as those seen in legal contexts, but may be less suited to identifying causal reasoning in informal deliberative discussions where the same relationship may be expressed in a variety of ways.

Ramey, 2016; Romer and Romer, 2023; Bauer and Swanson, 2023). An influential strand of this literature (e.g., Cieslak and Schrimpf, 2019; Jarociński and Karadi, 2020) has made important progress in separating monetary policy shocks from information effects (emphasized by Nakamura and Steinsson, 2018) by studying the co-movement between monetary shocks and other variables. We propose a complementary approach that builds on the seminal work by Romer and Romer (1989, 2023), which distinguishes different reasons for interest rate movements based on the way policymakers describe their decisions. Our text-based approach allows us to be guided directly by policymakers' own articulated reasoning in distinguishing different kinds of monetary policy shock. Consequently, we are able to identify multiple dimensions of heterogeneity that extend beyond what has previously been studied.

2. Data and Methodology

2.1. Data

Our empirical analysis uses historical transcripts of the Federal Open Market Committee (FOMC) of the Federal Reserve System. The data consists of 531 transcripts from 1967 to 2011, corresponding to approximately one meeting every six weeks. From 1967 to 1976, the data take the form of *Memoranda of Discussion*, which provide detailed summaries of the Committee's deliberations. Beginning in 1977, the records transition to verbatim *Transcripts* of the meetings.

FOMC transcripts are particularly valuable for identifying the reasons behind policy rate decisions because they provide a verbatim record of the Committee's deliberations. Their structured format—beginning with staff presentations on major topics, followed by member-by-member discussions—makes it possible to observe not only the information available to policymakers at the time of the meeting, but also how they interpret it. Table 1 reports summary statistics for this dataset.

We supplement the FOMC transcripts with five additional data sources. First, when prompting the LLM to capture causal links in FOMC transcripts, we make use of the economic relationships contained in the DSGE model developed by the Federal Reserve Bank of New York (Del Negro et al., 2013). Second, before our empirical analysis, we residualize our narrative measure with respect to contemporaneous Greenbook staff forecasts of key

Table 1: Summary statistics: FOMC meeting records (1967–2011)

	# of	# of words					
	Documents	Mean	Median	P10	P90	Min	Max
All meetings	531	28,536	27,294	4,568	50,868	193	94,826
Memoranda of Discussion	123	24,322	24,338	18,648	31,948	3,112	39,914
Meeting transcripts	303	38,012	33,910	22,023	63,121	12,699	94,826
Conference calls	105	6,128	5,178	1,614	12,866	193	24,227

macroeconomic variables, which proxy for the information about current economic conditions available at each meeting. We use the digitized Greenbook forecast series compiled by Romer and Romer (2004) and updated by Wieland (2021). Lastly, we run local projections using CPI and industrial production data obtained from FRED (Federal Reserve Bank of St. Louis, n.d.a,n), and the narratively identified monthly monetary shocks constructed by Romer and Romer (2004) and extended by Wieland and Yang (2020).

2.2. Measuring central banks' models from transcripts

To measure narratives, we begin from the observation that economic narratives can be represented by graphs that formalize causal relationships. In these graphs, vertices represent economic entities (such as agents, variables, shocks, and frictions), and edges denote the causal relationships among them. This approach builds on the broader literature that models narratives as subjective models of the economy (Eliaz and Spiegler, 2020; Macaulay and Song, 2022).

Our method consists of three steps. First, we prompt a large language model (LLM) to extract causal relationships between economic entities from FOMC transcripts. Then, we use these extracted relationships to construct directed graphs that represent policymakers' implicit models. Third, we apply tools from network theory to analyze these graphs, including constructing differences across meetings using graph distances. The following subsections describe each step in detail.

2.2.1. Relationships Extraction

We use the frontier, pretrained LLM GPT-5.1 by OpenAI to identify economic entities and causal relationships among them as described in FOMC transcripts.

We prompt the LLM to identify four types of entities present in economic frameworks: (*i*) actors (such as firms and households), (*ii*) variables (such as inflation, labor supply etc), (*iii*) shocks, and (*iv*) frictions. The complete entity list we specify is based on a widely used, medium-scale DSGE model developed by the Federal Reserve Bank of New York (Del Negro et al., 2013). As a workhorse model used by the central bank to produce regular forecasts, it provides a comprehensive list of the main entities relevant for economic analysis.

A well-known challenge with named entity recognition is that large text corpora often yield an intractably large number of identified entities. By pre-specifying a theory-based entity list in our prompt, we guide the LLM to focus on economically relevant entities, leveraging the model's ability to compare semantic similarities between the input text and the provided entity list. This approach offers two advantages: first, it improves relevance by ensuring that only economic entities are extracted; second, it reduces dimensionality by mapping entities with slightly different wording to the single entity specified in the prompt.

Our full prompt is provided in Figure 1. In addition to entities, we prompt the LLM to extract the causal relationships between the entities by identifying the source and target entities, as well as describing how the source entity affects the target entity.

We drop any relationships that include entities that are not on the prespecified list (0.93% of the extracted relationships). The final output contains 35 distinct entities and 54,930 causal relationships between them, averaging approximately 120 causal relationships per transcript. Appendix Table 2 contains summary statistics of relationships extracted from the transcripts, and Appendix Figure 9 displays the 15 most frequent entities, with output, monetary policy and inflation are the top three frequently mentioned entities.

Figure 1: Prompt for Identifying Causal Relationships

You will be given an excerpt from the FOMC minutes/transcript.

EXCERPT:
{`EXCERPT`}

Task: Extract causal relationships that are stated or clearly implied in the excerpt.

Causality rule (IMPORTANT):
Treat a statement as causal if it indicates a directional influence, including hedged or qualitative language such as: "because", "led to", "resulted in", "contributed to", "put upward/downward pressure on", "boosted", "reduced", "supported", "weighed on", "tended to", "would/might/could lead to", "likely to", "implies that". These are EXAMPLES, NOT an exhaustive list.

Treat "suggests/indicates/consistent with" as causal only when the text also states a directional effect (A affects B).

Set source_entity to the driver/cause and target_entity to the affected outcome, following the excerpt's wording.

Ontology + inclusion rule:
Output a relationship if BOTH source_entity and target_entity can be mapped to one of the standardized items below (Actors/Variables/Shocks/Frictions), using phrases that appear in the excerpt (including synonyms/paraphrases). Do NOT output anything outside the standardized items. Do NOT invent entities not grounded in the excerpt. If multiple standardized items could fit, choose the best fit from context.

The specified entities and their standard names are:

- **Actors**: firms (for firms/entrepreneurs/employers), households (for households/employees/workers/consumers), government (for government/state), central banks, commercial banks
- **Variables**: taxes, labor demand, labor supply, inflation (for price level/inflation), loans (for loans/credit conditions), capital, output, consumption, investment, deposits, government debt, investment tools, regulation and supervision (for regulatory tools/banking supervision/financial regulation), monetary policy (for monetary policy or interest rates), fiscal policy
- **Shocks**: investment shock, demand shock, spread shock, policy shocks, productivity shocks (for TFP shocks or productivity shock), labor supply shock, price shock (for markup shock or price shock), financial crisis, debt crisis, banking crisis, currency crisis, economic crisis
- **Frictions**: price rigidity, wage rigidity, credit frictions (for credit frictions or financial frictions)

Additional mapping guidance (use only when the phrase appears in the excerpt; NON-EXHAUSTIVE examples, not a whitelist):

- "interest rates", "rates", "federal funds rate", "policy rate" → monetary policy
- "prices", "price level", "price inflation", "inflation pressures", "price pressures" → inflation
- "economic activity", "growth", "real GDP", "output growth" → output
- "consumer spending", "household spending", "spending" → consumption
- "business investment", "capital spending", "capex" → investment
- "credit", "lending", "bank lending", "lending standards", "credit conditions", "financial conditions", "financing conditions" → loans
- "credit spreads", "risk premiums" → spread shock
- "liquidity strains", "funding pressures" → banking crisis OR credit frictions (choose the best fit from context)

Output format:
Return ONLY a JSON array (no prose, no markdown). Each element must be an object with `"source_entity"`, `"target_entity"`, and `"description"` (brief paraphrase of the causal direction). Return `[]` only if there are truly no causal relationships under the rules above.

Figure 2: Example: GPT Output and Constructed Graph

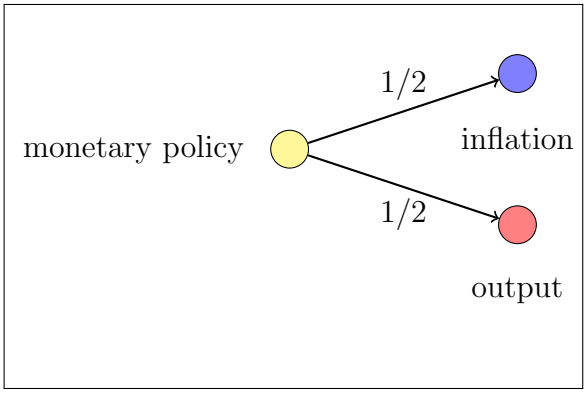
(a) GPT Output

```
{"causal_relationships":  [

{"source_entity":"monetary policy",
"target_entity":"output"},

{"source_entity":"monetary policy",
"target_entity":"inflation"}
] }
```

(b) Constructed Graph



Notes: Panel (a) reports the GPT output of Alan Greenspan's speech from the FOMC meeting dated March 22, 1994, quoted in the main text (illustrative). Panel (b) displays the graph constructed based on the GPT output.

To demonstrate our methodology, we use a paragraph from Alan Greenspan's speech from the FOMC meeting on the March 22, 1994:

> "One risk is that we might be overestimating the strength of the economy, and a less accommodative monetary policy might damp its growth too much. [...] It also may be that this point is significantly different from other points historically, but my guess is that if we push on this economy, we will get inflation and we will end the growth."

This meeting marked the early phase of the Greenspan tightening cycle, a series of rate hikes after a period of low interest rates following the 1990-01 recession. In the excerpt, Alan Greenspan, chair of the FOMC, emphasized that despite a potential negative effect on output of tighter monetary policy ("less accommodative monetary policy might damp its growth too much"), the lack of tightening will increase inflation ("if we push on this economy, we will get inflation").

Panel (a) of Figure 2 displays the GPT output.[6] It extracted two causal relationships from the text: from monetary policy to output, and from monetary policy to inflation. The

[6]While the prompt instructs the LLM to include a brief textual description of each relationship, this is only used for ex post verification and not the graph construction step. It is therefore omitted from the illustrative output shown here.

output accurately capture the economic entities and causal relationships contained in the speech. Panel (b) displays the graph constructed based on the GPT output, which we detail in the next subsection.

2.3. Graph Creation

Based on the identified relationships, we construct directed graphs to represent central bank models. Each graph is denoted by $G = (V, E, W)$, where $V = \{1, ..., n\}$ represents the set of vertices and $E \subseteq V \times V$ represents the set of edges.

To capture the relative importance of relationship, we assume that more frequently mentioned relationships carry greater significance. We capture this importance using a weighting function $W : E \to \mathbb{R}$ that assigns each edge $e_{i,j} \in E$ a weight $w_{i,j} \in \mathbb{R}^+$ equal to the proportion of times that relationship appears relative to the total number of relationships identified in that policy meeting. This weighting also renders transcripts with different lengths comparable. The graph construction is implemented using an off-the-shelf Python package `NetworkX`. We apply this method to construct a graph for each transcript in our dataset, yielding a set of graphs $\mathcal{G}$.[7]

Figure 2b illustrates the graph creation using the relationships extracted from Greenspan's speech. In this simple example, there are two relationships, both originating from the node "monetary policy". Therefore, each edge is weighted 1/2 under the weighting function.

Some of the weighted edges captured so far reflect the Committee's discussions of contemporaneous economic conditions and systematic monetary policy responses, rather than narratives that justify the unsystematic component of monetary policy changes. Because our focus is on narratives, we purge each graph of links that reflect contemporaneous macroeconomic conditions, analogous to the way Romer and Romer (2004) residualize interest-rate changes with respect to current data. Since most of the additional data we use is measured at a monthly frequency, we first aggregate the meeting-level graphs within each month by averaging edge weights across all transcripts that period. This yields a single, weighted graph per period, with edge weights w_{ijt} denoting the average importance of the relationship from

[7]This measurement approach therefore captures the presence and direction of links, but not the sign of the causal effect described: statements describing tighter monetary policy causing inflation to go up or down are treated identically. This is consistent with the formal definition of narratives set out in the theoretical literature (Eliaz and Spiegler, 2020), though unlike that literature we do not impose that the resulting graphs must be acyclic.

node i to node j in time t.

Then, for each pair of nodes i and j, we estimate

$$w_{ijt} = c + \Gamma' Z_t + \omega_{ijt} \tag{1}$$

where Z_t is the vector of Greenbook staff forecasts for current and next-quarter macroeconomic conditions.[8] We use the residualized weight, ω_{ijt}, as the edge weight. The residualized graphs capture aspects of central-bank narratives that are unrelated to the current state of the US economy.

2.4. Graph Distance Matrix

Once narratives are represented as graphs, we can apply standard methods from network theory to analyze them. One aspect that we focus on is the distance *between* narratives, which we measure using the cosine distance, a common distance metric in text and network analysis (e.g. Hoberg and Phillips, 2010; Girardi et al., 2021; Goetzmann et al., 2022).[9]

For two graphs $G_1 = (V_1, E_1, W_1)$ and $G_2 = (V_2, E_2, W_2)$, the cosine similarity is defined as:

$$S_c(G_1, G_2) = \frac{\sum_{e \in E} W_1(e)\, W_2(e)}{\sqrt{\sum_{e \in E} W_1(e)^2}\, \sqrt{\sum_{e \in E} W_2(e)^2}}, \tag{2}$$

where $\tilde{W}_{G_i}(e) = W_i(e)$ if $e \in E_i$ (the weight on edge e) and 0 otherwise. The cosine distance $D_c(G_1, G_2)$ is then defined as:

$$D_c(G_1, G_2) = 1 - S_c(G_1, G_2). \tag{3}$$

This distance ranges from 0 to 2, where 0 indicates identical weighted edges (up to a positive scalar multiple), 1 corresponds to unrelated patterns in edge weights (the vector of edge weights are orthogonal), and approaches 2 as the patterns of edge weights point in

[8] For each FOMC meeting we use Greenbook projections for the unemployment rate and the growth rates of real GDP, nominal GDP and the GDP deflator, at the current-quarter and the one-quarter-ahead horizon.

[9] The cosine distance is preferable to other graph distance measures (e.g. Jaccard distance) here, in particular because the residualization step described by equation (1) implies our graphs may have both positive and negative weights. Other alternatives, such as L1 and L2 distances, are more sensitive to changes in scale, for instance, due to few relationships per meeting.

opposite directions.[10]

3. Policymaker Narratives and Monetary Transmission

In this section, we use the graphs extracted from the FOMC transcripts to analyze the motivations and justifications for monetary policy shocks, and the relationship between the narrative given for a shock and its subsequent transmission. We find that narratives differ greatly between periods, and that these distinct types of shock have distinct effects: the power of monetary transmission varies with the narrative that accompanies it.

3.1. Validation of graph construction methodology

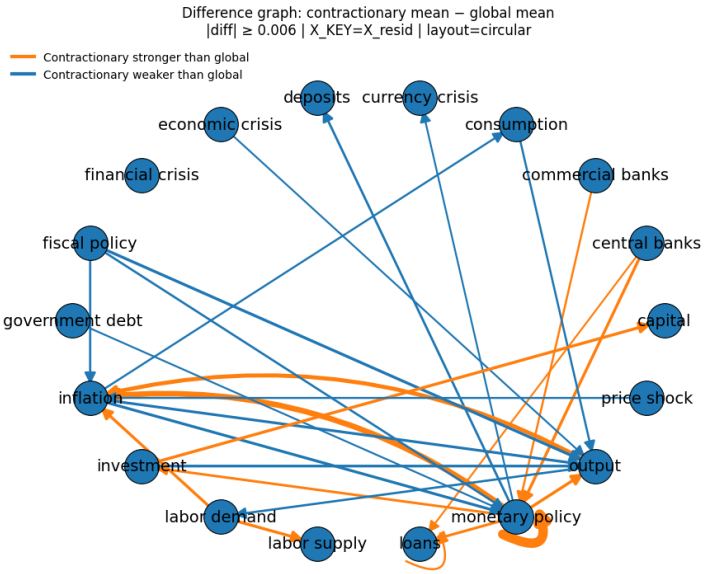
We begin by verifying whether the graphs we extract from FOMC transcripts accurately capture monetary narratives.

External validation using Romer and Romer narratives. We first compare our extracted graphs to a commonly used, narratively identified monetary shock series developed by Romer and Romer (2023), who read the FOMC transcripts and identify six episodes of unambiguous contractionary monetary policy changes. These six months were "*times when monetary policymakers felt the economy was roughly at potential (or normal) output, but decided that the prevailing rate of inflation was too high.*"(Romer and Romer, 2023, pp. 1399). Since the transcripts in these months have a consistent well-understood narrative, they provide a clear test of whether our algorithm can extract the relevant mechanisms discussed in policy meetings.

Figure 3 reports the graph aggregating all causal relationships identified by our algorithm from FOMC transcripts across the six episodes identified by Romer and Romer (2023). The graph displays edge weights that have been residualized with respect to contemporaneous macroeconomic conditions. Positive weights (shown in orange) indicate causal links which are stronger during the six episodes than would be predicted by contemporaneous macroeconomic conditions, while negative weights (shown in blue) indicate links that are weaker than would be predicted. Line thickness represents the strength of the link.

[10]Recall that our analysis will use residualized edge weights (Section 2.3), so edge weights can be positive or negative.

Figure 3: Aggregated graph from FOMC transcripts during Romer and Romer (2023) contractionary episodes



Notes: Figure plots the residualized edge weights for the graphs extracted from the 6 episodes identified as monetary policy shocks by Romer and Romer (2023) within our sample period (1967-2011). Orange arrows denote links are more frequently present in those months than is predicted by macroeconomic conditions, blue arrows denote links that are less frequently present. Line width reflects the magnitude of the residualized weight.

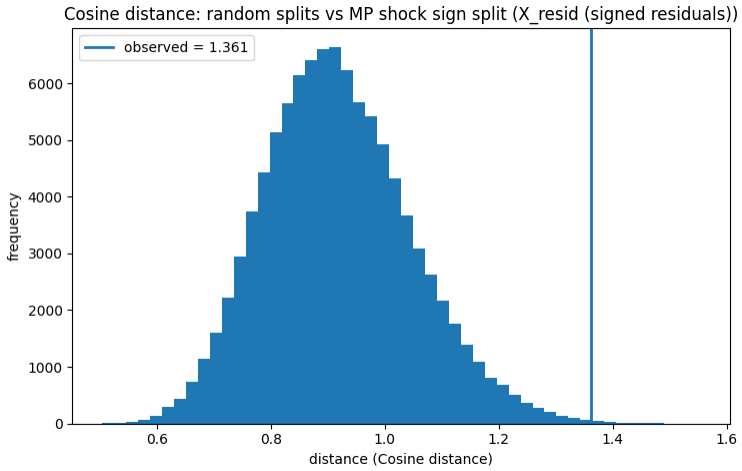
Consistent with Romer and Romer (2023), the link from monetary policy to inflation is substantially more prominent in these episodes than average behavior from other periods would suggest, and the link from monetary policy to output is also positive. At the same time, all other determinants of output receive less attention than contemporaneous macroeconomic conditions would predict, consistent with the notion that policymakers judged output to be at potential.Interestingly, links from monetary policy to loans and to capital (via investment) also exhibit substantially greater weights in these episodes relative to the remainder of the sample, patterns not emphasized by Romer and Romer (2023). This finding illustrates how our AI-driven approach can complement narrative identification and uncover new causal relationships from even well-studied documents.

Comparison between contractionary and expansionary narratives. Our second test verifies whether our constructed graphs differ between contrationary and expansionary

periods, as defined by Romer and Romer (2004) and extended by Wieland and Yang (2020).[11]

This test serves two purposes. First, it further validates our graph extraction method. If policy shocks are indeed justified in FOMC meetings and our method successfully detects those justifications, we should observe substantial differences in graphs extracted from periods with shocks of opposite signs. Second, it provides insights into how contractionary shocks are justified relative to expansionary shocks.

Figure 4: Cosine distance between contractionary and expansionary shock periods in the Wieland and Yang (2020) shock series, compared with the distribution of cosine distances between randomly drawn graphs.



Notes: Figure plots the histogram of cosine distances between the graphs from random splits of the transcripts 1967-2011 (blue bars). The solid blue line is the cosine distance between the average graphs computed from all periods with positive (contractionary) and negative (expansionary) shocks in the Wieland and Yang (2020) extension of the Romer and Romer (2004) monetary policy shock series.

We begin by building two graphs: one that aggregates links across all transcripts from months with positive Romer and Romer (2004) shocks, and another that does the same for months with negative shocks. The cosine distance between these graphs is 1.36. To gauge whether this difference is meaningful, we construct a reference distribution by repeatedly and randomly splitting the same set of transcripts (those in months with non-zero Romer and Romer (2004) shocks) into two groups, without regard to shock sign. For each of 100,000 random splits, we construct graphs for the two groups and compute their cosine distance.

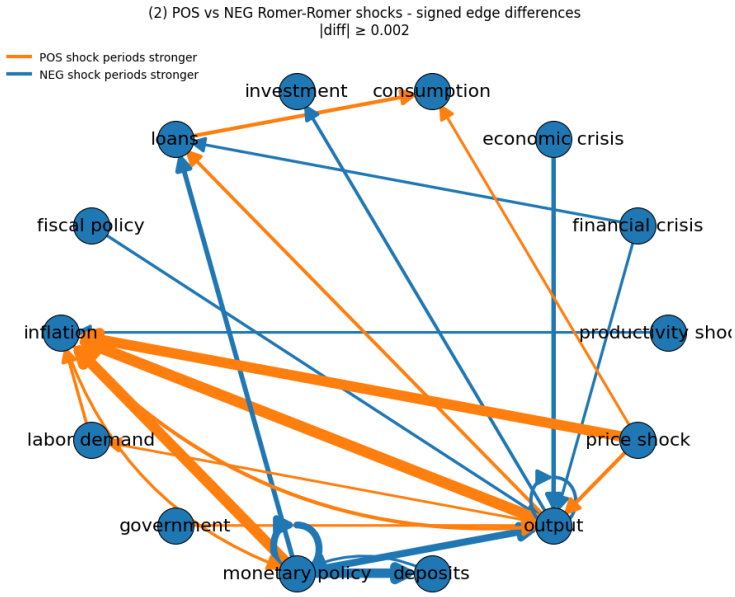
Figure 4 summarizes these results. The histogram shows the distribution of distances

[11]The shocks by Romer and Romer (2004) are constructed differently from the episode-based study by Romer and Romer (1989, 2023). Specifically, the shocks by Romer and Romer (2004) occur every month, and are constructed by regressing the intended change in the Federal Funds Rate at each FOMC meeting on a vector of internal Greenbook forecasts, and then taking the residual. There is, therefore, no selection based on the narratives or causal arguments made in the FOMC meetings.

from the random splits, with the distance obtained from the shock-sign split marked with a solid line. The observed distance between graphs from contractionary and expansionary periods lies far in the upper tail: it exceeds 99.86% of distances generated under random partitioning. This suggests that policymakers use systematically different causal arguments when delivering contractionary and expansionary shocks, consistent with them providing narrative justifications for interest rates that deviate from what prevailing macroeconomic data would predict.

Having confirmed that graphs indeed differ between contractionary and expansionary periods, we now examine the specific nature of these differences. Figure 5 reports the causal mechanisms that differ most strongly across periods with contractionary versus expansionary shocks. Nodes and links that appear more frequently during contractionary shock periods are shown in orange, while those more prevalent during expansionary shock periods are shown in blue. Line thickness indicates the magnitude of differences in link prevalence.

Figure 5: Edge weights differences: contractionary vs. expansionary shock periods



Notes: Figure plots the difference in edge weights between graphs extracted from periods with positive (contractionary) and negative (expansionary) monetary policy shocks in the Romer and Romer (2004) series extended by Wieland and Yang (2020). Orange arrows denote a link that is more frequently present in months with contractionary shocks, blue arrows denote a link more frequently present in months with expansionary shocks. Line width reflects the magnitude of the difference.

Figure 5 highlights a clear divide between the causal narratives accompanying contractionary and expansionary shocks. Contractionary shocks are accompanied by causal narratives emphasizing how price shocks, output, and monetary policy affect inflation, indi-

cating a prominent role for supply-side considerations.[12] In contrast, expansionary shocks are accompanied more by narratives involving financial forces, such as the effect of monetary policy on deposits and loans, as well as direct effects of monetary policy on output.

The Romer and Romer (2004) shocks used here are constructed by purging interest rate changes of the component predictable from the Federal Reserve's internal forecasts of output and inflation. The prominence of inflation in policymaker graphs during contractionary shocks is not therefore driven by inflation being higher during those months. The differences we observe in Figure 5 are specific to the sign of monetary policy *shocks*, not to the prevailing state of the economy at the time.

3.2. Clustering central bank narratives

The preceding analysis demonstrates large systematic differences between the narratives accompanying contractionary and expansionary changes in monetary policy. However, this does not imply that all contractionary periods, or all expansionary periods, have identical narratives. The results in Figures 4 and 5 reflect average graphs aggregated across all periods of a given shock sign. We now examine the heterogeneity in narratives underlying those averages.

Narrative variation between and within shock types. We compare narrative variation across contractionary and expansionary shocks to variation within each category. To do so, we construct the "central" graph for each shock sign by averaging all graphs from periods with contractionary or expansionary Romer and Romer (2004) shocks.

We then compute two sets of cosine distances. The between-type distance measures the distance between the two central graphs, equivalent to the solid line in Figure 4. This summarizes the difference between the typical contractionary and expansionary narratives. The within-type distance measures the average distance between each individual monthly graph and its group's central graph, capturing how dispersed narratives are within periods of the same shock sign. Appendix C.1 provides further details on the calculation of these distances.

[12] This interpretation aligns with the six specific contractionary episodes studied by Romer and Romer (1989, 2023), which were selected as instances where the FOMC unexpectedly shifted preferences towards prioritizing inflation control, despite our analysis incorporating all contractionary shocks from the econometrically identified series in Wieland and Yang (2020).

The within-type distance is substantial in both contractionary (0.89) and expansionary (0.90) periods—nearly two-thirds the magnitude of the between-type distance (1.36). This indicates that while narratives accompanying contractionary and expansionary shocks differ systematically on average, there is also considerable heterogeneity within each shock sign.

This heterogeneity has important implications. If the extracted graphs reveal the narratives justifying interest rate decisions that depart from those suggested by contemporaneous macroeconomic conditions, then not all contractionary shocks are alike, nor are all expansionary shocks. The scalar shock series by Romer and Romer (2004) thus aggregates many distinct shocks, enacted for different reasons and justified through different causal narratives.

Clustering of narratives. To study this heterogeneity, we partition interest rate decisions by the monetary narratives that are invoked and study whether the transmission of monetary policy differs across these narratives. This section describes our procedure for partitioning the narratives. Rather than imposing an a priori classification, we adopt a data-driven approach. Specifically, we cluster the extracted graphs using spectral clustering (Ng et al., 2002). Among clustering methods, spectral clustering provides the cleanest separation between clusters, as measured by the silhouette coefficient (Rousseeuw, 1987), and exhibits high stability across random initializations. Appendix C.2.3 shows that our results are robust to other standard clustering approaches, including k-medoids and agglomerative clustering.

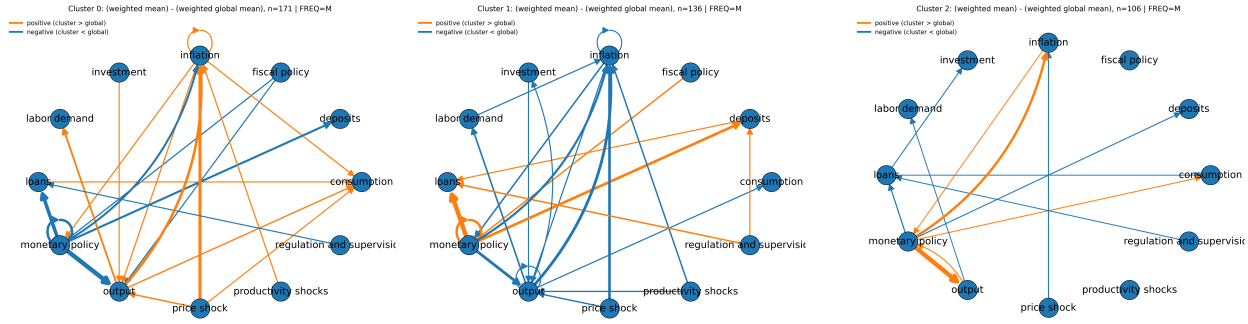
We choose the optimal number of clusters, K, using the gap statistic of Tibshirani et al. (2001a).[13] The gap statistic is positive for all values of K considered, indicating the presence of clustering, but applying the standard Tibshirani decision rule reveals no meaningful improvement beyond $K = 3$. Accordingly, we choose to divide the narratives in three clusters. This choice is further supported by stability measures, silhouette scores, and separation metrics, all of which favor $K = 3$ Appendix C.2.2 contains further details on the selection of K.[14]

Figure 6 reports the graphs associated each of the three clusters. Each panel shows the difference between the average graph constructed from transcripts in a given cluster and the

[13]The gap statistic compares the observed within-cluster dispersion, W_K, to the expected dispersion, W_K^{ref}, obtained from reference data generated by permuting edge weights across months to eliminate temporal structure.

[14]In Appendix D.3 we repeat the analysis for $K = 5$. Our main result that different narratives are associated with different transmission to output and inflation remains robust to this alternative.

Figure 6: Clusters of monetary policy narratives



(a) The inflation narrative **(b)** The finance narrative **(c)** The textbook narrative

Notes: Each panel plots the difference in edge weights between graphs extracted from periods within the cluster described in the panel label and the average graph extracted from all transcripts in the dataset. Orange arrows denote a link that is more frequently present in months within the cluster, blue arrows denote a link less present in the cluster. Line width reflects the magnitude of the difference.

average graph constructed from all transcripts. Orange arrows denote links that are more prevalent within the cluster, while blue arrows denote links that are less prevalent. Line thickness reflects the magnitude of the difference.

Cluster (a) exhibits stronger causal links involving inflation than average transcripts, with several orange arrows connected to the inflation node. This narrative emphasizes two aspects of inflation. First, it highlights the determinants of inflation: real economic activity (shown by orange arrows from output and productivity shocks to inflation) and cost pressures (shown by the link from price shocks to inflation). Second, it frames inflation as the key driver of monetary policy decisions. The link from inflation to monetary policy is the only relatively prominent link involving the policy node. Links from monetary policy to other nodes, including inflation, output, loans, and deposits, appear less frequently than in the baseline. This pattern suggests that monetary policy is primarily discussed as reactive, rather than as an independent driver of economic outcomes. Inflation, or inflation risk, is presented as the central justification for policy tightening or loosening. By contrast, alternative channels and justifications for policy changes, such as financial-sector conditions or output dynamics, are mentioned less frequently than in the average transcript, indicating that they play a secondary role in motivating policy decisions within this narrative. These features indicate that inflation serves as the central organizing node of this cluster, motivating our label of Cluster (a) as the *inflation narrative*.

Cluster (b) shifts the causal framing toward financial intermediaries and balance-sheet

channels. The graph reveals two key mechanisms. First, a strong link from monetary policy to loans suggests that policy is framed as operating through credit conditions more frequently than average transcripts. Second, deposits emerge as a central node with multiple orange links pointing toward it, indicating heightened discussion of deposit dynamics and related funding conditions relative to average transcripts. Regulation and supervision are also salient, with prominent orange arrows involving both deposits and loans. In contrast, inflation-centered causal chains emphasized in Cluster (a) and standard transmission channels from monetary policy to inflation and output appear less important. These patterns indicate that, relative to the baseline, meetings in Cluster (b) frame policy decisions primarily through financial-sector conditions, especially credit supply and funding, with inflation-output tradeoff language playing a comparatively smaller role. We therefore label Cluster (b) the *finance narrative*.

Finally, in Cluster (c) the two most prominent orange arrows run from monetary policy to inflation and from monetary policy to output, indicating that these meetings disproportionately emphasize monetary policy's effects on both variables. At the same time, orange arrows from inflation and output toward monetary policy reveal that these variables are also discussed more frequently as drivers of monetary policy changes. In contrast, financial intermediation channels and alternative explanations for inflation dynamics are de-emphasized. These patterns indicate that Cluster (c) frames policy decisions primarily through a conventional inflation-output tradeoff, motivating our label of the *textbook narrative*.

Notably, these clusters do not overlap with existing attempts to distinguish between different varieties of monetary policy shocks. Appendix D.5 shows that our clusters are distributed across periods with both powerful and weak "information effects" as classified by Jarociński and Karadi (2020), and across periods in which shocks are most powerful at different points in the yield curve identified by Swanson (2021).

This proliferation of narrative framings we identify is interesting in itself, particularly for research on central bank decision-making (Hansen et al., 2014, 2018). We now show, however, that it also matters for understanding how monetary policy affects the aggregate economy.

3.3. Do narratives affect shock transmission?

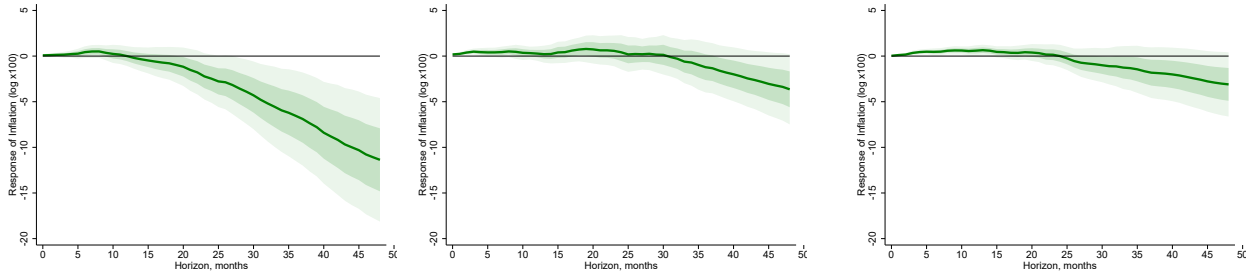
We study whether the narratives that accompany monetary policy shocks affect the transmission of those shocks to the macroeconomy by estimating local projections

$$\log(y_{t+h}) - \log(y_{t-1}) = \sum_{k=1}^{K} \alpha_{kh} \mathbb{1}(cluster_t = k) + \sum_{k=1}^{K} \beta_{kh} (\mathbb{1}(cluster_t = k) \times MP_t) + \Gamma' Z_{t-1} + u_{t+h}, \quad (4)$$

where y_{t+h} is the outcome of interest (log CPI or log industrial production depending on the specification) in month $t+h$; MP_t is the monetary policy shock from Romer and Romer (2004) extended in Wieland and Yang (2020); and Z_{t-1} is a vector of controls including twelve lags of log differences in inflation and industrial production and the first difference of the federal funds rate. This specification mirrors the standard local projection for estimating monetary policy shock effects (e.g. Jorda and Taylor, 2025), but allows the intercept and effects of the monetary policy shock to vary by narrative cluster.

The results from estimating (4) are reported in Figures 7 and 8, which show the impulse responses of cumulative CPI inflation and industrial production growth to a one-percentage-point contractionary monetary policy shock under a given narrative. The results reveal substantial heterogeneity in monetary policy transmission depending on the narrative that accompanies the shock, both in magnitude and timing.

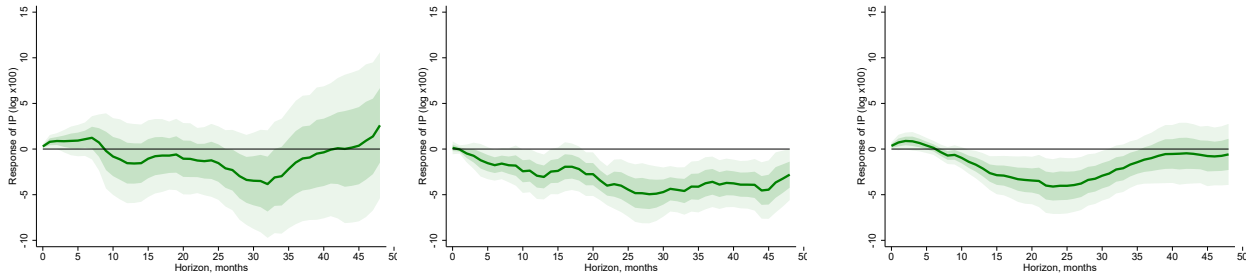
Figure 7: Impulse responses for CPI inflation.



(a) The inflation narrative **(b)** The finance narrative **(c)** The textbook narrative

Figure 7 reports the cumulative responses of inflation to contractionary monetary policy shocks. The disinflationary effect is substantially larger for shocks accompanied by the inflation narrative than for those accompanied by the finance or the textbook narratives.

Figure 8: Impulse responses for Industrial Production.



(a) The inflation narrative **(b)** The finance narrative **(c)** The textbook narrative

Notes: Each panel represents the impulse response of CPI inflation (Figure 7) or industrial production (Figure 8) to a one-percentage-point monetary policy shock (Wieland and Yang, 2020) that occurs alongside a narrative from a given cluster defined in Figure 6, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

When shocks are accompanied by the finance and textbook narratives, inflation initially rises significantly, declining only at longer horizons (approximately two years after the shock). In contrast, shocks associated with the inflation narrative lead to no significant inflation increase on impact, an earlier decline in inflation (around one year after the shock), and a substantially more pronounced disinflationary effect at medium horizons.

This pattern suggests that the "price puzzle" commonly observed in estimates of monetary policy shocks (see, e.g., Ramey, 2016) is concentrated in Clusters (b) and (c). The puzzle is most pronounced and persistent for shocks accompanied by the textbook narrative, even though ex ante, one might expect this cluster to capture the most "pure" monetary policy shocks.[15]

Figure 8 displays the cumulative responses of industrial production growth, which also differ substantially across narrative clusters. The heterogeneity shows up not only in the peak magnitude of each response, but also in the persistence and short-run dynamics. Monetary policy shocks associated with the inflation narrative initially produce the largest and most persistent short-run increase among all clusters, and generate the weakest contraction at longer horizons. Shocks associated with the textbook narrative exhibit a similar short-run rise; however, the response turns negative at 6 months after the shock, before recovering. In

[15]The one- and two-standard-deviation confidence intervals for the inflation response lie entirely above zero for 17 and 11 months under the textbook narrative, compared to 9 and 4 months under the finance narrative. Under the inflation narrative, the two-standard-deviation interval never lies fully above zero, and the one-standard-deviation interval does so only for months 6–8.

contrast, shocks accompanied by the finance narrative generate the most persistent decline in industrial production. The contraction magnitude is marginally larger than that of Cluster (c) at medium horizons, and it remains negative and pronounced over a longer period and lacks the short-run increase observed in Clusters (a) and (c).

3.4. Discussion

Taken together, our findings in Figures 7 and 8 show that monetary policy shocks exhibit heterogeneous transmission patterns depending on the narrative accompanying the policy decision. These differences are visible both in inflation and real activity and are not limited to peak magnitudes, but also involve timing, persistence, and short-run dynamics. This highlights that the heterogeneity is multidimensional and cannot be reduced to a single scalar measure of "strength" or "size" of policy.

Notably, no single narrative cluster delivers the canonical joint response predicted by standard models, namely a clean and monotonic decline in both inflation and real activity following a contractionary shock. The inflation narrative is associated with strong disinflationary effects of monetary policy shocks, but no contraction in output. In contrast, shocks accompanied by the finance and textbook narratives cause significant real contractions, but have significant price puzzles at short horizons, and only weak disinflationary effects in the longer run. This suggests that estimates recovering responses consistent with standard theories may be averaging over heterogeneous episodes motivated by different reasons for changing interest rates, with different impacts.

This heterogeneity in transmission stands in sharp contrast with what we would observe in simple models with full information and rational expectations (FIRE), in which a given monetary policy shock yields identical impulse responses regardless of the reasons policymakers have for their actions. Our results therefore suggest an important role for expectations in monetary transmission, and for the differential effects that narratives have on expectations. In particular, the costless disinflation we observe after monetary policy shocks accompanied by the inflation narrative is consistent with those narrative-shock pairs causing inflation expectations to fall rapidly, as Hazell et al. (2022) and others have argued occurred in the Volcker disinflation. Exploring this hypothesis, and other potential explanations for the transmission heterogeneity we observe, is an important direction for future work.

3.5. Robustness

Appendix D contains four sets of robustness checks. First, we repeat the analysis using alternative distance measures. Appendix D.1 shows that replacing our baseline cosine distance with the $L1$ (Manhattan) or $L2$ (Euclidean) distance does not qualitatively change the findings: periods with positive and negative shocks continue to exhibit clearly different average narratives, and the $K = 3$ clustering continues to recover qualitatively similar narrative types to those in Figure 6. Moreover, re-estimating the local projections in (4) with the new clusters generates impulse responses that are very similar to the baseline across narratives. Relative to our baseline, however, the clusters we find with these alternative distance measures are somewhat less well separated. This is consistent with the fact that $L1/L2$ distances are more sensitive to overall magnitude differences in residual weights, for instance in low-information months in which few relationships are discussed and the residual graphs are correspondingly sparse.

Second, we repeat the clustering step using alternative clustering methods (agglomerative hierarchical clustering and k-medoids). Appendix D.2 shows that the extracted narrative types are likewise qualitatively similar to the baseline, with the same broad inflation-, finance-, and textbook-oriented narratives (though, again, separation is somewhat weaker). The corresponding narrative-conditioned impulse responses are again similar to our baseline estimates.

Third, we increase the number of clusters. Increasing the number of clusters to $K = 5$ continues to recover narrative types that closely align with our baseline clusters, alongside two additional clusters that capture other justifications for policy changes. Re-estimating the local projections conditional on narratives yields the same qualitative conclusion: the transmission of monetary policy differs substantially across narrative episodes, and the baseline patterns for inflation and output responses remain present. For $K > 5$, spectral clustering becomes unstable, which is consistent with the gap statistic evidence that clustering gains are limited beyond a small number of groups.

Fourth, we test if inclusion in each narrative cluster is correlated with other known sources of state-dependence or non-linearity in monetary transmission identified in the literature. Inclusion in each cluster is spread across months in which the monetary policy

shocks had different sizes and signs, and across booms and recessions.[16] This suggests that the narrative-conditioned results do not merely re-label existing state-dependent patterns emphasized in the literature (Tenreyro and Thwaites, 2016; Ascari and Haber, 2022).

4. Conclusion

Monetary policymakers change interest rates for a wide variety of reasons. Some of these changes are endogenous responses to macroeconomic conditions. In this paper, we document that even within exogenous interest-rate shocks, policymakers invoke a range of distinct justifications that vary over time, and that these "shock narratives" are associated with strikingly different macroeconomic outcomes.

To measure shock narratives, we use recent advances in AI to construct graphs representing the causal reasoning employed by the FOMC in its deliberations. We then show that the transmission of an externally identified monetary policy shock varies substantially depending on the graph—or narrative—used to justify it. In particular, we identify three recurring narrative types that do not map cleanly into existing shock classifications: an inflation narrative, a finance narrative, and a textbook narrative. These narratives also matter economically. Narrative-conditioned impulse responses show that the disinflationary effect of a contractionary shock is much larger and faster for shocks accompanied by the inflation narrative, while it is more muted for shocks accompanied by the finance narrative. In contrast, the finance narrative is associated with a more pronounced and persistent contraction in industrial production, whereas real effects are substantially weaker under the inflation narrative. The textbook narrative lies between these two cases along both dimensions. Moreover, the price puzzle is concentrated in the finance and textbook narratives. As a result, no single narrative cluster delivers the canonical clean joint decline in inflation and real activity following a contractionary shock. This suggests that textbook average impulse responses may reflect aggregation over heterogeneous policy episodes motivated by different reasons for changing interest rates.

These results matter for how monetary policy is evaluated and interpreted. If the

[16] Note that this finding is not at odds with the result in Figure 4 that on average graphs differ by shock sign. This is because, as highlighted in Section 3.2, there is a large degree of heterogeneity underlying those average graphs, such that some narratives appear in both contractionary and expansionary shock periods.

transmission of a given policy shock depends systematically on the narrative accompanying the decision, then it is more informative to assess policy through narrative-conditioned responses than using a single benchmark impulse response. This suggests that from an ex post perspective, conditioning on narratives provides a more informative basis for interpreting historical episodes and attributing subsequent movements in inflation and real activity to policy changes. More broadly, our results support a risk-management view of policy analysis, where policy assessment may benefit from scenario-dependent ranges of outcomes associated with different policy episode types rather than relying on one average path.

The results also point to several directions for future work. First, this paper remains agnostic about the source of the association between narratives and policy transmission. Relating narrative clusters to intermediate outcomes, including credit conditions, inflation expectations, and financial market reactions, may help disentangle which channels drive the cluster-specific responses to policy shocks, and whether narratives mainly proxy for information effects or for differences in policymakers' information sets. Second, studying the relationship between narrative-conditioned transmission and macroeconomic and policy regimes, such as periods of elevated inflation, episodes of financial stress, or changes in the policy framework, could clarify whether narrative effects are regime-invariant or whether narratives interact with broader state dependence in the transmission of monetary policy. Third, while our analysis relies on FOMC transcripts that are made public with a five-year lag, the same approach could be applied to faster-release communications (policy statements, minutes, and press conferences) to assess whether narrative information can be used for real-time interpretation and forecasting. Finally, applying the framework to other central banks, historical periods, or policy instruments would shed light on the external validity of narrative-based heterogeneity in monetary transmission.

References

Andre, Peter, Ingar Haaland, Christopher Roth, and Johannes Wohlfart, "Narratives about the Macroeconomy," *CEPR Discussion Papers*, 2022.

Arold, Benjamin W., Elliott Ash, W. Bentley MacLeod, and Suresh Naidu, "Worker Rights in Collective Bargaining," *NBER Working Paper Series*, March 2025, *33605*.

Aruoba, S. Borağan and Thomas Drechsel, "Identifying Monetary Policy Shocks: A Natural Language Approach," *American Economic Journal: Macroeconomics*, 2025, *forthcoming*.

Ascari, Guido and Timo Haber, "Non-Linearities, State-Dependent Prices and the Transmission Mechanism of Monetary Policy," *The Economic Journal*, January 2022, *132* (641), 37–57.

Ash, Elliott and Stephen Hansen, "Text Algorithms in Economics," *Annual Review of Economics*, September 2023, *15* (Volume 15, 2023), 659–688.

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

Auclert, Adrien, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, September 2021, *89* (5), 2375–2408.

Bauer, Michael D. and Eric T. Swanson, "A Reassessment of Monetary Policy Surprises and High-Frequency Identification," *NBER Macroeconomics Annual*, May 2023, *37*, 87–155.

Blinder, Alan S., "Central Banking in Theory and Practice: Lecture I: Targets, Instruments, and Stabilization," *Marshall Lecture, Presented at University of Cambridge, Cambridge, England*, May 1995.

Cieslak, Anna and Andreas Schrimpf, "Non-monetary news in central bank communication," *Journal of International Economics*, 2019, *118*, 293–315.

de Groot, Oliver and Alexander Haas, "The Signalling Channel of Negative Interest Rates," *Journal of Monetary Economics*, September 2023, *138*, 87–103.

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

Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl, *Cluster Analysis*, 5 ed., Wiley, 2011.

Federal Reserve Bank of St. Louis, "Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL)," FRED, Federal Reserve Bank of St. Louis. Source: U.S. Bureau of Labor Statistics. Accessed: 2025-12-15.

_ , "Industrial Production Index (INDPRO)," FRED, Federal Reserve Bank of St. Louis. Source: Board of Governors of the Federal Reserve System (US). Accessed: 2025-12-15.

_ , "Real Gross Domestic Product (GDPC1)," FRED, Federal Reserve Bank of St. Louis. Source: U.S. Bureau of Economic Analysis.

Friedman, Milton and Anna Jacobson Schwartz, *A Monetary History of the United States, 1867-1960*, Princeton University Press, 1963.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as Data," *Journal of Economic Literature*, September 2019, *57* (3), 535–574.

Girardi, Giulio, Kathleen W. Hanley, Stanislava Nikolova, Loriana Pelizzon, and Mila Getmansky Sherman, "Portfolio Similarity and Asset Liquidation in the Insurance Industry," *Journal of Financial Economics*, October 2021, *142* (1), 69–96.

Goetzmann, William N., Dasol Kim, and Robert J. Shiller, "Crash Narratives," *NBER Working Paper Series*, July 2022, *30195.*

Haaland, Ingar, Christopher Roth, Stefanie Stantcheva, and Johannes Wohlfart, "Measuring What Is Top of Mind," *NBER Working Paper*, 2024.

Hansen, Stephen and Michael McMahon, "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication," *Journal of International Economics*, 2016, *99* (S1), S114–S133.

_ , _ , **and Andrea Prat**, "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach," *The Quarterly Journal of Economics*, 2018, *133* (2), 801–870.

_ , _ , **and Carlos Velasco**, "How Experts Decide: Preferences or Private Assessments on a Monetary Policy Committee?," *Journal of Monetary Economics*, October 2014.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from U.S. States," *The Quarterly Journal of Economics*, August 2022, *137* (3), 1299–1344.

Hoberg, Gerard and Gordon Phillips, "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis," *The Review of Financial Studies*, October 2010, *23* (10), 3773–3811.

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

Jorda, Oscar and Alan M. Taylor, "Local Projections," *Journal of Economic Literature*, March 2025, *63* (1), 59–110.

Jr., Henry W. Chappell, Thomas M. Havrilesky, and Rob Roy McGregor, "Monetary Policy Preferences of Individual FOMC Members: A Content Analysis of the Memoranda Of Discussion," *The Review of Economics and Statistics*, August 1997, *79* (3), 454–460.

Kaufman, Leonard and Peter J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley, 1990.

Kuttner, Kenneth N, "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market," *Journal of monetary economics*, 2001, *47* (3), 523–544.

Macaulay, Alistair and Wenting Song, "Narrative-Driven Fluctuations in Sentiment: Evidence Linking Traditional and Social Media," *University of Oxford Department of Economics Discussion Paper Series*, 2022, *973*.

Malmendier, Ulrike, Stefan Nagel, and Zhen Yan, "The Making of Hawks and Doves," *Journal of Monetary Economics*, January 2021, *117*, 19–42.

Melosi, Leonardo, "Signalling Effects of Monetary Policy," *Review of Economic Studies*, 2017, *84* (2), 853–884.

Miranda-Agrippino, Silvia and Giovanni Ricco, "The Transmission of Monetary Policy Shocks," *American Economic Journal: Macroeconomics*, July 2021, *13* (3), 74–107.

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

National Bureau of Economic Research, "U.S. Business Cycle Expansions and Contractions," https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions 2025. Accessed: 2025-12-12.

Negro, Marco Del, Stefano Eusepi, Marc Giannoni, Argia Sbordone, Andrea Tambalotti, Matthew Cocci, Raiden Hasegawa, and M Henry Linder, *The FRBNY DSGE Model*, Federal Reserve Bank of New York, 2013.

Ng, Andrew Y., Michael I. Jordan, and Yair Weiss, "On Spectral Clustering: Analysis and an Algorithm," in "Advances in Neural Information Processing Systems" 2002.

Ramey, Valerie, "Macroeconomic Shocks and Their Propagation," in "Handbook of Macroeconomics," Vol. 2, Elsevier, January 2016, pp. 71–162.

Romer, Christina D. and David H. Romer, "Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, January 1989, *4*, 121–170.

_ **and** _ , "A New Measure of Monetary Shocks: Derivation and Implications," *American Economic Review*, September 2004, *94* (4), 1055–1084.

_ **and** _ , "Presidential Address: Does Monetary Policy Matter? The Narrative Approach after 35 Years," *American Economic Review*, June 2023, *113* (6), 1395–1423.

Rousseeuw, Peter J, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, 1987, *20*, 53–65.

Shapiro, Adam Hale and Daniel J Wilson, "Taking the Fed at Its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis," *The Review of Economic Studies*, October 2022, *89* (5), 2768–2805.

Shiller, Robert J, "Narrative Economics," *American Economic Review*, April 2017, *107* (4), 967–1004.

Swanson, Eric T., "Measuring the Effects of Federal Reserve Forward Guidance and Asset Purchases on Financial Markets," *Journal of Monetary Economics*, March 2021, *118*, 32–53.

Tenreyro, Silvana and Gregory Thwaites, "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions," *American Economic Journal: Macroeconomics*, October 2016, *8* (4), 43–74.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the royal statistical society: series b (statistical methodology)*, 2001, *63* (2), 411–423.

_ , _ , and _ , "Estimating the Number of Clusters in a Data Set via the Gap Statistic," *Journal of the Royal Statistical Society: Series B*, 2001, *63* (2), 411–423.

U.S. Bureau of Economic Analysis, "Interactive Data Application (iTable): National Income and Product Accounts," BEA.

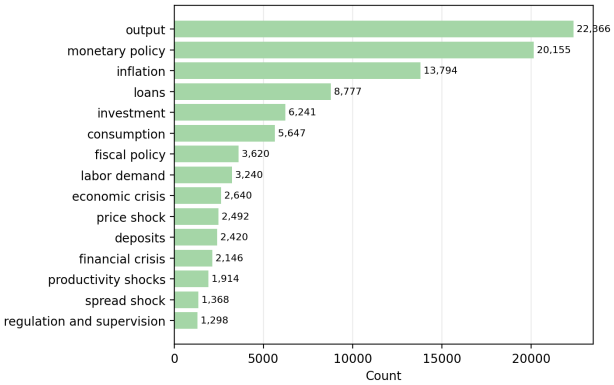
von Luxburg, Ulrike, "A Tutorial on Spectral Clustering," *Statistics and Computing*, 2007, *17* (4), 395–416.

Ward, Joe H., "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, 1963, *58* (301), 236–244.

Wieland, Johannes, "Updated Romer-Romer Monetary Policy Shocks," openICPSR 2021. Includes the Romer–Romer Greenbook forecast dataset (RRimport) and cross-check files. Accessed 2025-12-11.

Wieland, Johannes F. and Mu-Jeung Yang, "Financial Dampening," *Journal of Money, Credit and Banking*, 2020, *52* (1), 79–113.

Figure 9: Most frequent entities (source or target).



A. Data

Table 2 shows the summary statistics for the causal relationships extracted from the FOMC transcripts using an LLM. We drop 0.93% of relationships because they contain entities beyond our prespecified list of entities (see Section 2), resulting in an average of 102 relationships per document. Figure 9 shows the 15 entities that occur most of the relationship data set.

Table 2: Summary statistics of extracted causal relationships (1967–2011)

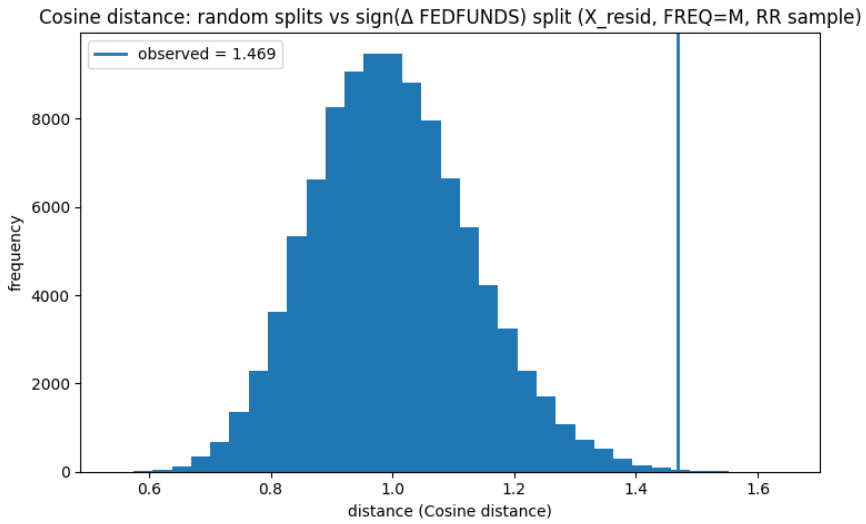
Statistic	Value
Number of extracted relationships	55,443
Dropped (non-whitelisted endpoints)	513
Share dropped	0.93%
Number of extracted relationships	54,930
Number of documents with ≥ 1 relationships	445
Relationships per document (mean)	123.44
Relationships per document (median)	102
Relationships per document (p10)	20
Relationships per document (p90)	249
Unique entities	35
Unique directed pairs	753

Notes: Relationships are extracted from FOMC transcripts and conference calls. The analysis sample retains only relationships whose source and target entities lie in the prespecified ontology (Actors/Variables/Shocks/Frictions). "Documents" refers to individual meeting/call records. Percentiles are computed over the distribution of relationships per document.

B. Further validation exercises

Figure 10 plots the histogram of cosine distances between the graphs of random sample divisions as in Figure 4 using residualized graphs, with a solid line for a split based on the sign of the interest rate change that month. For comparability with Figure 4, we restrict the sample for this histogram to all periods where we also have data in the shock series Romer and Romer (2004) (extended by Wieland (2021)). Figure 11 plots the cosine distance of random divisions of the observations as well as dividing the observations according to positive and negative shocks in Romer and Romer (2004) (extended by Wieland (2021)) as in Figure 4, but using the raw, non-residualized graphs.

Figure 10: Cosine distance between periods with positive and negative interest rate changes, compared with the distribution of cosine distances between randomly drawn graphs.

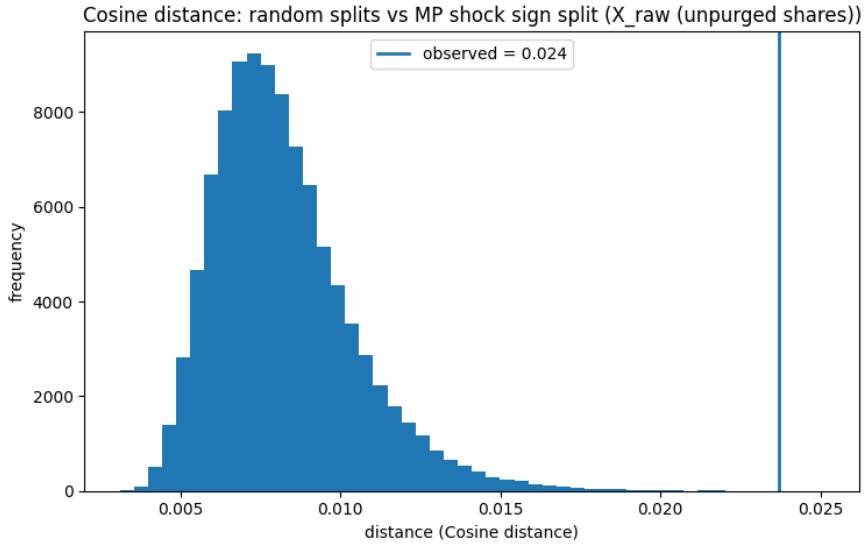


Note: Figure plots (in blue) the histogram of cosine distances between the average graphs from random splits of the months between 1969-2007.The blue line is the cosine distance between the average graphs computed from all periods with positive and negative changes in the Federal Funds Rate. For comparability, we restrict the sample to periods in which the Romer and Romer (2004) shock series (extended by Wieland (2021)) is available.

C. Dispersion and clustering

This section provides further details on our empirical method. Section C.1 describes the calculation of the between vs. within-group dispersion. Section C.2 provides details about how we choose the clustering method and number of clusters K.

Figure 11: Cosine distance between contractionary and expansionary shock periods in the Wieland and Yang (2020) shock series, without residualizing edge weights with respect to macroeconomic conditions, compared with the distribution of cosine distances between randomly drawn graphs.



Note: Figure plots (in blue) the histogram of cosine distances between the graphs from random splits of the months between 1969-2007. The blue line is the cosine distance between the average graphs computed from all periods with positive (contractionary) and negative (expansionary) shocks in the Wieland and Yang (2020) extension of the Romer and Romer (2004) monetary policy shock series. Unlike Figure 4, this figure is constructed using graphs that have not been residualized with respect to contemporaneous macroeconomic conditions as described in Section 2.3.

C.1. Further details for Section 3.2

Between vs within-group dispersion details. Let $\{x_t\}_{t=1}^{T} \subset \mathbb{R}^{E}_{\geq 0}$ be monthly graph vectors (typically $\sum_e x_{t,e} = 1$). Given a partition into groups $g \in \mathcal{G}$ with index sets I_g, define the centroid

$$c_g \;=\; \frac{1}{|I_g|} \sum_{t \in I_g} x_t \quad \text{(optionally renormalized to } \sum_e c_{g,e} = 1\text{)}.$$

Using the cosine distance $D(u, v)$ defined in equation (3), the within-group dispersion is

$$W \;=\; \sum_{g \in \mathcal{G}} \frac{|I_g|}{T} \frac{1}{|I_g|} \sum_{t \in I_g} D(x_t, c_g),$$

and the between-group separation (for two groups g, h) is

$$B \;=\; D(c_g, c_h).$$

The within-group dispersion intuitively describes how far away observations per group are from their average graph, the between-group separation characterizes how far apart the groups are from another.

C.2. Further details on the clustering method

This section describes how we choose between different clustering methods and decide on the number of clusters K. After summarizing how each clustering method works, we proceed in two steps. First, for each method, we decide on the optimal number of clusters. Second, we compare the performance of the different algorithms using the the silhouette score, as well as stability properties. Spectral clustering with $K = 3$ is our preferred method, but we show in Appendix D that our results are robust to the choice of K and clustering method.

C.2.1. Clustering methods

Given our non-Euclidean cosine distance measure, we consider three standard clustering algorithms (using the cosine distance measure defined in Section 2): k-medoids, spectral clustering and agglomerative spectral clustering. Note that we cannot use k-means since the cosine distance is non-Euclidean.

- **k-medoids:** k-medoids selects K representative observations ("medoids") from the sample and assigns each observation to the cluster of its nearest medoid (under the chosen distance measure). The medoids are chosen to minimize the total within-cluster dissimilarity, i.e. the sum of distances from observations to their assigned medoid (Kaufman and Rousseeuw, 1990).

- **Agglomerative hierarchical clustering:** Agglomerative clustering starts with each observation in its own cluster and repeatedly merges the two closest clusters under a specified linkage criterion until K clusters remain (Ward, 1963; Everitt et al., 2011). We use average linkage, which operates directly on the precomputed distance matrix. Given a fixed distance matrix and linkage rule, the procedure is deterministic.

- **Spectral clustering:** Spectral clustering transforms the distance matrix into an affinity (similarity) matrix (optionally sparsified using a k-nearest-neighbors graph). It then computes a low-dimensional embedding from the leading eigenvectors of a graph

Laplacian constructed from this affinity matrix, and finally applies a standard algorithm (here: k-means) to the embedded points (Ng et al., 2002; von Luxburg, 2007). This approach is particularly useful when cluster boundaries are non-convex in the original feature space.

C.2.2. Choosing between different K

We select the number of clusters using the gap statistic of Tibshirani et al. (2001b). For a given clustering method and number of clusters K, let $D(\cdot, \cdot)$ denote the cosine distance. Given a partition into clusters with index sets $I_1, \ldots, I_K$, we measure within-cluster dispersion by the average pairwise distance within clusters,

$$W_K \;=\; \sum_{g=1}^{K} \frac{1}{2|I_g|} \sum_{i \in I_g} \sum_{j \in I_g} D(x_i, x_j).$$

For k-medoids, we instead use its objective value, i.e. the sum of distances from each observation to its assigned medoid.

We generate B reference datasets by independently permuting each edge weight across periods, which preserves each edge's marginal distribution but destroys temporal structure, and recompute W_K^{ref} for each draw. The gap statistic is

$$\text{gap}(K) \;=\; \mathbb{E}_{\text{ref}}\big[\log W_K^{\text{ref}}\big] - \log W_K,$$

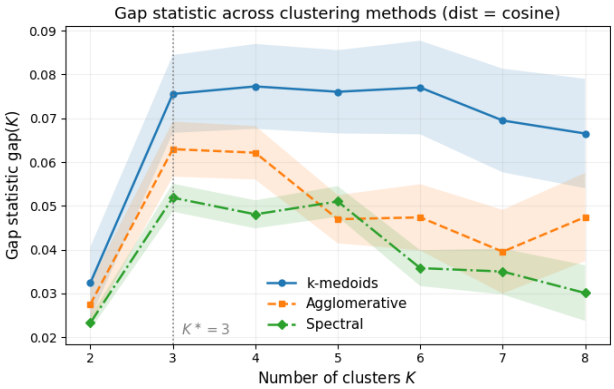
with standard error

$$s_K \;=\; \sqrt{1 + \frac{1}{B}}\, \text{sd}\big(\log W_K^{\text{ref}}\big).$$

We choose K using the Tibshirani rule, i.e. the smallest K such that $\text{gap}(K) \geq \text{gap}(K+1) - s_{K+1}$.

Intuitively, the gap statistic compares the within-cluster dispersion in the observed data to that in reference datasets that have the same marginal distribution of edge weights but no cluster structure. Larger values of $\text{gap}(K)$ indicate stronger evidence of clustering at K. The Tibshirani rule selects the smallest K such that moving from K to $K+1$ does not yield a meaningful improvement beyond sampling variation in the reference draws, i.e. $\text{gap}(K) \geq \text{gap}(K+1) - s_{K+1}$. Figure 12 summarizes the results. Across all three clustering

Figure 12: Gap statistic across $K \in \{2, \ldots, 8\}$.



Note: The figure plots the gap statistic of Tibshirani et al. (2001b) for each clustering method and number of clusters K, computed using the cosine distance. Shaded areas indicate $\pm s_K$, the standard error of the gap statistic based on the reference distributions. Gap values are comparable within a given clustering method across different K, but not directly comparable across methods, since each method induces a different notion of within-cluster dispersion. The dashed vertical line marks the selected number of clusters, $K^* = 3$, according to the Tibshirani rule.

methods, the gap statistic increases sharply from $K = 2$ to $K = 3$ and then flattens (or declines) for larger K, so we set $K = 3$.

C.2.3. Choosing different clustering methods

To choose between clustering methods at the common $K = 3$ suggested by the gap statistic, we compare k-medoids, agglomerative, and spectral clustering using the average silhouette score across 30 independent initializations.[17] In addition, we quantify the stability across runs using the mean and median adjusted Rand index (ARI)[18] . The results are summarized in Table 3. Since all three methods show good stability properties, we use spectral clustering as our primary specification, and consider k-medoids and agglomerative clustering as robustness checks.

The Heatmap 13 confirms that spectral clustering leads to well-separated clusters.

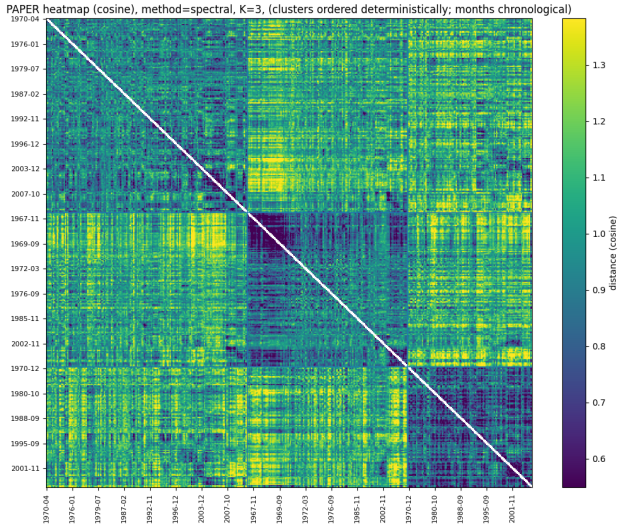
[17]For agglomerative clustering the runs are identical by construction and the stability metrics are perfect, as this procedure is deterministic. The stability metrics from repeating over different initializations are however informative for spectral clustering and k-medoids.

[18]The adjusted Rand index (ARI) measures the agreement between two clusterings based on pairwise co-membership of observations. The ARI is corrected for chance, taking the value 1 for identical partitions, values near 0 for similarity no better than random, and negative values for agreement worse than random.

Table 3: Comparison of clustering methods at $K = 3$ using the cosine distance matrix. $s_{\max}$ denotes the best silhouette score over all seeds, $\bar{s}$ the mean silhouette across 30 runs, and ARI_{mean} / $\text{ARI}_{\text{median}}$ the mean/median adjusted Rand index across all pairs of runs for a given method.

Method	K	$s_{\max}$	$\bar{s}$	sd(s)	ARI_{mean}	$\text{ARI}_{\text{median}}$
Spectral	3	0.1414	0.1414	0.0000	0.9963	1.000
k–medoids	3	0.1350	0.1315	0.0012	0.9247	1.000
Agglomerative	3	0.1176	0.1176	0.0000	1.0000	1.000

Figure 13: Heatmap visualizing spectral clustering for $K = 3$.



Note: Months are ordered by cluster, and chronologically within clusters. Blue corresponds to lower cosine distances across graphs, yellow higher distances. The darker squares along the diagonal indicate that spectral clustering finds meaningful clusters whose observations are closer to another than to the rest of the dataset.

D. Robustness Checks

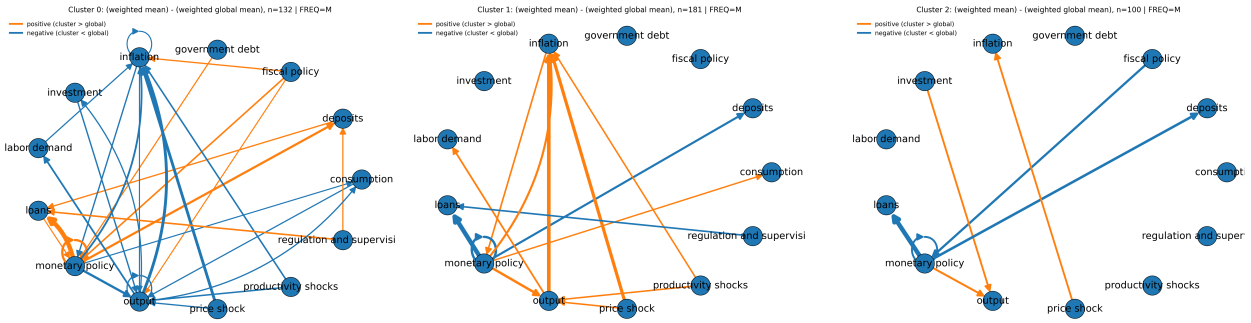
This section shows that our results are robust to (*i*) the choice of distance measure (Section D.1), (*ii*) the clustering method (Section D.2), (*iii*) the number of clusters K (Section D.3) and (*iv*) whether we use quarterly or monthly data (Section D.4). Section D.5 shows that our clusters do not simply reflect existing results on state-dependent monetary transmission by testing correlations between existing transmission regime classifications and our clusters.

D.1. Results with different distance measures

Figure 14 shows the representative graphs for spectral clustering using L1 (Manhattan distance) as the distance measure, and Figures 15 and 16 show the corresponding local projection results; Figures 17-19 show the results using L2 (Euclidean distance) as the distance

measure. The extracted clusters are qualitatively very similar to the ones using the cosine distance in the main text.

Figure 14: Clusters using the L1 distance.



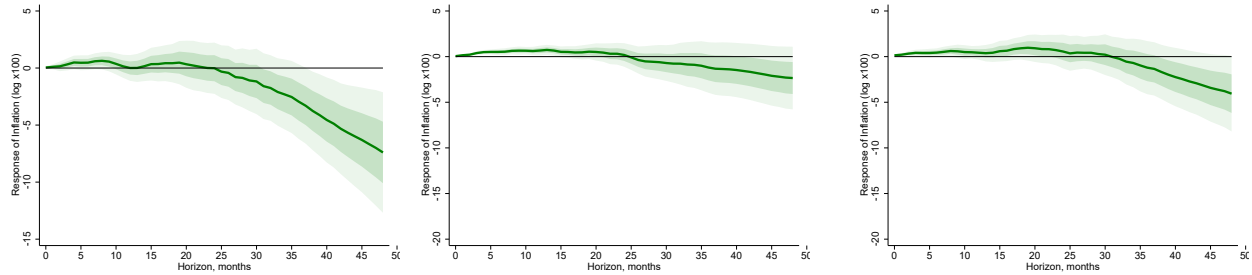
(a) Cluster 1: the "finance narrative"

(b) Cluster 2: the "textbook narrative"

(c) Cluster 3: the "inflation narrative"

Note: Each panel plots the difference in edge weights between graphs extracted from periods within the cluster described in the panel label and the average graph extracted from all transcripts in the dataset. The clustering method (spectral clustering) uses the distance measure L1.

Figure 15: IRFs for CPI inflation - L1 distance.



(a) Cluster 0: the finance narrative

(b) Cluster 1: the textbook narrative

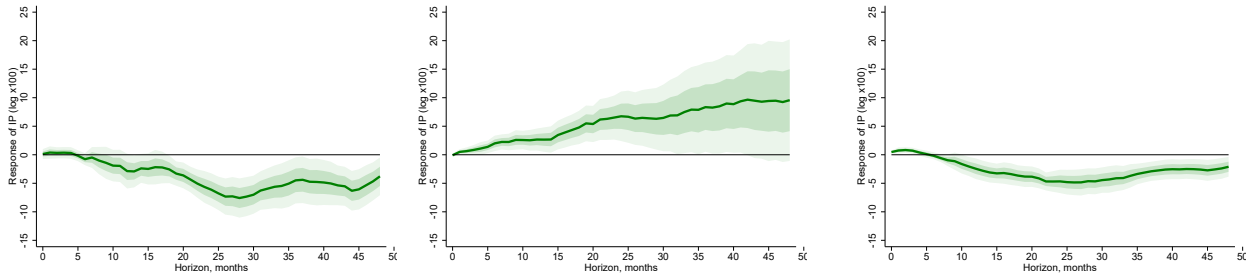
(c) Cluster 2: the inflation narrative

Note: Each panel plots the impulse response of CPI inflation to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 14, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

D.2. Results with different clustering methods

This section illustrates that the extracted clusters using spectral clustering are robust to the choice of clustering method. Figure 20 shows the results for agglomerative hierarchical clustering; Figure 23 for k-medoid clustering, both using the cosine distance and $K = 3$. The representative graphs for each cluster are again very similar to the ones extracted using spectral clustering.

Figure 16: IRFs for Industrial Production - L1 distance.



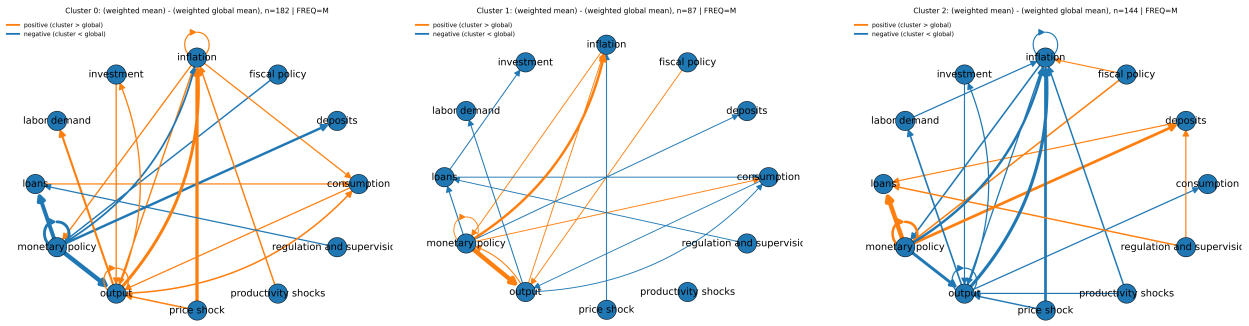
(a) Cluster 0: the finance narrative

(b) Cluster 1: the textbook narrative

(c) Cluster 2: the inflation narrative

Note: Each panel plots the impulse response of industrial production to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 14, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

Figure 17: Clusters using L2 distance.



(a) Cluster 1: the "inflation narrative"

(b) Cluster 2: the "textbook narrative"

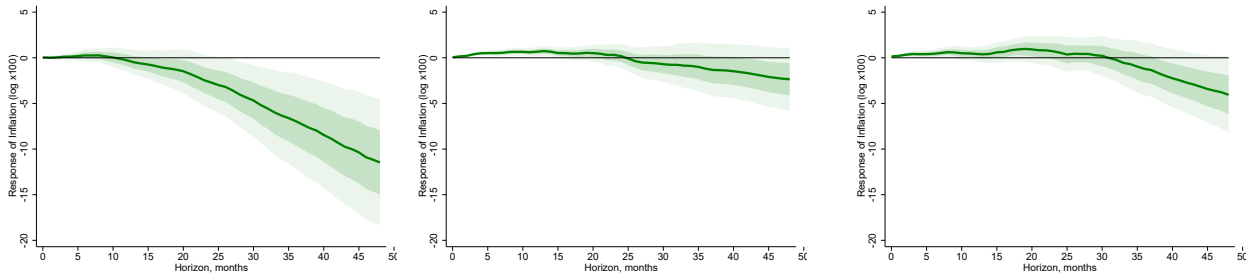
(c) Cluster 3: the "finance narrative"

Note: Each panel plots the difference in edge weights between graphs extracted from periods within the cluster described in the panel label and the average graph extracted from all transcripts in the dataset. The clustering method (spectral clustering) uses the distance measure L2.

D.3. Results with different numbers of narrative clusters

This section demonstrates that the qualitative clusters for $K = 3$ are still visible with a higher number of cluster. We choose $K = 5$ since stability of spectral clustering substantially drops beyond $K = 5$ (i.e., different runs return different clusters). Figure 26 shows representative graphs for each cluster. Figures 27 and 28 show the impulse response functions for monetary policy shocks conditioned on the narrative of inflation and industrial production, respectively.

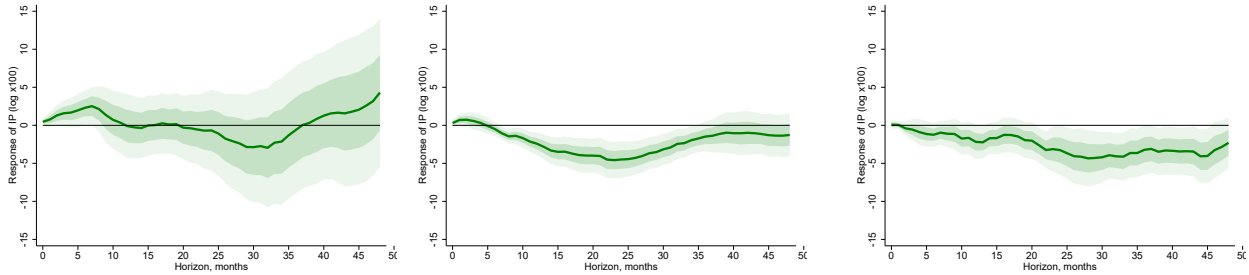
Figure 18: IRFs for CPI inflation using clusters of distance measure L2.



(a) Cluster 0: the inflation narrative **(b)** Cluster 1: the textbook narrative **(c)** Cluster 2: the finance narrative

Note: Each panel plots the impulse response of CPI inflation to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 17, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

Figure 19: IRFs for Industrial Production using clusters of distance measure L2.



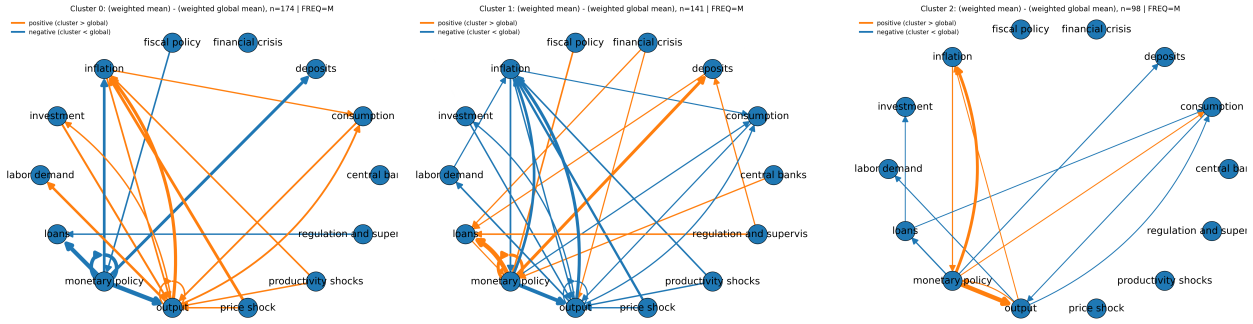
(a) Cluster 0: the inflation narrative **(b)** Cluster 1: the textbook narrative **(c)** Cluster 2: the finance narrative

Note: Each panel plots the impulse response of industrial production to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 17, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

D.4. Results with quarterly data

This section repeats the baseline exercises from Section 3, but using quarterly data. Figure 29 displays the average graph of Romer and Romer (2023) shock episodes; Figure 30 the cosine distance histogram of random splits in comparison to splitting the data according to Romer and Romer (2004) shocks; Figure 31 the narrative difference between positive and negative Romer and Romer (2004) shocks; Figure 32 illustrates the identified narratives from spectral clustering; and Figures 33 and 34 show impulse response functions for real GDP growth and inflation, respectively.

Figure 20: Clusters using agglomerative clustering.



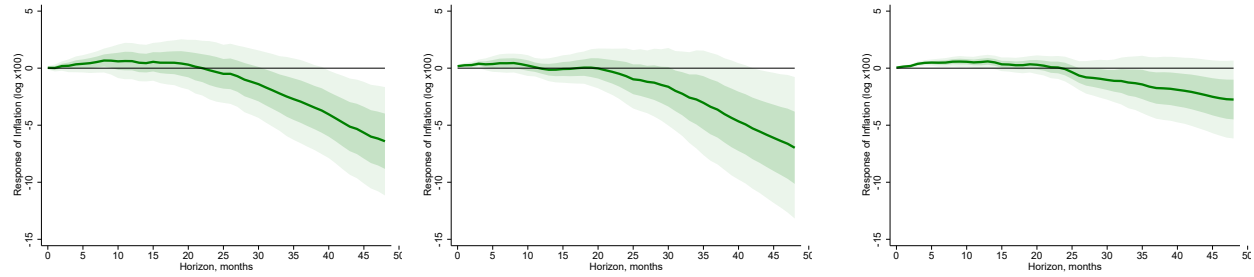
(a) Cluster 1: the "inflation narrative"

(b) Cluster 2: the "finance narrative"

(c) Cluster 3: the "textbook narrative"

Note: Each panel plots the difference in edge weights between graphs extracted from periods within the cluster described in the panel label and the average graph extracted from all transcripts in the dataset. The clustering method (agglomerative hierachical clustering, average linkage) uses the cosine distance measure.

Figure 21: IRFs for CPI inflation with clusters using agglomerative hierarchical clustering.



(a) Cluster 0: the inflation narrative

(b) Cluster 1: the finance narrative

(c) Cluster 2: the textbook narrative

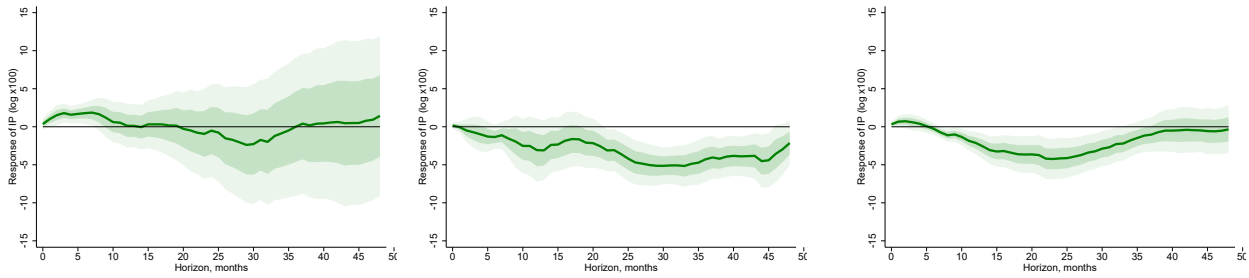
Note: Each panel plots the impulse response of CPI inflation to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 20, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

D.5. Overlap of clusters with other sources of transmission heterogeneity.

Tables 4 and 5 show the fraction of periods in each of the clusters identified in Section 3.2 that correspond to other ways in which monetary policy shocks have been split in the previous literature, including boom and recession periods, high-shock periods, and shock categorizations by Jarociński and Karadi (2020) and Swanson (2021).

In general, the clusters we identify do not map closely to any of these series. Cluster 1 (the finance narrative) contains a somewhat larger proportion of recession periods than the other clusters, but despite this it does not seem that this cluster maps closely to the business cycle overall: the majority of NBER recession months fall outside of Cluster 1, and the

Figure 22: IRFs for Industrial Production with clusters using agglomerative hierarchical clustering.



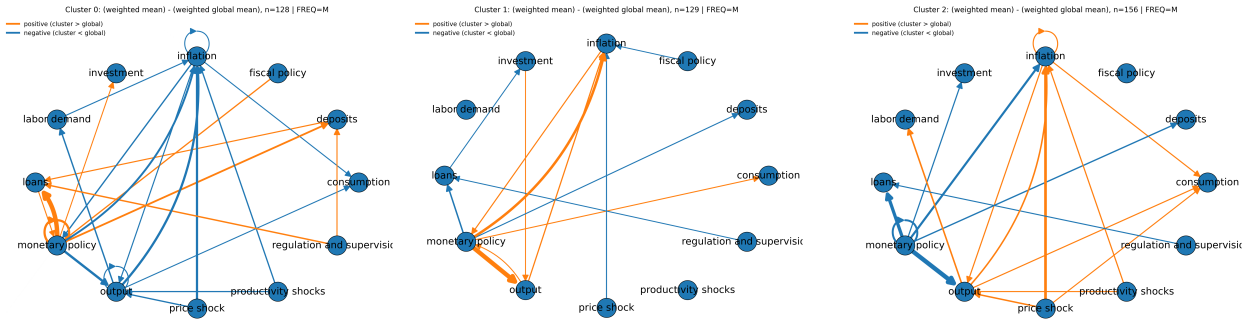
(a) Cluster 0: the inflation narrative

(b) Cluster 1: the finance narrative

(c) Cluster 2: the textbook narrative

Note: Each panel plots the impulse response of industrial production to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 20, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

Figure 23: Clusters - k-medoids clustering.



(a) Cluster 1: the "finance narrative"

(b) Cluster 2: the "textbook narrative"

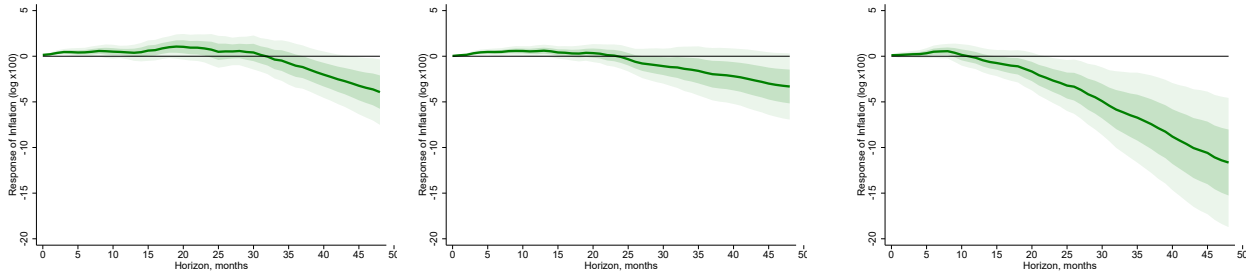
(c) Cluster 3: the "inflation narrative"

Note: Each panel plots the difference in edge weights between graphs extracted from periods within the cluster described in the panel label and the average graph extracted from all transcripts in the dataset. The clustering method (k-medoids) uses the cosine distance measure.

number of "boom" months is similar in all clusters. That cluster is also more concentrated in the early part of the sample, but still almost half of the observations in that cluster come after the change in document style in 1977.

Table 6 then reports the results of three linear probability models, where indicator variables for membership of each cluster are regressed on the monetary policy and central bank information shocks from Jarociński and Karadi (2020), and indicators for which of the Swanson (2021) factors are largest (omitted category: Federal Funds Rate factor). Although some coefficients are significantly different from 0 at conventional levels, the R^2 is very low in all columns, suggesting that existing shock series cannot explain the clusters we identify.

Figure 24: IRFs for CPI inflation with k-medoid clusters.



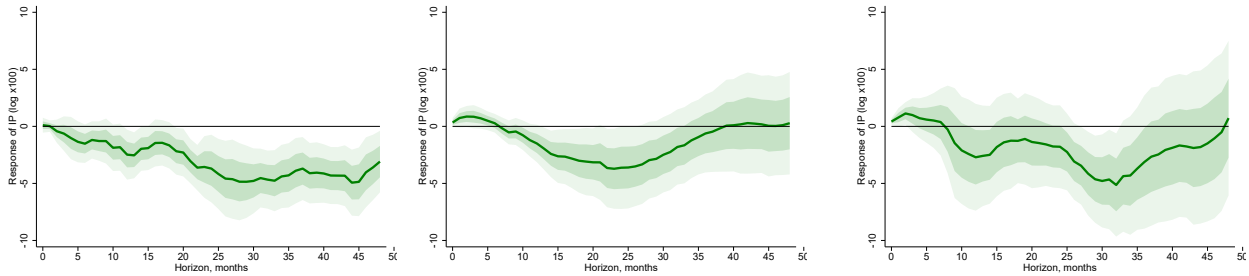
(a) Cluster 0: the finance narrative

(b) Cluster 1: the textbook narrative

(c) Cluster 2: the inflation narrative

Note: Each panel plots the impulse response of CPI inflation to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 23, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

Figure 25: IRFs for Industrial Production with k-medoid clusters.



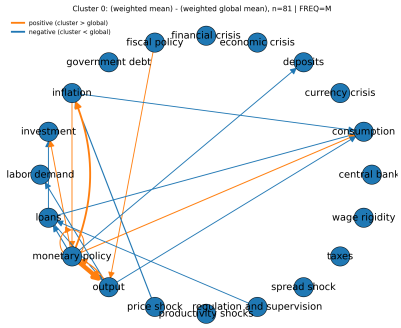
(a) Cluster 0: the finance narrative

(b) Cluster 1: the textbook narrative

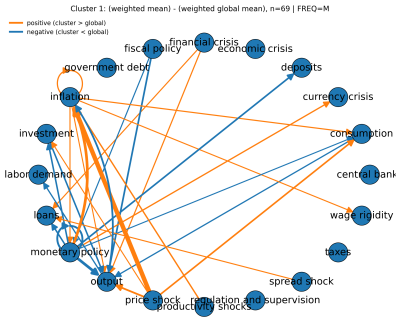
(c) Cluster 2: the inflation narrative

Note: Each panel plots the impulse response of industrial production to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 23, estimated using the local projection in equation (4). Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

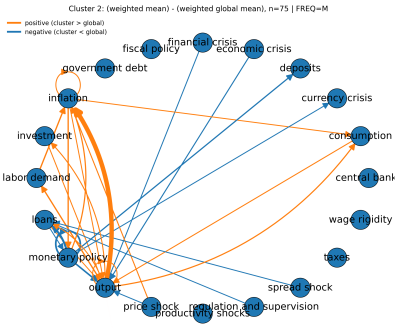
Figure 26: Narrative clusters obtained using spectral clustering with $K = 5$.



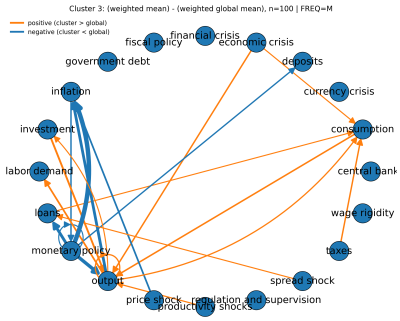
(a) Cluster 1: the "textbook narrative"



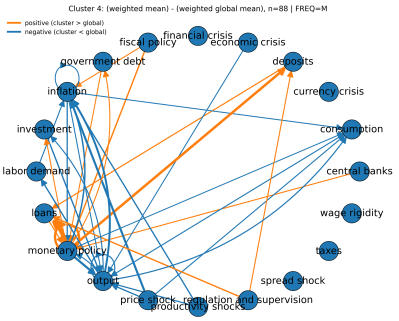
(b) Cluster 2: the "inflation narrative"



(c) Cluster 3: additional narrative 1



(d) Cluster 4: additional narrative 2



(e) Cluster 5: the "finance narrative"

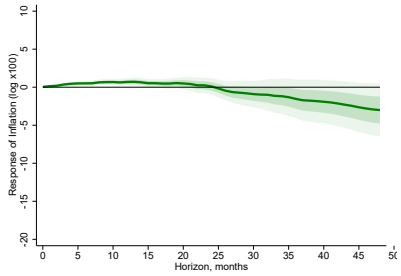
Table 4: Overlap between monthly clusters, Romer–Romer shocks, recessions, and booms

	Cluster 0	Cluster 1	Cluster 2
n	171	136	106
Fraction high |shock|	0.019	0.069	0.087
Fraction recession	0.164	0.221	0.132
Fraction boom	0.281	0.243	0.226
Fraction pre-1977	0.146	0.537	0.151
Avg relationships	120.32	163.97	113.62
Avg RR shock	0.033	-0.023	-0.020

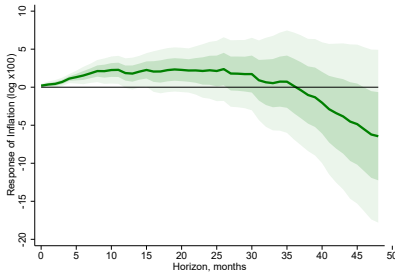
Notes: "High |shock|" indicates months flagged with high absolute Romer and Romer (2004) shocks (2 std from mean). "Boom" is defined as months in which log real GDP growth is in the top quartile of the observation periods in our sample.[a] "Recession" is defined using the NBER business-cycle chronology (National Bureau of Economic Research, 2025). Fraction pre-1977 reports the share of months dated 1976 or earlier. Average relationships is the mean number of extracted causal relationships per month. Avg RR shock is the mean Romer–Romer monetary policy shock. For comparison, the standard deviation of the Romer–Romer shock is 0.234.

[a]Real GDP is from the U.S. Bureau of Economic Analysis (NIPA) (U.S. Bureau of Economic Analysis, n.d.) as provided by FRED (series GDPC1) (Federal Reserve Bank of St. Louis, n.d.c).

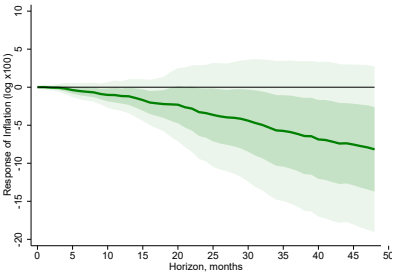
Figure 27: Local projections for CPI inflation using clusters for $K = 5$.



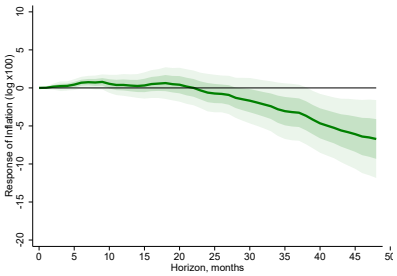
(a) Cluster 0: the textbook narrative



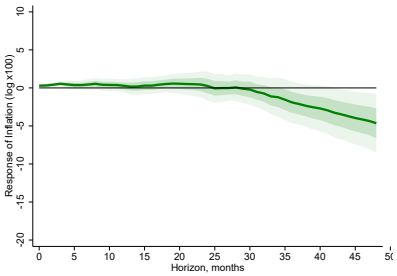
(b) Cluster 1: the inflation narrative



(c) Cluster 2: additional narrative 1

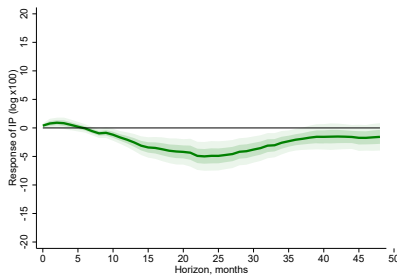


(d) Cluster 3: additional narrative 2

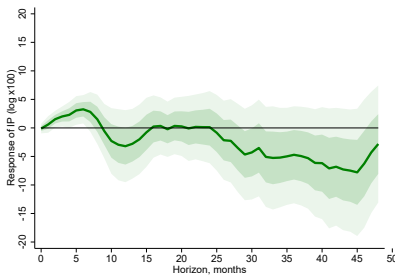


(e) Cluster 4: the finance narrative

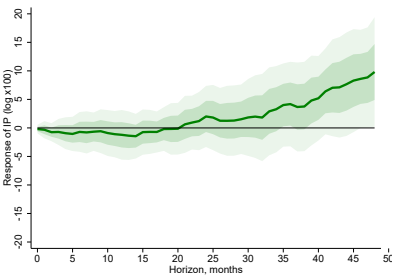
Figure 28: Local projections for industrial production using clusters for $K = 5$.



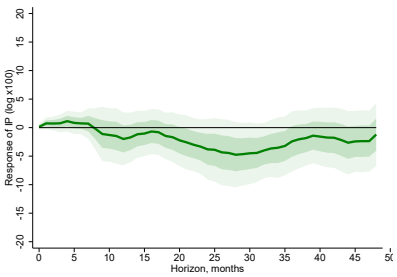
(a) Cluster 0: the textbook narrative



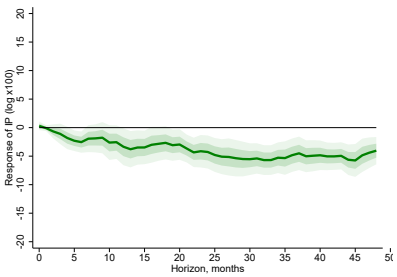
(b) Cluster 1: the inflation narrative



(c) Cluster 2: additional narrative 1



(d) Cluster 3: additional narrative 2



(e) Cluster 4: the finance narrative

Figure 29: Edge weights in causal graph extracted from the 6 periods identified as contractionary monetary policy shocks in Romer and Romer (2023). Panel A: total edge weights; Panel B: edge weights in comparison to the average graph over all months.

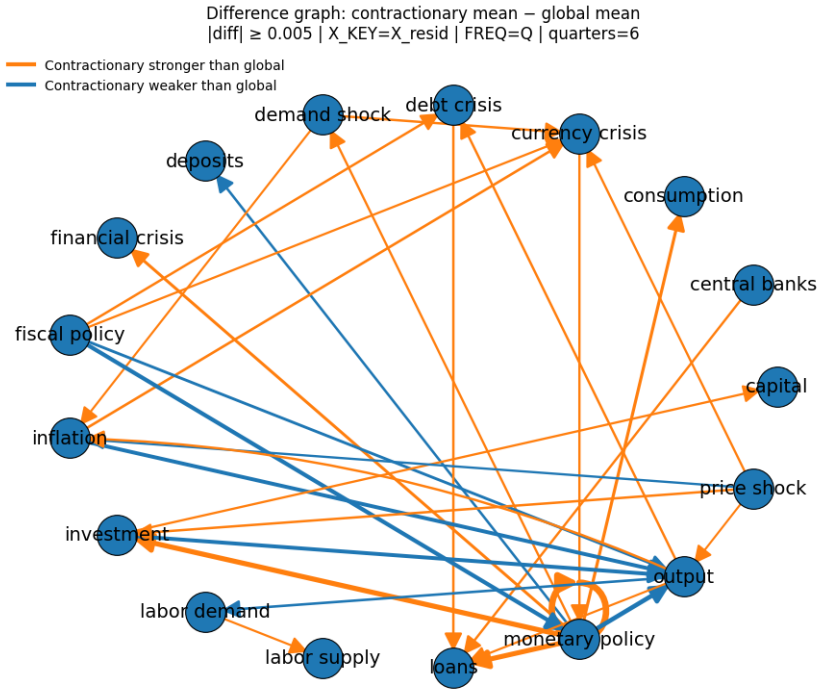
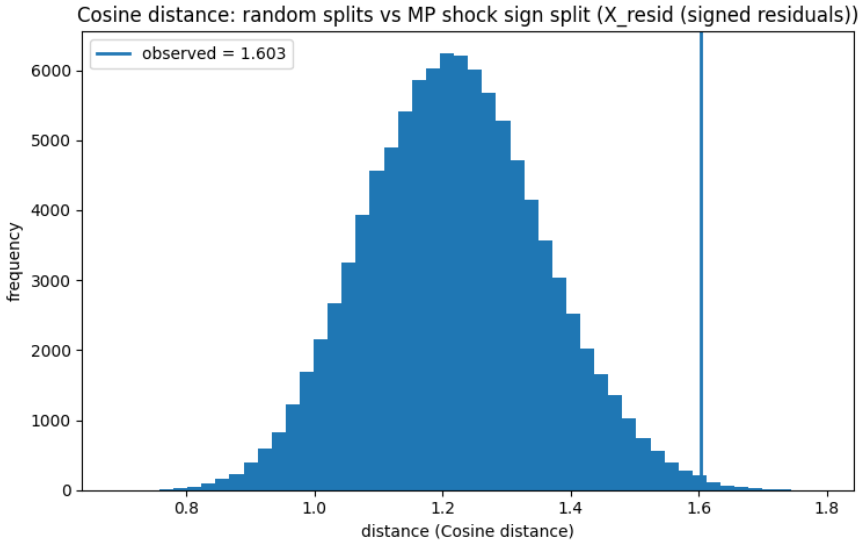
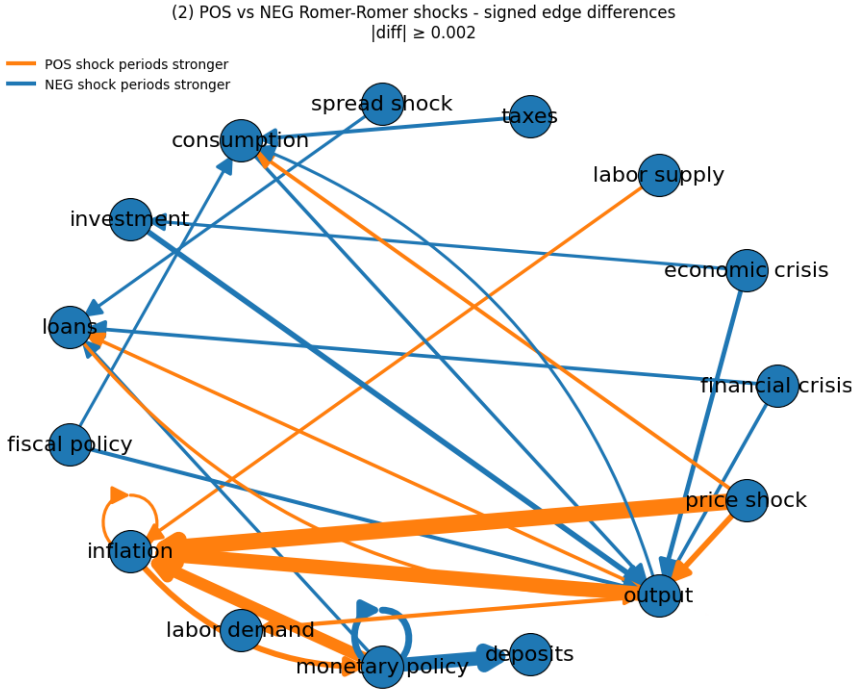


Figure 30: Cosine distance between contractionary and expansionary shock periods in the Wieland and Yang (2020) shock series, compared with the distribution of cosine distances between randomly drawn graphs.



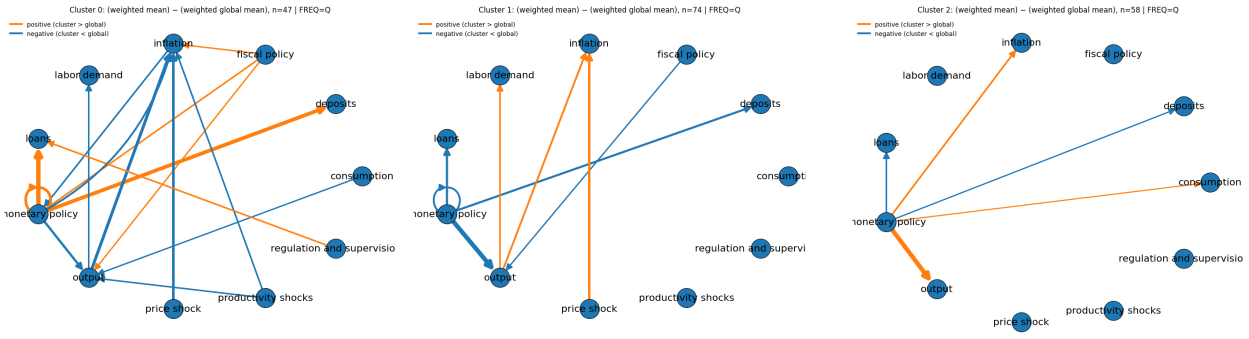
Note: Figure plots the histogram of cosine distances between the graphs from random splits of the transcripts 1969-2007 (blue bars). The solid blue line is the cosine distance between the average graphs computed from all periods with positive (contractionary) and negative (expansionary) shocks in the Wieland and Yang (2020) extension of the Romer and Romer (2004) monetary policy shock series.

Figure 31: Differences in edge weights between contractionary and expansionary shock period graphs.



Note: Figure plots the difference in edge weights between graphs extracted from periods with positive (contractionary) and negative (expansionary) monetary policy shocks in the Romer and Romer (2004) series extended by Wieland and Yang (2020). Orange arrows denote a link that is more frequently present in months with contractionary shocks, blue arrows denote a link more frequently present in months with expansionary shocks. Line width reflects the magnitude of the difference.

Figure 32: Clusters using spectral clustering.



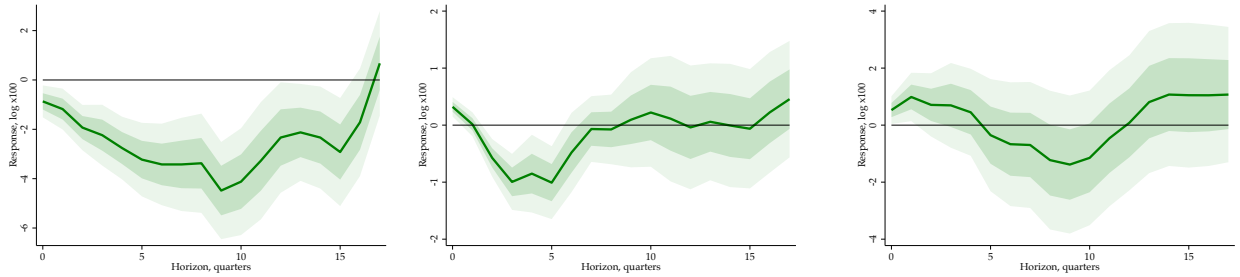
(a) Cluster 0: the finance narrative

(b) Cluster 1: the inflation narrative

(c) Cluster 2: the textbook narrative

Note: Each panel plots the difference in edge weights between graphs extracted from periods within the cluster described in the panel label and the average graph extracted from all transcripts in the dataset. Orange arrows denote a link that is more frequently present in months within the cluster, blue arrows denote a link less present in the cluster. Line width reflects the magnitude of the difference.

Figure 33: IRFs for real GDP, quarterly data, $K = 3$.



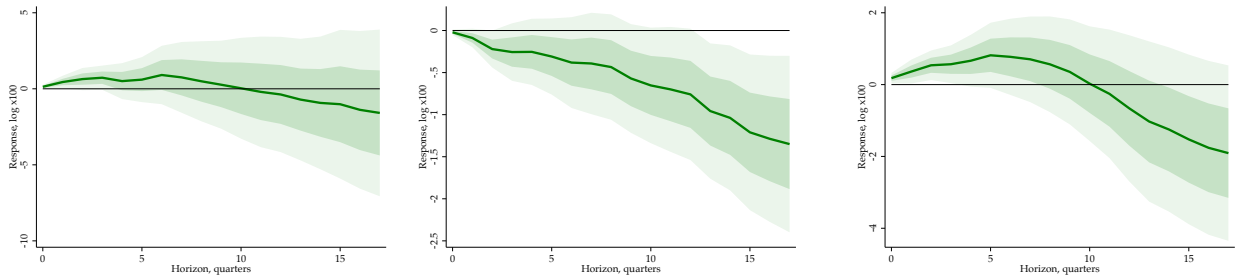
(a) Cluster 0: the finance narrative

(b) Cluster 1: the inflation narrative

(c) Cluster 2: the textbook narrative

Note: Each panel plots the impulse response of real GDP growth to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 32, estimated using the local projection in equation (4), with 4 lags in Z_{t-1} rather than the 12 in the monthly specification. Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

Figure 34: IRFs for Inflation, quarterly data, $K = 3$.



(a) Cluster 0: the finance narrative

(b) Cluster 1: the inflation narrative

(c) Cluster 2: the textbook narrative

Note: Each panel plots the impulse response of CPI inflation to a one percentage-point monetary policy shock from the series by Wieland and Yang (2020) that occurs alongside a narrative from a given cluster defined in Figure 32, estimated using the local projection in equation (4), with 4 lags in Z_{t-1} rather than the 12 in the monthly specification. Dark and light shaded areas present one and two standard deviation confidence intervals, constructed using heteroscedasticity robust standard errors. Sample: 1969-2007.

Table 5: Overlap between monthly clusters, "Poorman" shocks, and factor classifications

	Cluster 0	Cluster 1	Cluster 2
Panel A: Within cluster: share of months that are indicator months			
MP shock (Poorman)	0.44	0.50	0.59
CBI shock (Poorman)	0.27	0.28	0.19
FFR largest	0.33	0.32	0.27
FG largest	0.54	0.41	0.67
LSAP largest	0.17	0.29	0.08
Panel B: Within indicator: share of indicator months in each cluster			
MP shock (Poorman)	0.43	0.22	0.35
CBI shock (Poorman)	0.54	0.24	0.22
FFR largest	0.52	0.21	0.27
FG largest	0.48	0.15	0.37
LSAP largest	0.50	0.36	0.14

Notes: Unit of observation is a month. Poorman shock indicators equal one if the corresponding monthly shock series is nonzero in the "poor man's sign restrictions" exercise in Jarociński and Karadi (2020) (available from February 1990). Factor indicators equal one if, in at least one FOMC meeting during the month, the stated factor has the largest absolute magnitude among the {FFR, FG, LSAP} factors identified by Swanson (2021) (available from July 1991). Denominators vary by indicator due to data availability; months may have multiple indicators equal one.

Table 6: Explaining cluster assignment: baseline specification

	Cluster 0	Cluster 1	Cluster 2
MP shock (Poorman)	1.206	-1.321	0.115
	(0.410)	(0.605)	(0.641)
CBI shock (Poorman)	0.661	-0.867	0.207
	(1.904)	(1.440)	(1.097)
FG (largest)	-0.067	-0.035	0.102
	(0.088)	(0.068)	(0.083)
LSAP (largest)	-0.044	0.206	-0.161
	(0.123)	(0.112)	(0.090)
Constant	0.548	0.174	0.278
Observations	167	167	167
R^2	0.026	0.083	0.042
Adj. R^2	0.002	0.060	0.018

Notes: Each column reports a separate linear probability model where the dependent variable equals one if the month is assigned to the corresponding cluster. MP and CBI are monthly Poorman monetary policy shocks entering linearly. The dominant policy instrument is captured by categorical indicators for FG and LSAP, with FFR as the omitted category. Robust (HC1) standard errors are shown in parentheses.