Reading between Lines: Measuring Macroeconomic Narratives from Texts using Large Language Models

Chenyu (Sev) Hou* Simon Fraser University

Jiannan (Jay) Jiang† UT Austin

Tao Wang‡ Bank of Canada

March 14, 2025
(a preliminary draft)

Abstract

Macroeconomic texts like Fed's Beige Book contain rich descriptive information of real-time economic conditions based on first-hand surveys of diverse decision makers. This paper uses a large language model (LLM) to systematically transform such narrative assessments of the economy into a structuralized format that can be used for further economic analysis. We not only differentiate discussions regarding distinctive economic variables, such as quantity, price, supply, and demand embedded in context-specific phrases, but also measure their **directions of change** that can be compared to macroeconomic time series. We then identify the **causal statements** regarding such variables made by authors, e.g., an increase in consumer demand leads to an increase in the goods prices. We show that (1) economic phrases with directional information track the corresponding economic variables well. (2) The extracted supply and demand narratives line up with economic theory. Supply narratives focus on price dynamics, while demand narratives emphasize the consequences on real activity. (3) Focusing on the pandemic period, we find that input prices were the key driver of inflation, with demand pressures dominating right after lockdowns ended, followed by input prices becoming the main factor as core inflation exceeded the 2% target. The discussions about price change became more balanced after late 2022.

Keywords: Macroeconomic Expectations, Subjective Models, Narratives, LLM

JEL Codes: D84, E30, E71

*Chenyu Hou: Department of Economics, Simon Fraser University, chenyu_hou@sfu.ca.

†Jiannan (Jay) Jiang: Department of Economics, University of Texas at Austin, jiannanjiang@utexas.edu.

‡Tao Wang: Bank of Canada, TaoWang@bank-banque-canada.ca.

1 Introduction

Economic information exists beyond what is measured in statistics. Often, our understanding of economic reality is limited by the frequency and scope of the economic measures we track through statistical agencies and market data. Economic analysis also shrinks the hypothetically high-dimension state spaces of the economy into a selected number of dimensions. However, real-time economic conditions are frequently discussed in various forms of textual data, such as central bank documents, market analysis, and news media reports, all of which provide valuable qualitative insights into not only the actual economic conditions but also how agents assess and perceive the economy.

Take the U.S. Federal Reserve Banks' Beige Books, released by the 12 bank districts at least eight times a year since the 1970s, as one example. Such texts are written by Fed staff economists who conduct in-person or online surveys of economic contacts, such as firm managers, retailers, and bank lenders. These surveys gather insights on economic conditions in their respective districts and across different sectors. This type of information contains rich content that goes beyond what standard statistical measures can capture. Actually, the Federal Reserve Bank system openly acknowledges the value of such qualitative assessment based on the so called "anecdotal" evidence.[1]

The goal of this paper is to create an automatic procedure based on large language models to extract rich narrative information from macroeconomic texts such as Beige Books and turn them into structured data that can be used for further macroeconomic analysis. Our extracted narratives consist of various components.

- The detected economic variables, such as supply, demand, price, and quantity, which are also classified by various dimensions congenial to economic analysis, such as its sector and temporal flag (current versus future).
- The indicated direction of dynamics of each variable, such as significant increase, moderate increase, no change, to moderate and significant decreases.
- Determined causal link between the detected variables according to the narrative, e.g. an increase in energy prices $\rightarrow$ lower production where $\rightarrow$ indicates a causal relationship.

Achieving these three goals all together has traditionally been challenging, especially the second and the third. Standard natural language processing (NLP) tools such as topic modeling

[1] For instance, in the Beige Book' official webpage, it states that "...The qualitative nature of the Beige Book creates an opportunity to characterize dynamics and identify emerging trends in the economy that may not be readily apparent in the available economic data."

can identify economic topics using unsupervised learning algorithms based on bag-of-words representation of texts. However, these methods rarely assign clear directional meaning to changes in economic variables. They mostly rely on sentiment scores or frequency counts to approximate directional information about the economy. Furthermore, even when a newspaper article contains multiple economic topics, it isn't easy to extract how these topics, variables, and concepts are interconnected as perceived by the author.

The rise of large language models, such as OpenAI's ChatGPT-4 (GPT thereafter), has created exciting possibilities for addressing these challenges. We leverage transformer-based language models built in the GPT to systematically analyze textual data with a careful treatment of their dynamic and complex contexts. This approach allows us to not only identify economic topics and financial variables but also intuitively capture the direction of changes as discussed in the text. For each of these variables, we also determine their temporal flag, i.e., reflecting past realization, current realization or future expectations.

We use an excerpt from August 1992's Beige Books to illustrate our methodology. A paragraph like this reports the change in an array of variables in different sectors and draws some causal links between the variables under discussion.

> "*Despite ongoing challenges, consumer spending has remained robust in recent months. According to recent reports, employment levels have shown moderate growth, with particular strength in the manufacturing sector. However, rising energy prices and increased raw material costs have put pressure on production levels. The Federal Reserve's recent interest rate hikes have contributed to increased uncertainty in the financial markets. Meanwhile, export activity has seen a significant uptick, driven by demand from Asia.*"

When this chunk of text is fed into an LLM, in addition to our carefully constructed prompt with instructions and the pre-fixed structural format for the output, our procedure yields the information more or less in the following format.

- Economic variables and their directions: good sales ++, overall employment level ++, manufacturing employment level ++, raw material and commodity price ++, good production level –, monetary policy ++, trade (export) +++, good demand ++.
- Logic chains: (a). energy prices ++/raw material costs++ → production level - -; (b). demand ++ → trade (export) +++

We apply such a procedure to all 3536 Beige Books published between January 1990 and December 2023. We first validate that such an automatic procedure yields outputs in quality comparable to human labeling. For instance, we confirm that the diverse, dynamic, and context-specific phrases are indeed reasonably labeled as a certain economic variable. Because of the LLM's capability of "understanding" the contexts of language, the exact same phrases may be determined to represent different economic concepts depending on the exact discussion. Second, the LLM-determined dynamic scores of each variable, when aggregated into time series, are found to be highly correlated with actual macroeconomic data, i.e. the majority has a correlation coefficient of at least 40%.

With the identified micro causal links in local narratives, we are able to study the overall prevalence of distinctive macroeconomic narratives and their time-varying patterns. It is found that overall, the majority of the causal narratives are consistent with common and intuitive economic rationale, such as the law of supply and law of demand. Throughout the sample, we found there are more demand narratives than supply narratives. Supply narratives focus more on price changes rather than consequences in real activities, but demand narratives focus more on consequences in real activities.

Fixing an outcome variable, we can also examine how the perceived causal drivers of an economic outcome change over time. For instance, one of the heated macroeconomic debates during the pandemic has been the causes of the initial surge in inflation and their subsequent dynamics. According to our measured narratives, input prices were perceived to be the major contributor to the initial rise in inflation since 2021. The demand factor was indeed considered most important right after the pandemic lockdown, but input price soon dominated, right when core inflation started shooting above the 2% target. Throughout the discussion of the inflation surge, wage is another important factor that contributes to the price hikes. What contributed to wage dynamics, according to the narratives, were labor shortage and labor demand. We find that narrative inflation started to decline after September 2022, followed by the decline of discussions about wage increases shortly after. Starting in late 2022, descriptions of wage and inflation dynamics became more balanced.

Related Literature

Most closely related, Bybee (2023) uses the large-language model GPT 3.5 to transform newspaper articles in *The Wall Street Journal* into quantitative variable/horizon-specific macroeconomic forecasts. The paper found that the machine's macroeconomic expectations exhibit similar deviations from the full-information rational benchmark as survey forecasts. In terms of focus, our paper is less about "generating" future expectations and more about "measuring" described economic conditions. We do not rely on LLM to make forecasts directly. Instead, we extract structured information from the narrative assessment of the current economy, such as economic variables, directions, temporal tag, and their logic links, and further aggregate them into quantifiable information based on such structuralized information.

We also contribute to the study of economic narratives, especially those that involve leveraging textual analysis tools. Although various studies attempt to concretize and measure economic narratives, echoing the call for research on narrative economics by Shiller (2017), different studies vary greatly in their exact definition of economic narratives.[2] (Ash et al., 2021; Larsen and Thorsrud, 2019; Andre et al., 2022; Macaulay and Song, 2022; Flynn and Sastry, 2022; Yang et al., 2020). We operationalize on the definition of narratives as one that consists of economic variables, their described changes, and the mentioned causal relationships among them. Our definition of narratives is perhaps closest to that in Yang et al. (2020), which uses unsupervised learning and human feedback to extract economic variables causing inflation and output dynamics. Their procedures involve multi-step unsupervised learning, while our procedures directly deploy LLMs.

Our use of LLM also builds on other applications of natural language processing (NLP) tools to macroeconomic questions prior to LLMs.[3] For instance, Bybee et al. (2021) applies Latent Dirichlet allocation (LDA) topic models to historical news archives and explores their correlation with macroeconomic and financial variables. Angelico et al. (2022) measures inflation expectations using Tweets. In the recent central bank communication literature, Byrne et al. (2023b,a) classify information in monetary policymakers' speeches by its temporal dimension. Gáti and Handlan (2022) studies the Fed's communication rule via estimating the mapping between the Fed's numeric macroeconomic forecasts and language use of textual statements.[4]

Lastly, our work relates to the previous studies of the information content of Beige Book using different methods.(Zavodny and Ginther, 2005; Armesto et al., 2009) Aruoba and Drechsel (2024) uses sentiment scores associated with frequently discussed economic concepts in Beige Books as part of the information set of the Fed before its monetary policy decisions. Balke et al. (2017) turn Beige Books into a quantitative index of economic activities that are found to contain information beyond contemporaneous economic statistics at the date of release. Gascon and Martorana (2024) constructs a sentiment index based on Beige Book texts to help identify real-time economic recessions. Kliesen and Werner (2022); Soto (2023) use methods such as word embeddings and unsupervised sentiment classifiers to measure supply chain conditions

[2]There is also a separate literature that identifies subjective models using open-ended instruments Andre et al. (2022) and observational surveys. Kamdar and Walker (2024); Hou and Wang (2024).

[3]Surveying the proliferating literature on NLP applications in macroeconomic research goes beyond the scope of our paper. See Gentzkow et al. (2019) and Ash and Hansen (2023) for a comprehensive survey.

[4]Related to our focus on measuring directions from texts, one of the approaches they adopt is the following. First, assign topics, i.e. unemployment and inflation, to a particular sentence, and then calculate the fraction of "increase" and "decrease" in the nearby window of the relevant statement. They also experimented with the use of ChatGPT to generate numeric economic forecasts based on inputs from Fed statements.

based on Beige Book texts. Compared to these papers, our use of LLM allows us to go beyond simple quantitative measures based on word counts or dictionary-based sentiment and instead directly measure the directions of the dynamics and causal logic.

2 Methodology

2.1 A Conceptual Framework

We treat the state of the economy at any point of the time as an infinite-dimensioned object. The state space is rich because it contains distributional information relevant to different sectors, type of agents, those involving past and regarding future, the structural causes, and the exogenous events. While conventionally, such a high-dimension state is represented and analyzed along a much reduced number of dimensions, such as those corresponding to measurable economic variables. e.g. inflation, output, consumption, etc, much more information characterizing the state of the economy is left unmeasured and non-analyzed. We treat the textual data, i.e. articles summarizing economic conditions, as a high-dimensional representation of the state of the economy. Our approach, in its essence, is to extract from such a high-dimension representation various forms of structured information and summarize them along dimensions that may be or may not be orthogonal to aggregate time series.

In its concrete form, an article is composed of multiple sentences, each contributing to the overall structure and meaning of the text. Within each sentence, there are phrases–groups of words that work together to convey specific ideas or concepts. These phrases are connected through logical relationships such as cause and effect, contrast, or parallel, which help to establish a narrative to be delivered by the article. In light of this structure, we will first identify phrases that are related to economics in each sentence and classify them into different economic concepts such as inflation, employment, and monetary policy, etc. This step is similar to the standard topic modeling task. However, a phrase in the context of a sentence may contain more information than just its economic meaning. For example, the sentence "the consumers expect the grocery prices to increase further in the future" contains both directional information (price goes up) and information about time stance (in the future) about prices. With the help of LLM, we will also extract directional and time-stance information associated with each economic concept we identified in the text.

After extracting phrases about economic concepts from the texts, we will establish the logical relations between these phrases. The possible relations include causal, simple, and paralleled relations. A causal relation is when the texts suggest that one concept leads to another. Simple relation is extracted when several concepts occur together, but no apparent causal chains are present. For example, they might be caused by the same factor, or they might be expressions of the same phenomenon. Parallel relation happens when two concepts are jointly mentioned, but no clear relation can be inferred. Our focus in this project is causal relations. With both classified economic concepts and the relations identified between them, we will construct logic chains which can later be summarized as a macroeconomic narrative.

2.2 The choice of texts

Our baseline choice of macroeconomic texts is the Beige Book by the Federal Reserve Bank system of the United States. The Beige books are qualitative summaries of district economic conditions based on interviews with business contacts, economists, and market experts by teams from the 12 Federal Reserve Banks. Since 1970, Beige books have been published 8 times a year by the Federal Reserve branches, which yields 6,095 articles, and each article will contain multiple paragraphs regarding different aspects of the economy.

There are several advantages to such types of texts. They are written by professional economists with dense and structured language standard to economic commentaries. The texts reflect rich narrative evidence on different districts' overall and sector-specific economic conditions. Across time and space, the Beige Book maintains a relatively stable and consistent tone and writing style. They usually consist of short, self-contained sentences indicating directional changes in the assessed topic, with occasional causal/relational statements linking different descriptions. It is also worth noting that "contacts are not selected at random; rather, Banks strive to curate a diverse set of sources that can provide accurate and objective information about a broad range of economic activities." This means that they are not a perfect substitute for representative surveys. It is also stated in the Fed Board's webpage that "the Beige Book is not a commentary on the views of Federal Reserve officials."

2.3 The choice of large language model

Our baseline LLM is GPT-4o (gpt-4o-2024-08-06). This model has leading performance among generative language models. The knowledge cutoff for the model is October, 2023. We use JSON schema to force LLM to generate structural output. JSON schema also strictly requires LLM to select categories from designed categories. We choose Temperature=1 and Top-P=0.95 for all current results.

2.4 The Detailed Procedure for LLM Processing

Economic Variable and Event Extraction Schema

1. General Classification - Determine whether a phrase is a **Variable** or **Event**. Then follow either 2. Economic Variable Dimensions or 3. Economic Event Dimensions.

2. Economic Variable Dimensions If the entity is a **Variable**, classify it using the following dimensions:

1. **Sector:** *overall, manufacturing, retail, services, energy, agriculture, real estate, transportation*

2. **Economic Concept:** *good demand, good price, good sales, etc. Please refer to Table A.1 in Appendix A.1*

3. Economic Event Dimensions If the entity is an **Event**, classify it using the following dimensions:

1. **Economic Concept:** *fiscal income side, fiscal expenditure side, trade import side, trade export side, monetary policy, weather, etc. Please refer to Table A.1 in Appendix A.1.*

4. Shared Dimensions for Variables and Events Both Variables and Events share the following dimensions:

1. **Realized or Expected:** *past realized, current realized, producer expectation, consumer expectation, expert expectation.*

2. **Dynamic and Value:**

   - *+++* (significant increases/expansion), *++* (moderate increases/expansion), *+* (slight increases/expansion),
   - - - - (significant decreases/shrinking), - - (moderate decreases/shrinking), - (slight decreases/shrinking),
   - = (no change or unclear information), *?+* (uncertainty enlarges), *?+* (uncertainty reduces).

3. **Linguistic Relation:**

   - *is* (exact match), *is part of* (the phrase is a part of the economic concept), *is considered as a measure of* (the phrase is considered a measure of the economic concept).

Our method of using LLM to extract logic chains from text involves two steps. First, we prompt the LLM to perform a classification task, identifying economics-related phrases at the sentence level and categorizing them according to a multi-dimensional labeling system. The labeling system is designed to contain multiple dimensions to provide not only the economic meaning of the phrases but also information about their dynamics and time tenses. In the second step, the LLM determines the logical relationships between these phrases based on the text. Both steps rely on carefully designed prompts. We now outline this two-step procedure in detail.

Step 1: Classification

For a given unstructured text article, LLM is required to produce both an annotated article containing the detected **phrases** and multi-dimensional labels for each of the phrases. We call this task the **classification step**. We manually define all economic-related phrases as either **economic variables** or **economic events**. Economic variables refer to the observed statistics or indicators appearing in economic narratives. They can be categorized by both their corresponding economic concept (such as prices, employment rate, production activity, etc) and the sector to which they belong (such as agriculture, manufacturing, etc). Economic events refer to events not particularly related to specific sectors, such as policies, exogenous events like weather and pandemics, etc. To convert nearly infinitely many kinds of economic phrases into a finite amount of interpretable standard terminologies, we define a multi-dimensional labeling system for LLM to follow during this step. Since the labeling system is pre-defined, we allow the LLM to classify economic phrases that do not clearly fit into any existing category under the label 'Others'. This labeling system is formalized by the ***schema*** shown in the textbox above.

To illustrate the classification step, we present a very short example sentence from the Beigebook: "District merchants noted a slight negative effect on consumer spending from the resumption of the full Social Security tax.", the corresponding annotation is "District merchants noted [*a slight negative effect on consumer spending*] {V1} from [*the resumption of the full Social Security tax*]{V2}." The string "*slight negative effect on consumer spending*" and "*the resumption of the full Social Security tax*" are identified as two **phrases**, V1 and V2. The table below shows the multi-dimensional labels for these two phrases.

Step 2: Logic Extraction

In step 2, LLM receives the source article, the annotated article, and the multi-dimensional labels, and is asked to find all causal statements about the extracted economic variables and events. For a causal statement, LLM should report the related source text, interpretations,

Dimensions \ Phrase	**a slight negative effect on consumer spending(V1)**	**the resumption of the full Social Security tax(V2)**
General	variable	event
Sector	retail	NA
Economic Concept	goods sales	fiscal income side
Realized or Expected	current realized	current realized
Dynamics	-	++
Linguistic Relation	is	is

Table 1: Multi-dimensional labeling for short sample article entities.

and a list (generally only contains one phrase) of phrases as **cause** and another list of phrases **result** to describe the causal relationship. For the short example annotated article "District merchants noted [*a slight negative effect on consumer spending*] {V1} from [*the resumption of the full Social Security tax*]{V2}.", the extracted cause is [V2] and result is [V1].

3 Analysis with Outputs

We apply our method described in Section 2 to Beige books published from January 1990 to December 2023 to demonstrate the use of our methodology. As described by the **schema** above, our results include the phrases classified into economic concepts and the causal logical chains between them. We first present some summary statistics about the economic phrases we extracted and their time series patterns. Then we present our results of logic chains between these economic concepts.

3.1 Phrases and Economic Concepts

In our classification exercise, each phrase will be assigned a multi-dimension label with the economic concept and sector it belongs to, and its directional and time-stance information, as mentioned in **schema**. We include a total of 36 different economic concepts and 8 sectors described in Table 2. We also group the economic concepts by the market they belong to. If the economic concepts are about the financial market, Policy, or Exogenous Events, we do not assign sector labels to them. Among the economic concepts, we also include a category called "other event". If LLM believes the economic phrase does not belong to any of the listed economic concepts, it can classify that phrase as "other event". Reassuringly, only less than 0.5% of all the economic phrases are classified into this category.

We first show the frequency of these economic concepts by counting the number of phrases classified into each of the concepts. Within each economic concept, we further decompose it into

Table 2: Key Economic Concepts

Market	Economic Concepts		Sectors
Goods Market (Consumer)	Consumption	Consumer Demand	
	Price		
Goods Market (Producer)	Profit	Production	
	Inventory	Investment	Service
	Input Price (Raw Material)	Other Input Price	
	Input Quantity (Raw Material)	Other Input Quantity	
Labor Market	Employment	Wage	Agriculture
	Labor Market Condition	Labor Demand	
	Labor Shortage		Manufacturing
Financial Market	Loan Quantity	Interest Rate	Energy
	Loan Demand	Loan Risk	
	Deposit Quantity	Loan Supply	Real Estate
	Stock Price	Stock Quantity	
	Stock Risk	Bond Price	Retail
	Bond Quantity	Bond Risk	
Condition	General Conditions		Transportation
Policy	Trade Policy	Monetary Policy	
	Fiscal Policy	Exchange Rate	Overall
Exogenous Events	Pandemic	Weather	
	Political Event	Other Event	

For ease of presentation we group some labels into one economic concept. For example, Fiscal Policy includes both the income and expenditure side of policies. In our classification method, we separate them but when we present the economic concept they belong to, we group them under "Fiscal Policy". For a detailed list of labels, please refer to Table A.1 in Appendix A.1

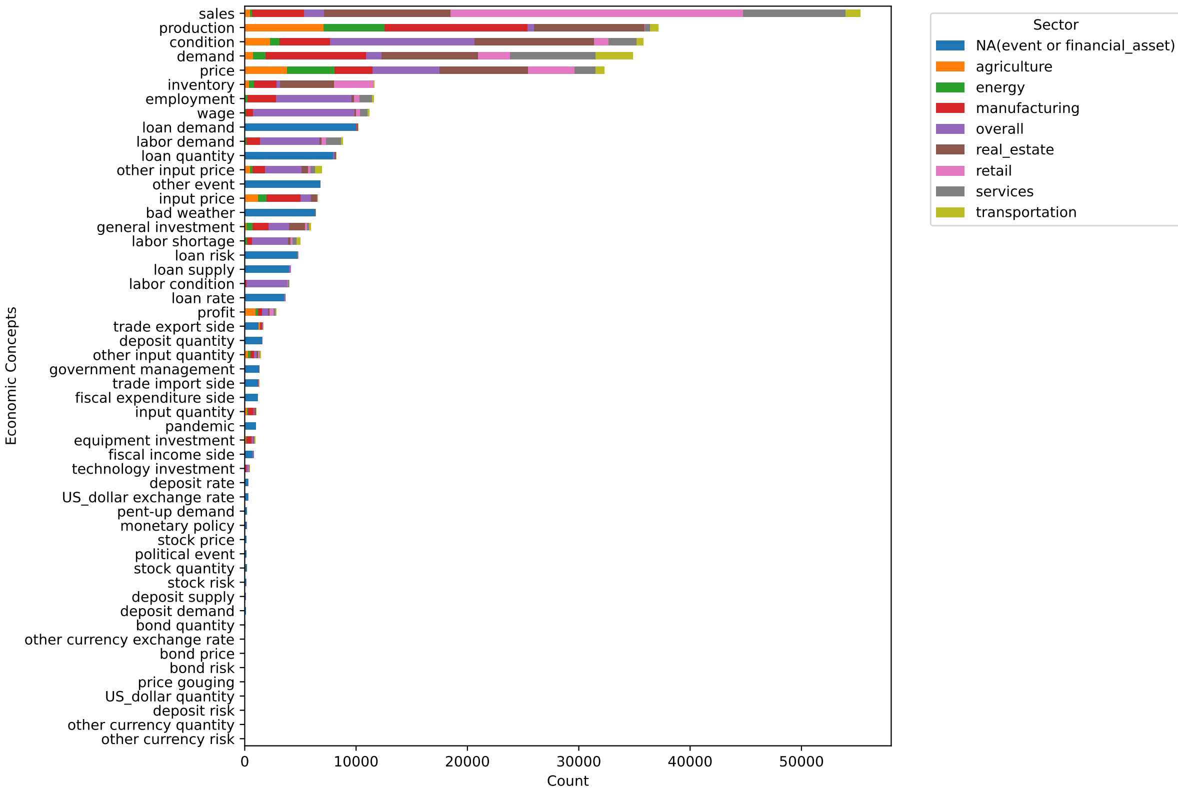
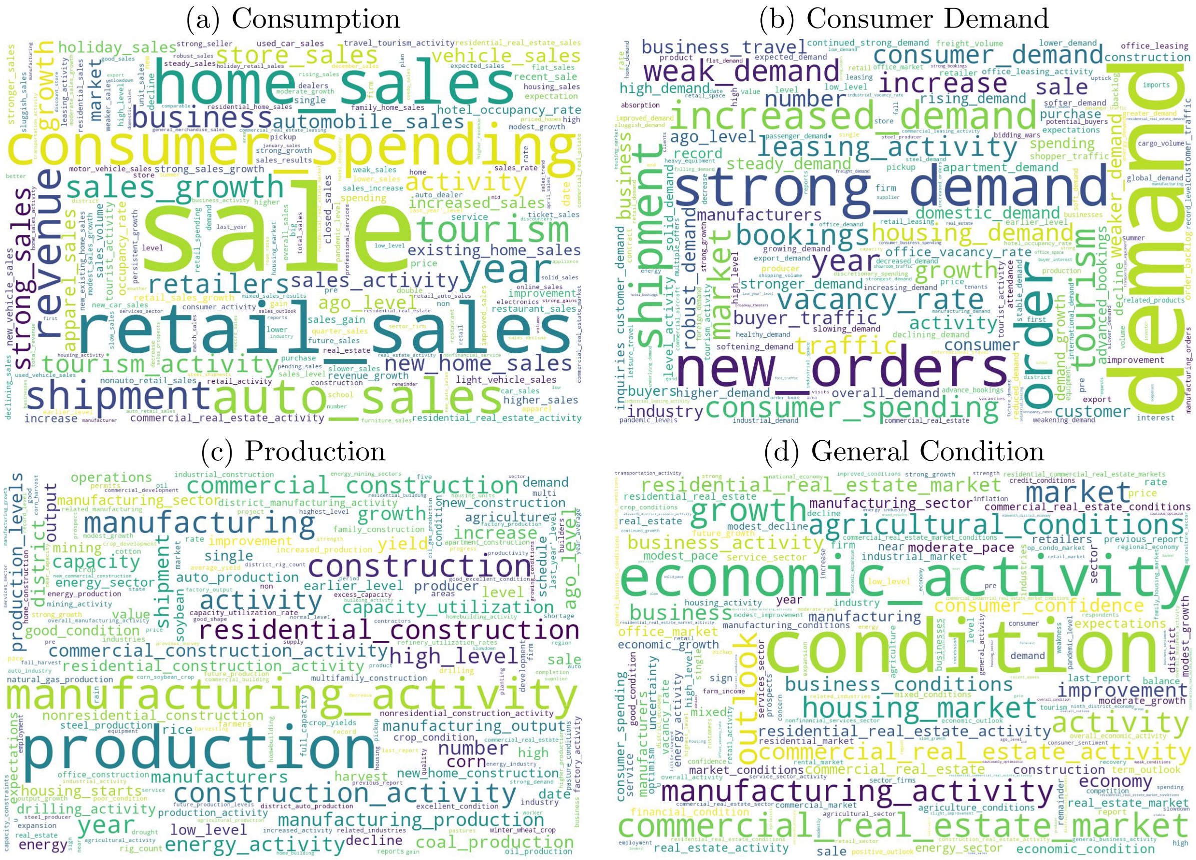


Figure 1: Frequency of Economic Concepts by Sector

different sectors that the phrases belong to. These results are shown in Figure 1. Among these concepts, 25 of them account for more than 96 percent of all the economic phrases classified. The top 5 most frequent economic concepts are consumption (sales), production, general condition, demand, and prices. Together they account for more than 55% of all phrases.

Wordclouds of high-frequency concepts What are the typical phrases used to describe a particular economic concept? To illustrate this we present the word clouds of the phrases being classified into each economic concept by LLMs. In Figure 2 we present the word clouds for consumption, consumer demand, production, and general condition. One can see the keywords used to describe consumption involve sales, consumer spending, retails, etc, which differ from those for consumer demand usually involves words like "demands" or "orders", and the keywords for production-related topics, which typically involve "production" or the specific output produced by a specific sector. In contrast, the "condition" category contains more general words about the condition or outlook for a specific sector. Moreover, the word clouds show that what keywords are used to describe economic concept depends heavily on which specific sector the text is about. For example, "housing starts" or "construction activity" are used to describe production in housing market whereas "manufacturing activities" refer to production in the manufacturing sector.

Figure 2: Word Clouds for Consumption, Demand, Production, and Condition



Notes: the word clouds of phrases labeled in Beigebook as one of the four identified economic variables.

In Figure 3 we report the word clouds for categories about prices and interest rates. We see that the keywords for prices depend less on the specific sector as they usually have "price" as a key indicator. However, the keywords for interest rates are quite rich, including terms like "loan rates", "prime rates", and "mortgage rates" which are related to the financial market.

Figure 3: Word Clouds for Price and Interest Rate

(a) Price (b) Interest Rate

Notes: the word clouds of phrases labeled in Beigebook as one of the two identified economic variables.

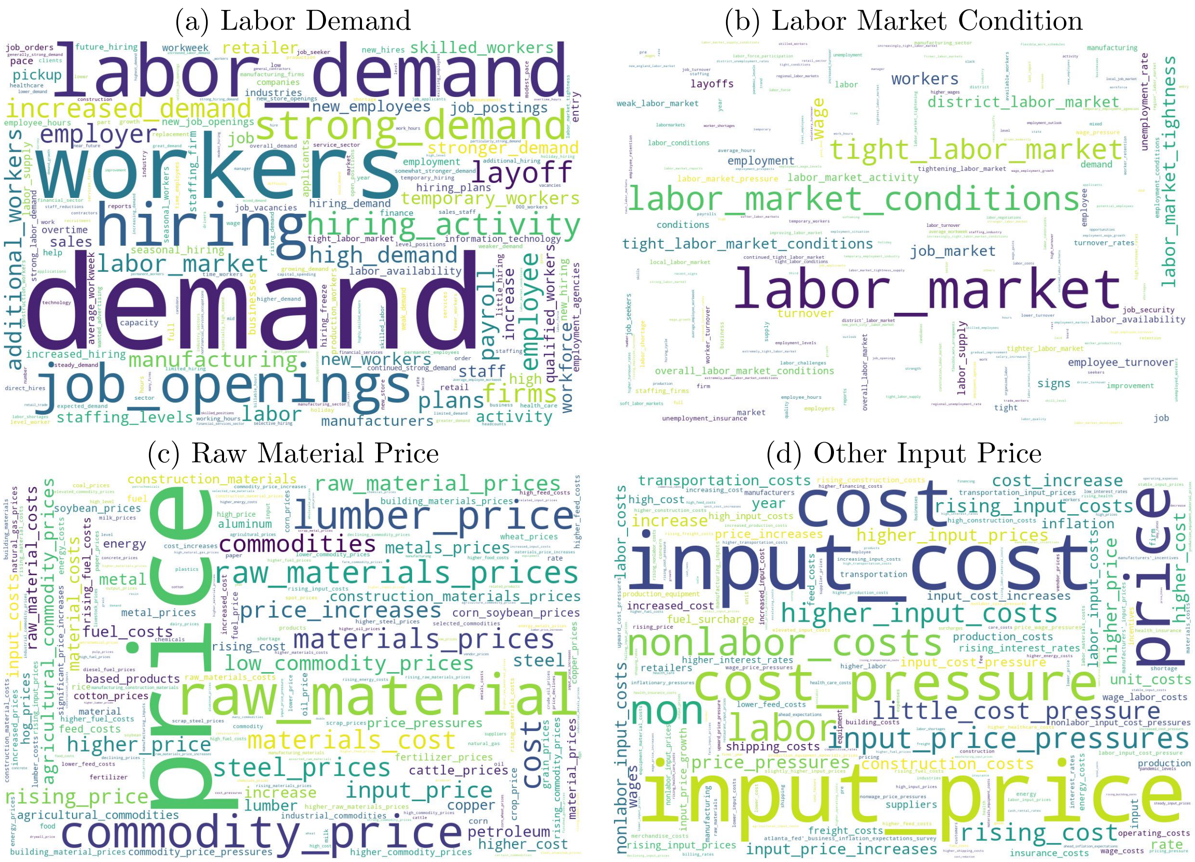
Finally, our method can also classify phrases into categories with specific economic meaning, such as labor demand, labor market conditions, and input prices. Figure 4 describes the word clouds for these economic concepts.[5]

3.2 Directional changes

One important benefit of our method is that it can also measure the direction of changes in economic variables described in the texts. For example, when a sentence discusses changes in prices, our method can capture whether the price is going up or down. When one particular economic concept is more frequently mentioned in the texts, it can mean there are conflicting dynamics in this economic variable, which highlights the uncertainty. Alternatively, the texts may repeatedly describe the same dynamics which means the corresponding economic statistics are moving in one specific direction. The directional information captured by our method can help to distinguish these two possibilities. Moreover, it also helps us to examine whether there are biases towards a specific direction when the texts describe different economic concepts. Figure 5 shows the fraction of phrases that describe increase, decrease, unchanged, or unknown dynamics of the corresponding economic concept.

[5]We include the word clouds for top 25 most frequent economic concepts in the Appendix A.2.

Figure 4: Word Clouds for Concepts with Specific Economic Meaning



Notes: the word clouds of phrases labeled in Beigebook as one of the four identified economic variables.

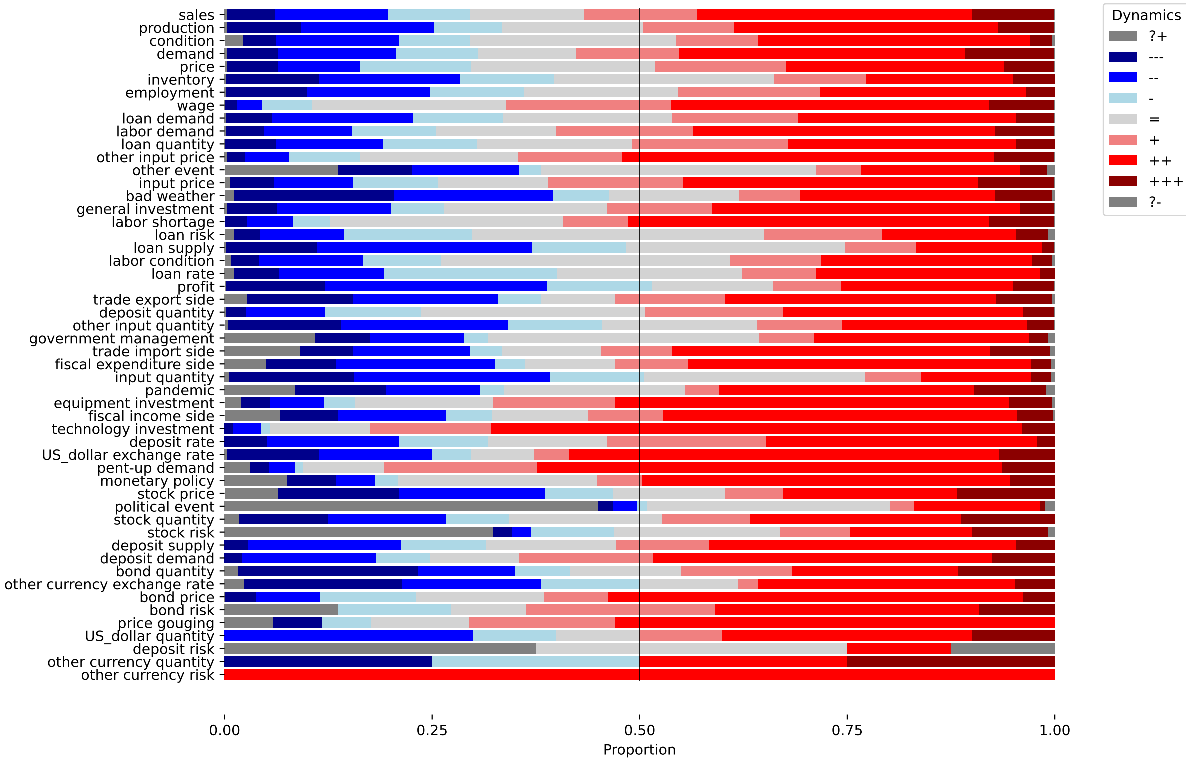


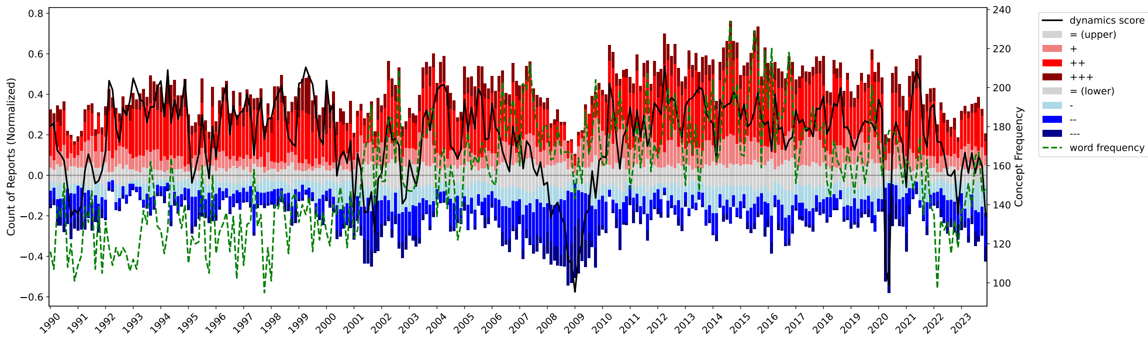
Figure 5: Positive or Negative Dynamics for each Economic Concept

From Figure 5, we see that on average the Beige books describe more positive dynamics, in particular for concepts like wage, input prices, and investment. This may be due to the more frequent positive fluctuations of these economic variables. Obviously, this decomposition of positive or negative dynamics will vary a lot across time – we expect to see more descriptions of price hikes during inflation episodes than at other times. To examine this, we plot the time series pattern of the same decomposition for various economic concepts. As we have many categories, we focus on the most frequent concepts in Figure 1. In Appendix A.4, we include some extra concepts we considered.

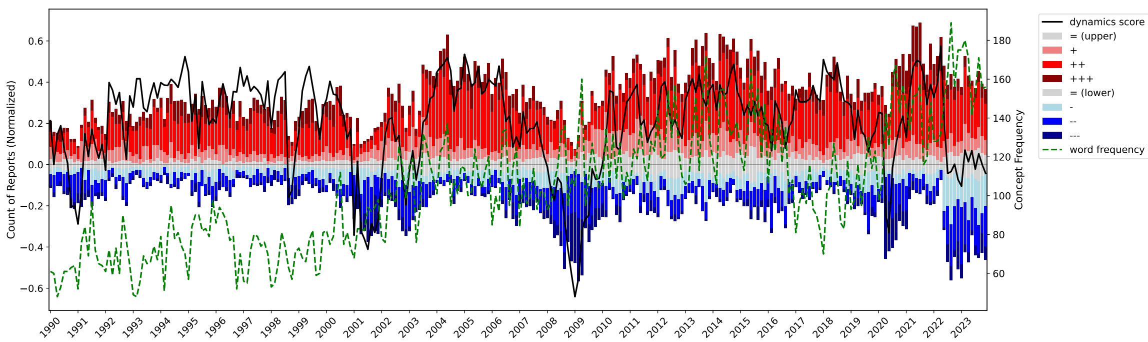
Figure 6 presents phrases related to four economic conceptsconsumption, consumer demand, production, and general conditionsalong with their directional trends over time. In each panel, red indicates a positive dynamic, blue represents a negative dynamic, and gray represents no change for the corresponding concept. Darker shades signify greater intensity (significant, moderate, and slight as described in our **schema**). We then assign values to the dynamic information and compute the average score of each concept. This corresponds to the solid black line in each panel. We also plot the total count of the phrases mentioned in each point of time in dashed green lines. In Figure 6, we see that the scores of all four concepts are pro-cyclical and capture the 2001, 2008, and 2020 recessions well. Whereas simple word counts typically

Figure 6: Dynamics of Economic Concepts

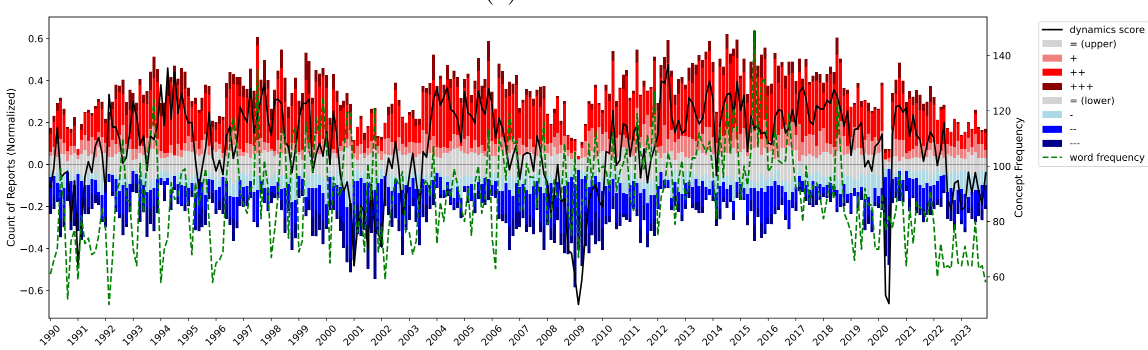
(a) Consumption



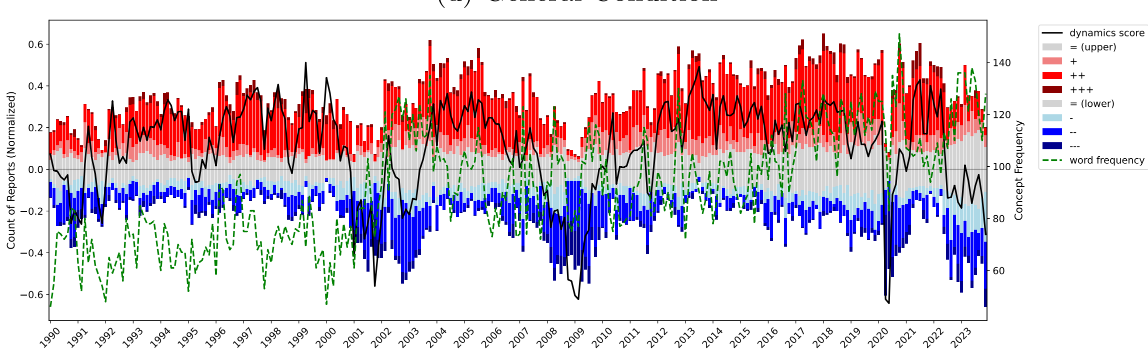
(b) Consumer Demand



(c) Production



(d) General Condition



Notes: the figure plots the dynamic score of each variable, and their breakdown of different directions, in addition to the total frequency of each variable identified in the Beigebooks.

capture the frequency of a concept being mentioned but cannot provide information about the dynamics of that concept. Across time, positive dynamics are more frequently mentioned than negative dynamics, which are mostly used during economic downturns.

Next, we examine the dynamics of concepts related to prices. Figure 7 displays the dynamic scores for price, wage, input prices (including raw materials and commodities), and other input costs (such as supply chains). The discussion of general prices, wages, and input prices fluctuates over time, with a notable surge in recent years, predominantly reflecting positive dynamics. In particular, between March 2020 and June 2022, discussions about rising general prices, wages, and input costs all increased sharply. After July 2022, mentions of input prices declined, while discussions on general prices and wages remained elevated. However, both positive and negative dynamics of prices and wages are being reported.

How do soft narratives compare to hard numbers? We compare the narrative-based directions identified via LLM with the corresponding economic time series data. This serves as a validation of LLM-based measures for empirical analysis in the spirit of Ludwig et al. (2025). Figure 8 plots the narrative dynamics of a few selective variables along with their corresponding economic statistics with reported in-sample correlation coefficients. Across a large array of variables, the correlation coefficients are well above 0.4. This implies that our extraction procedures yield a large part of the meaningful changes in economic conditions as measured by statistics, but meanwhile contain additional information that is probably by construction orthogonal to aggregate series.

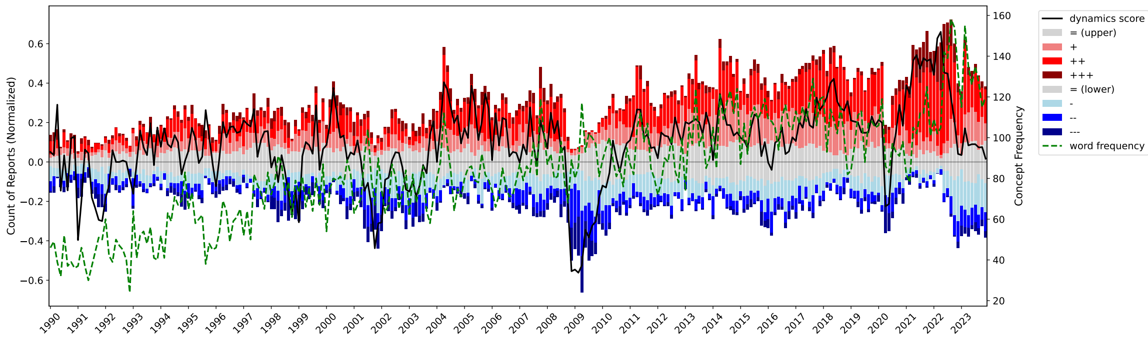
For comparability, wherever feasible, we calculate high-frequency changes or growth rates of the time series to be consistent with the approximately one-to-two month window of Beigebook release. But for the majority of considered variables, monthly changes are actually less correlated with the Beigebook scores than those over a slightly lower frequency, i.e. quarterly. The most natural explanation for such a pattern is probably that although the Beige book's discussions are in general relative to the previous release, narrative discussions and anecdotal evidence from survey contacts ultimately capture dynamics over a longer horizon. Another reason is that our procedures extract very decentralized dynamics in different sectors and different districts. The simple sum of such dynamic scores may not fully reflect the aggregate patterns of the economic conditions.

3.3 Logic Chains

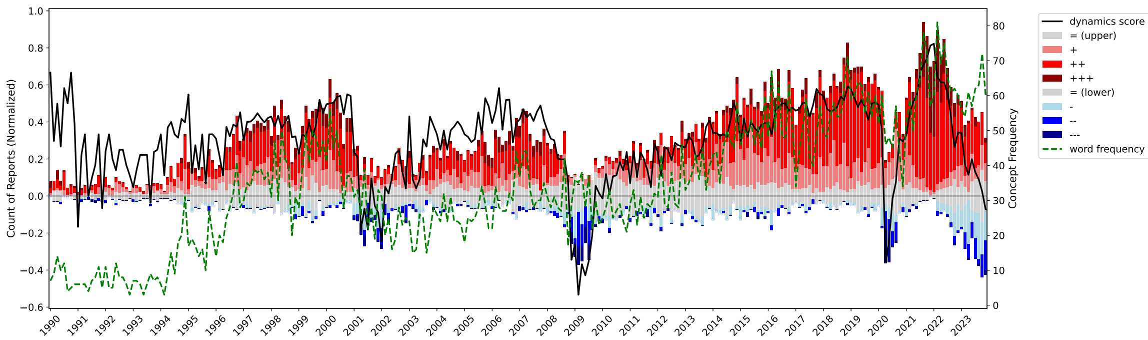
With various economic phrases classified into economic concepts, we constructed logic chains described in the Beige Book between these concepts. A logic chain consists of two economic phrases and the text indicates one leads to or is the consequence of the other. As a result,

Figure 7: Dynamics of Prices

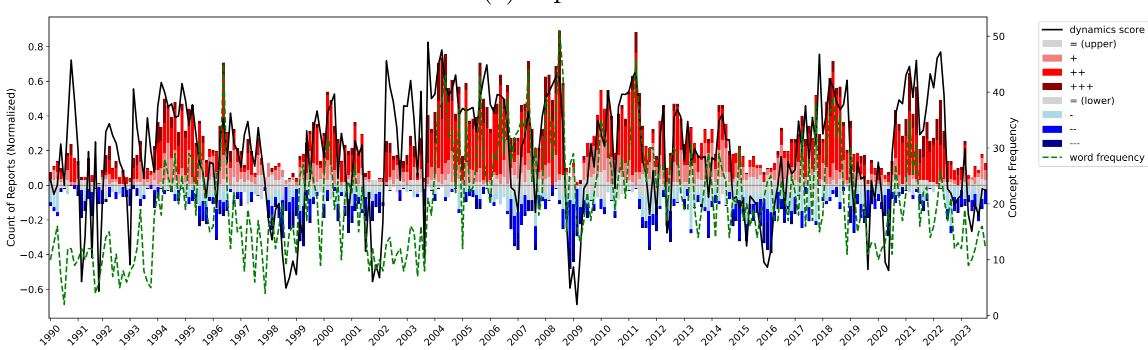
(a) Price



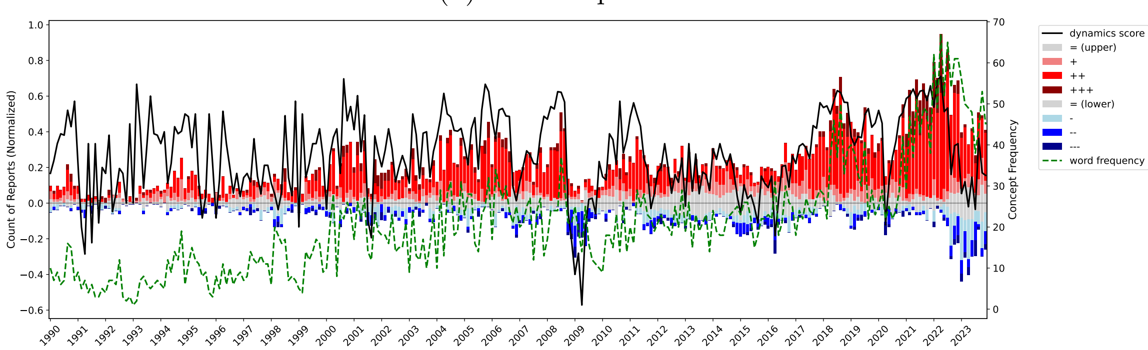
(b) Wage



(c) Input Price

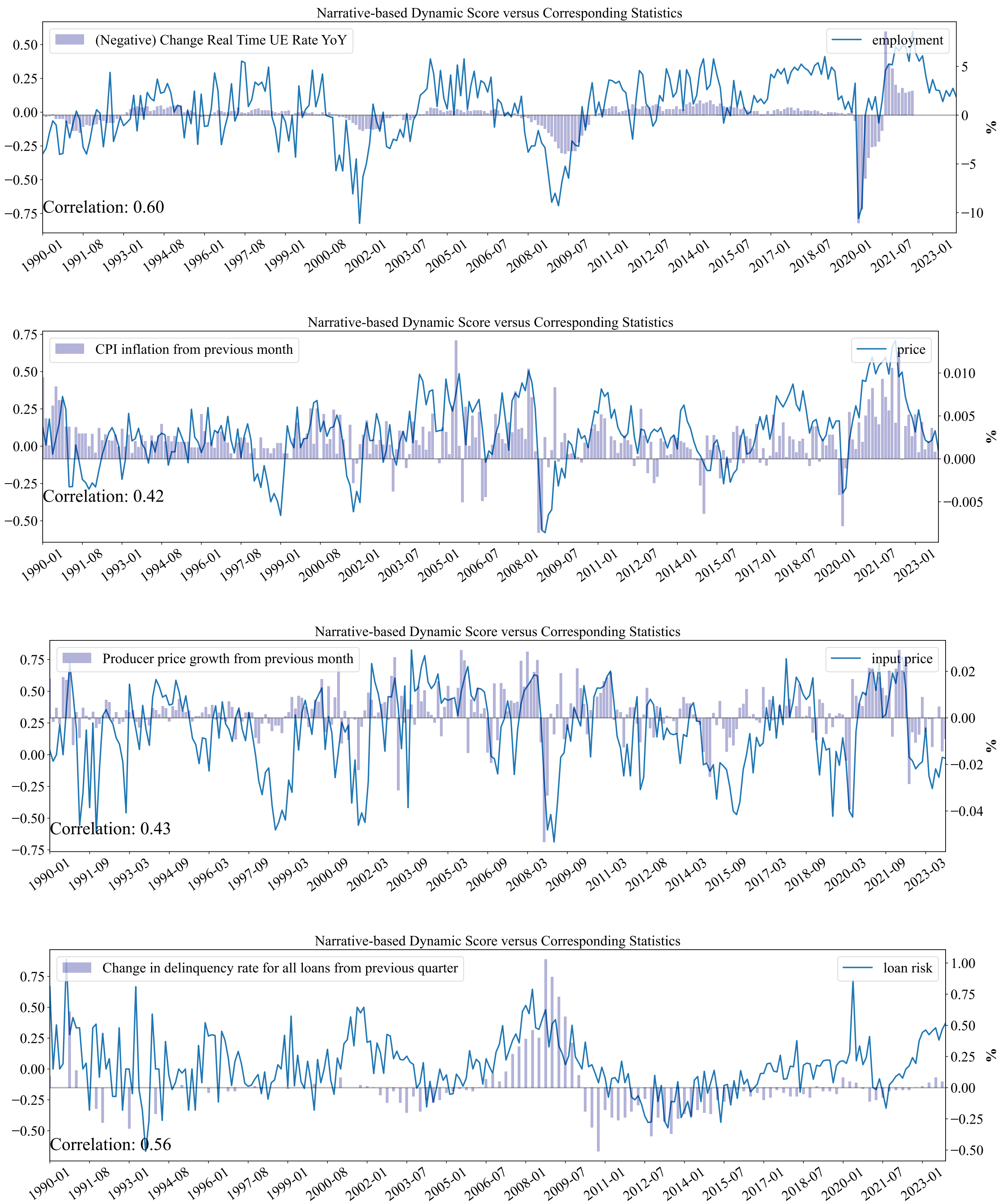


(d) Other Input Price



Notes: the figure plots the dynamic score of each variable, and their breakdown of different directions, in addition to the total frequency of each variable identified in the Beigebooks.

Figure 8: Narrative Directions versus Economic Statistics



each logic chain has a *cause* and a *consequence*. To illustrate how a logic chain is constructed, consider a short sentence example from the Beige Book:

> The [prices of raw materials for manufacturing, construction and food services rose]{V13}, with [some pass-through to finished goods prices]{V14}.

In this sentence, two economic phrases are extracted: *[prices of raw materials for manufacturing, construction, and food services rose]{V13}* is classified as input prices, while *[finished goods prices]{V14}* is labeled as goods prices. The extracted logic chain indicates that a *modest increase in input prices* leads to a *slight increase in goods prices.*

As we have a large set of economic concepts, we first present a heatmap matrix that summarizes the causes and consequences in all the logic chains we extracted. In Figure 9, the vertical axis includes 36 economic concepts that are the causes of extracted logic chains, and the horizontal axis includes the consequences. The number in each entry of the matrix represents the number of logic chains with the corresponding cause and consequence. To give a better visualization, we group the economic concepts according to the market they belong to, following Table 2.[6]

In Figure 9, we see that most logic chains are concentrated within the specific market to which the economic concept belongs. This is expected as most of the causal descriptions focus on direct consequences and within-market dynamics. However, with sentence-level logics, we can further connect short logics and construct a longer logic chain that describes a complete narrative including both within-market and cross-market dynamics. This requires that we at least have a significant amount of logic chains across different markets. Reassuringly, Figure 9 does show that there is a substantial number of logic chains that are across different markets. Notably, the most frequently mentioned concepts (such as "price", "sales", "production", and "condition") are also the most common consequences in logical chains. In contrast, exogenous events (such as "weather", "political events", and "pandemics") rarely appear as consequences in these chains. Conceptually, we can think of the within-market logic chains as "micro-level" logics and across-market logic chains as "macro-level" logics. With our directional information on economic phrases, we can first examine whether the logic chains line up with economic theory.

We start by examining the logic chains within the goods market, leveraging our method's ability to extract concepts related to demand and supply. Specifically, we use phrases about consumer demand (word cloud shown in panel b of Figure 2) as indicators of demand, while

[6] We skip the diagonal as most logic chains with causes and consequences of the same economic concept are those rephrasing the same within-market dynamics.

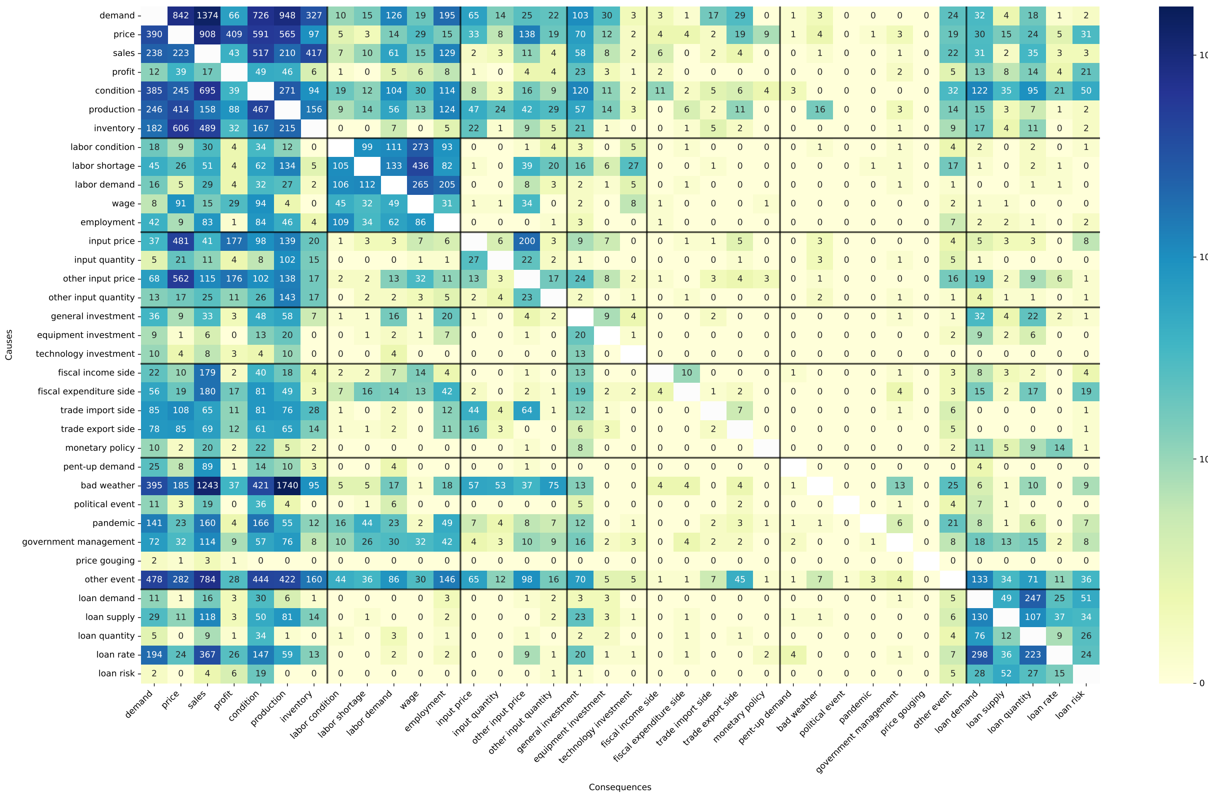


Figure 9: Matrix of Causes and Consequences of Logic Chains

phrases related to input prices (word clouds in panels c and d of Figure 4) and energy prices serve as indicators of supply factors.

Figure 10 illustrates how the dynamics of demand and supply factors influence key equilibrium variables in the goods market, including price, sales (consumption), production, and inventory. Each panel presents a cause-consequence matrix, where the y-axis represents changes in the cause variable, and the x-axis represents changes in the outcome variable. The matrix elements indicate the number of extracted logic chains, with darker colors representing a more frequent appearance.

From Figure 10, we observe that positive demand shocks lead to increases in both prices and quantities. In contrast, rising input and energy prices drive up goods prices but reduce sales and production.[7] Interestingly, higher demand generally leads to lower inventory levels, though in some cases, firms respond by increasing production and building up inventory instead.

Similarly, we can analyze the logic chains of demand and supply factors within the labor market. Figure 11 examines how labor demand and labor shortages influence wages, employment, and labor market conditions.

[7]Some exceptions appeared when the logic extracted indicates that higher input price leads to higher production and consumption. This is due to either (1) the sentence is describing production substitution; (2) the sentence is describing high holding cost leads to earlier production. See Appendix A.3 for two examples.

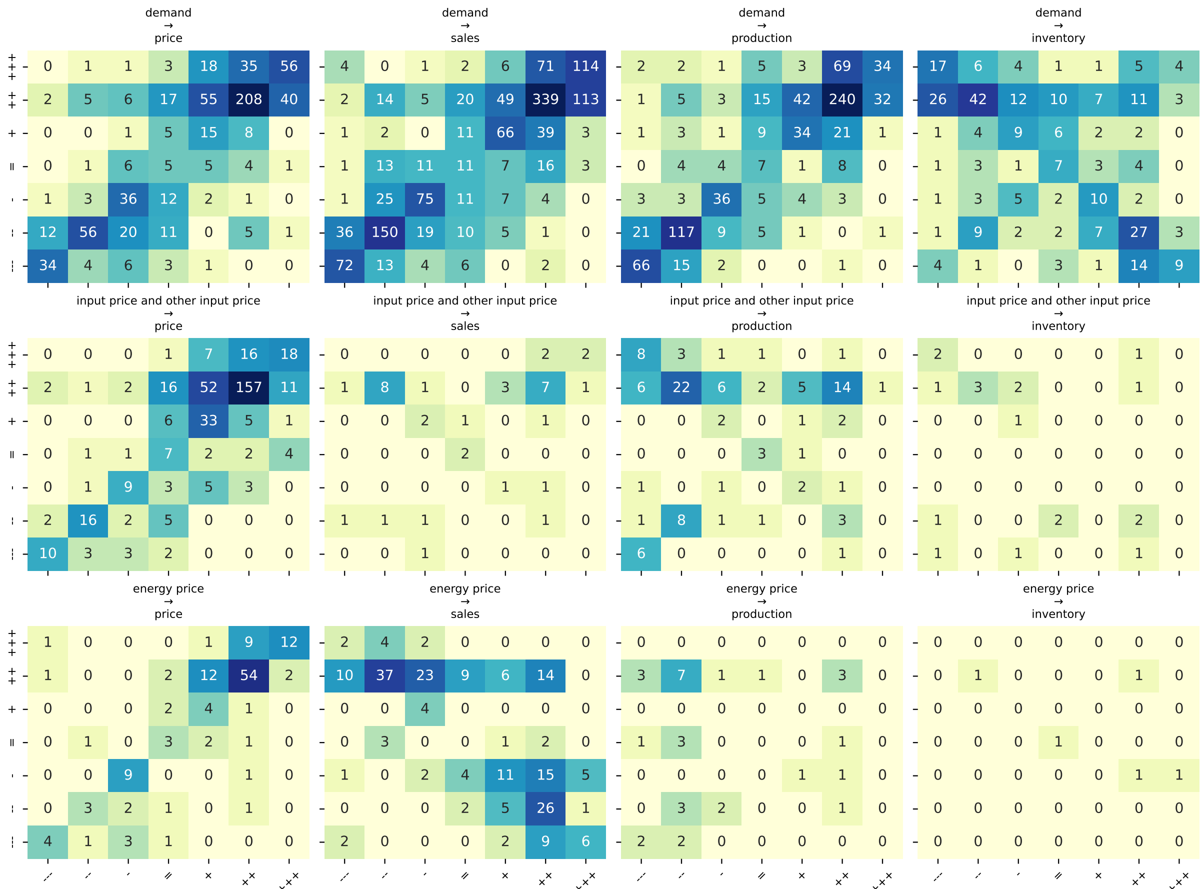


Figure 10: Consequences of Demand v.s. Supply factors in Goods Market

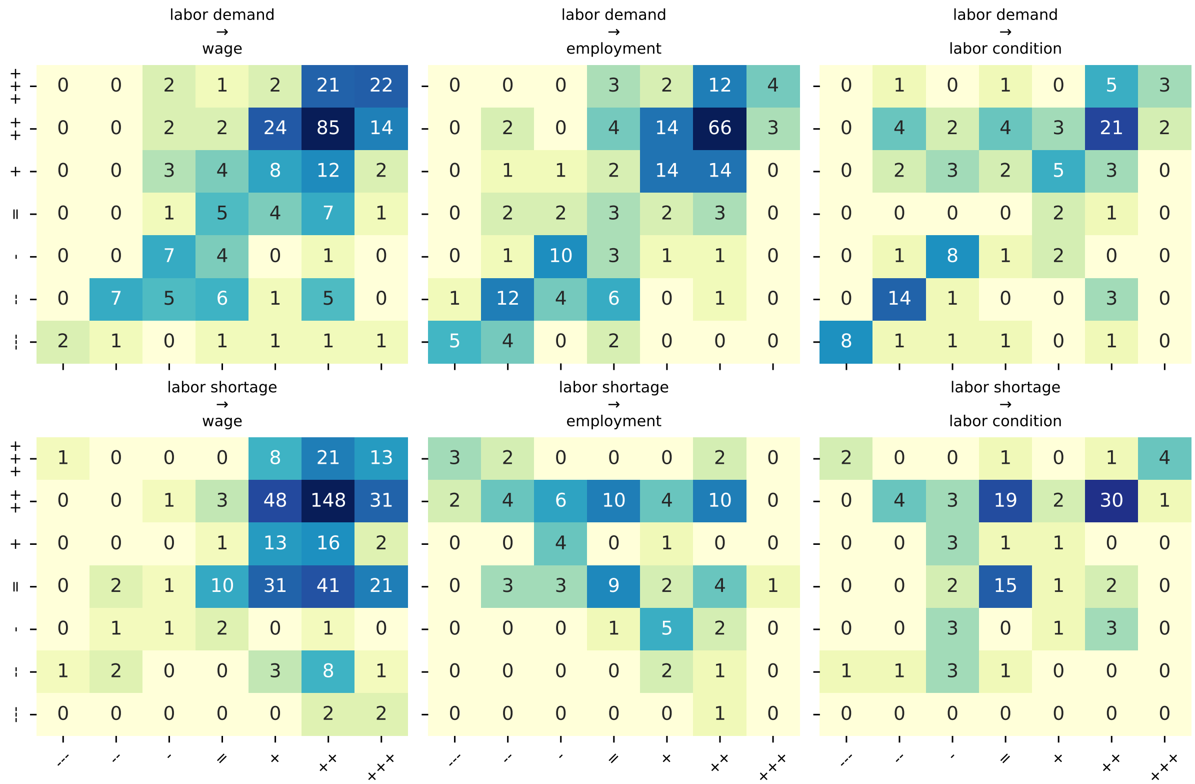


Figure 11: Consequences of Labor Demand v.s. Supply factors in Labor Market

Since the Beige Book primarily gathers information from interviews with key business contacts, economists, market experts, and other sources, most labor market descriptions reflect the perspective of businesses and firms. As a result, labor supply is often represented through discussions of labor shortages.

From Figure 11, we observe that higher labor demand drives increases in wages, employment, and labor conditions. In contrast, labor shortages tend to push wages up but reduce employment. Interestingly, greater labor shortages are also associated with improvements in labor conditions. As shown in the labor conditions word cloud, this category primarily captures labor market tightness, suggesting that labor shortages contribute to a tighter labor market.

After demonstrating that our extracted logic chains align with economic intuitions within specific markets, we now explore whether the Beige Book discusses supply and demand narratives commonly found in macroeconomics. For example, input prices such as raw materials and energy are typically considered "supply factors" in macroeconomic theory, likely creating inflationary pressure and leading to lower production, consumption, employment, and profits. In contrast, consumer demand is a standard demand shock that drives up both prices and real economic activity. In Figure 12, we examine how the supply and demand narratives are reflected in our extracted logic chains.

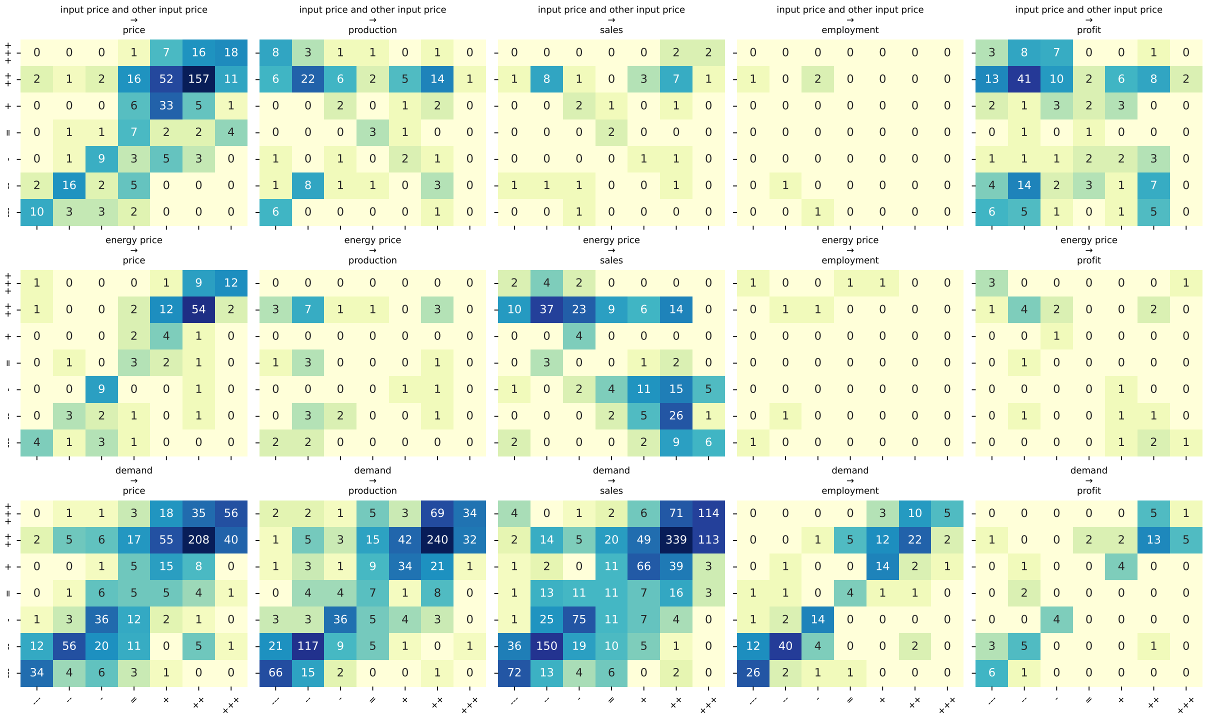


Figure 12: Supply v.s. Demand Narratives in Macroeconomics

As with the market-specific logic chains, the supply narrative generally suggests that higher input prices lead to increased inflation, reduced firm profits, and dampened real activities such as production and consumption. However, the Beige Book rarely mentions the impact of supply dynamics on employment. In contrast, the demand narrative indicates that higher demand results in higher prices, greater profits, and increased economic activity. Both narratives align with insights from a standard New Keynesian model.

Figure 12 clearly shows that demand narratives are more prevalent than supply narratives. More importantly, the supply narratives predominantly focus on price changes as their consequence, with limited discussion of their impact on real activities. In contrast, demand narratives emphasize their effects on real economic activity. Figure 13 compares the number of logic chains with price or real activity as the outcome for supply and demand narratives, confirming that supply dynamics are typically linked with price changes, while demand dynamics are more often associated with consequences for real activities.

Another interesting logic chain to explore is the impact of labor market factors on the goods market. In Figure 14, we present logic chains that originate from labor market dynamics and examine their effects on prices and real economic activity. We observe that higher wages predominantly drive up prices and reduce profits, while their impact on real activity is rarely mentioned. High labor demand and a tight labor market work similar to a demand shock, contributing to increases in sales and production, though their effects on these real activities

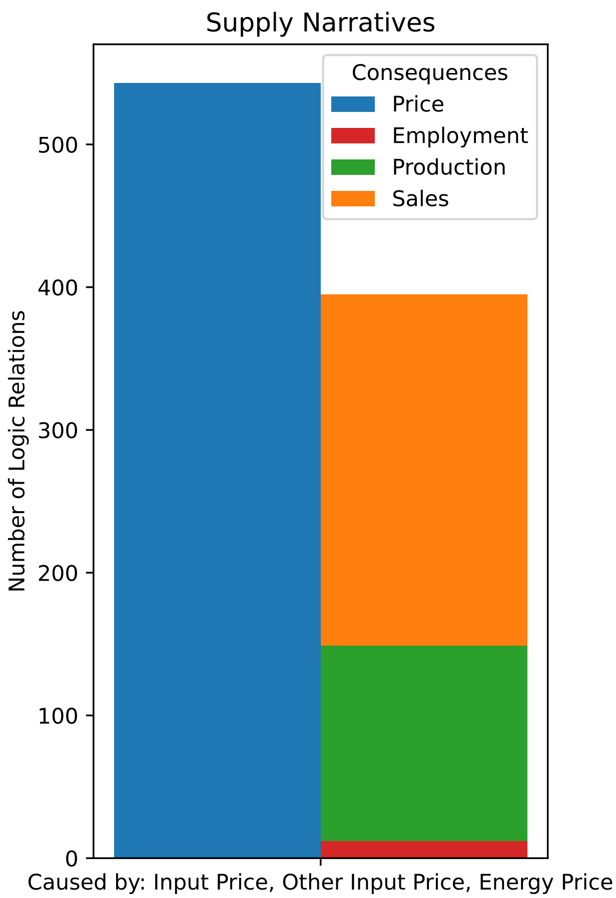


Figure 13: Supply v.s. Demand Narratives

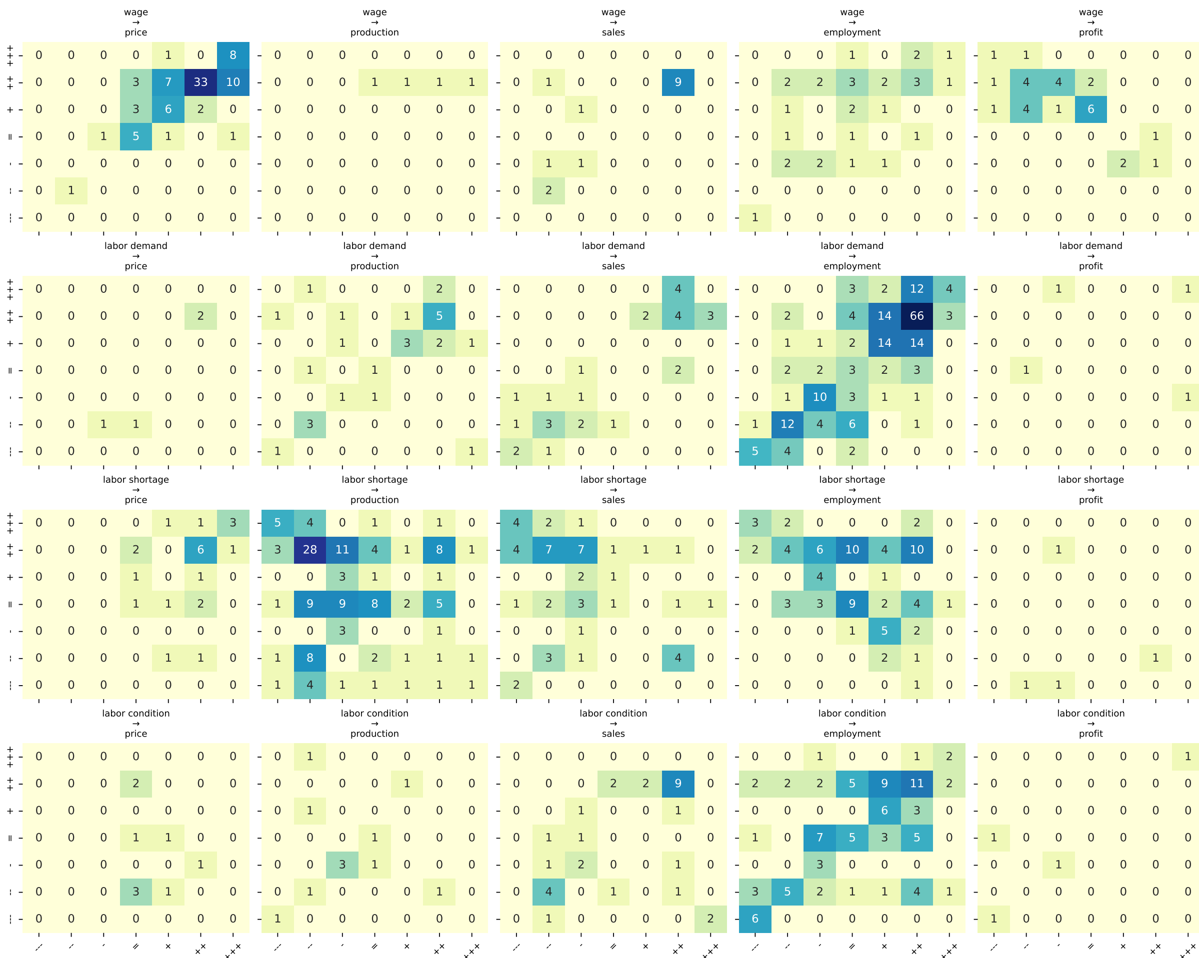


Figure 14: Macroeconomics Narratives about Labor Market Factors

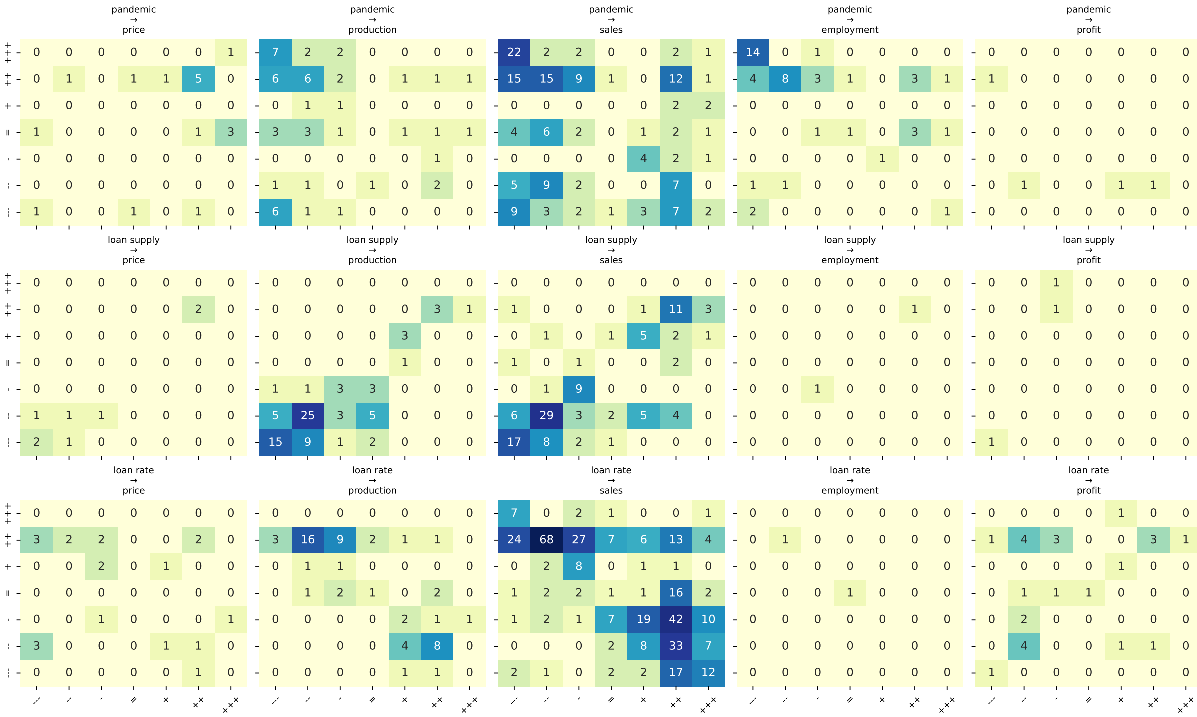


Figure 15: Macroeconomics Narratives about Labor Market Factors

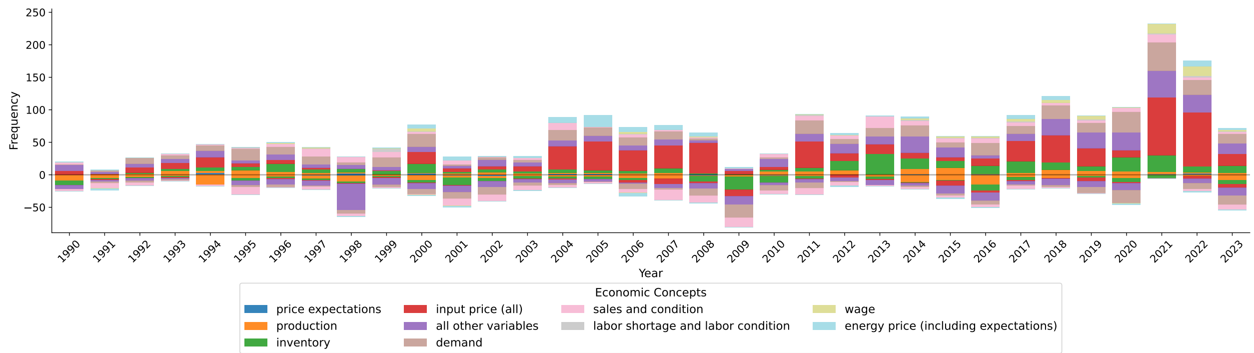
and prices are seldom discussed. Meanwhile, labor shortages tend to reduce production and sales, mirroring the effects of a decline in labor supply. The results suggest that wages are primarily discussed as a key factor influencing prices, while labor shortages play a significant role in shaping production and sales dynamics. In contrast, other labor market-related factors in the Beige Book are generally examined within the context of their own market rather than their broader economic impacts.

Finally, we examine the impact of key economic events and financial market variables. In Figure 15, we present the consequences of the pandemic, loan supply, and interest rates. As expected, the pandemic negatively affects real economic activities such as production, sales, and employment. Meanwhile, easier credit conditionswhether through higher loan supply or lower interest rateshave similar effects, increasing production and consumption while also contributing to inflation. However, much like demand narratives, discussions tend to focus more on their impact on real economic activity rather than on prices.

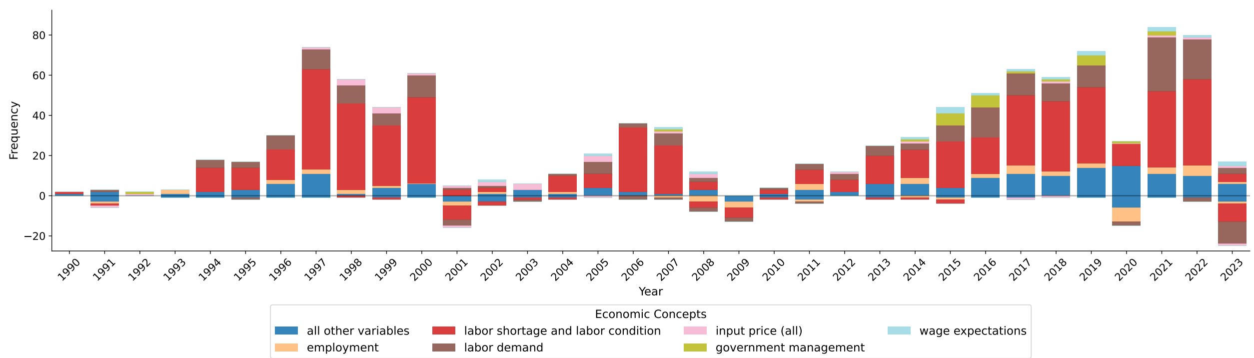
Causes of Inflation, Wage, and Employment The extracted logic chains provide insights into the immediate drivers of key macroeconomic variables frequently mentioned in the Beige Book. As an example, we analyze the most commonly cited causes of inflation, wages, and employment over time, categorizing them based on whether they contribute to positive or negative dynamics in these variables. These findings are presented in Figure 16.

Figure 16: Reasons for Inflation, Wage, and Employment Dynamics

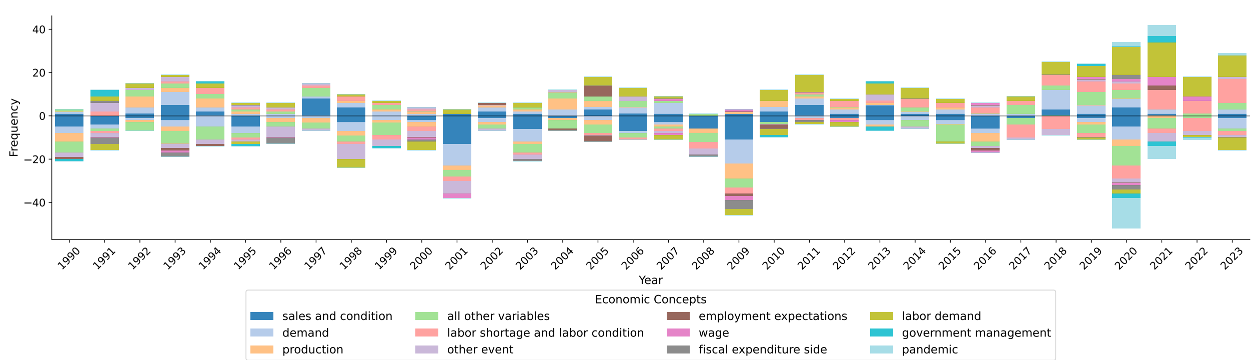
(a) Causes of price dynamics (excluding energy sector)



(b) Causes of wage dynamics



(c) Causes of employment dynamics



Panel (a) in Figure 16 illustrates the increasing discussion of inflation drivers over time, with a sharp rise following the 2020 pandemic, when inflation reached historically high and persistent levels. These discussions primarily focus on the causes of price hikes, with high input costs emerging as the most significant factor throughout this period, followed by demand and wages. Another period with extensive discussion on high input prices as a driver of inflation occurred between 2005 and 2008, when energy prices rose sharply. The reported drivers of inflation dynamics also exhibit asymmetry. Input prices and wages are primarily cited as reasons for price increases, while demand is most often mentioned as a factor in price declines.

In Figure 16, panel (b) shows that labor shortages and labor demand are the most frequently cited reasons for wage increases. Discussions about wage hikes were particularly prominent before and after the 2020 pandemic, but they also received similar attention between 1997 and 2000. This aligns with the fact that during 19972000, the phrase 'labor shortage' appeared with a frequency comparable to that of the periods before and after the 2020 pandemic. Another notable aspect of the logic chains related to wages is that discussions on wage dynamics predominantly focus on wage increases. Consequently, labor demand and labor shortages are frequently cited as key drivers of wage hikes.

Finally, panel (c) presents the breakdown of factors cited as drivers of employment changes. Compared to wage and inflation dynamics, the logic chains for employment show a more balanced distribution between positive and negative changes. As expected, labor market-related factors, such as labor shortages and demand, are the most commonly cited immediate causes of employment fluctuations.

Next, we focus on the pandemic period to examine the key drivers of inflation and wage changes at a monthly frequency. Figure 17 presents the decomposition of inflation and wage drivers mentioned in the Beige Book from January 2019 to November 2023.

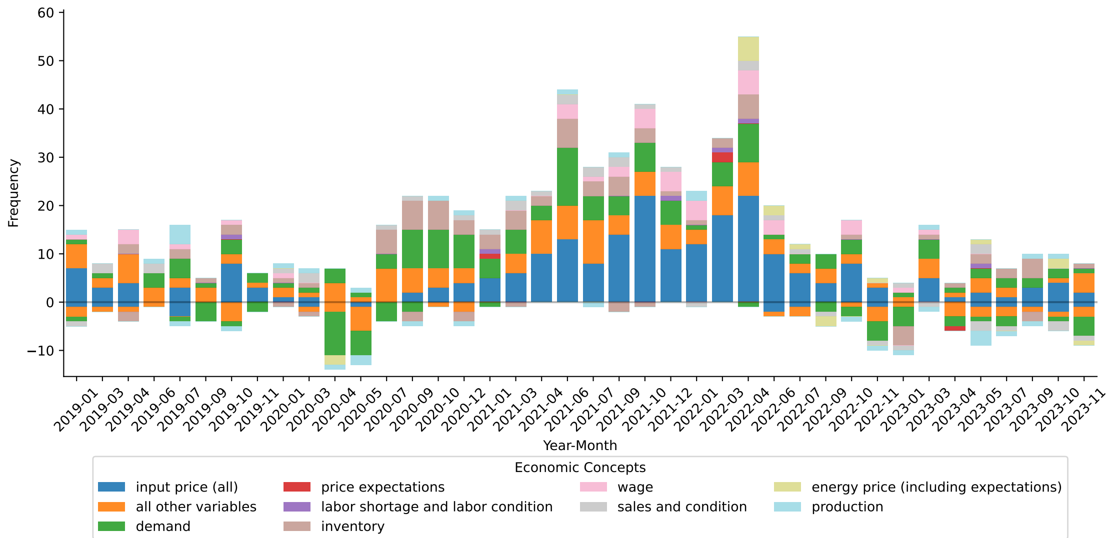
In March and April 2020, when strict quarantine measures were imposed, discussions largely centered on declines in inflation and wages, primarily attributed to weak demand and layoffs. This was followed by a rebound in demand, which emerged as a key driver of price increases once restrictions were lifted.

Starting in March 2021, input prices replaced demand as the primary driver of price surges. Around the same time, core CPI began rising above its 2% target. Shortly after the surge in inflation discussions, wage pressure became a prominent topic, beginning in June 2021, with labor demand and shortages cited as the primary causes for wage increases.

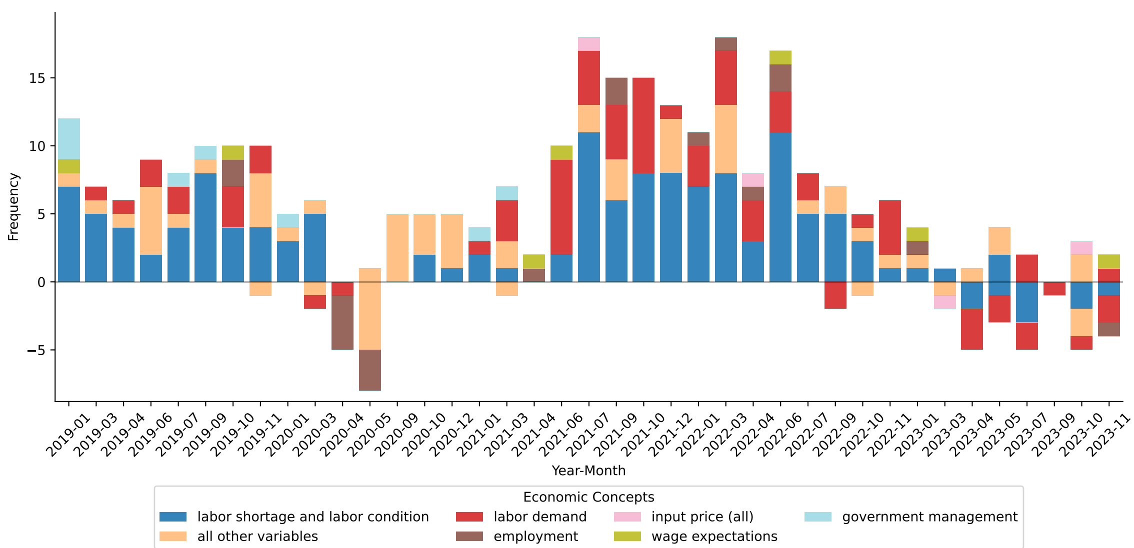
Mentions of inflation and wage dynamics began to decline after June 2022. By November 2022, discussions on inflation increases and decreases had become more balanced, and several months later, discussions about wage pressure followed a similar pattern.

Figure 17: Monthly reasons for inflation and wage

(a) Causes of price dynamics (excluding energy sector)



(b) Causes of wage dynamics



This analysis highlights a timing difference in the factors driving inflation. Demand initially received the most attention immediately after COVID restrictions were lifted. However, core CPI only began to surge when input prices became a major concern. Discussions about wage pressure lagged behind the rise in inflation, which is consistent with findings from Lorenzoni and Werning (2023).

4 Conclusion

In this paper, we develop a method to use a large language model (LLM) to systematically transform texts about the assessment of the economy into a structuralized format that can be used for further economic analysis. We first classify economic phrases in each sentence of the article according to the economic concepts they belong to. We distinguish key variables such as quantity, price, supply, and demand and at the same time capture their directional changes and differentiate between realized conditions and future expectations. We then construct logic chains to capture the causal relationships between these economic concepts as described in the texts being analyzed.

We apply our method to read reports from the Beige books. Our classification method successfully labeled 99.5% of the identified economic phrases, effectively distinguishing similar concepts such as input prices, energy prices, and general price levels. With the directional information extracted with our method, we reveal that economic narratives tend to emphasize positive over negative dynamics. The resulting dynamic score aligns well with macroeconomic data, validating the effectiveness of our methodology. Furthermore, we construct logic chains that capture both within-market dynamics as well as cross-market interactions between different economic variables. We show that supply and demand narratives about the macroeconomy are broadly consistent with economic theory. Moreover, we find that demand narratives are more prevalent than supply narratives, with supply discussions focusing on price changes while demand discussions emphasize real activity consequences.

Focusing on the pandemic period, our analysis reveals the evolutions of the primary drivers of the post-pandemic inflation surge. We find that demand pressures dominate immediately after the strict lockdowns are lifted, followed by input prices becoming the main factor as core inflation exceeded the 2% target. Meanwhile, wage dynamics were largely driven by labor shortages and labor demand. The discussions of inflation hikes start to decline after September 2022. Narratives about wage increases also fall shortly after that. By December 2022, discussions about price and wage changes became more balanced. These findings underscore the value of our method in systematically extracting structured economic insights from textual data.

References

Andre, Peter, Carlo Pizzinelli, Christopher Roth, and Johannes Wohlfart, "Subjective models of the macroeconomy: Evidence from experts and representative samples," *The Review of Economic Studies*, 2022, *89* (6), 2958–2991.

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

Armesto, Michelle T, Rubén Hernández-Murillo, Michael T Owyang, and Jeremy Piger, "Measuring the information content of the beige book: A mixed data sampling approach," *Journal of Money, Credit and Banking*, 2009, *41* (1), 35–55.

Aruoba, S Borağan and Thomas Drechsel, "Identifying monetary policy shocks: A natural language approach," Technical Report, National Bureau of Economic Research 2024.

Ash, Elliott and Stephen Hansen, "Text algorithms in economics," *Annual Review of Economics*, 2023, *15* (1), 659–688.

_ , Germain Gauthier, and Philine Widmer, "Text semantics capture political and economic narratives," *arXiv preprint arXiv:2108.01720*, 2021.

Balke, Nathan S, Michael Fulmer, and Ren Zhang, "Incorporating the beige book into a quantitative index of economic activity," *Journal of Forecasting*, 2017, *36* (5), 497–514.

Bybee, J Leland, "The ghost in the machine: Generating beliefs with large language models," *arXiv preprint arXiv:2305.02823*, 2023.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, "Business news and business cycles," Technical Report, National Bureau of Economic Research 2021.

Byrne, David, Robert Goodhead, Michael McMahon, and Conor Parle, "The central bank crystal ball: Temporal information in monetary policy communication," Technical Report, Central Bank of Ireland 2023.

_ , _ , _ , and _ , "Measuring the temporal dimension of text: an application to policymaker speeches," Technical Report, Central Bank of Ireland 2023.

Flynn, Joel P and Karthik Sastry, "The Macroeconomics of Narratives," *Available at SSRN 4140751*, 2022.

Gascon, Charles S and Joseph Martorana, "The Beige Book and the Business Cycle: Using Beige Book Anecdotes to Construct Recession Probabilities," Technical Report 2024.

Gáti, Laura and Amy Handlan, "Monetary communication rules," 2022.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, "Text as data," *Journal of Economic Literature*, 2019, *57* (3), 535–74.

Hou, Chenyu and Tao Wang, "Uncovering Subjective Models from Survey Expectations," Technical Report, mimeo 2024.

Kamdar, Rupal and Ray Walker, "Attention-Driven Sentiment and the Business Cycle," *Manuscript. https://walkerdray.github.io/pdfs/retroexpect.pdf*, 2024.

Kliesen, Kevin L and Devin Werner, "Using Beige Book text analysis to measure supply chain disruptions," *Economic Synopses*, 2022, (18).

Larsen, Vegard and Leif Anders Thorsrud, "Business cycle narratives," 2019.

Lorenzoni, Guido and Iván Werning, "Wage-Price Spirals," *Brookings Papers on Economic Activity*, 2023.

Ludwig, Jens, Sendhil Mullainathan, and Ashesh Rambachan, "Large language models: An applied econometric framework," Technical Report, National Bureau of Economic Research 2025.

Macaulay, Alistair and Wenting Song, "Narrative-Driven Fluctuations in Sentiment: Evidence Linking Traditional and Social Media," *Available at SSRN 4150087*, 2022.

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

Soto, Paul E, "Measurement and effects of supply chain bottlenecks using natural language processing," 2023.

Yang, Yucheng, Yue Pang, Guanhua Huang et al., "The knowledge graph for macroeconomic analysis with alternative big data," *arXiv preprint arXiv:2010.05172*, 2020.

Zavodny, Madeline and Donna K Ginther, "Does the Beige Book move financial markets?," *Southern Economic Journal*, 2005, *72* (1), 138–151.

A Online Appendix

A.1 Categories of Economic Concepts

Table A.1: List of categories for economic concepts

Variable Type	Categories for Economic Concepts
Economic Variable	good demand (*consumer demand*), good price (*price*), good sales (*consumption*), good profit margins (*profit*), good market general condition (*condition*), labor demand, labor shortage, labor wage/compensation (*wage*), employment level, labor market general condition (*labor market condition*), good production level (*production*), good inventory level (*inventory*), raw material and commodity price (*input price*), raw material and commodity input amount (*input quantity*), other input price for production (*other input price*), other input amount for production (*other input quantity*), general investment, equipment investment, technology investment
Economic Event	fiscal income side, fiscal expenditure side, trade import side, trade export side, monetary policy, pent-up demand, bad weather, political event, pandemic, government management, price gouging, other event, loan demand, loan supply, loan quantity, loan rate, loan risk, deposit demand, deposit supply, deposit quantity, deposit rate, deposit risk, stock quantity, stock price, stock risk, bond quantity, bond price, bond risk, US dollar quantity, US dollar exchange rate, US dollar risk, other currency quantity, other currency exchange rate, other currency risk, other: it is a reasonable economic concept, but no category (including other event) works for it.

A.2 Extra word clouds

Figure A.1: Word clouds for most frequent 25 economic concepts

A.3 Example Logic Chains

Here we show two examples of logic chains extracted as a higher input price leads to more production.

Example 1: Production substitution

> [High natural gas prices]{V7} have pushed up [fertilizer and irrigation costs]{V8}, leading contacts to expect farmers in the Texas Panhandle to switch away from [corn production]{V9} to crops that require less water, such as [cotton]{V10}, [sorghum]{V11} or [sunflowers]{V12}.

The logic chain extracted involves *[High natural gas prices]{V7}* leads to more production in *[cotton]{V10}, [sorghum]{V11} or [sunflowers]{V12}* and lower production of *[corn pro-*

duction]{*V9*}. This creates three logic chains that indicate high input price will lead to high production.

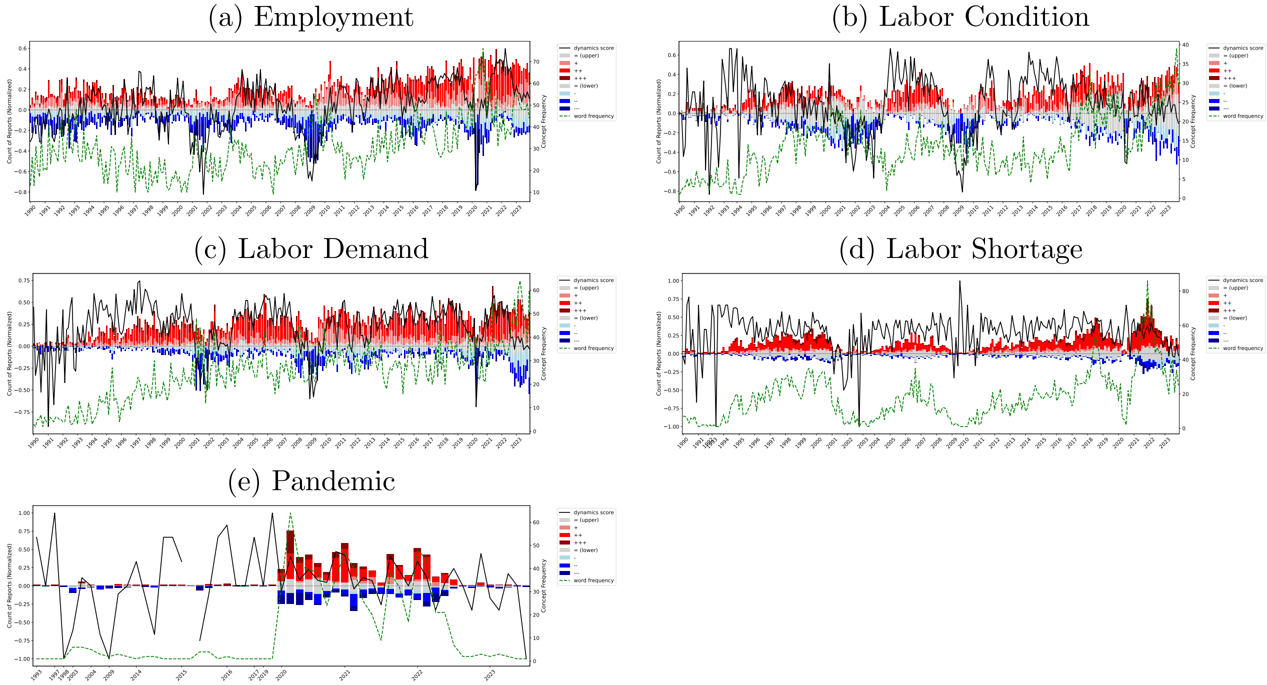
Example 2: Higher holding cost leads to early production

> [Continuing high feed prices]{V4} and [drought conditions in the Southwest]{V5} are leading [cattle ranchers to ship large quantities of beef to market earlier than preferred]{V6}.

The logic chain extracted involves *[Continuing high feed prices]*{*V4*} leads to more production in *[large quantities of beef]*{*V6*}. This creates a logic chain that indicates high input price will lead to high production. This problem is only prominent in Agriculture sector, as input costs there sometimes involve maintaining cost, the increase of which may lead to early production.

A.4 Extra Dynamic Scores

Figure A.2: Dynamics Scores of Economic Concepts



(a) Employment

(b) Labor Condition

(c) Labor Demand

(d) Labor Shortage

(e) Pandemic