Measuring Shortages Since 1900*

Dario Caldara Matteo Iacoviello David Yu

June 12, 2024

VERY PRELIMINARY DRAFT

Abstract

This paper introduces a monthly shortage index for the United States from 1900 to 2023, constructed using a sample of 25 million newspaper articles. The index measures the intensity of shortages of labor, materials, goods, and energy by calculating the proportion of articles discussing shortages each month. The resulting index reveals significant variation in shortage intensity over time, with notable peaks during periods of economic turmoil and wars. We explore the relationship between the shortage index and key economic indicators, discussing potential applications for researchers, policymakers, and businesses. Increases in shortages that are orthogonal to standard demand, supply and commodity price shocks generate persistent inflationary effects.

KEYWORDS: Shortages; Textual Analysis.

JEL CLASSIFICATION: C43, E32, N11, N12.

*Corresponding author: Matteo Iacoviello (matteo.iacoviello@frb.gov). All authors are affiliated with the Federal Reserve Board. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of anyone else associated with the Federal Reserve System.

1 Introduction

Shortages, defined as a lack of sufficient supply to meet demand, have been a recurring feature of economic life throughout the 20th and early 21st centuries. The hallmark of economics is the study of how to allocate scarce resources. It is therefore to be expected that shortages could have significant impacts on consumers, businesses, and the overall functioning of the economy. Yet, despite their importance, there has been limited research on the long-term trends and patterns of shortages across various sectors.

In the first part of this paper, we aim to fill this gap by constructing a monthly shortage index for the United States, using newspaper articles from 1900 to 2023. The shortage index is a news-based indicator of the intensity of shortages of labor, materials, goods, and energy in the United States. To construct the index, we use news articles from six major U.S. newspapers, amounting to about 20,000 articles per month (approximately 25 million articles over the entire sample). The index covers a wide swath of history (both domestic and global events), and is much higher during periods of increased economic turmoil, such as the World Wars and the 1970s oil crises. It spikes considerably during the COVID-19 pandemic, reaching its highest level in the last 40 years, however, there are some other past spikes of comparable or larger size. We validate that our index is plausible with a narrative investigation and an AI-based text analysis.

In the second part of this paper, we show that shortages lead to persistently higher inflation and lower economic activity, using both predictive regressions and a structural vector autoregressive (VAR) model.

We start with a series of predictive regressions in which we estimate the effect of shortages on measures of inflation and real growth. We find that shortages place downward pressure on GDP and upward pressure on prices, particularly in durable consumption and private fixed investment. Our results are robust to the inclusion of variables that are associated with inflation—for instance commodity prices and wages—indicating that our shortage index provides additional information not captured by traditional indicators.

To examine the variation in the effects of shortages over time, we estimate our regressions on a rolling sample. We find that shortages have consistently contributed to higher inflation over time. The effects of shortages on economic growth has been modest but negative for most of the sample, with two periods of higher growth associated with shortages: World War II and the COVID-19 pandemic. We interpret these results as a sign that both supply and demand forces are relevant for analyzing shortages.

To further investigate the interplay between shortages and the business cycle, we estimate a structural vector autoregressive (VAR) model. In the model, we identify "traditional" demand and supply shocks, and shocks to shortages capturing, for instance, atypical market adjustment in response to sudden shifts in economic conditions, to regulation—such as mandated price ceilings or quantity rationing—, or weather and geopolitical events that impede the regular flow of goods in an economy. The advantage of a fully identified model compared to the predictive regressions is that it leads to an interpretation of movements in GDP, inflation, and shortages that is comprehensive and economically palatable.

We find that not all shortages are created equal. An important part of the shortages in the 1950s are driven by demand shocks, associated with the Korean war, pent-up demand after World War II, and the rapid economic recovery that accompanied the transitioning from a wartime to a peacetime economy. The oil embargo of OPEC producers caused an unusual rise in shortages in the 1970s, over and above the direct effect of commodity price shocks. During the Covid pandemic, shortages emerged as a confluence of demand, supply, and shortage shocks, causing a persistent increase in inflation.

Our approach builds on previous work using news-based measures to track economic phenomena, such as the Economic Policy Uncertainty Index (Baker et al., 2016) and the Geopolitical Risk Index (Caldara and Iacoviello, 2022). However, to our knowledge, this is the first attempt to create a comprehensive shortage index for the United States spanning over a century. By examining the evolution of the index over a long period of time in conjunction with other economic indicators, we can gain insights into the causes and consequences of shortages and inform policy responses.

A number of other studies have also used news sources to construct indicators of shortages such as ours. As early as 1997, Lamont (1997) created a hand-coded news-based indicator of shortages using Wall Street Journal headlines. This was followed by additional attempts to measure shortages or related concepts via news-based indices (Chen and Houle (2023) for Canada, Burriel et al. (2023) for several advanced economies), as well as a more direct measure of supply chain pressures based on a variety of factors such as transportation costs (Benigno et al., 2022). Recently, Pitschner (2022) and Bernanke and Blanchard (2023) have tackled the intersection of shortages and the behavior of inflation during COVID-19 pandemic. In particular, Bernanke and Blanchard (2023), who use Google Trends to identify shortages, lay down a model which seeks to explain the causes of pandemic-era inflation, and contend that shortages have a "strong but temporary effect" on inflation.

Relative to this literature, we make two main contributions. Our index is the first comprehensive measure of shortages spanning over 125 years, covering periods as diverse as the World Wars, the oil shocks of the 1970s and the 1980s, and the COVID-19 pandemic. Furthermore, we use both univariate regressions and a structural VAR analysis and find that shortages have persistent effects on inflation, more so than documented by some studies. Finally, we also find that while shortages have generally reflected a mix of demand and supply shocks, the realized effects have been more in line with those of supply shocks.

The remainder of the paper is structured as follows. Section 2 discusses the construction of the index, presents it, and discusses its evolution over time. Section 3 validates the index. Sections 4 and 5 explore the relationship between the shortage index and economic activity and inflation in the United States, using predictive regressions and a VAR, respectively. Section 6 concludes and discusses potential applications and future research directions.

2 The Shortage Index

In this section, we first discuss the construction of the shortage index. We then show how the index captures key episodes of shortages in U.S. history.

2.1 Construction of the Index

The shortage index is a monthly news-based indicator of the intensity of shortages of materials, goods, labor, and energy in the United States. It is constructed using a sample of about 20,000 news articles per month from 1900 through the end of 2023—for a total of about 25 million articles—published in the *Boston Globe*, the *Chicago Tribune*, the *Los Angeles Times*, *The New York Times*, *The Wall Street Journal*, and *The Washington Post*. Each month, the index counts the number of articles discussing energy, food, industry or labor shortages—the set $\mathcal{S}$ depicted in Figure 1—normalized by the total number of articles, the set $\mathcal{A}$. Higher values of the index indicate higher intensity of shortages. In what follows, we discuss the steps that led to the search query used to isolate $\mathcal{S}$.

We start by constructing the set of articles $\mathcal{B}$, represented by the pink area in Figure 1. The articles in this set contain at least one mention of a shortage word—namely 'shortage', 'scarcity', 'bottleneck', or 'rationing'—in conjunction with one or more economics words—such as the words 'economy', 'market', or 'commerce'.[1] The 'shortage' words above are those more frequently associated with economically-relevant constraints to the production capacity of a country or to the availability of goods to consumers. The inclusion of at least one economics-related word in the search reduces the likelihood of false positives, that is, articles which mention shortage words but are unrelated to the economic phenomenon that we intend to measure.[2]

We then draw from $\mathcal{B}$ a random sample of just over 3,300 articles, which we use to construct a list of the 1,000 most frequent collocates within five words of the 'shortage' words above. We select from this list words highlighting shortages in particular sectors of the economy or of particular goods. The most common words in the list (excluding stopwords) are oil, water, war, time, coal, days, food, cars, people, government, million, labor, state, home, steel, and fuel. Then, we remove low-information words such as 'time' or 'days', as well as other words that convey little content about the topic at hand, such as 'people' or 'government'. The resulting list of collocates constitute our set of topic words that, for ease of exposition, we group into four categories: food, industry, labor, and energy.

With these three lists of words (shortage words, economics words, and topic words), we construct the search query shown in Table 1. An article contributes to the shortages index—the set $\mathcal{S}$ in Figure 1—when two conditions are met: first, a shortage word appears within five words ($N/5$) of a topic word; second, the article contains at least one economics word. If an article meets these two conditions and contains topic words from two categories, it is counted twice in the index. Thus, the number of shortage articles is then the sum of the articles in each of the four categories in that period. This definition allows us to weight articles that discuss multiple types of shortages more heavily in the overall index. In Section 3, we show that requiring shortage words to appear in close proximity to topic words is a critical part of our methodology. This step reduces the number of false positives and improves the accuracy of our search.

The classification into four topics is supported by the Latent Dirichlet Allocation (LDA)

[1] The full list of economics words are in Table 1.

[2] Other potential synonyms of shortage, such as "lack" or "paucity" or "insufficiency," either have a wider range of meanings, or are less likely to be exclusively associated with economic shortages.

analysis that we perform ex-post on a sample of articles satisfying our criterion for inclusion in the index. LDA is a popular unsupervised machine learning technique used for topic modeling in natural language processing, and is designed to discover hidden topics within a collection of documents by analyzing the co-occurrence of words in these documents. There are two inputs to the algorithm. The first is the corpus of documents of text to be analyzed. Here, the corpus is a random sample of 13,623 abstracts of newspaper articles mentioning shortages and satisfying our criterion for inclusion in the index. This sample is about 4 percent of the approximately 330,000 articles that are included in our index from 1900 to 2023.[3] The second input is the number of clusters, or topics, which we set to four.

The results of the LDA analysis are illustrated and summarized in Figure 2. The most recurrent words for each topic are presented as word clouds in the panels to the left. Topic 1 focuses on energy, topic 2 on water, food and agricultural products, topic 3 on industrial products such as coal, steel, railroads, cars, and topic 4 on jobs. The stacked bar chart to the right visualizes the topic mixtures for each abstract, for a total of 13,623 bars, sorted by year. Each article straddles different topics. Early in the sample, industry-related and, to a lesser extent, food and water shortages dominate the conversation. Energy-related shortages are the most common topic throughout most of the 1970s. Labor-related shortages become more prevalent in the post-pandemic years.

2.2 Shortages In History

We now present the shortage index, examine spikes, and consider the historical context in which the spikes occurred. Figure 3 plots the shortage index at a monthly frequency from 1900 through the present. The index is calculated by taking the monthly share of articles discussing economic shortages ($\mathcal{S}/\mathcal{A}$) and indexing it to an average value of 100 over the period 1900-2023.[4] In Table 2, we list the thirty largest spikes in the index, accompanied by a description of the key events surrounding each episode.[5]

[3] Following standard practice, before using the words in the LDA analysis we remove stopwords and numbers. We also stem the words reducing them to a common root.

[4] Specifically, we divide the share by its mean over the period 1900-2023 and then multiply by 100.

[5] To calculate the spikes listed in the table, we first extract the residuals from a regression of the shortage index h_t against its value two months before and two months after. The largest residuals form our list of 'candidate' spikes. Second, we drop a candidate spike if it occurs in a month t in which the index itself is not a local maximum over the 13-month period $[t-6, t+6]$. The second step ensures that the identified spikes are not just large, but also (locally) peak values. Of the remaining observations, the ones with the thirty largest residual values are reported in Table 2.

Over its long history, the index exhibits considerable variation, and the episodes behind the spikes in the shortage index are primarily related to three themes: geopolitical events, labor shortages and energy shocks. First, adverse geopolitical events, especially wars, are associated with severe shortages. Most notably, the index rises dramatically during World War I. It surges during World War II, reaching a value above 1,000 (10 times larger than the sample mean) in January 1943. Second, many spikes are associated with labor shortages, especially those caused by strikes. These tend to occur towards the beginning of our sample (for example, coal-related strikes in 1903, 1919, and 1922). Strike-related shortages are less frequent in more recent years. Third, post-World War II spikes are often associated with energy shortages, especially many arising in response to geopolitical turmoil in the Middle East. Oil supply issues crop up repeatedly, occurring at the same time as jumps in the index during the Suez Crisis, the oil shocks of the 1970s—associated with the only other reading of the index above 1,000—, and around the Iraqi invasion of Kuwait in 1990.

Figure 4 breaks down the shortage index by topic. This decomposition illustrates that the aforementioned jumps in the index largely reflect a spike in the energy component of our index, consistent with the oil events of the time. Fourth, the index increases multiple times around the COVID-19 pandemic. The first spike corresponds to shortages of medical equipment and nursing staff at the onset of the pandemic. The second spike is larger and occurs at the beginning of 2022 when global supply bottlenecks emerged, as many economies reopened following prolonged mobility restrictions in 2020 and 2021. This second spike, as seen in Figure 4, is mostly composed of increases in the labor and industry components.

Finally, we compare our index to a 'global' counterpart. The global counterpart includes articles which meet the requirements to be included in the original index, but also requires that an international word (such as the name of a country, major city, or related term) appears in the article. Accordingly, the global component places weight either on foreign events, or on global events with U.S. repercussions. Figure 5 plots both the original index and the component accounted for by global counterpart. Both measures move in lockstep, with peak values broadly occurring around the same time.

3 Assessing the Accuracy of the Shortage Index

We conduct two separate exercises to assess the accuracy of the index. In the first, we verify that the newspaper articles included in the index indeed mention concerns about shortages. In the second, we verify that the index is aligned with alternative measures of or proxies for shortages for the limited time periods in which these alternatives exist and overlap with ours.

3.1 Validation of the Shortage Index

We verify that our index accurately measures shortages by minimizing Type-I and Type-II error. We sample the abstract of 872 articles belonging to the shortage set $\mathcal{S}$.[6] By construction, each of these articles contains at least one business-related word as well as one mention of scarcities, shortages or bottlenecks in proximity of a topic word such as energy, food, industry and labor. For each article, we extract the first snippet of text that contains references to shortages. We center the snippet around the shortage word, and set its length to 110 characters, drawing inspiration from Twitter's original 140-character limit, and striking a balance between brevity and computational and cognitive burden. For instance, examples of recent and old snippets in our sample are respectively "[A]lthough demand remains strong. ... the resulting supply shortage of german manufacturing goods could also...", from 2021; and "[...] men interested in the industries affected by shortage of steel are anxious to see the strike settled," from 1901.

We then use the Claude AI assistant (Anthropic, 2024) to determine whether each snippet mentions current or prospective shortages, rationing, scarcity, or bottlenecks related to goods, labor, materials, food, or water. Claude was instructed to return a table of results coding articles either as 1—shortage mentioned—, or zero—shortage not mentioned—or 99—unsure whether the existence of shortages was mentioned. In addition to the classification, Claude was asked to provide a brief explanation for the coding of each snippet.

Before the classification, we provided Claude with some examples of how we would code the snippets, and made sure that the training sample included false positives, mentioning, for

[6] Our search query to calculate the index relies on the six U.S. newspapers listed in Section 2. For technical reasons, abstracts are not available for these newspapers since 2015. For the period 2015-2023, we sample the abstract we used for the validation from a larger set of newspapers, including some U.K., Canadian, and Australian ones. A detailed list of the 872 abstracts and sources used for validation is available upon request from the authors. Abstract are short portions of text often containing the initial sentences or the initial paragraph of the entire article.

instance, lack or end of shortages.[7] Use of AI for validation is not foolproof, but we found Claude did as good a job as a human, for instance by extrapolating the context of a particular sentence to a particular country or person. For instance, for the sentence "economy may be slowing but Lowe is banking on labour shortages gradually leading to an increase...", Claude classified the text as 1 and added that "Reserve Bank [of Australia] expecting labour shortages to lead to wage growth."

The results of the audit are in Table 3. Out of 872 articles belonging to the set $\mathcal{S}$, only 6.3 percent were deemed by ClaudeAI as false positives.[8]

We then repeat the audit for a sample of 298 articles not belonging to the set $\mathcal{S}$. Out of these 298 articles, only one appears to mention shortages but is not captured by our search query ("recycling of newsprint was held back by a shortage of deinking plants"). Of note, our search query deliberately did not include the word "plants" since in preliminary attempts we found instances of false positives associated with this word.[9]

Finally, we confirm that restricting the search to include shortage words in proximity of words indicating goods, labor, food or energy improves the accuracy of the search. The share of articles mentioning economically-relevant shortages in the set that allows for, but does not require the presence of shortage words is 84.2 percent, corresponding to a Type I error of 15.8 percent, much larger than in our preferred search query. False positives included in this set—that are not captured by our preferred search query—include articles mentioning shortages of political campaign funds, lack of good baseball photos, legislative bottlenecks, and shortage of sunshine.

[7] The specific prompt was as follows. "I give you 872 snippets of text each about 110 characters long. For each of them, can you tell me whether they mention current or perspective shortages or rationing or scarcity or bottlenecks related to goods, labor, materials, food, water? Just return a table with yes=1, no=0, unsure=99, and a brief explanation. For instance. Article 1 mentions perspective shortages since it mentions that steel shortages will prevail in the near future, so it is a 1. Article 2 says steel shortages caused a plant closure, so it is coded 1. Article 3 says shortage of cars is crimping coal production, so 1. Article 4 mention shortage of cars, so 1. Article 10 mention shortage of workspace, so not really a work shortage, so 0. Article 329 says no shortage of cars has been experienced, so 0."

[8] For instance, Claude classified the snippet "...says that while there is not necessarily a shortage of people wanting to work in management" as 0 with the explanation "No shortage of people wanting to work in management." Similarly, the snippet "a motive for mr. newt gingrich's knife job, had no shortage of conspiracy theories, most leading to the..." was classified as 0 with the explanation "Speaker's ouster sparked many conspiracy theories but not actual shortages." In some cases, Claude classified as 0 articles that we would have probably classified as 1. For instance, the snippet "canada's action today in temporarily suspending meat rationing" was classified as 0 since Claude gave more weight to the temporary suspension of the rationing rather than its existence.

[9] See for instance the article "Brighten Up Indoors With Colorful Plants" (Los Angeles Times, Feb. 4, 1996) that states "there's no shortage of plants with brightly colored foliage to liven up your kitchen."

3.2 Comparison with Other Indicators of Supply Constraints

In this section, we compare our index with other, related measures of supply constraints.

Figure 6 plots our shortage index alongside the New York Fed Global Supply Chain Pressure (GSCPI) Index, the Supplier Delivery Index (SDI), and the Burriel et al. (2023)'s Supply Bottlenecks Index (SBI) for the US. To enhance the comparison, we standardize each of the variables to have mean 0 and variance 1 over the period in which they overlap. While all these indexes contain useful information to analyze the extent of broader supply chain pressures, an advantage of our index relative to these indicators is that our index is available over a much longer period of time, thus being particularly amenable to historical applications for research.[10]

Figure 6 shows that our index shares similar features to these three indicators. Over the entire existence of the GSCPI, the correlation between our index and the GSCPI is 0.73. Both measures increase sharply at the onset of COVID-19 in early 2020, and again in the beginning of 2022 as supply chain bottlenecks took hold.

There is a lower correlation between our index and the SDI. This comparison begins in 1976, when the SDI starts. Our measure has a correlation of 0.25 with the SDI. Both measures spike around the 1979 oil crisis and COVID-19. One possible explanation is that both events brought about severe delays in transportation (via fuel costs and supply bottlenecks, respectively), which may have led to shortages for manufacturers captured by the rapid increases in both indices. Finally, our index has a very high correlation (0.90) with the U.S. SBI.

These high correlations suggest that our index is capturing movements that are similar to those of other indicators, although most of the correlation is obviously driven by the extraordinary rise in supply chain pressures as other related supply chain variables.

4 Shortages as Predictors of Inflation and Activity

Figure 7 illustrates the positive relationship between U.S. inflation and our indicator of shortages for a sample starting in 1940. In this section, we formally explore the relationship

[10] The GSCPI is published by the Federal Reserve Bank of New York and is designed to measure supply chain conditions around the world, using data on both manufacturing and transportation costs. The SDI is published by the Institute of Supply Management (ISM), and captures the result of a monthly survey which asks firms whether they are experiencing longer or shorter wait times compared to the previous month. The value of the SDI is the share of respondents reporting longer wait times plus half the share of respondents reporting no change. The SBI is produced by Burriel et al. (2023) and uses a text-based newspaper search to quantify the level of supply chain issues.

between shortages, near-term inflation, and economic activity. Specifically, we estimate the following predictive regression:

$$\Delta Y_{t+h} = \alpha + \beta SHORTAGE_t + \sum_{i=0}^{p} \mathbf{X}_{t-p} + \varepsilon_{t+h}, \tag{1}$$

where $\Delta Y_{t+h} = \frac{400}{h} \ln(\frac{Y_{t+h}}{Y_t})$ is the annualized log change of a variable of interest Y_t between period t and forecast horizon h and $SHORTAGE_t$ denotes the level of the shortage index. $\mathbf{X}$ is a vector of control variables.

We use quarterly data from 1950 through 2023 from the National Income and Product Accounts (NIPA) for real per-capita GDP, personal consumption expenditure, and private fixed investment. We measure inflation for each category using the associated price deflators. For GDP and its price deflator, we extend the sample back to 1900 using data from Ramey and Zubairy (2018). Data on total population are also from Ramey and Zubairy (2018), which we extend through 2023 using the POP series from FRED.

For each price and economic activity indicator, we estimate regression (1) separately by OLS. As control variables, we include quarterly changes of the dependent variable and the associated economic indicator or price deflator, contemporaneously and with three lags. For instance, the predictive regression for real GDP growth includes contemporaneous and lagged values of both real GDP growth and GDP inflation.

We tabulate results for the one-year ahead regressions in Table 4, while results for the one- and eight-quarter horizons are reported in Table A.1 and Table A.2, respectively. To facilitate comparison across variables, we report standardized estimates of the coefficients β. A standardized coefficient represents the movement of the dependent variable (in standard deviation units) in response to a one standard deviation change in the explanatory variable.

The first two columns of Table 4 report estimates for the full sample. An increase in shortages is associated with a rise in inflation (first column) and a decline in economic activity (second column), the typical effects of supply-side disruptions. The inflationary effects of shortages are evenly distributed across GDP components, with the price of services consumption being the least impacted. While consumption of durable goods and private fixed investment decline, there is no statistically significant effect on the consumption of nondurable goods and of services.

To quantify the economic effects of shortages, take for instance the standardized coefficients of 0.25 and -0.25 estimated for durable consumption inflation and real growth, respectively. A one-standard deviation increase in the shortage index is associated with an increase in durable goods inflation of 0.75 percentage points and a decline in durable goods consumption growth of 1.75 percentage points. The coefficients for GDP imply an increase in inflation of 0.5 percentage point and a reduction in growth of 0.3 percentage points.

The remaining columns of Table 4 show that the results for the full sample hide notable time variation in the relationship between shortages, inflation, and economic growth. In a sample running from 1950 through 2014, we find that the effects of shortages on inflation and activity are more precisely estimated. In addition, the reduction in economic growth is widespread, sparing only private consumption of services. In a sample starting in 2015 and encompassing the COVID-19 pandemic, the effects on inflation are substantially larger, albeit less precisely estimated. The coefficients on economic growth become *positive*, although mostly not statistically significant. We interpret the estimates as evidence of a more prominent role of demand forces in driving shortages during and in the aftermath of the COVID-19 pandemic compared to the pre-COVID sample.

The estimates for the sample starting in 2015 are consistent with the U.S. experience since the start of the pandemic. Strong demand and weak supply both contribute to the emergence of shortages. Both demand and supply forces push prices in the same direction—and have contributed to the rise of inflation—while having opposite effects on economic activity, thus contributing to its resilience.

A potential concern with these results is whether shortages can provide information beyond what is already embedded in other macroeconomic variables. For example, oil prices may move similarly to the energy component of our index. In Table 5, we test this possibility by adding in several potential variables: oil prices, commodity prices, wage growth, and inflation expectations. The top row of Table 5 reports the same baseline GDP estimate seen in the top row of Table 4. Each of the next rows adds the named variable to the baseline as an additional control.[11] We find that our shortage index is informative, as its coefficient typically remains significant. Columns 1 and 2 show that the effect of shortages is somewhat attenuated but still similar in magnitude to the effects seen in Table 4. In the next four columns, we again partition the sample into two periods. As in Table 4, the pre-COVID effects are stable and highly significant. The effects in the later part of the sample are less significant but remain

[11] For details, refer to the notes for Table 5.

largely in line with the baseline results. Overall, these results suggest that our index encodes additional information over and above what is in these variables.

One might wonder whether the COVID-19 experience has been unique, or whether other historical episodes of shortages were also associated with a role for private demand. We investigate this question running regression (1) on a rolling sample using a 30-year window. We focus on real GDP growth and inflation to leverage data prior to World War II. Figure 8 shows the results for inflation (top panel) and real GDP growth (bottom panel).

Starting with World War II, shortages have consistently exerted inflationary effects. The 10 years prior to the pandemic are a notable exception, due to the fact that in the three decades prior to the pandemic, the U.S. did not experience episodes of severe shortages against a background of remarkable stable inflation.

Shortages have exerted economically meaningful and statistically significant adverse effects on economic activity starting in the 1970s up until the COVID-19 pandemic. In contrast, an increase in our index is associated with higher activity around just two historical episodes: the pandemic and World War II. Rationing, which was commonplace during the war, implies that demand outstripped supply. Thus both episodes, through their own unique circumstances, were times in which high demand propped up economic activity. Furthermore, the imposition of price caps during the war may have contributed to keeping inflation more muted than expected.

In the next section, we take a different approach in investigating how shortages interact with demand and supply factors, estimating a structural VAR model.

5 Shortages and Activity: A VAR Analysis

In this section, we present a simple structural VAR model of the U.S. economy that incorporates shortages. The goal is to decompose movements in shortages and other macroeconomic variables into mutually orthogonal components that have an economic interpretation, and to look at the contribution of shortages to inflation and business cycles during the pandemic and in other historical episodes. An additional goal is to provide insights into what forces drive shortages over time.

5.1 The Model

Consider a model based on quarterly data for economic activity, inflation, commodity prices, and shortages, summarized by the vector $\mathbf{X}_t = (y_t, \pi_t, c_t, h_t)'$. Economic activity is measured by four-quarter change in log GDP, inflation is measured by the four-quarter change in the log of the GDP deflator, commodity prices are measured by the four-quarter change in the Reuters–CRB Spot Commodity Price Index for raw industrials and foodstuffs, and shortages are expressed in levels (and standardized). All series are demeaned.

The economy has a vector autoregressive representation given by:

$$\mathbf{A}\mathbf{X}_t = \sum_{j=1}^{p} \boldsymbol{\alpha}_j \mathbf{X}_{t-j} + \mathbf{B}\mathbf{u}_t, \tag{2}$$

where $\mathbf{u}_t = (u^S, u^D, u^C, u^H)'$ is a vector of structural shocks described below, p is the lag length, and $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{\alpha}_j$ for $j = 1, ..., p$ are matrices of structural parameters. The structural shocks have zero mean and variance–covariance matrix $E[\mathbf{u}_t \mathbf{u}_t'] = \boldsymbol{\Sigma}_u$. Without loss of generality, we normalize one entry on each row of $\mathbf{A}$ to 1 and we assume that $\boldsymbol{\Sigma}_u$ is a diagonal matrix.

Abstracting from lagged terms, the following equations describe the joint modeling of shortages and economic activity, and summarize the contemporaneous restrictions that we impose on the parameters in the matrices $\mathbf{A}$ and $\mathbf{B}$:

$$\pi = \kappa y + u^S, \tag{3}$$

$$y = -\delta \pi + u^D, \tag{4}$$

$$c = u^C + \phi_D u^D + \phi_S u^S, \tag{5}$$

$$h = u^H + \theta_S u^S + \theta_D u^D + \theta_C u^C. \tag{6}$$

Equation (3) is an aggregate (inverse) supply equation: inflation is positively related to output ($\kappa > 0$), and subject to adverse supply shocks u^S. Equation (4) is an aggregate demand equation: it states that aggregate demand y is negatively related to inflation ($\delta > 0$), and subject to demand shocks u^D. Equation (5) describes the evolution of commodity prices, that are assumed to respond to demand and supply shocks within the quarter, and to be driven by commodity-market specific shocks. Taken together, equations (3) to (5) imply that the model does not allow shortage shocks to impact demand, supply or broader commodity markets within a quarter. Thus, the effects of shortages on activity and prices happen with at least a one-quarter delay, and depend only on the lagged feedback from shortages to GDP, inflation and commodity prices captured by the coefficients $\boldsymbol{\alpha}_j$ for $j = 1, ..., p$.

Equation (6) describes the relationship between shortages and the remaining block of the model. We assume that shortages reflect "regular" business cycle movements caused by supply and demand shocks, shocks in commodity markets, and an "exogenous" component that proxies for newsworthy disruptions to the regular flow of goods, services and factors of production in an economy. Under this interpretation, innovations to u^H stand in for any unusual combination of unpredictable events that cause demand to temporarily exceed supply.[12] Examples may include: atypical market adjustment in response to sudden shifts in economic conditions, for instance demand reallocation towards one sector to another causes bottlenecks that cannot be absorbed without some temporary rationing; a shock to regulation—such as mandated price ceilings or quantity rationing—that disrupts the regular functioning of a market economy; panic buying that suddenly causes rationing of certain goods when social norms prevent large price adjustment; extreme weather events or sudden geopolitical shocks that impede the smooth flow of goods in an economy. We assume that these "exogenous" shortages should have the same macroeconomic effects: ultimately, the specific nature and context of shortages shocks may produce varying effects on the economy, but we leave this issue for another time.[13]

The advantage of a fully identified model compared to a model that only identifies shocks to shortages—or compared to the predictive regressions discussed in Section 4—is that it leads to an interpretation of movements in GDP, inflation, commodity prices, and shortages that is comprehensive and economically palatable. While the model could be enriched with additional variables, we prefer its transparency and simplicity: essentially, shortage (news) shocks are every movement in shortages that cannot be explained by current and past activity, inflation and commodity prices.

The matrices summarizing the system are therefore:

[12] Our assumption is that shortages reflect within a quarter movements in economic conditions caused by traditional demand and supply shocks. This assumption differentiates our analysis of the effect of shortages on activity and inflation from the work of Burriel et al. (2023), who order shortages first in a monthly VAR from 1990 through 2020; and from the work of Bernanke and Blanchard (2023), who use a quarterly VAR from 1990 through 2023 and also assume that shortages affect inflation within a quarter.

[13] Kahneman et al. (1986) is a classic article describing how fairness considerations may lead to shortages and inefficient market functioning in the presence of unexpected shifts in demand or in supply.

$$\mathbf{A} = \begin{bmatrix} 1 & -\kappa & 0 & 0 \\ 1 & \delta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \phi_S & \phi_D & 1 & 0 \\ \theta_S & \theta_D & \theta_C & 1 \end{bmatrix}, \tag{7}$$

with the impact matrix relating variables at time t to structural shocks given by:

$$\mathbf{C} \equiv \mathbf{A}^{-1}\mathbf{B} = \begin{bmatrix} \frac{1}{1+\kappa\delta} & \frac{\kappa}{1+\kappa\delta} & 0 & 0 \\ \frac{-\delta}{1+\kappa\delta} & \frac{1}{1+\kappa\delta} & 0 & 0 \\ \phi_S & \phi_D & 1 & 0 \\ \theta_S & \theta_D & \theta_C & 1 \end{bmatrix}. \tag{8}$$

Following Baumeister and Hamilton (2019), we view identification and estimation of this structural VAR model as a special case of a Bayesian VAR with prior restrictions imposed on the parameters of the matrices $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{\alpha}_j$. To see why, consider the example in which one assumes a flat within-quarter Phillips curve, so that $\kappa = 0$. The matrix $\mathbf{C} = \mathbf{A}^{-1}\mathbf{B}$ relating variables to structural shocks becomes lower triangular and the VAR model has a familiar recursive interpretation with variables ordered as (y_t, π_t, c_t, h_t). However, rather than fixing some of the structural parameters of the model to exactly identify the shocks, we estimate directly the parameters of the matrices $\mathbf{A}$, $\mathbf{B}$ and $\boldsymbol{\alpha}_j$ by imposing prior distributions. In estimating the model, we set the number of lags at two.

Prior distributions for the key parameters are plotted in red in Figure 9. The priors for κ and δ are normally distributed with mean 0.5 and 1, respectively, and standard deviation 0.25 and 0.5. At the prior mean, the slope of the aggregate supply curve implies that in the short-run inflation rises 0.5 percentage points in response to a demand-driven, 1 percent increase in production. Additionally, the slope of the aggregate demand curve implies that on impact a 1 percent contraction in GDP raises inflation by 1 percentage point. We also assume that commodity prices increase in response to expansionary demand shocks and to contractionary supply shocks (that is, the support of ϕ_S and ϕ_D is positive).[14] The support of the prior densities is wide, encompassing substantially smaller and larger elasticity values. As discussed in the empirical results, the data and the model structure lead to revising both in

[14] We assume that ϕ_S and ϕ_D are drawn from an inverse gamma distribution with mean equal to 1 and standard deviation equal to 0.5.

the location and spread of these distributions.

We set the priors for θ_S, θ_D, θ_C—the parameters dictating how shortages response to macroeconomic disturbances—so that they are drawn from an inverse gamma distribution with mean 0.25 and standard deviation 0.5. This choice of priors embeds the assumption that shortages and associated news can reflect standard demand and supply shocks, as well as shocks to commodity markets, within a quarter. Of note, implicit in our choice of priors is the assumption that expansionary demand shocks are associated with an increase in shortages, that expansionary supply shocks are associated with a decrease in shortages, and that shocks that lead to higher commodity prices are associated with an increase in shortages (that is, the support of θ_S, θ_D and θ_C is positive).

Finally, the prior mean for the coefficients $\boldsymbol{\alpha}_j$ is set so that variables have a first-order autocorrelation of 0.8, and that all lagged indirect effects have a prior mean of zero.[15] In Figure 9 we show prior densities for the lagged coefficients of the shortage index on inflation.

5.2 Results

We estimate the model on quarterly data from 1950 through 2023 using standard Bayesian techniques. To construct the estimates of interest, we take 50,000 draws from the posterior distribution using a Random Walk Metropolis-Hastings algorithm.

Parameter Estimates

Figure 9 shows the prior distributions, in red, and the posterior distributions, in blue, for some of the key parameters of the model. The location and spread of all posterior distributions is substantially updated compared to the priors, revealing that the data and the structure of the model are informative about the structural parameters.

The posterior median for the parameter δ in the demand equation is around 2, which implies that a supply shock that reduces GDP growth by 1 percentage point leads to an increase in inflation of about 0.5 percentage point. The posterior mean for the parameter κ is

[15] We rewrite the VAR system as $\mathbf{X}_t = \mathbf{R_1}\mathbf{X}_{t-1} + \mathbf{R_2}\mathbf{X}_{t-2} + \mathbf{A^{-1}}\mathbf{B}\mathbf{u}_t$, where $\mathbf{R_1} = \mathbf{A^{-1}}\boldsymbol{\alpha}_1$ and $\mathbf{R_2} = \mathbf{A^{-1}}\boldsymbol{\alpha}_2$. The elements of lagged coefficient matrices $\mathbf{R_1}$ and $\mathbf{R_2}$ are assumed to be normally distributed. The elements of the main diagonal of $\mathbf{R_1}$ are assumed to have a prior mean of 0.8. The other elements of $\mathbf{R_1}$ and all the elements of $\mathbf{R_2}$ have zero prior mean. We assume tighter standard deviation the higher the lag order, as in the VAR priors discussed in Doan et al., 1984: we use 0.25 for the first lag and of 0.20 for the second lag. The prior on the standard deviation of the structural shocks is an inverse gamma with mean 1 and standard deviation 0.5.

0.22—which implies that a demand shock that raises GDP growth by 1 percentage point leads on impact to an increase in inflation of around 0.22 percentage point. All told, the posterior distributions reveal lower slopes both for the aggregate supply and for the aggregate demand curves.

Estimates for θ_D, θ_S and θ_C are consistent with the notion that shortages respond to the business cycle and to movements in commodity prices, albeit to a smaller degree than assumed by our prior. The posterior estimate for both parameters is positive but relatively small. The parameters ϕ_D and ϕ_S are estimated to be positive, meaning that commodity prices increase in response to inflationary demand and supply shocks.

The last two panels of Figure 9 illustrate the underlying force driving the inflationary effects of shortages in our model. After controlling for current and lagged movements in inflation and GDP, the posterior density measuring the effect of shortages in period t on inflation in period $t+2$ is positive, indicating a high likelihood of lagged shortages having a positive effect on inflation.

We conclude this subsection with a brief discussion on the importance of directly estimating the structural form of the model by imposing restrictions through prior distributions. Consider two alternative versions of the model that are closer to the recursive identification schemes of a VAR, with parameters estimated without imposing informative prior restrictions on their magnitude or sign.

In the first version of the model, we assume that short-run supply curve is almost vertical, by fixing $\kappa = 10$. The resulting estimate of δ is large, very close to 10, so that the aggregate demand curve is very flat. In this alternative model, contractionary supply shocks generate large declines in output and negligible increases in inflation. By contrast, expansionary demand shocks generate large increases in inflation and negligible movements in output. Additionally, inflationary demand shocks lead to a counterfactual decline in output after one year. Thus, both demand and supply shocks generate a negative comovement between output and inflation. Even if the effect of shortage shocks is similar to that of the baseline version, this model has the unpalatable property that no shock creates a meaningful Phillips curve, that is, positive conditional comovement between output and inflation.

In the second alternative version, which assumes a flat Phillips curve by setting $\kappa = 0$, the resulting estimate of δ is -0.31, implying a counter-factually upward sloping aggregate demand curve. As a result, both estimated demand and supply shocks generate positive comovement between output and inflation, so that no primitive economic shock—aside from the shortage shock—can move output and inflation in opposite directions.

Impulse Responses

Figure 10 reports the impulse responses to the estimated shocks in the model. The solid line depicts the posterior mean, while the shaded areas represent the 80 percent posterior credible sets.

The impulse response analysis shows that shortages can arise as an endogenous response to business cycles. However, not all news about shortages can be explained with reference to traditional demand, supply, and commodity price shocks. As shown in panel (a), an exogenous increase in shortages leads to a small but negligible decline in GDP and to a rise in inflation, with inflation peaking two years after the shock and remaining significant for a over five years. However, an important part of movements in shortages also owes to demand and supply forces. An expansion in aggregate demand and a contraction in aggregate supply, shown in panels (b) and (c), are both associated with higher shortages. A commodity price shock, of the kind seen in the 1970s and the 1980s, also leads to higher shortages, as shown in panel (d).

The comparison across shortages, demand and supply shocks illustrates how shortages shocks combine both elements of higher demand and of lower supply, so that shortages shocks are not just a traditional cost-push shock that decreases output while increasing inflation. Of note, the increase in inflation following a shortage shock is larger and more persistent than would be predicted by a "regular" expansionary demand shock. Additionally, the decline in output is smaller than would be predicted by a "regular" contractionary supply shock. The negligible effects on output of shortage shocks differentiate our findings from those of Burriel et al. (2023). Comparing our results to theirs, our shortage shock has smaller negative effects on activity for a given increase in inflation, resulting in a smaller sacrifice ratio. This happens because the timing assumptions of our VAR model attribute some of the countercyclical movements in shortages to adverse supply shocks.

With all shocks being scaled to take value of one standard deviation, we can compare their relative importance for output and inflation. In particular, shortage shocks exert only a moderate effect on GDP compared to standard demand and supply shocks. By contrast, their effect on inflation is about half compared to traditional demand and supply shifters.

The Effects of Shortages throughout History

Figure 11 shows the historical decomposition of the model. A large part of the shortages in the 1950s are driven by demand shocks, associated with the Korean war, pent-up demand after World War II, and the rapid economic recovery that accompanied the transitioning from a wartime to a peacetime economy. The oil embargo of OPEC producers caused an unusual rise in shortages in the 1970s, over and above the direct effect of commodity price shocks.

Figure 12 reports the role of shortage shocks between 2020 and 2023. The COVID and post-COVID years can be separated into three distinct periods. The first period, which runs from 2020:Q1 through 2021:Q1, witnessed a simultaneous reduction in demand and tighter supply conditions. Shortages, primarily driven by lack of workers and supply chain issues, started to build up, but were offset by weak demand. The second period runs from the second quarter of 2021 through the third quarter of 2022 and is characterized by unusual shortage shocks stemming from from supply chains strains and tight labor market conditions, in a context of strong aggregate demand and rising commodity prices. Throughout this period, shortage shocks start exerting upward pressure on inflation. The third period runs from 2022:Q4 through the end of sample in 2023:Q4 and is characterized by a slow decline in shortages, which nonetheless remain above their pre-pandemic level. Importantly, as the estimated effect of shortage shocks on inflation is delayed and long-lived, shortages continue to be a prominent driver of inflation all the way through the end of 2023.

6 Conclusions

This paper introduces a new monthly, newspaper-based shortage index for the United States spanning more than a century, from 1900 through present. The index captures the intensity of shortages across various sectors of the economy, including labor, materials, goods, and energy. The index exhibits significant spikes during periods of heightened economic turmoil, such as the World Wars, the oil crises of the 1970s, and the COVID-19 pandemic.

Validation exercises confirm the accuracy of the index in measuring shortages and reveal strong correlations with other indicators of supply constraints. Notably, our index covers a substantially longer time horizon than existing alternatives, making it a valuable tool for historical analysis.

Predictive regressions demonstrate that increases in the shortage index are associated with persistently higher inflation and lower economic activity. These effects are particularly pronounced for durable goods consumption and private investment. Furthermore, the inflationary impact of shortages appears to be much stronger in the post-2015 period, which includes the COVID-19 pandemic.

The paper has also presented structural VAR analysis that decomposes shortage movements into the endogenous response to traditional business cycle shocks, on the one hand, and exogenous shortage shocks, on the other. The results indicate that shortage shocks exhibit characteristics of both higher demand and lower supply. Historically, the 1950s saw shortage movements primarily driven by demand factors, while the 1970s experienced unusual shortage shocks stemming from the OPEC oil embargo, beyond the direct effects of commodity prices. During the COVID-19 pandemic, shortages emerged as a combination of supply, demand, and exogenous factors, leading to a persistent inflationary impact.

The new shortage index developed in this paper provides a comprehensive and long-run perspective on the prevalence, drivers, and economic consequences of shortages in the United States over the past century. Our findings highlight the complex interplay of forces behind shortages and their tendency to boost inflation. This index serves as a valuable addition to the toolkit of policymakers and researchers, enabling a deeper understanding of the role of shortages in shaping macroeconomic outcomes.

References

Anthropic (2024): "Claude," Conversational AI model developed by Anthropic, version from March 30, 2024. Available at https://www.anthropic.com.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593.

Baumeister, C. and J. D. Hamilton (2019): "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks," *American Economic Review*, 109, 1873–1910.

Benigno, G., J. Di Giovanni, J. J. Groen, and A. I. Noble (2022): "The GSCPI: A new barometer of global supply chain pressures," *FRB of New York Staff Report*.

Bernanke, B. and O. Blanchard (2023): "What caused the US pandemic-era inflation?" *Peterson Institute for International Economics Working Paper*.

Burriel, P., I. Kataryniuk, C. Moreno Pérez, and F. Viani (2023): "A new supply bottlenecks index based on newspaper data," *Banco de Espana Working Paper*.

Caldara, D. and M. Iacoviello (2022): "Measuring geopolitical risk," *American Economic Review*, 112, 1194–1225.

Chen, L. and S. Houle (2023): "Turning Words into Numbers: Measuring News Media Coverage of Shortages," Tech. rep., Bank of Canada.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1–100.

Kahneman, D., J. L. Knetsch, R. Thaler, et al. (1986): "Fairness as a constraint on profit seeking: Entitlements in the market," *American economic review*, 76, 728–741.

Lamont, O. (1997): "Do" shortages" cause inflation?" in *Reducing Inflation: Motivation and Strategy*, University of Chicago Press, 281–306.

Newey, W. K. and K. D. West (1987): "Hypothesis testing with efficient method of moments estimation," *International Economic Review*, 777–787.

Pitschner, S. (2022): "Supply chain disruptions and labor shortages: COVID in perspective," *Economics Letters*, 221, 110895.

Ramey, V. A. and S. Zubairy (2018): "Government spending multipliers in good times and in bad: evidence from US historical data," *Journal of Political Economy*, 126, 850–901.

Table 1: Search Query

Search Name	Search Query	Peak Month
Energy Shortages	(*shortages* N/5 *energy*) AND *economics*	Dec-1973
Food Shortages	(*shortages* N/5 *food*) AND *economics*	Mar-1943
Industry Shortages	(*shortages* N/5 *industry*) AND *economics*	Aug-1942
Labor Shortages	(*shortages* N/5 *labor*) AND *economics*	Jan-2022
Articles	*articles*	–

Topic Sets	Components
shortages	`shortage* OR bottleneck* OR scarcit* OR rationing*`
energy	`oil OR gas OR coal OR fuel OR fuels OR gasoline OR energy OR heating OR petroleum OR electricity OR refinery OR pipeline OR petrol`
food	`food OR wheat OR meat OR milk OR crop OR crops OR grain OR farm OR agriculture OR famine OR feed OR farmer OR farmers OR water OR fertilizer OR drought`
industry	`steel OR copper OR iron OR metal* OR automotive OR textile OR machinery OR equipment OR transportation OR railway OR airline OR freight OR shipping OR transit OR deliveries OR shipment* OR ships OR chip* OR semiconductor* OR infrastructure OR materials OR distribution OR car OR cars OR parts OR goods OR material OR auto OR computer OR ‘supply chain’ OR components`
labor	`labor OR workers OR job* OR work OR employment OR manpower OR worker OR staff OR professional* OR technician* OR staffing OR skills OR workforce OR` personnel OR strike* OR union*
economics	`economic OR industr* OR production OR manufactur* OR economy OR trade OR commerce OR business OR budget OR tax OR fiscal OR corporation OR market OR price OR capacity OR company OR demand OR sales OR factory OR wages OR suppl*`
articles	`the AND be AND to AND of AND and AND at AND in`

Note: The truncation character (*) denotes a search including all possible endings of a word, e.g. “scarcit*” includes “scarcity” and “scarcities”.

Table 2: Largest Shortage Surprises, 1900-2023

Month	**Index**	**Surprise**	**Event**
Jan-1903	166	118	Nationwide coal shortages
Dec-1906	182	108	Shortage of coal and freight cars in Midwest
Nov-1916	225	106	Nationwide coal shortages
Jan-1918	552	319	Fuel and coal shortages
Dec-1919	286	115	Fuel and coal shortages due to war, strikes
Jul-1920	303	97	Freight car shortage affects coal and steel transportation
Aug-1922	268	135	Coal shortage due to strikes
Aug-1930	89	53	Drought leads to food and water shortages
Jul-1934	128	71	Strike by Teamsters unions in the West Coast
Sep-1939	145	76	Steel shortage due to the beginning of WW2
Aug-1941	493	138	War-related energy, materials and labor shortages
Jan-1943	1036	203	War-related oil, labor and food shortages
Jan-1945	538	174	War-related widespread shortages
Aug-1945	531	144	Labor shortages at the end of war
May-1946	563	180	Strikes by coal workers and fuel shortages
Jan-1948	439	167	Metal, fuel and food shortages
Feb-1950	216	96	Coal shortages amid strikes
Jan-1951	273	102	Labor shortages due to demand from defense industries
Jan-1952	259	120	Nationwide and worldwide shortages
Dec-1956	164	75	Oil shortages due to Suez crisis
Dec-1973	1036	471	Gasoline shortages due to 1973 oil crisis
Jan-1975	255	84	Concerns about gasoline rationing
Feb-1977	403	284	Carter's appeal on energy conservation
Feb-1978	151	66	Concerns about energy shortages
May-1979	553	201	Concerns about energy shortages
Aug-1981	146	49	Gasoline shortages due to 1979 oil crisis
Aug-1990	153	61	Concerns about energy shortages
Apr-2020	284	191	Medical shortages due to COVID-19 pandemic
Jan-2022	529	112	Labor shortages
Oct-2023	205	51	UAW strike and food shortages in Gaza

Note: The table lists the largest shocks to the shortage index. For this table, the shocks are constructed using the residuals of an autoregression and a condition on local maxima (see footnote 5).

Table 3: Validation of the Shortage Index

Set	Share of Total Newspaper Articles in Set	Validation Sample	Articles Mentioning Actual Shortages	Type I Error	Type II Error
Shortages AND Topic Words $\mathcal{S}$	1.58%	872	817	6.30%	–
Not Shortages $\mathcal{A} \setminus \mathcal{S}$	98.42%	298	1	–	0.33%
All Shortages $\mathcal{B}$	2.93%	334	284	14.97%	–

Note: Validation of the Shortage Index using a sample of newspaper articles used or not used to construct the index.

Table 4: Predicted Effect of Shortages on Prices and Quantities (4-quarters ahead)

	(1) 1950Q1–2023Q4		(2) 1950Q1–2014Q4		(3) 2015Q1–2023Q4	
	Prices	Quantities	Prices	Quantities	Prices	Quantities
GDP	0.19**	-0.13	0.22***	-0.21***	1.02***	0.15
	(2.04)	(-1.32)	(3.54)	(-3.13)	(4.12)	(0.57)
PCE Durables	0.25*	-0.25***	0.37***	-0.33***	1.10	1.61*
	(1.70)	(-3.13)	(8.14)	(-3.99)	(1.31)	(1.73)
PCE Nondurables	0.31***	-0.15	0.37***	-0.23***	0.83*	0.23
	(2.94)	(-1.61)	(3.64)	(-2.78)	(1.85)	(0.52)
PCE Services	0.18***	-0.02	0.19***	-0.05	1.17***	0.90**
	(4.14)	(-0.24)	(4.79)	(-0.50)	(6.85)	(2.47)
Investment	0.28***	-0.25***	0.32***	-0.31***	1.27***	0.05
	(2.70)	(-3.51)	(5.73)	(-5.15)	(3.21)	(0.08)
Obs.	292	292	260	260	32	32

Note: The table reports standardized coefficients of predictive regressions of economic activity and inflation on the shortage index. The dependent variable for each regression is the log difference between $t+4$ and t of the variable listed in each row, both its real quantity and its associated price deflator. Each regression includes as controls the quarterly changes of the dependent variable and the associated economic indicator or price deflator, contemporaneously and with three lags. Data are quarterly. The full sample, reported in the first two columns, runs from 1950Q1 to 2023Q4. We also partition the sample into two periods: a "pre-Covid" period that runs from 1950Q1 to 2014Q4, and a "Covid" period which runs from 2015Q1 to 2023Q4. Heteroskedasticity and autocorrelation consistent t-statistics are reported in parentheses and computed according to Newey and West (1987). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Predicted Effect of Shortages on Prices and Quantities (with controls)

	(1) 1950Q1–2023Q4		(2) 1950Q1–2014Q4		(3) 2015Q1–2023Q4	
	Prices	Quantities	Prices	Quantities	Prices	Quantities
GDP	0.19**	-0.13	0.22***	-0.21***	1.02***	0.15
	(2.04)	(-1.32)	(3.54)	(-3.13)	(4.12)	(0.57)
Oil	0.16	-0.22**	0.22***	-0.35***	0.92***	0.03
	(1.57)	(-2.33)	(5.17)	(-4.17)	(3.12)	(0.14)
Commodities	0.17*	-0.25***	0.24***	-0.37***	1.06***	0.13
	(1.69)	(-2.67)	(6.38)	(-4.39)	(3.65)	(0.58)
Wages	0.17*	-0.19**	0.24***	-0.29***	0.29	-0.06
	(1.68)	(-2.19)	(7.45)	(-4.21)	(0.67)	(-0.12)
Inf. Exp. (10 Yr.)	0.67***	0.23	-0.06	-0.40***	1.15***	0.14
	(5.52)	(1.21)	(-0.47)	(-3.54)	(4.29)	(0.17)

Note: The table reports standardized coefficients of predictive regressions of economic activity and inflation on the shortage index. The dependent variable for each regression is the log difference between $t + 4$ and t of GDP, both real GDP and GDP deflator. For 10-year inflation expectations, the dependent variable for the price regressions is detrended by 10-year inflation expectations, and so we do not use this on the right-hand side. Each regression includes as controls the quarterly changes of the dependent variable and the associated economic indicator or price deflator, contemporaneously and with three lags. In each row, we add the listed variable as an additional control. Data are quarterly. The full sample, reported in the first two columns, runs from 1950Q1 to 2023Q4. We also partition the sample into two periods: a "pre-Covid" period that runs from 1950Q1 to 2014Q4, and a "Covid" period which runs from 2015Q1 to 2023Q4. Due to data limitations, the wage regressions start in 1964Q1, and the inflation expectation regressions start in 1982Q1. Heteroskedasticity and autocorrelation consistent t-statistics are reported in parentheses and computed according to Newey and West (1987). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Grouping of Newspaper Articles for the Construction of the Shortages Index.

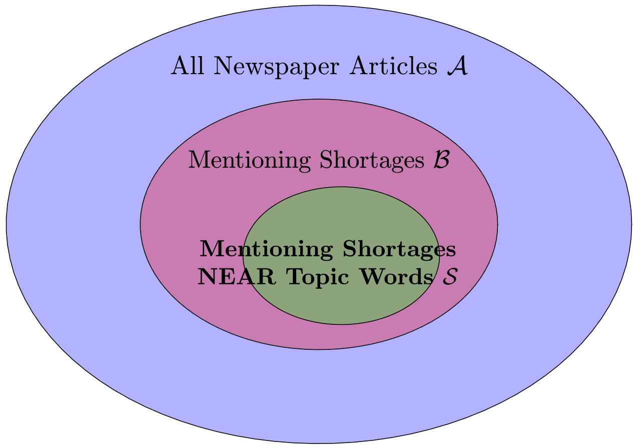
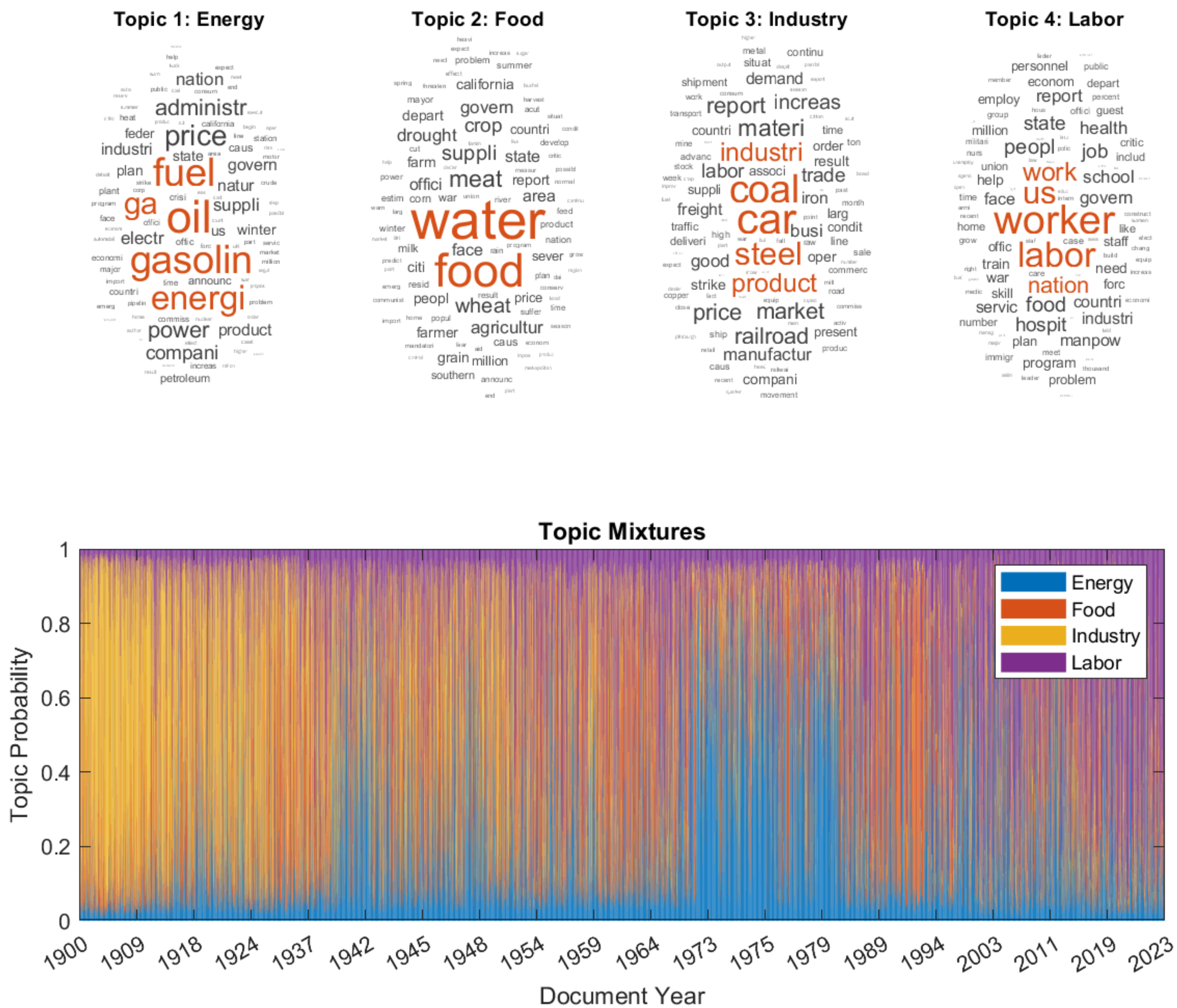
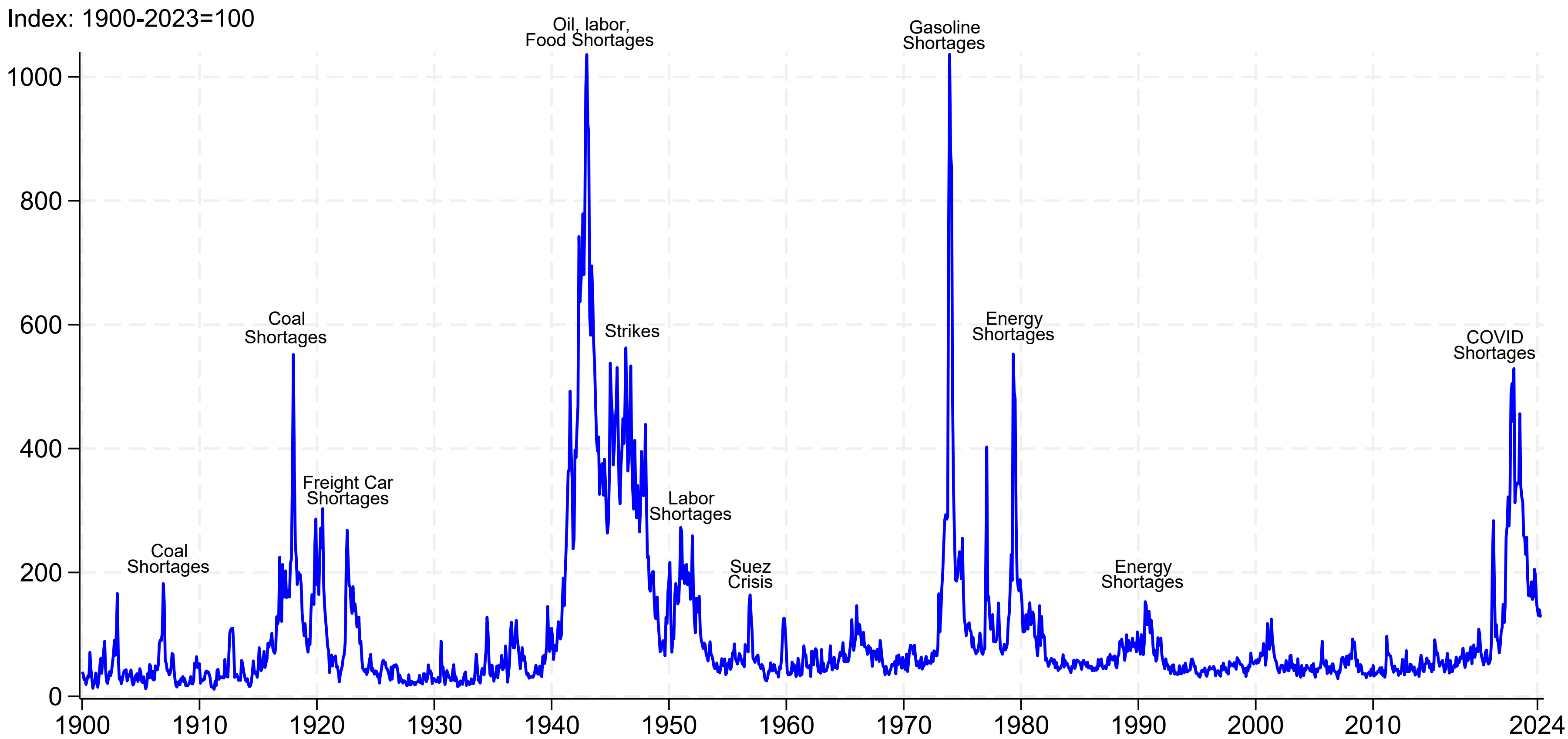


Figure 2: Topic Classification for the Index



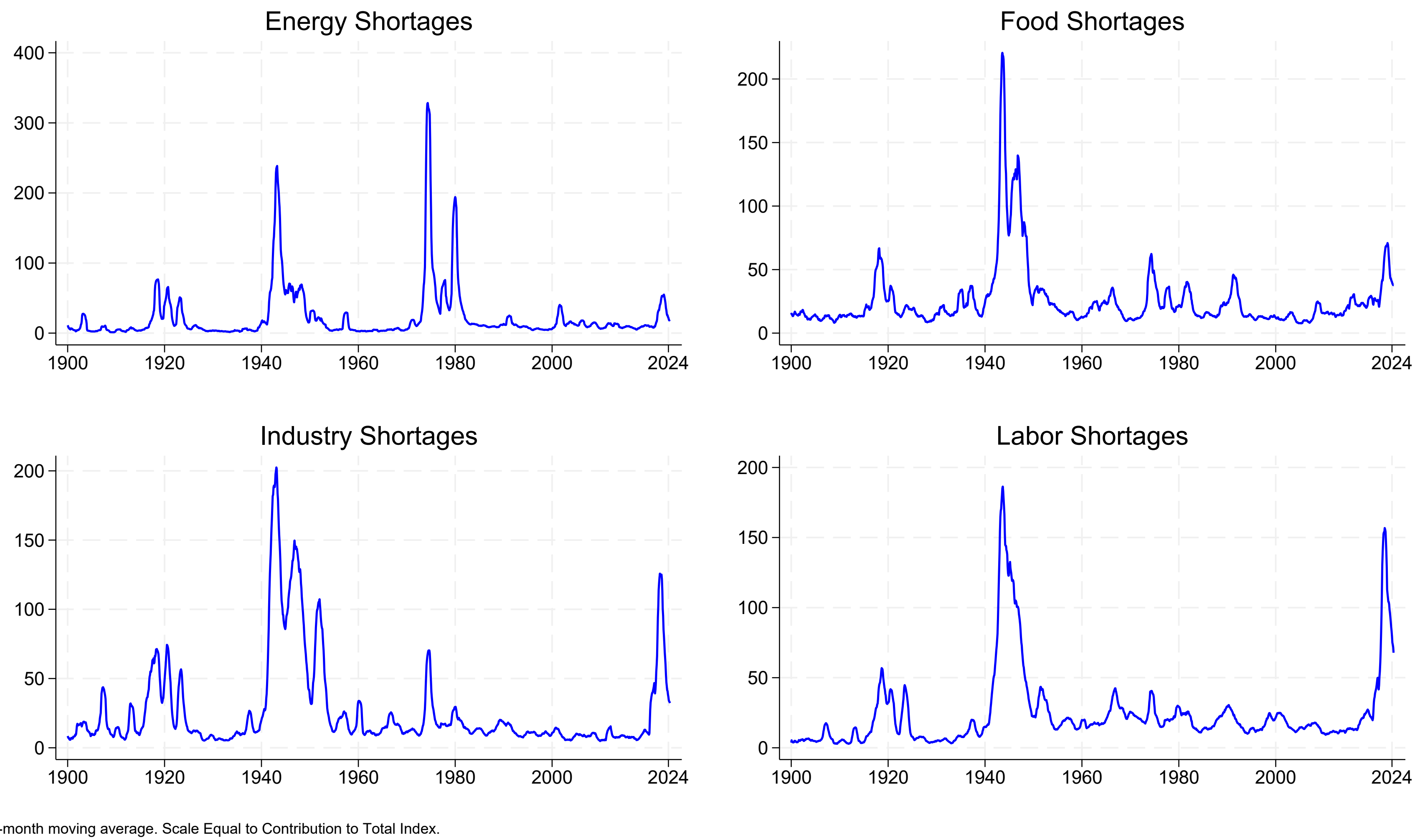
Note: Topic classification for a large sample of articles used in the construction of the Shortage Index. The top four rows indicate the four most influential topics chosen by LDA analysis on the articles. The bottom rows show topic probability by article, sorting the articles by time.

Figure 3: The Shortage Index



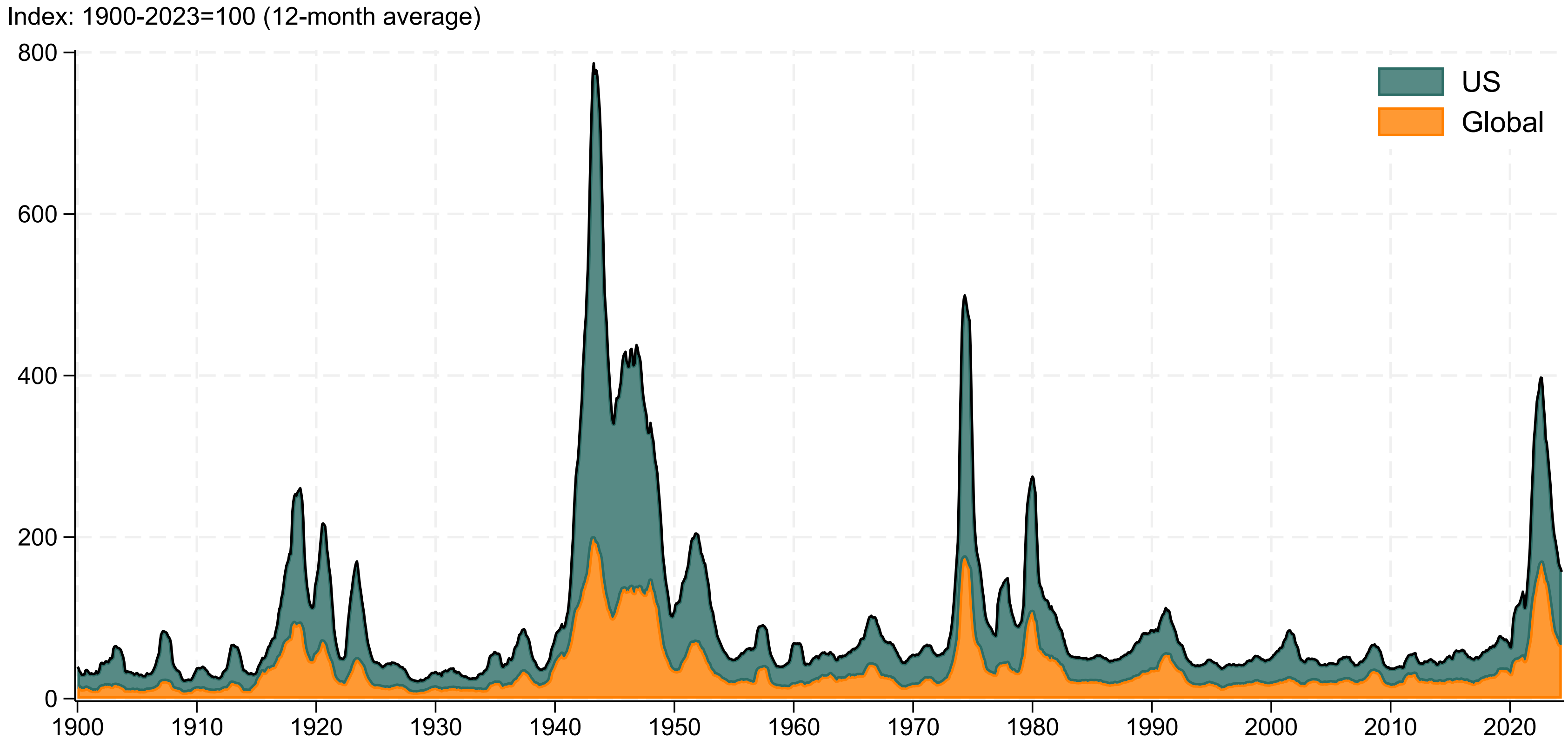
Note: The figure shows the shortage index from January 1900 through May 2024.

Figure 4: Decomposition of the Shortage Index by Category



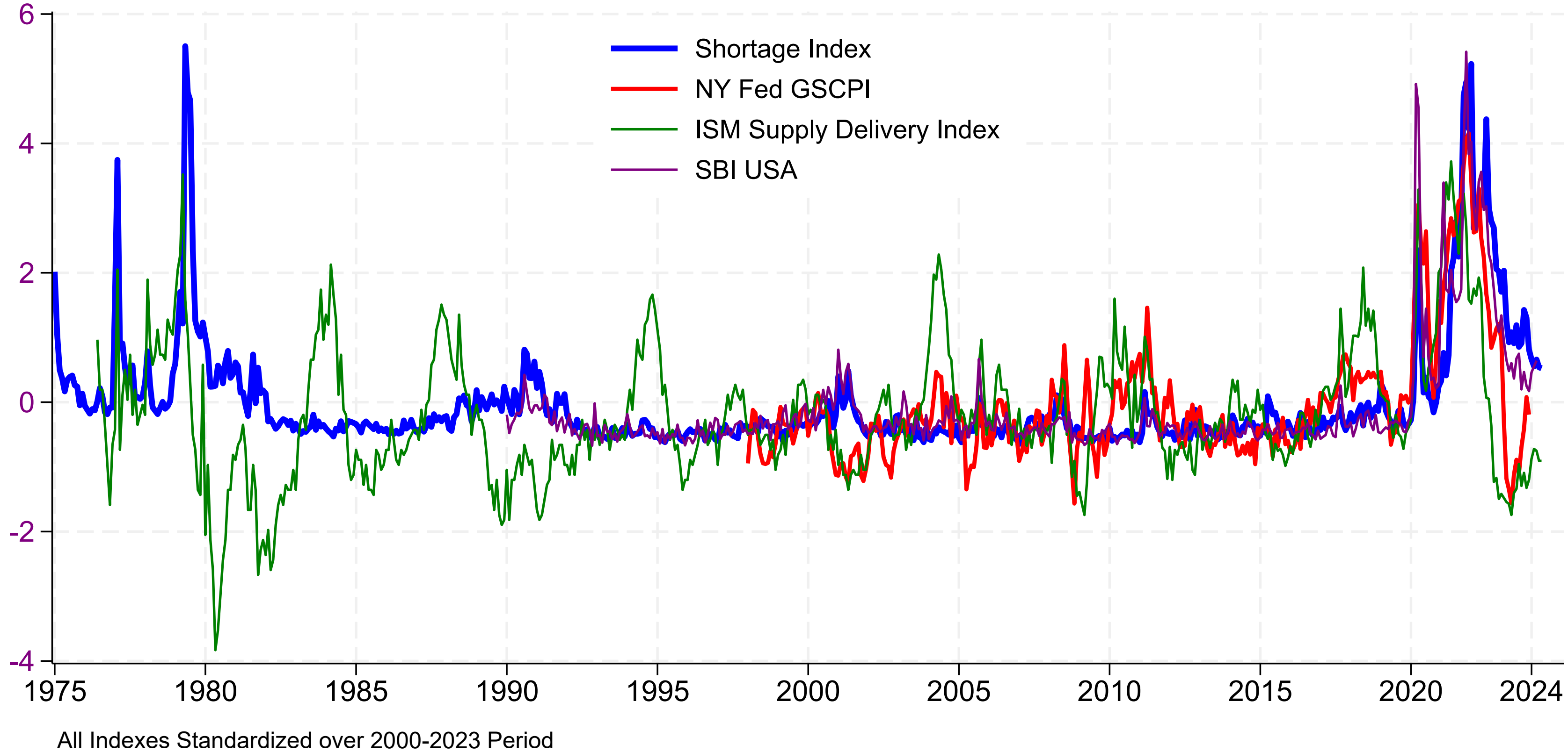
Note: The figure illustrates the four categories adding up to the total shortage index.

Figure 5: U.S. vs. Global Shortages



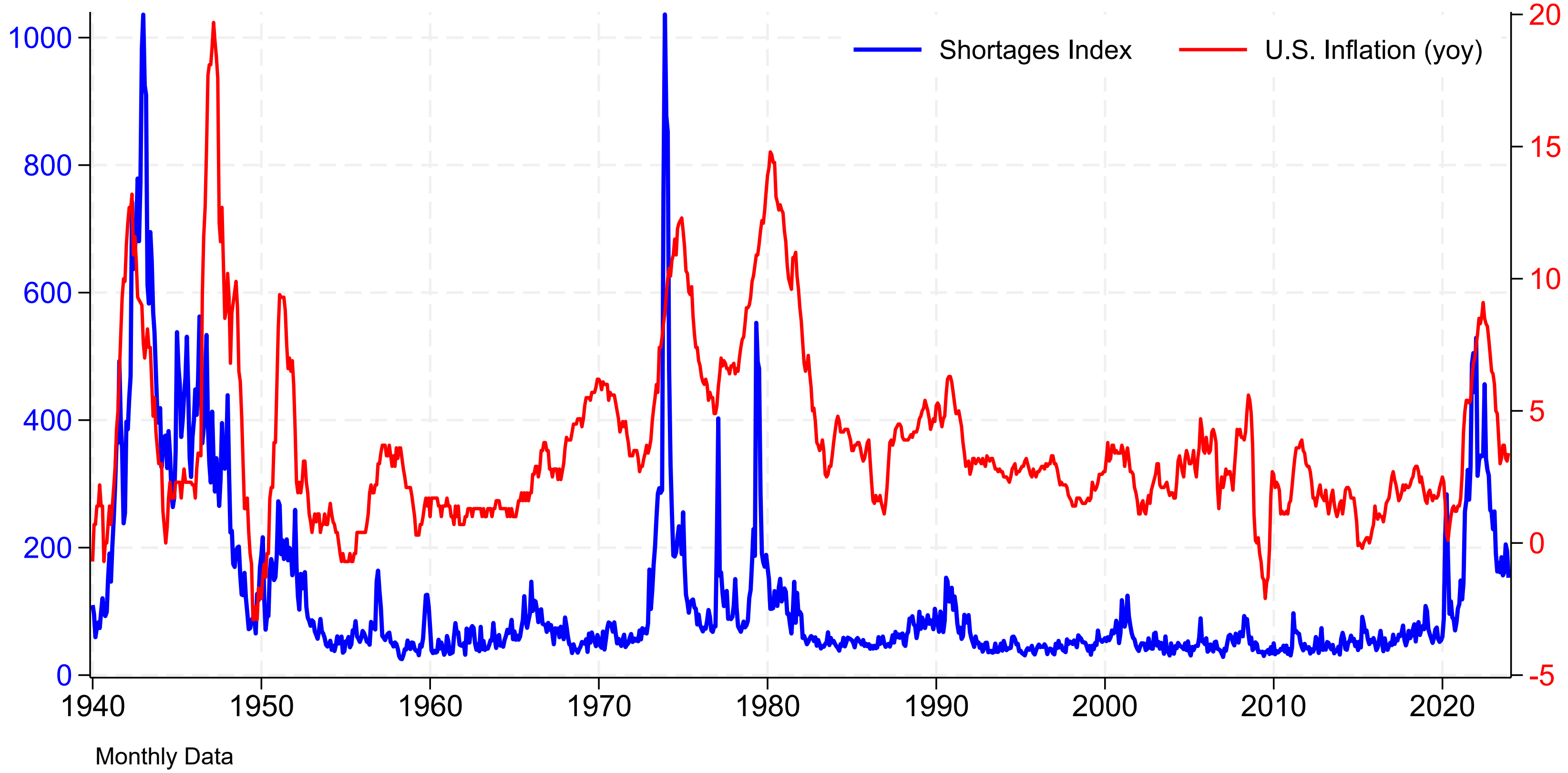
Note: The figure breaks down shortages into a domestic (US) and a global component.

Figure 6: Comparison to Other Supply Constraints Measures



Note: The figure compares the shortage index to alternative measures of supply constraints from January 1975 through May 2024. The ISM Supply Delivery Index is computed as the share of respondents reporting longer delivery times plus half the share of respondents reporting no change in delivery times. The SBI USA index is the U.S. Supply Bottlenecks Index from Burriel et al. (2023).

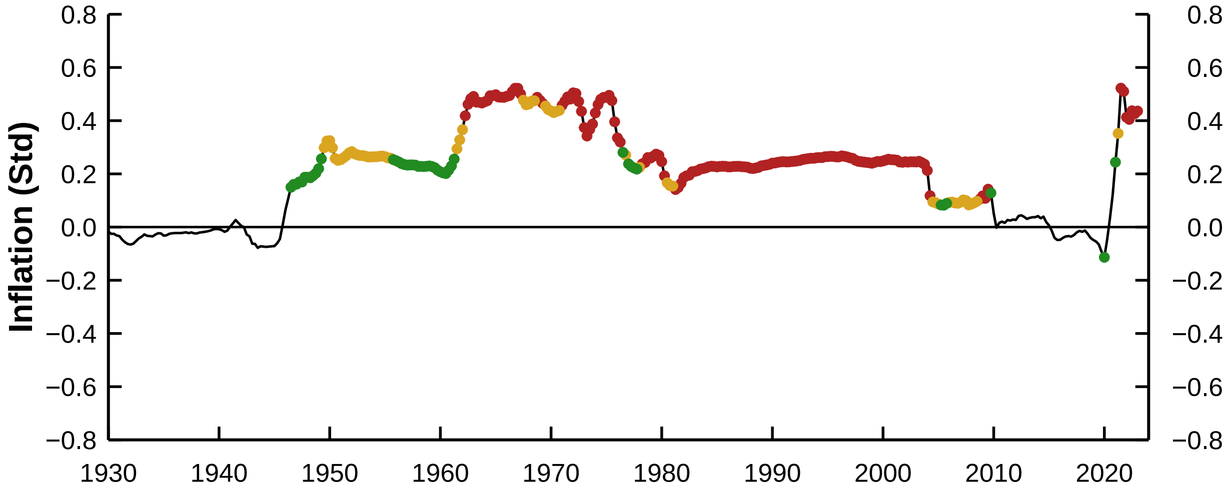
Figure 7: Shortages and U.S. Inflation



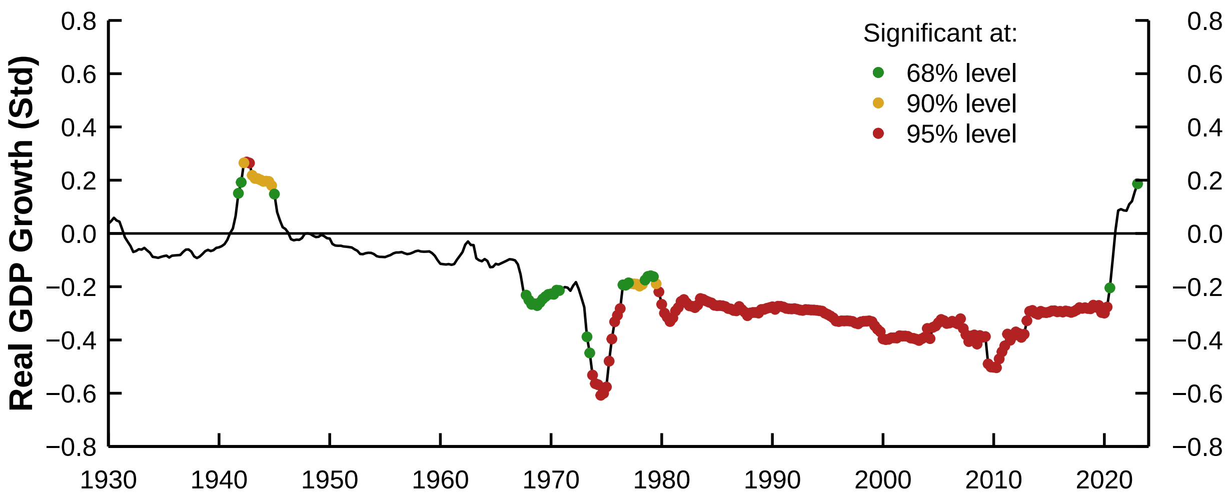
Note: The figures compares the shortage index (left) with U.S. inflation (right), from January 1940 through May 2024.

Figure 8: Effect of Shortages on Inflation and Real GDP Growth (30-Year Window)

(a) One-Year Ahead Inflation

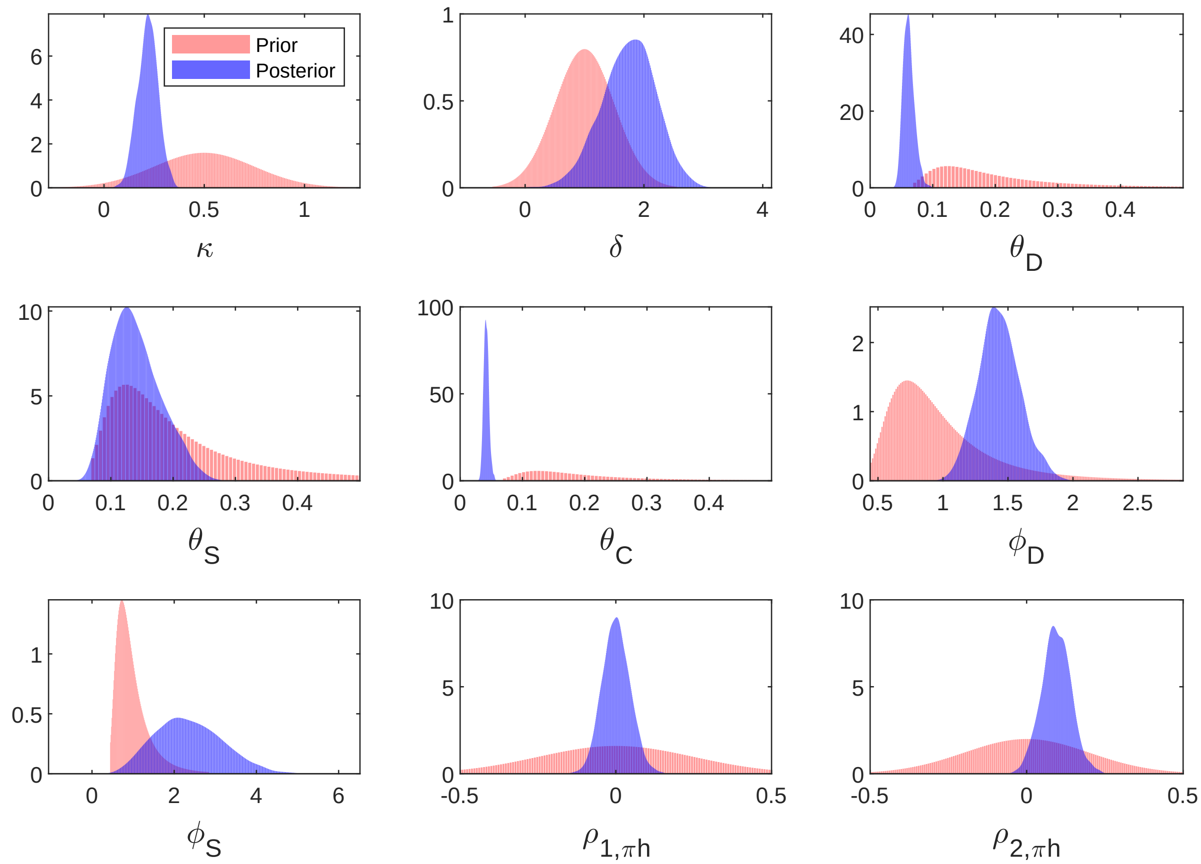


(b) One-Year Ahead Real GDP Growth



Note: The solid black line shows the time-varying estimated effect of shortages on inflation (top panel) and real GDP growth (bottom panel). The estimates are based on regressions using rolling 30-year windows. In each regression, the dependent variable is the 4-quarter ahead difference in log real GDP per capita or the 4-quarter ahead difference in log GDP deflator. On the right-hand side, the main explanatory variable is our shortage index. As controls, we include the one-quarter change in both log real GDP per capita and log GDP deflator, in quarter t plus three lags, and allow for up to three quarters of autocorrelation.

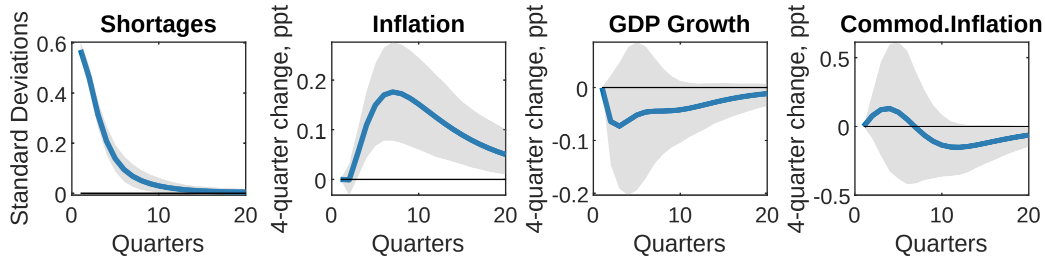
Figure 9: Prior and Posterior Densities of Structural VAR Parameters



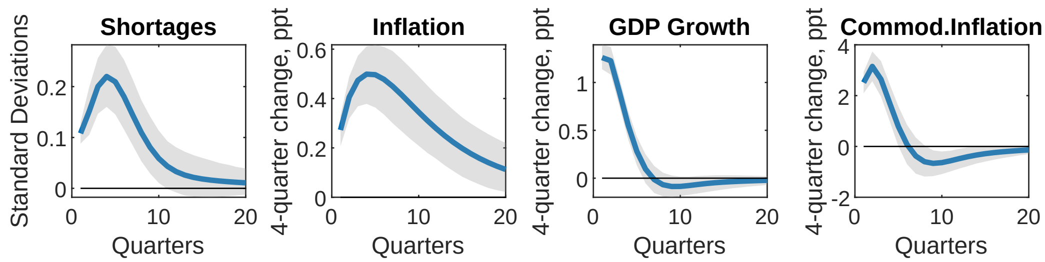
Note: Prior and posterior densities of the parameters of the structural VAR model.

Figure 10: Effects of Shortages on US Activity and Inflation in the Estimated VAR Model

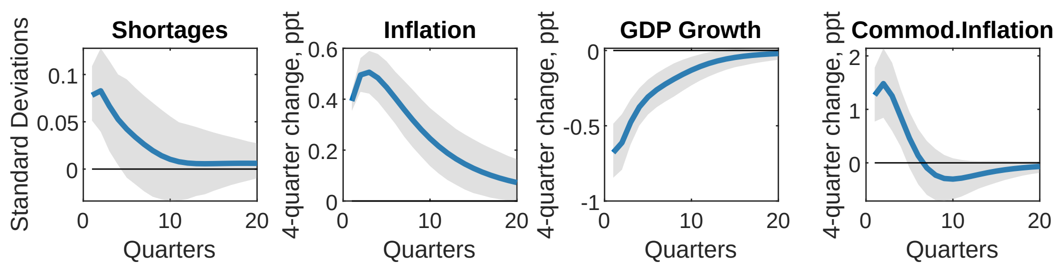
(a) Shock to Shortages



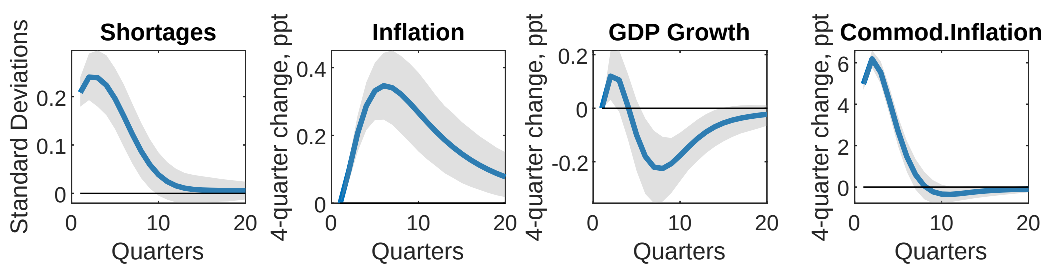
(b) Shock to Demand



(c) Shock to Supply

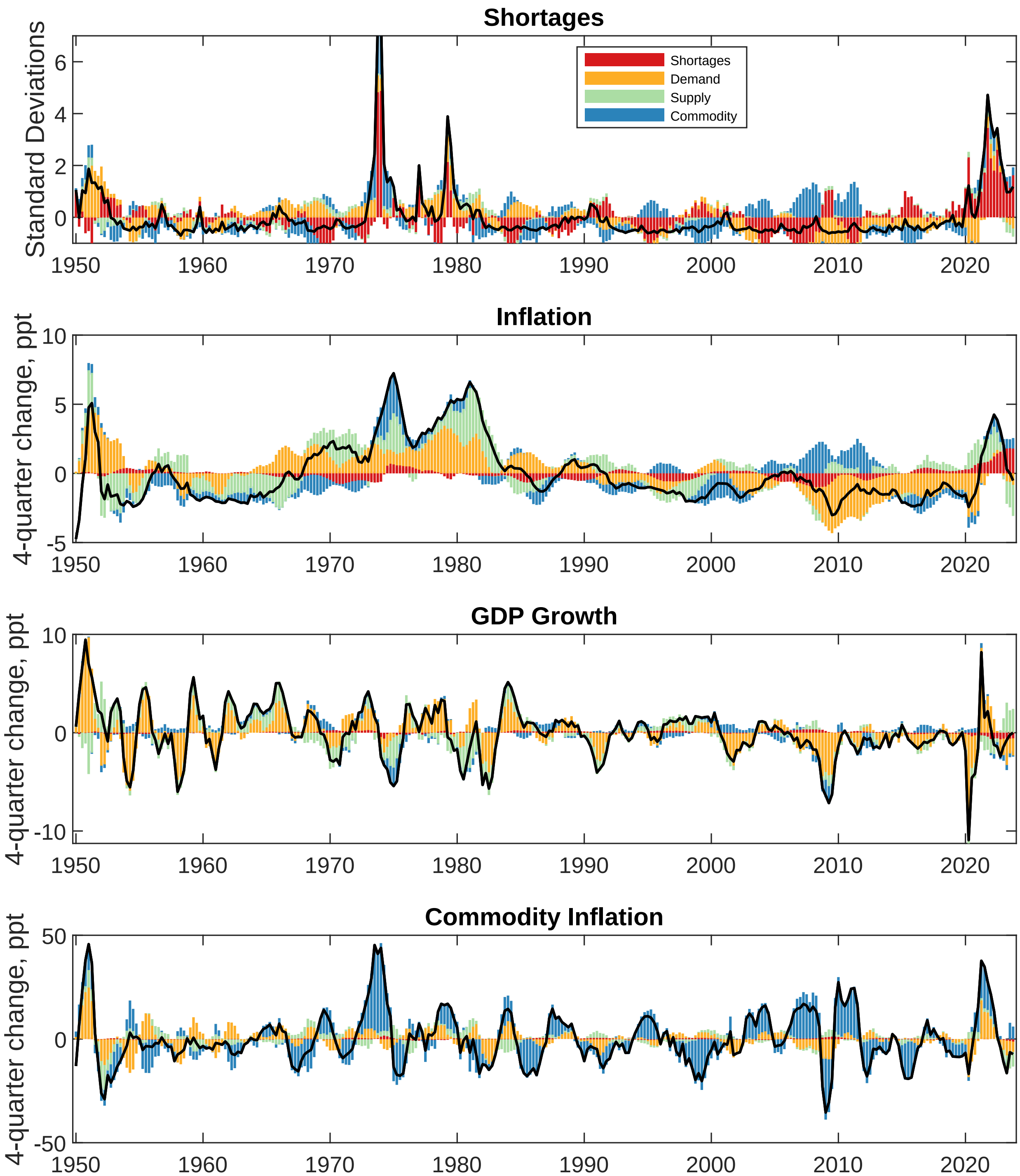


(d) Shock to Commodity Prices



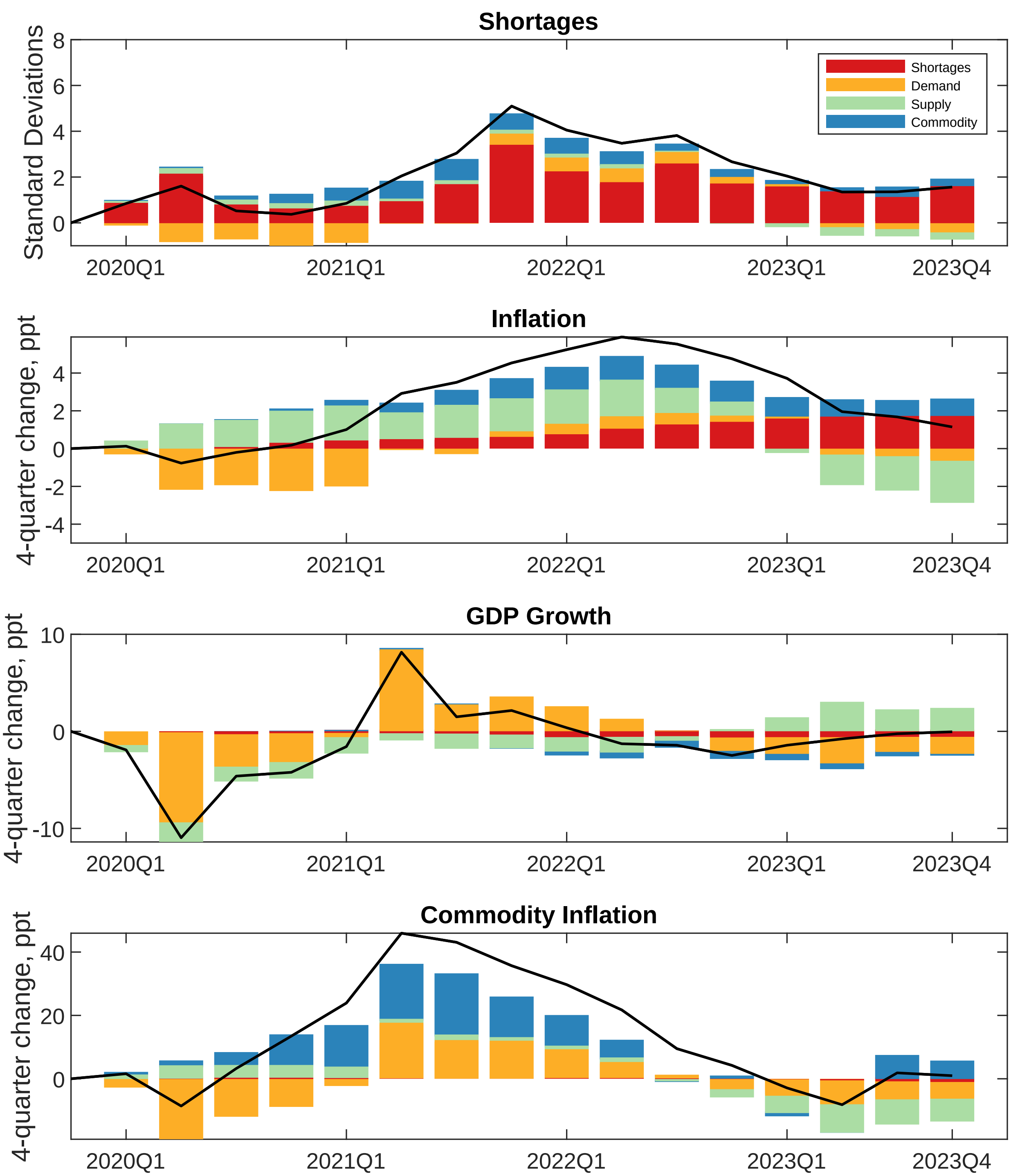
Note: Impulse Response in VAR to Estimated Shocks. All shocks are one-standard deviation in size. Solid lines denote the response at the posterior mean. Shaded areas denote 80 percent confidence intervals.

Figure 11: Historical Decomposition of Estimated VAR Model: Full Sample



Note: The figure decomposes movements in shortages, economic activity and inflation in the contribution of shortage shocks, demand shocks, supply shocks, and commodity price shocks. All variables are expressed in deviation from their sample mean.

Figure 12: Historical Decomposition of Estimated VAR Model: 2020-2023



Note: The figure decomposes movements in shortages, economic activity and inflation since 2020 in the contribution of shortage shocks, demand shocks, supply shocks, and commodity price shocks. All variables are expressed in deviation from their 2019Q4 value.

Appendix

A Appendix Tables

Table A.1: Predicted Effect of Shortages on Prices and Quantities (1-quarter ahead)

	(1) 1950Q1–2023Q4		(2) 1950Q1–2014Q4		(3) 2015Q1–2023Q4	
	Prices	Quantities	Prices	Quantities	Prices	Quantities
GDP	0.10**	-0.01	0.11***	-0.06	0.91***	-0.10
	(2.36)	(-0.18)	(3.22)	(-1.09)	(5.60)	(-0.45)
PCE Durables	0.12	-0.06	0.19***	-0.12**	0.07	1.21
	(1.58)	(-1.13)	(4.32)	(-2.13)	(0.15)	(1.28)
PCE Nondurables	0.28***	-0.06	0.31***	-0.09	0.52*	0.27
	(3.67)	(-0.71)	(3.83)	(-1.13)	(1.85)	(0.80)
PCE Services	0.13***	-0.02	0.14***	-0.06	0.71***	0.57**
	(4.35)	(-0.33)	(4.70)	(-0.91)	(4.77)	(2.14)
Investment	0.17***	-0.02	0.18***	-0.05	0.89***	-0.03
	(3.79)	(-0.38)	(5.13)	(-1.14)	(5.04)	(-0.07)
Obs.	295	295	260	260	35	35

Note: The table reports standardized coefficients of predictive regressions of economic activity and inflation on the shortage index. The dependent variable for each regression is the log difference between $t+1$ and t of the variable listed in each row and its associated price deflator. Each regression includes as controls the quarterly changes of the dependent variable and the associated economic indicator or price deflator, contemporaneously and with three lags. Data are quarterly. The full sample, reported in the first two columns, runs from 1950Q1 to 2023Q4. We also partition the sample into two periods: a "pre-Covid" period that runs from 1950Q1 to 2014Q4, and a "Covid" period which runs from 2015Q1 to 2023Q4. Heteroskedasticity and autocorrelation consistent t-statistics are reported in parentheses and computed according to Newey and West (1987). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Predicted Effect of Shortages on Prices and Quantities (8-quarters ahead)

	(1) 1950Q1–2023Q4		(2) 1950Q1–2014Q4		(3) 2015Q1–2023Q4	
	Prices	Quantities	Prices	Quantities	Prices	Quantities
GDP	0.17**	-0.08	0.15**	-0.12*	0.46	0.96
	(2.53)	(-1.14)	(2.54)	(-1.72)	(0.96)	(1.37)
PCE Durables	0.28***	-0.16**	0.32***	-0.21*	1.41	-0.29
	(3.06)	(-1.98)	(6.70)	(-1.94)	(1.31)	(-0.30)
PCE Nondurables	0.25***	-0.02	0.25***	-0.08	1.29***	0.12
	(2.74)	(-0.27)	(2.84)	(-1.04)	(3.24)	(0.36)
PCE Services	0.19***	-0.00	0.18***	0.01	1.13***	1.14**
	(3.52)	(-0.05)	(3.83)	(0.10)	(5.28)	(2.68)
Investment	0.26***	-0.19***	0.25***	-0.22***	0.76	0.44
	(4.42)	(-2.96)	(4.71)	(-3.63)	(1.52)	(0.72)
Obs.	288	288	260	260	28	28

Note: The table reports standardized coefficients of predictive regressions of economic activity and inflation on the shortage index. The dependent variable for each regression is the log difference between $t+4$ and t of the variable listed in each row and its associated price deflator. Each regression includes as controls the quarterly changes of the dependent variable and the associated economic indicator or price deflator, contemporaneously and with three lags. Data are quarterly. The full sample, reported in the first two columns, runs from 1950Q1 to 2023Q4. We also partition the sample into two periods: a "pre-Covid" period that runs from 1950Q1 to 2014Q4, and a "Covid" period which runs from 2015Q1 to 2023Q4. Heteroskedasticity and autocorrelation consistent t-statistics are reported in parentheses and computed according to Newey and West (1987). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$