Identifying OPEC News Shocks: The Impact of OPEC Announcements Using Textual Data

Gabriel A. Abeyie*

October 24, 2024

Click here for the most recent version

Abstract

This paper presents a novel method for analyzing the macroeconomic effects of oil price shocks by integrating textual analysis of news articles related to OPEC announcements. Using text data from major newspapers, I extract features that capture shifts in oil price expectations driven by OPEC news. These features function as external instruments in a proxy SVAR model. The analysis reveals that 91.3% of oil price surprises are attributable to OPEC-related factors, with 85.6% stemming from actual supply decisions and 5.7% from market sentiment and demand expectations. Historical decomposition indicates that supply news dominates during geopolitical crises, such as the Iranian Revolution and the Gulf War, while demand-driven sentiments play a larger role during economic downturns like the Global Financial Crisis and the COVID-19 pandemic. Additionally, experiments with synthetic news data confirm that real news narratives have a significant impact on market behavior, unlike artificially generated content. These findings suggest that OPEC news shocks primarily reflect both supply-side decisions and demand-side expectations.

Keywords: Oil Price Shocks, OPEC Announcements, Textual Analysis, News Sentiment, Large Language Models, Synthetic News Data

***Correspondence:** Gabriel Appau Abeyie, Department of Economics, The University of Tennessee, Knoxville, USA; Email: gabeyie@vols.utk.edu

1 Introduction

The dynamics of crude oil prices, particularly around announcements by the Organization of the Petroleum Exporting Countries (OPEC), have long been a topic of significant interest. OPEC is a major player in the crude oil market; its actions and decisions play a significant role in influencing global oil supply, which, in turn, affects crude oil prices.[1]

Traditionally, researchers have relied on futures market data to understand these shocks, assuming that price movements around OPEC meetings are solely the result of OPEC's actions and decisions (Demirer and Kutan (2010), Känzig (2021)). While this may seem plausible, such an approach overlooks the broader context. For instance, major macroeconomic events, such as a Federal Open Market Committee (FOMC) meeting occurring on the same day as an OPEC meeting, could influence crude oil prices as well. Additionally, the narratives, expectations, and sentiments conveyed through media coverage about OPEC's meetings play a crucial role in shaping market behavior. It is therefore necessary to consider these factors when analyzing the impact of OPEC's decisions.

This paper addresses this gap by incorporating textual data from newspaper articles about OPEC announcements to provide a better understanding of oil price movements. Rather than treating price data as a standalone measure of market response, I decompose the total surprise (price movement) into OPEC-driven and non-OPEC-driven components, using textual data to estimate the influence of narratives and sentiment. With this approach, I am able to obtain an estimate of the expected market reaction to OPEC's decisions, which represents the surprise component attributable to those decisions.

The novelty of this research lies in its use of textual features derived from newspaper articles to both estimate the surprise component attributable to OPEC's decisions and serve as external instruments in a structural Vector AutoRegressive (SVAR) model.[2] By doing so,

[1] OPEC is an international body which currently includes Algeria, Angola, Equatorial Guinea, Gabon, Iran, Iraq, Kuwait, Libya, Nigeria, the Republic of the Congo, Saudi Arabia, the United Arab Emirates and Venezuela as its members. OPEC is responsible for more than 35% of worldwide crude oil production.

[2] To the best of my knowledge, this is the first paper in the crude oil literature to incorporate textual features as an external instrument in an SVAR model.

I demonstrate that news narratives and sentiments provide significant explanatory power and offer a more complete picture of oil price dynamics than price data alone. This contribution bridges a gap in the literature by showing that the surprise (price movement) that occurs during OPEC announcement days is not solely driven by OPEC's decisions. Other possible non-OPEC factors could also be responsible for the price movement.

Preview of results. This study demonstrates that incorporating news text data significantly enhances the understanding of oil price movements around OPEC announcements. By analyzing both the sentiment and content of news articles, I successfully separated oil price changes directly driven by OPEC's actions from those influenced by other factors.

The findings reveal that a substantial 91.3% of oil price surprises can be attributed to OPEC-related factors. Of this, the majority (85.6%) comes from OPEC's actual supply decisions, while a smaller portion (5.7%) is influenced by market sentiment and demand expectations captured through news narratives.

Additionally, historical analysis highlights that supply news were more prominent during geopolitical crises, such as the Iranian Revolution and the Gulf War. In contrast, demand-driven sentiments played a larger role during economic downturns like the Global Financial Crisis and the COVID-19 pandemic. This distinction emphasizes how different contexts influence the impact of OPEC announcements on the oil market.

To ensure the robustness of the results, I conducted an experiment using synthetic news data generated by a language model. These synthetic instruments failed to replicate the effects observed with real news, reinforcing the importance of genuine news narratives in influencing market behavior.

Related literature. Crude oil is a major energy source for industrialized countries, and fluctuations in its price significantly impact these countries' economies. Therefore, it is important to identify these disturbances and measure their effects accurately. Since the introduction of the SVAR model by Sims (1980), it has been the primary tool for identifying these shocks and their impact on the macroeconomy.[3] The SVAR model has evolved in three stages: the use of short-run exclusion restrictions, long-run exclusion restrictions, and sign restrictions.

The short-run exclusion restriction approach involves imposing constraints on how certain shocks can contemporaneously affect variables in the system. This method has been used extensively in the financial, fiscal, and monetary policy literature (Blinder et al. (1973), Christiano (1999), Clarida et al. (1999)). In the crude oil literature, this approach was pioneered by Kilian (2008, 2009) to disentangle oil demand and supply shocks. Alquist and Kilian (2010) used this method to derive oil-specific demand shocks. Despite its popularity, the short-run exclusion restriction approach faces several challenges and criticisms regarding the validity and robustness of the imposed restrictions (Uhlig (2005), Ramey (2011)).

Long-run exclusion restrictions are used to identify structural shocks by imposing constraints on the long-term relationships between variables. Unlike short-run restrictions, which focus on contemporaneous (immediate) impacts, long-run restrictions assume that certain shocks have no permanent effect on particular variables (Blanchard and Quah (1993)). Long-run exclusion restrictions also face identification issues as they suffer from model sensitivity. This means that a small change in the restrictions can lead to significantly different conclusions (Lütkepohl and Velinov (2016)).

Uhlig (2005) introduced the concept of sign restrictions in SVAR models as an alternative to traditional zero exclusion restrictions. This approach allows for more flexible identification by imposing only the expected signs of impulse responses rather than exact zeros. In the crude oil literature, sign restrictions have also been used to identify oil price shocks (Peersman and Van Robays (2009), Lippi and Nobili (2012), Kilian and Murphy (2012, 2014), Inoue and Kilian (2013), Antolín-Díaz and Rubio-Ramírez (2018)).

Beyond these three approaches, other methods have been introduced to address their restrictive and strong theoretical assumptions. One popular method is the use of external

[3] Sims (1980) argued that imposing zero restrictions based on economic theory allows for the disentanglement of structural shocks from reduced-form residuals.

instruments (Hansen and Singleton (1982), Stock and Watson (2012), Mertens and Ravn (2013)), or proxies, to identify specific structural shocks. This approach is known as the proxy SVAR.[4] Proxy SVAR incorporates additional information through instruments that are correlated with the structural shocks of interest but uncorrelated with other shocks in the system.

In the crude oil literature, the event study approach has been used to derive external instruments for use in a proxy SVAR to identify structural shocks. For instance, Känzig (2021) uses observed changes in crude oil futures prices during OPEC meeting announcement days as proxies to identify structural shocks. His method involves aggregating daily surprises in oil futures prices by summing the daily surprises within each month to create a monthly series. In this paper, while I also utilize OPEC meeting announcement days, **I diverge from Känzig's approach in three key ways**.

First, instead of relying on observed changes in futures prices, I employ textual data from newspaper articles published about OPEC on these announcement days as external instruments. This novel approach allows for capturing the narratives and market sentiments expressed in the news, which may not be fully reflected in price changes alone.

Second, rather than aggregating daily surprises by simple summation, I adopt the method proposed by Kilian (2024), which accounts for the timing of the daily surprises within the month when constructing the monthly proxy. Specifically, Kilian's method weights the daily surprises based on their occurrence within the month, recognizing that surprises early in the month have a larger impact on the monthly average than those occurring later. This adjustment provides a more accurate representation of the information content in the news articles at a monthly frequency.

Furthermore, unlike Känzig (2021), who describes the shocks obtained as adverse supply news shocks, I describe the shocks in this paper as **OPEC news shocks**—shifts in oil price expectations driven by OPEC news. This term encompasses both demand and supply

[4] Also known as Instrumental Variable-SVAR (IV-SVAR). I use them interchangeably in the paper.

factors because, upon reviewing the newspaper articles in my study, I found that many of them discuss both aspects. Therefore, it would not be accurate to label these shocks only as supply-related. To further clarify, I break down these OPEC news shocks into OPEC demand news shocks, which reflect market expectations about oil demand, and OPEC supply news shocks, which relate to actual production or supply changes by OPEC. This distinction provides a clearer understanding of the different forces driving oil price changes.

The use of textual data to identify economic shocks has gained significant traction in recent years, driven by the increasing availability of textual data and the decreasing cost of computational resources. Numerous studies have leveraged this trend. For example, Aruoba and Drechsel (2024) utilized documents from the Federal Reserve to identify and extract monetary policy shocks. Similarly, Ochs (2021) extracted a time series of directional sentiment from keywords identified in Federal Open Market Committee (FOMC) meeting minutes, using these sentiments as regressors to explain variations in the federal funds rate, with the residuals labeled as monetary policy text shocks. Additionally, Handlan (2020) employed FOMC statements alongside machine learning methods, specifically neural networks, to extract text-based shocks and assess their macroeconomic implications. These studies underscore the potential of textual data in uncovering economic shocks.

In the oil literature, while several studies (Li et al. (2019), Bai et al. (2022), Jiao et al. (2022), Abeyie (2024)) have successfully used text data to forecast crude oil prices, there has been limited research on using text data to identify oil shocks. Plante (2019) introduced an OPEC newspaper index based on article counts to measure interest in OPEC over time. He investigates how unexpected changes in this index affect oil price volatility, finding that increases in media attention are strongly associated with higher levels of both realized and implied oil price volatility. Caldara et al. (2019) utilized news articles from the Oil Daily, published by the Energy Intelligence Group (EIG), and the Oil Market Report, published by the International Energy Agency (IEA), to identify exogenous events that disrupt oil production. However, Caldara et al. (2019)'s approach is subjective; the events classified as exogenous are based on judgment calls, leading to a risk of misclassification, as highlighted by Kilian and Zhou (2020). Datta and Dias (2019) advanced the literature by introducing a novel automated approach to identifying oil price shocks with text data. They constructed oil supply and demand indexes from oil-related news articles by generating time series of standardized counts or frequencies of the words supply and demand in proximity to increase and decrease. These indexes are described as shocks, and their macroeconomic implications were measured. However, this approach has also been questioned by Kilian and Zhou (2020, 2023).

Contribution to the literature. This paper makes several important contributions to the understanding of oil price dynamics and their impact on the broader economy. First, it introduces a novel approach by using text-based features from news articles as external instruments in a Structural Vector Autoregression (SVAR) model to identify oil price shocks. Unlike traditional methods that rely solely on futures market data, this study incorporates media sentiment and narratives, providing a more comprehensive view of how market expectations influence oil prices around OPEC announcements.

By breaking down the total oil price surprise into OPEC-driven and non-OPEC-driven components, the paper clearly distinguishes the direct effects of OPEC decisions from other market forces. Further, it separates OPEC news into sentiment-driven demand expectations and objective supply-side adjustments. This detailed decomposition reveals that about 85.6% of the OPEC-driven surprise comes from supply news, while 5.7% is due to demand news. This analysis highlights the significant role that both supply decisions and market sentiment play in shaping oil price movements.

The study also validates the effectiveness of text-based features by comparing them with traditional instruments like observed price changes. The results show that text-based features perform as well as, if not better than, traditional methods, underscoring the value of incorporating media narratives into economic models. Additionally, the paper includes a unique experiment using GPT-4 to generate synthetic news articles. This experiment demonstrates that synthetic news has much lower explanatory power compared to real news, emphasizing the importance of using authentic media content to capture genuine market signals.

Moreover, the paper employs historical decomposition and forecast error variance decomposition to analyze how OPEC demand and supply news shocks have influenced oil prices and economic indicators over time. These analyses provide deeper insights into the temporal impact of different types of shocks during key historical events, such as geopolitical crises and economic downturns.

This research advances the field by integrating textual analysis into macroeconomic modeling, offering a more detailed and accurate framework for understanding the drivers of oil price changes. It not only enhances the identification of oil price shocks but also provides valuable implications for policymakers and market participants in making informed decisions based on both supply-side actions and market sentiment.

Structure of the paper. The paper is organized as follows: Section 2 presents the conceptual framework for understanding oil price surprises. Section 3 introduces the data and methodology, including the process of extracting features from newspaper articles. Section 4 details the dimensionality reduction techniques applied to the extracted text features. The empirical results are discussed in Section 5, followed by robustness checks in Section 6. Finally, Section 7 concludes the paper, summarizing key findings and suggesting directions for future research.

2 Conceptual Framework

Understanding what drives oil price changes, particularly during OPEC announcements, has often relied on price data alone.[5] While this approach captures immediate market reactions, it overlooks the broader context in which these price movements occur. The market is not isolated; it is shaped by the sentiment, expectations, and narratives that surround OPEC decisions, and these are often reflected in media coverage and public discourse. As a result, price changes around OPEC announcements can be driven by a combination of OPEC-related factors and external influences.

2.1 Traditional Approach: Price as Surprise

Previous studies, such as Känzig (2021), have typically calculated oil price "surprises" by observing price changes around OPEC meetings and interpreting any deviations from expected prices as being driven by OPEC.[6] This approach assumes that all relevant information is embedded within the price data itself, with oil futures acting as a proxy for market expectations. While this method provides valuable insights into immediate price movements, it overlooks the role of external factors, as well as the narratives that shape market expectations before and after OPEC announcements. Particularly in speculative environments like oil markets, narratives and sentiment can heavily influence price dynamics.

2.2 A New Lens: Incorporating Text Data

To overcome these limitations, this paper incorporates text data into the analysis. Rather than assuming that the entire price change is driven by OPEC alone, I recognize that other forces are at play. I decompose the total "Surprise" into two distinct components: the portion driven by OPEC's actions ($Surprise^{OPEC}$) and the portion driven by other factors

[5] Degasperi (2023), Bruns (2021), Plagborg-Møller and Wolf (2022), Bruns and Lütkepohl (2023), Gagliardone and Gertler (2023), Caravello and Martinez-Bruera (2024), Patzelt and Reis (2024) are examples of studies that have built on Känzig (2021)

[6] Another major macroeconomic event occurring at the same time could also drive oil prices.

($Surprise^{Non\text{-}OPEC}$).

$$\text{Surprise} = \text{Surprise}^{\text{OPEC}} + \text{Surprise}^{\text{Non-OPEC}} \tag{1}$$

To estimate $Surprise^{OPEC}$, I incorporate text data to capture the sentiment and expectations surrounding OPEC announcements as expressed in news articles and reports. By doing this, I can estimate the part of the price movement that's directly attributable to OPEC's decisions, while acknowledging that other factors like geopolitical tensions or market speculation also influence the price.

2.3 Defining "Surprise" in This Framework

Within this framework, an oil price "Surprise" is the result of two driving forces: OPEC-driven surprises and other factors. The aim is to use text data to estimate the OPEC-driven surprises, while treating the residuals as those factors unrelated to OPEC.

To quantify the change in oil prices around OPEC announcements, let $\Delta P_{d,t}^h$ represent the log change in the settlement price of the h-month-ahead oil futures contract on OPEC announcement day d at time t:

$$\text{Surprise}_{d,t}^h = \Delta P_{d,t}^h = \ln\left(\frac{P_{d,t}^h}{P_{d,t-1}^h}\right) \tag{2}$$

Here, $P_{d,t}^h$ is the settlement price of the oil futures contract at time t, and $P_{d,t-1}^h$ is the price from the previous trading day.

To estimate the portion of the price change attributable to OPEC's announcement, I model the expected price change as a result of OPEC's announcement as a function of the *text-based features* ($\mathbf{X}_t$) derived from news articles[7]:

[7]These features include sentiment scores, the prevalence of key themes, and the frequency of specific

$$\text{Surprise}_{d,t}^{\text{OPEC},h} = \mathbb{E}[\Delta P_{d,t}^{h}|\mathbf{X}_t] = \alpha + \beta\mathbf{X_t} \tag{3}$$

In this equation, α represents the intercept, while β quantifies the impact of each text-based feature on the expected price change. By doing so, the model isolates the effect of OPEC's actions as reflected in the media from other influencing factors.

The residual component, capturing price movements not explained by OPEC-related text data, is defined as:

$$\text{Surprise}_{d,t}^{\text{Non-OPEC},h} = \Delta P_{d,t}^{h} - \mathbb{E}[\Delta P_{d,t}^{h}|\mathbf{X}_t] \tag{4}$$

This $\textit{Surprise}^{\textit{Non-OPEC}}$ represents the influence of external factors unrelated to OPEC, such as geopolitical events, demand shocks, or speculative trading behavior.

Combining these components, the total *oil price surprise* on OPEC announcement day d at time t is expressed as:

$$\text{Surprise}_{d,t}^{h} = \text{Surprise}_{d,t}^{\text{OPEC},h} + \text{Surprise}_{d,t}^{\text{Non-OPEC},h} \tag{5}$$

This decomposition allows us to isolate the portion of the price change that can be directly attributed to OPEC while recognizing that other forces such as geopolitical events, demand-side shocks, or speculative market behavior may also play a role in the remaining unexplained price movements.

By decomposing the total price surprise into OPEC-related and non-OPEC-related components, I show that these narratives offer additional explanatory power beyond traditional price data alone. This reflects the reality that price movements are often driven not just by

keywords related to OPEC decisions.

immediate supply-demand changes, but by how market participants expect those changes to unfold based on the surrounding narratives.

3 Data and Variables

This analysis relies on two key data sources: oil futures prices and textual data from news articles covering OPEC announcements. The oil futures data spans from April 1989 to December 2020, covering 1-month to 12-month WTI crude oil futures contracts.[8]

The choice of April 1989 as the starting point is deliberate and based on considerations regarding the availability and reliability of oil futures data. Prior to this date, trading in oil futures markets, especially for longer maturities, was limited and intermittent. As Kilian (2024) points out, issues such as incomplete futures contracts, price stagnation due to low trading volumes, and data reliability concerns undermine the accuracy of oil price surprises measured before April 1989. By starting from April 1989, we ensure that our analysis is based on consistent and active trading periods, enhancing the reliability of our results.

During this period, there were 110 OPEC meetings, and the analysis focuses on the price changes around these specific announcement dates. The text data comes from the ProQuest repository, focusing on articles related to OPEC and crude oil. The following sections provide a detailed discussion of these data.

3.1 Oil Price Data

For the analysis, I use futures contracts for West Texas Intermediate (WTI) crude oil, covering 1-month to 12-months ahead contracts. Following the methodology outlined in Kanzig (YEAR), I compute the log change in prices for each of these futures contracts. The formula for the log change in price is given by:

[8]WTI crude oil is a major benchmark in oil pricing, representing oil extracted from the U.S. Midwest.

Table 1: Summary Statistics of Futures Contracts (1-Month to 12-Months) and First Principal Component (PC1) Around OPEC Announcement Days

	1M	2M	3M	4M	5M	6M	7M	8M	9M	10M	11M	12M	PC1
Count	110	110	110	110	110	110	110	110	110	110	110	110	110
Mean	-0.50	-0.43	-0.36	-0.32	-0.29	-0.26	-0.23	-0.21	-0.19	-0.18	-0.16	-0.15	-0.11
Std.	3.10	2.96	2.86	2.76	2.67	2.59	2.51	2.45	2.39	2.33	2.27	2.22	2.65
Min	-10.72	-10.58	-10.34	-10.07	-9.87	-9.69	-9.50	-9.34	-9.18	-9.05	-8.93	-8.80	-9.81
25%	-1.94	-1.76	-1.67	-1.62	-1.58	-1.48	-1.43	-1.39	-1.25	-1.24	-1.24	-1.23	-1.27
50%	-0.16	-0.15	-0.05	0.00	0.06	0.05	0.04	0.04	-0.00	-0.01	0.01	-0.02	0.18
75%	1.50	1.50	1.49	1.42	1.34	1.30	1.25	1.26	1.20	1.13	1.06	1.08	1.47
Max	8.73	8.58	8.38	8.13	7.89	7.65	7.43	7.24	7.05	6.88	6.69	6.53	7.97

$$\Delta P_{d,t}^{h} = \ln\left(\frac{P_{d,t}^{h}}{P_{d,t-1}^{h}}\right)$$

where $P_{d,t}^{h}$ represents the settlement price of the h-month-ahead futures contract on OPEC announcement day d at time t, and $P_{d,t-1}^{h}$ is the settlement price on the previous trading day.

To capture the overall price movement, I extract the first principal component (PC1) from the 12 futures contracts.[9] This approach condenses the information from the full term structure of futures prices into a single variable that reflects the broader price dynamics across all maturities.

Table 1 presents summary statistics for the log changes in the 1-month to 12-month futures contracts, as well as the first principal component. As seen, the mean log changes are negative across all maturities, with decreasing volatility as the contract length increases. PC1, which explains the majority of the variance, has a standard deviation of 2.65, indicating significant variability in the futures price movements during OPEC announcement periods.

[9]Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction, transforming a large set of variables into a smaller one that still contains most of the information in the large set.

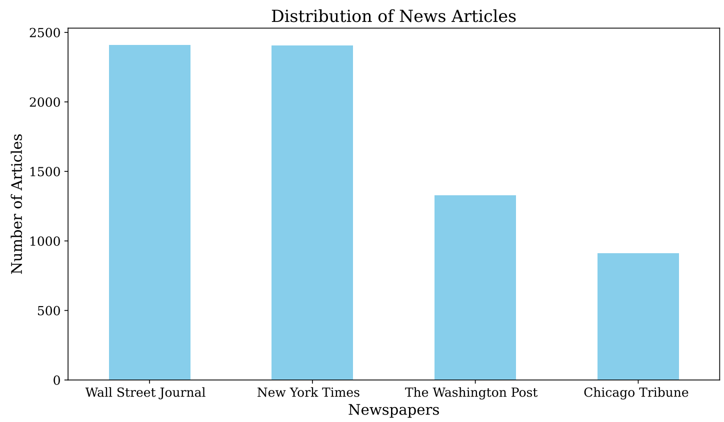


Figure 1: Distribution of news articles related to OPEC across four major newspapers

3.2 Textual Data

The text data for this analysis comes from the ProQuest Repository, using articles from four major newspapers: *The New York Times (NYT)*, *The Wall Street Journal (WSJ)*, *The Washington Post (WP)*, and *The Chicago Tribune (CT).* I collected articles from April 1989 to December 2020 using the search terms "*OPEC*" or "*Organization of the Petroleum Exporting Countries*". This ensures that the articles are focused on OPEC and its role in the oil market during this period.[10] Out of the 7,055 articles collected during this period, 2,410 were from the WSJ, 2,406 from the NYT, 1,329 from the WP, and 910 from the CT. Figure 1 shows the distribution of articles across the four newspapers.

Figure 2 provides sample news articles collected in the corpus. As shown in Figure 2, the last sentence of the third paragraph of the first newspaper article reads:

> *"This increase, which is less than 1 percent of the global oil market, comes at a time when demand is still under pressure from the impact of the coronavirus pandemic".*

Similarly, the first paragraph of the second newspaper article in Figure 2 reads:

> *"The Organization of the Petroleum Exporting Countries proposed on Thursday that oil outputs be kept at 1.5 million barrels a day, or 1.5 percent of world oil supplies, to*

[10] The search terms were selected to capture both the acronym and the full name of the organization to maximize the relevance of the collected articles.

OPEC and Russia Agree to Increase Oil Production by 500,000 Barrels a Day

By STANLEY REED

OPEC and other oil-producing nations led by Russia, trying to gauge the strength of the global economy as the coronavirus continues to rage but with vaccines on the horizon, reached a compromise on Thursday to modestly increase production in January.

But the talks revealed strains in the unwieldy group, known as OPEC Plus, which has tried to manage the oil market since 2016. These tensions could make it more difficult for the producers to stay in line with production targets as the global economy recovers in the coming months.

Under the agreement, members of the Organization of the Petroleum Exporting Countries along with Russia and other countries will increase production by 500,000 barrels a day in January and, potentially, by a similar amount in the following months. The increase, less than 1 percent of the global oil market, comes while demand is still under pressure from the impact of the coronavirus pandemic.

The group will also hold monthly meetings to sign off on further adjustments.

The arrangement was a compromise between countries that wanted to proceed with a much larger increase of two million barrels a day, which had been agreed upon at an earlier meeting, and others, led by Saudi Arabia, that preferred to maintain current production cuts, estimated at 7.7 million barrels a day, given the uncertainties stemming from the pandemic.

Saudi Aramco's Ras Tanura oil refinery and oil terminal in Saudi Arabia. Thursday's agreement was a compromise, as officials struggled to reach a consensus on output.

Reaching a deal had been surprisingly difficult. The meeting on Thursday had been delayed for two days while officials struggled to reach a consensus.

The recent news about the efficacy of vaccines to ward off the coronavirus, which has caused oil prices to climb to their highest levels since they crashed in April, probably made it harder to reach agreement. Responding to those higher prices, some oil producers saw less need to keep supplies tight and wanted to increase pumping to try to make up for almost a year of dismal oil earnings.

"As prices rise, the willingness to restrain supplies withers," said Bhushan Bahree, an executive director at IHS Markit, a research firm.

What was striking was that the United Arab Emirates, long the closest ally of Saudi Arabia, OPEC's de facto leader, proved difficult to corral. Analysts say the emirates' ambitious leaders are irritated about several issues, including a tight quota that sharply crimps their plans to substantially increase oil production and the Saudi-Russian domination of oil decision-making in recent years.

Analysts, including Mr. Bahree, say the country, the third-largest producer in OPEC after Saudi Arabia and Iraq, may be considering going its own way in oil matters.

"They don't want to be sidelined and just be a follower," said Amrita Sen, head of oil analysis at Energy Aspects, a market research firm.

Out of frustration with the difficulty of forging a consensus, Prince Abdulaziz bin Salman, the Saudi oil minister and chairman of the OPEC Plus meetings, offered on Monday to resign the chair of an influential committee that prepares the way for decisions by the group.

According to an OPEC news release on Thursday, the oil producers urged him to continue leading the committee as well as OPEC Plus meetings, saying his efforts were "highly appreciated."

ENERGY

OPEC Proposes Slashing Oil Output Over Russian Resistance

By STANLEY REED

The Organization of the Petroleum Exporting Countries proposed Thursday that oil output be curbed by 1.5 million barrels a day, or 1.5 percent of world oil supplies, to deal with the effects of the spreading coronavirus outbreak on demand.

The proposed cuts are higher than most analysts expected but seem unlikely to change the gloomy sentiment in the oil market. After the announcement, prices for Brent crude, the international benchmark, fell about 0.8 percent to $50.71 a barrel.

The group wants the cuts, which would run through June 30, to be shared with non-OPEC allies. Under the proposal, the 14 members of OPEC would make one million barrels of trims, while Russia and other allies would cut 500,000 barrels.

The deal will need to be ratified at a meeting, scheduled for Friday in Vienna, of officials from OPEC, Russia, and other oil-producing countries like Kazakhstan and Oman. Uncertainty over an agreement may explain the market's negative reaction.

Amrita Sen, chief oil analyst at Energy Aspects, a market research firm, said that Saudi Arabia, OPEC's de facto leader, hoped the Russians would sign on, but that Saudi officials were becoming impatient with Moscow's tough negotiating and relatively slight contributions to cuts.

"The chances that the deal unravels is probably higher than the market currently believes," she said, based on soundings in Vienna. "If they don't get something," she said of the Saudis, "they will walk away."

If so, prices might crash, analysts say, as they did after a 2014 OPEC meeting. In that instance, the Saudis threw up their hands and stopped trying to regulate the market after failing to persuade Russia and other producers to make trims to compensate for big increases of production from the United States.

Analysts say OPEC is trying to accomplish two things: push a reluctant Russia into new cuts, while signaling to the market that it is serious about addressing the reduced demand for oil caused by the coronavirus outbreak.

The decision to announce a significant cutback without Moscow's prior approval of the deal was seen as a gutsy call.

"It was quite a power move from OPEC today," said Helima Croft, head of commodities strategy at RBC Capital Markets, an investment bank.

Saudi Arabia wants a significant cutback because prices, which have fallen about 23 percent this year, might keep sliding, analysts say. Already prices are well below the levels that countries like Saudi Arabia need to finance their budgets, analysts say.

Saudis lose patience with Moscow's tough negotiating stance.

"This is a crisis situation," said Robert McNally, president of Rapidan Energy Group, a market research firm. He said a cut in the range proposed Thursday was "the minimum they have to do to put a floor under prices now."

Mr. McNally said he thought Russia was likely to come on board "kicking and screaming, trying to reduce its share until the last minute."

In its proposal, OPEC said the coronavirus epidemic had "a major adverse impact" on economic and oil demand forecasts. IHS Markit forecast on Wednesday that demand for the first three months of 2020 would fall by 3.8 million barrels a day, the biggest quarterly drop that its analysts had seen, topping even what occurred in the financial crisis of 2008-9.

The meetings are taking place at OPEC's headquarters in Vienna.

OPEC also said voluntary cuts of an estimated 2.1 million barrels a day, already in place, should be continued for the rest of the year. If all the existing and suggested cuts are put into effect, OPEC and its allies will have removed close to 4 percent of supply from the market.

Along with the voluntary cuts, Iran, Libya and Venezuela — all OPEC members — are producing substantially less than their potential because of a mixture of sanctions and political turmoil.

Despite this large fall in production, oil prices have been under pressure since the emergence of the coronavirus epidemic in China this winter. Abdulaziz bin Salman, the oil minister of Saudi Arabia, has been pushing hard for emergency measures to head off a glut.

Most analysts think that the recent pattern of apportioning cuts will continue. The Saudis have been taking by far the largest share of trims, helped by their allies Kuwait and the United Arab Emirates.

Russia, however, is usually a tough negotiator with OPEC and has been reluctant to agree to new cuts, forcing showdowns at the Vienna meetings. Since joining forces with OPEC more than three years ago, the Russians have succeeded in pushing Saudi Arabia, the world's largest oil exporter, to absorb the brunt of production cuts.

"The Saudis are cutting more and more, and the Russians haven't cut much at all," Bhushan Bahree, senior director at IHS Markit, a research firm, said in an interview before the current series of meetings.

A Saudi Aramco facility. OPEC seeks to cut output by 1.5 million barrels a day.

Figure 2: Sample news articles in the corpus

address the impact of the spreading coronavirus outbreak on demand".

These excerpts demonstrate that the articles often discuss OPEC's supply decisions (e.g., increasing or maintaining production levels) in the context of demand conditions (e.g., the impact of the coronavirus pandemic on global demand). Additionally, I analyzed the frequency with which demand and supply-related keywords co-occur in the corpus of news articles. Using predefined demand-related keywords (*growth, expansion, demand, consumer, travel, vehicle, industrial production, heating oil*) and supply-related keywords (*output, production, OPEC, cut, sanctions, embargo, capacity, rig count, supply chain*), I calculated the proportion of articles each year that mention both types of keywords.

Figure 3 presents the proportion of OPEC news articles containing both demand and supply keywords over time. The figure shows that a significant and consistent proportion of articles include both demand and supply terms, indicating that discussions about OPEC often encompass both aspects simultaneously.

This duality shows that OPEC news articles contain intertwined supply and demand information, reinforcing the decision to describe the shocks as "**OPEC news shocks**" rather than solely supply shocks.

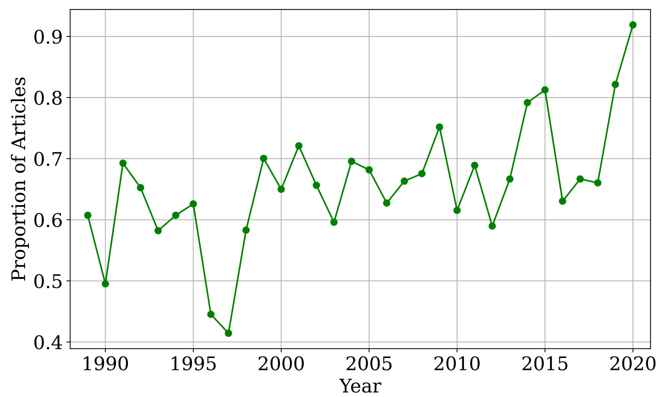


Figure 3: Proportion of OPEC News Articles Containing Both Demand and Supply Keywords Over Time

3.3 Text Preprocessing

To prepare the text data for analysis, several cleaning steps were undertaken:

1. **Removal of Noise:** HTML tags, URLs, email addresses, websites, numbers, and words consisting of only one or two characters were removed to eliminate irrelevant information.

2. **Normalization:** Accented characters, hyphens, apostrophes, and common stopwords that do not contribute significant meaning to the analysis were discarded.

3. **Stemming:** Words were stemmed to reduce them to their base forms, ensuring that

similar words like "prices" and "price" are treated equivalently.[11]

4. **Tokenization into Bigrams:** Words were tokenized into bigrams (pairs of words) to capture important phrases and contextual information.[12]

5. **Aggregation and Lagging:** All articles published on a given day were aggregated and lagged to account for any delay between publication and potential market impact.

Figure 4 presents a word cloud of the top 200 most frequent bigrams in the corpus, highlighting the prevalent themes and topics discussed in the news articles.[13]

Finally, the dataset was filtered to include only articles from the 110 OPEC meeting announcement days, ensuring that the analysis focuses on relevant events.

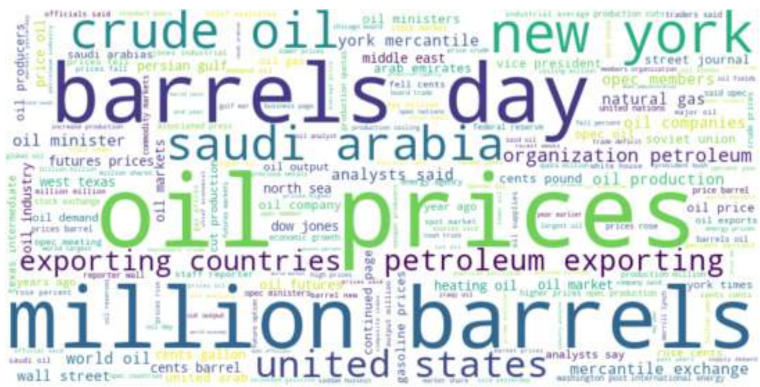


Figure 4: Word Cloud of the Top 200 Most Frequent Bigrams

[11] Stemming is a text normalization technique that reduces words to their root form to improve the efficiency of text analysis.

[12] Bigrams are pairs of consecutive words in a given text or sequence of words. Bigram tokenization helps in understanding the context by analyzing pairs of consecutive words, which can reveal meaningful phrases not apparent from single words.

[13] An Illustrative example of the text preprocessing is provided in Appendix A

4 Extracting Information from Text Data

To utilize the textual data for quantitative analysis, the news articles were converted into a structured format using the vector space model Salton et al. (1975). This model transforms each daily aggregated article into a vector, where each bigram in the article is treated as a unique term.

By applying this method to all daily aggregated articles in the corpus, a document-term matrix $\mathbf{X}_t(n, m)$ was created.[14] This matrix comprises n rows (one for each aggregated article on OPEC announcement days) and m columns (one for each unique bigram in the dataset). Each entry in the matrix indicates the frequency of a specific bigram in a given daily aggregated article.

In essence, this matrix quantifies the content of the news articles, enabling further analysis of textual patterns and their relationship with oil price movements.

4.1 Selection of Relevant Bigrams

The document-term matrix $\mathbf{X}_t$ initially yielded a total of 145,687 unique bigrams. Given the high dimensionality of this matrix, it was necessary to refine the selection to focus on the most relevant terms. To do this, I employed the Term Frequency-Inverse Document Frequency (TF-IDF) metric, a widely used method for assessing the significance of terms within individual documents relative to the entire corpus.[15]

TF-IDF prioritizes bigrams that appear frequently in individual documents but are less common across the entire corpus, thus highlighting terms that are more informative for analysis. By applying a minimum TF-IDF threshold of 0.05, the number of bigrams was reduced to 2,057.[16] This step ensured that the retained bigrams carried meaningful informational

[14] A document-term matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents.

[15] TF-IDF is commonly used in information retrieval and text mining to evaluate how important a word is to a document in a collection or corpus.

[16] The threshold of 0.05 was selected based on preliminary analyses to balance the inclusion of meaningful terms while excluding noise.

weight, allowing the analysis to focus on the most important textual patterns while minimizing noise. At this stage, the structured text data matrix $\mathbf{X}_t$ is ready for use in the estimation process.

4.2 Incorporating Text Data into the Model

The dependent variable in the model is the log change in oil prices around OPEC announcement days, denoted by $\Delta P^h_{d,t}$. The independent variables are the selected 2,057 bigrams, which represent the sentiment and narrative derived from news articles surrounding OPEC announcements. Hence, the regression equation takes the following form:

$$\Delta P^h_{d,t} = \alpha + \mathbf{X}_t \beta + \epsilon_t$$

where $\mathbf{X}_t$ represents the document-term matrix at time t, β is the vector of coefficients associated with the bigrams, and ϵ_t is the residual, representing the Non-OPEC surprise component.

4.3 Curse of Dimensionality and Elastic Net Regularization

Although the TF-IDF filtering process significantly reduced the number of features from the initial 145,687 bigrams to 2,057, the dimensionality is still high compared to the 110 OPEC meeting days in the dataset. This creates a situation where the number of features ($m = 2,057$) exceeds the number of observations ($n = 110$), commonly referred to as the curse of dimensionality ($m > n$).[17]

To address this issue, I employ the Elastic Net regularization technique (Zou and Hastie (2005)).[18] Elastic Net combines both Lasso (L_1) and Ridge (L_2) regularization methods, making it particularly well-suited for high-dimensional datasets where feature selection and

[17]The curse of dimensionality refers to various phenomena that arise when analyzing and organizing data in high-dimensional spaces that do not occur in low-dimensional settings.

[18]Elastic Net is a regularization and variable selection method that linearly combines the penalties of Lasso (L_1)(Tibshirani (1996)) and Ridge (L_2)(Hoerl and Kennard (1970)) regression.

multicollinearity might be concerns. Unlike Lasso, which may randomly select one feature among highly correlated features, Elastic Net tends to select groups of correlated features together, providing a more balanced approach.

The Elastic Net objective function minimizes the following expression:

$$\min_{\beta} \left(\frac{1}{2n} \sum_{i=1}^{n} \left(\Delta P_{d,t}^{h} - \mathbf{X}_t^{\top} \beta \right)^2 + \lambda \left(\alpha ||\beta||_1 + (1-\alpha) ||\beta||_2^2 \right) \right)$$

Here, β represents the coefficients of the bigrams, $\mathbf{X}_t$ represents the document-term matrix for observation t, λ is the regularization parameter, and α is the mixing parameter that determines the balance between Lasso and Ridge penalties. When $\alpha = 1$, the model is purely Lasso, while $\alpha = 0$ results in Ridge regularization. Elastic Net is particularly effective when $m > n$, as it encourages both sparsity (feature selection) and regularization, , thereby addressing multicollinearity and enhancing model interpretability.

4.4 Model Standardization and Cross-Validation

Prior to fitting the Elastic Net, the document term-frequency matrix $\mathbf{X}_t$ is standardized to ensure that all the features are on the same scale.[19] The model is then fitted using 5-fold time series cross-validation to determine the optimal λ and α parameters.[20] This cross-validation procedure helps ensure that the model generalizes well to unseen data by mitigating overfitting. The resulting model strikes a balance between feature selection and regularization, retaining only the most relevant bigrams for subsequent analysis.

[19]Standardization involves scaling the features so that they have a mean of zero and a standard deviation of one, which is essential for regularization techniques to perform effectively. The `StandardScaler` module from the scikit-learn library is used.

[20]Time series cross-validation accounts for the temporal ordering of data, ensuring that the model is validated on future data relative to the training set, thereby preventing look-ahead bias.

4.5 Post-Elastic Net Feature Reduction and Partial Least Squares

After fitting the Elastic Net model, I obtained a total of 405 non-zero features from the final set of estimated coefficients. While this reduction is significant, utilizing such a high number of features in the final model still presents challenges, particularly with multicollinearity. To further refine the feature set, I employ forward selection in combination with Partial Least Squares (PLS) regression (Wold et al. (2001)).[21]

4.6 Forward Selection with Partial Least Squares

Forward selection is employed to iteratively identify the most relevant features that contribute to explaining the variation in oil prices. At each step, the model assesses whether adding a particular feature improves the adjusted R^2.[22] If a feature positively enhances the model's ability to explain the dependent variable (i.e., the change in oil prices), it is retained. Conversely, features that do not significantly improve the model are excluded.

Once forward selection identifies the optimal combination of features, I apply Partial Least Squares (PLS) for dimensionality reduction. PLS is particularly advantageous in this context as it addresses multicollinearity by extracting latent variables (components) that capture the most variance in both the predictors and the dependent variable.[23] This approach facilitates a more stable regression model when numerous predictors are correlated, ensuring that the selected features possess the most explanatory power without redundancy.

This process resulted in a final set of 34 features.[24] From these, I extract the first principal component, denoted by $\mathbf{F}_t$, using Partial Least Squares (PLS). This component captures the largest amount of variance in both the predictors and the dependent variable, serving as the key predictor for the subsequent Ordinary Least Squares (OLS) regression model:

[21] PLS is a statistical method that finds the fundamental relations between two matrices (predictors and responses) by extracting latent variables that maximize covariance. See Appendix B for details.

[22] Adjusted R^2 accounts for the number of predictors in the model, providing a more accurate measure of model fit when comparing models with different numbers of predictors.

[23] PLS is beneficial when predictors are highly correlated, as it reduces dimensionality while retaining the most informative aspects of the data.

[24] See Appendix B for details.

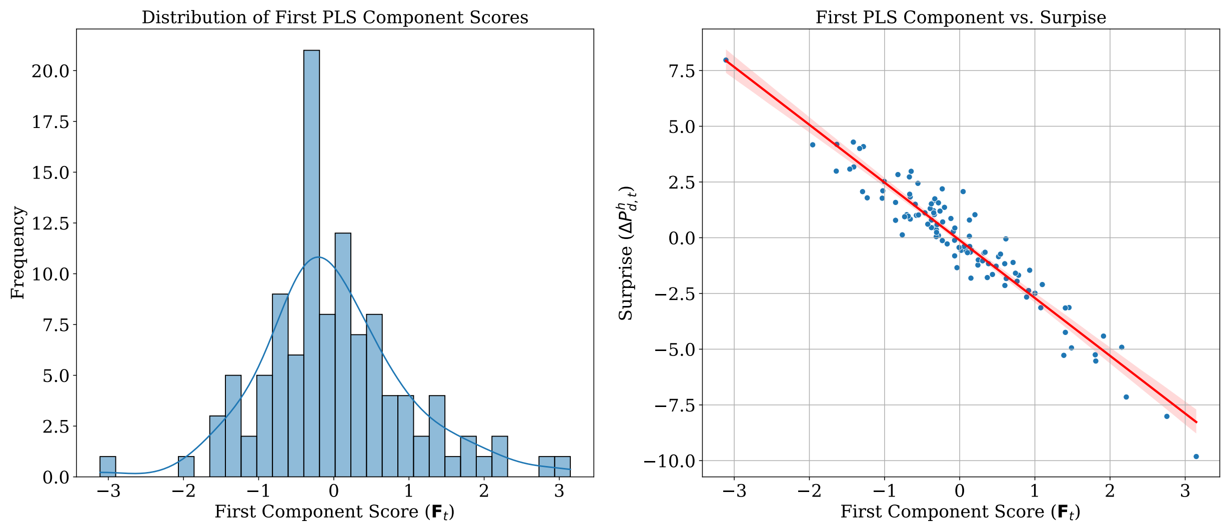


Figure 5: Distribution of the First Partial Least Squares (PLS) Component and Its Relationship with the Oil Price Surprise Series. The left panel shows the normal distribution of the first PLS component, while the right panel illustrates a linear relationship between the PLS component and the surprise series, indicating a significant inverse correlation.

$$\Delta P^h_{d,t} = \alpha + \mathbf{F}_t \beta + \epsilon_t$$

Here, $\mathbf{F}_t$ represents the PLS component derived from the selected textual features, and ϵ_t is the residual term capturing the unexplained variation in the model. Figure 5 displays the distribution of the first PLS component and its relationship with the surprise series, demonstrating that it is normally distributed and maintains a linear relationship with the surprise series.

4.7 Results

Table 2 displays the results of the OLS regression model, where the dependent variable is the log change in oil prices ($\Delta P^h_{d,t}$) around OPEC announcement days. The independent variable is the first PLS component ($\mathbf{F}_t$), which captures the most significant variance in the selected textual features.

The coefficient for the PLS component ($\mathbf{F}_t$) is negative and highly significant, indicating that

Table 2: Identification of Oil Surprise

	Surprise $(\Delta P^h_{d,t}) = \ln\left(\frac{P^h_{d,t}}{P^h_{d,t-1}}\right)$
Constant	-0.1146
	(0.074)
PLS Component ($\mathbf{F}_t$)	-2.5907***
	(0.077)
N	110
R-squared	91.4%
R-squared Adj.	**91.3%**

Notes:
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
N = Number of observations.
Standard Error in parenthesis.

an increase in the PLS component is associated with a decrease in oil prices around OPEC announcement days. The model exhibits a high explanatory power, with an R-squared value of 91.4% and an adjusted R-squared of 91.3%, suggesting that the PLS component accounts for the majority of the variability in oil price changes during OPEC announcement periods. Hence, the results can be interpreted as this:

> *"Of the total 100% surprise change in oil prices occurring on OPEC announcement days, 91.3% can be attributed to changes in oil price expectations resulting from OPEC's announcement, while the remaining 8.7% is due to other factors unrelated to OPEC."*

$$\underbrace{\text{Surprise}}_{100\%} = \underbrace{\text{Surprise}^{\text{OPEC}}}_{91.3\%} + \underbrace{\text{Surprise}^{\text{Non-OPEC}}}_{8.7\%}$$

4.8 Model Validation

To assess the robustness and generalizability of the regression model, a 5-fold time series cross-validation was conducted.[25] Table 3 presents the cross-validated R-squared scores for each fold, along with the average and standard deviation.

[25] Time series cross-validation respects the temporal order of data, ensuring that the model is trained on past data and validated on future data, thereby preventing look-ahead bias.

Table 3: Cross-Validated R-squared Scores

Fold	R-squared
1	85.1%
2	86.0%
3	80.3%
4	84.7%
5	95.9%
Average	**86.4%**
Standard Deviation	**0.024**

The cross-validation results demonstrate consistent performance across the different folds, with an average R-squared of 86.4% and a low standard deviation of 0.024. This indicates that the model maintains high explanatory power and generalizes well to unseen data, effectively avoiding overfitting. Figure 6 illustrates the time series of the three surprise components: total surprise, OPEC-attributed surprise, and non-OPEC surprise. This visual allows us to observe how each component has evolved over time, providing a clearer understanding of the relative contributions of OPEC-driven and non-OPEC factors.

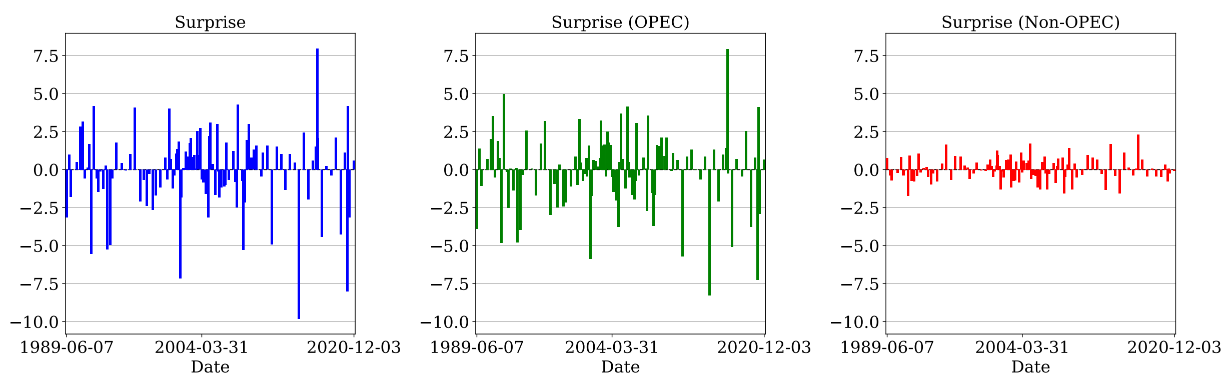


Figure 6: Time series plot of the three oil price surprise components around OPEC announcement days: (1) Total Surprise, (2) OPEC-Attributed Surprise, and (3) Non-OPEC-Attributed Surprise.

5 Text Features and Surprises as External Instruments

Having established that the news narrative is a reliable predictor of market behavior, this section demonstrates how the **text features** and the **surprise series** can be used as external instruments to identify oil price shocks.[26] By employing a Structural VAR (SVAR) model, I assess how well the textual data, alongside the OPEC and non-OPEC surprise components, explains the dynamics of oil price movements and their impact on the broader economy.

The goal here is to test whether the PLS component (Text Features) serves as a strong and reliable instrument, generating comparable results to more traditional instruments used in the literature.

Before incorporating the instruments into the SVAR model, the text features and the three surprise series are transformed into monthly frequencies, covering the period from June 1989 to December 2020. Following the transformation process outlined by Kilian (2024) which accounts for the timing of the daily surprises within the month, the transformed series ensures that the data reflects macroeconomic trends while accounting for potential delays in market reactions.

Figure 7 presents the time series plot of the two potential instruments: Text Features (PLS component ($\mathbf{F}$)) and Surprise series. A visual inspection reveals no significant signs of heteroscedasticity across these instruments.

5.1 Diagnostic Checks: Autocorrelation and Forecastability

To further validate the suitability of the instruments, I conduct several diagnostic checks. First, I assess the presence of autocorrelation by analyzing the autocorrelation function (ACF) of each instrument. Figure 8 shows the ACF plots, which confirm that none of the

[26] The total surprise series has been the widely used approach in the literature, proposed by Känzig (2021). The use of residuals as instruments is a very popular approach in the macroeconomic shocks identification literature. In the crude oil literature, Wu and Cavallo (2012) used the residuals directly as shocks. To the best of my knowledge, this is the first paper that proposes the use of text features as external instruments in a SVAR model. This is a significant contribution to the literature.

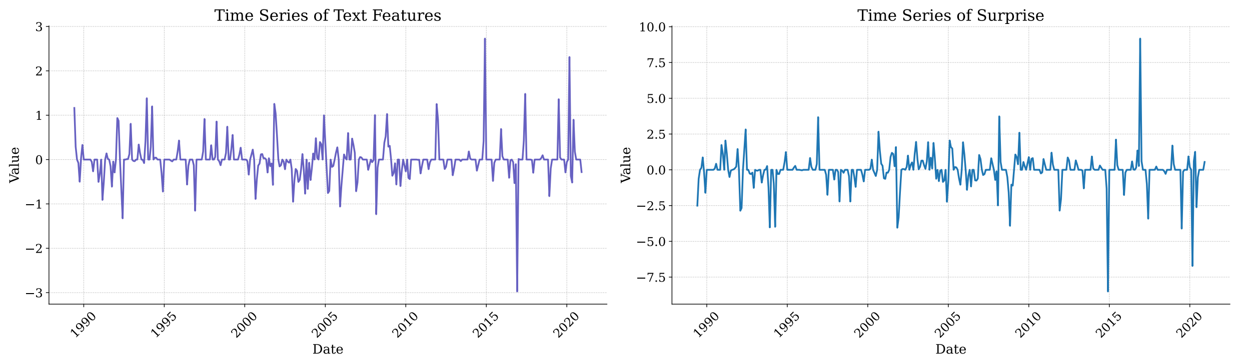


Figure 7: Time series plot of aggregated external instruments: Text Features and Surprise.

external instruments exhibit significant serial correlation. This ensures that the instruments are not temporally dependent, making them suitable for use in a VAR framework.

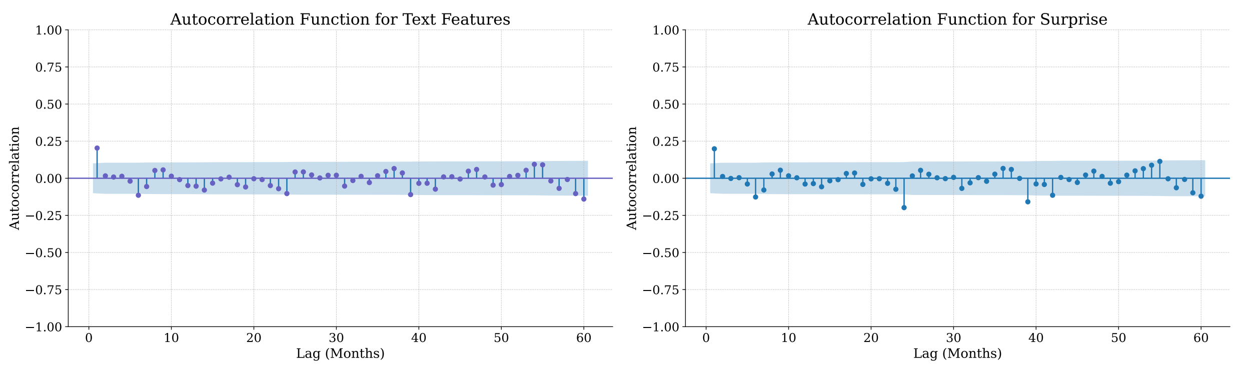


Figure 8: Autocorrelation plots for external instruments: Text Features and Surprise.

Next, I perform a causality test using the Granger Causality Test (GCT). The Granger Causality Test is a widely accepted method in econometrics for determining whether one time series can predict another. In this analysis, I apply GCT to examine whether any of the macroeconomic variables (consumer price index (CPI), world oil production, world oil inventories, producer price index (PPI), and U.S. industrial production) Granger-cause the external instruments (text features and surprise series). The objective is to ensure that these macroeconomic variables do not have predictive power over the instruments, which is essential for their validity as exogenous instruments in the IV-SVAR framework.

A significant Granger causality result (typically a p-value below 0.05) would imply that the macroeconomic variable can forecast the instrument, violating the exclusion restriction necessary for valid instruments. Therefore, it is crucial that none of the macroeconomic variables Granger-cause the external instruments. The results of the Granger Causality Tests, presented in Table 4, indicate that none of the tested macroeconomic variables have a statistically significant predictive relationship with the instruments. This means that the text features and surprise series are orthogonal to the macroeconomic variables, validating their suitability as external instruments in the SVAR model.

Table 4: Granger Causality Test Results: Macroeconomic Variables on External Instruments

	Text Features	Surprise
Real Oil Price	0.6364	0.3508
World Oil Production	0.4458	0.5364
World Oil Inventories	0.9869	0.9883
Producer Price Index	0.5412	0.3537
US Industrial Production	0.4707	0.4780
Consumer Price Index	0.7811	0.6446

Notes: The table displays the p-values from the Granger Causality Tests, assessing whether each macroeconomic variable Granger-causes the external instruments (Text Features and Surprise). A p-value below 0.05 indicates significant Granger causality. All p-values in this table are above 0.05, suggesting no significant causality.

The diagnostic checks confirm that the text features and surprise series possess the desirable characteristics to be used as external instruments. With no significant autocorrelation and no evidence of forecastability from macroeconomic variables, these instruments are well-suited for use in the structural VAR model.

5.2 Structural VAR Model with External Instruments

Consider a reduced-form VAR model, without exogenous variables, written as:

$$\mathbf{y}_t = \mathbf{C}_1\mathbf{y}_{t-1} + \cdots + \mathbf{C}_p\mathbf{y}_{t-p} + \mathbf{v}_t$$

where $\mathbf{y}_t$ is a $n \times 1$ vector of endogenous variables, $\mathbf{C}_1, \ldots, \mathbf{C}_p$ are $n \times n$ coefficient matrices, and $\mathbf{v}_t$ is a $n \times 1$ vector of residuals. The residuals $\mathbf{v}_t$ have a covariance matrix $\mathbf{\Omega}$, which is $n \times n$. Following standard SVAR practice, we assume that these residuals can be expressed as a linear combination of independent structural shocks ε_t:

$$\mathbf{v}_t = \mathbf{D}\varepsilon_t$$

where $\mathbf{D}$ is a $n \times n$ matrix, and ε_t is a $n \times 1$ vector of structural shocks. The covariance matrix of the reduced-form residuals is then related to $\mathbf{D}$ by:

$$\mathbf{\Omega} = \mathbf{D}\mathbf{D}'$$

where $\mathbb{E}[\varepsilon_t\varepsilon_t'] = \mathbf{I}_n$, with $\mathbf{I}_n$ being the identity matrix. While $\mathbf{D}$ has n^2 elements, the data only provide $n(n-1)/2$ pieces of information from the symmetric matrix $\mathbf{\Omega}$. To estimate the remaining elements of $\mathbf{D}$, we impose additional structure on $\mathbf{D}$. For example, we could assume that $\mathbf{D}$ is lower triangular, which allows the diagonal and lower-diagonal elements to be estimated.

In the case of an instrumental-variables SVAR (IV-SVAR) model, we aim to estimate certain columns of $\mathbf{D}$. Instead of relying solely on covariance restrictions, IV-SVAR models incorporate external instruments to provide extra information. Let z_t be an external instrument, which satisfies the following conditions:

$$\mathbb{E}[z_t\varepsilon_{1,t}] = \lambda_z \quad \text{and} \quad \mathbb{E}[z_t\varepsilon_{j,t}] = 0 \quad \forall j \neq 1$$

Here, $\varepsilon_{1,t}$ is referred to as the *target shock*, and z_t is the external instrument, which is correlated with the target shock $\varepsilon_{1,t}$, but uncorrelated with other shocks. This allows us to recover the first column of $\mathbf{D}$.

The relationship can be expressed as:

$$\mathbb{E}[\mathbf{v}_t z_t] = \mathbb{E}[\mathbf{D}\varepsilon_t z_t] = \mathbf{D}_1 \lambda_z$$

This equation shows that the column of $\mathbf{D}$ corresponding to the target shock $\varepsilon_{1,t}$ is identified up to the scale factor λ_z. The IV-SVAR model does not aim to identify the other columns of $\mathbf{D}$, which correspond to non-target shocks.

Finally, the IV-SVAR model identifies the first column of $\mathbf{D}$, and the final step is to normalize the impact effect on a specific target variable to 1.

This approach contributes to the literature by using text features derived from OPEC-related news as a novel external instrument. To my knowledge, this is the first time text data has been applied in such a capacity within a structural VAR framework.

5.3 Model Specification

The VAR model adopted in this paper includes six variables: the real price of oil, world oil production, world oil inventories, world industrial production, U.S. industrial production, and the U.S. consumer price index (CPI). The data spans from January 1974 to December 2020, estimated in log levels with a lag order of 12. The shocks are normalized to induce a 10% increase in the real oil price, which is calculated using WTI spot prices and deflated using U.S. CPI.[27]

This model allows us to trace the effects of oil price shocks identified using the PLS component and the surprise series, examining their impacts on the macroeconomic variables included.

[27]OPEC meetings held before 1989.04 are assigned a value of zero due to reasons explained in section 3.

5.4 Empirical Results

5.4.1 Strength of External Instruments

The strength of the external instruments is assessed using the first-stage F-statistics and R^2 values from the regressions of the oil price residuals on the PLS component (Text Features) and the surprise series.

Table 5: Instrument Strength: First-Stage Regression Results

	Text Features ($\mathbf{F}$)	Surprise (PC1)
F-stat	29.24	30.61
F-stat (robust)	18.66	19.54
R^2	5.05	5.27
R^2 (adjusted)	4.88	5.10
Observations	552	552

Notes: The table shows the results of the first-stage regressions of the oil price residual $\hat{v}_{1,t}$ on the PLS component (Text Features ($\boldsymbol{F}$)) and the surprise series. F-statistics/robust above 10 indicate strong instruments.

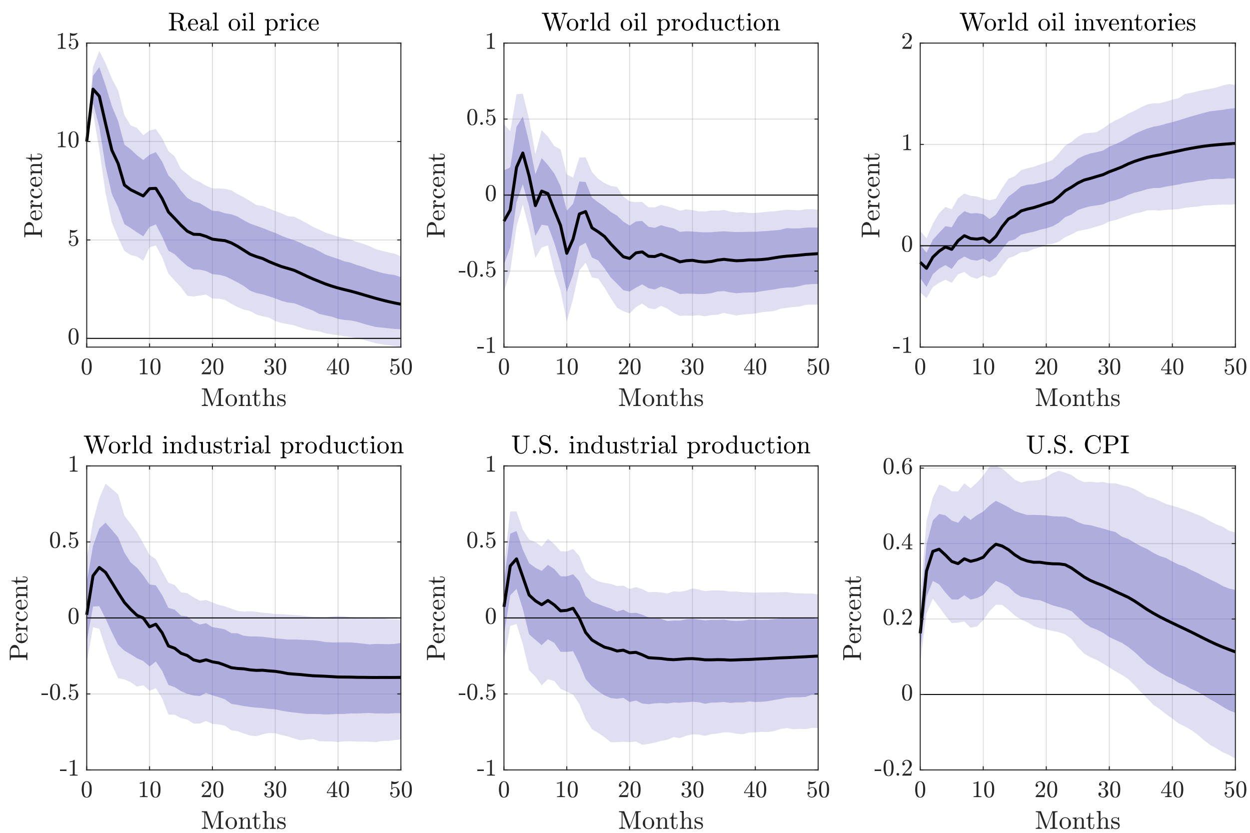
From Table 5, the F-statistics for the text features, and surprise series are all well above the conventional threshold of 10, indicating that these instruments are strong predictors of the oil price residuals. The robust F-statistics for these instruments also remain high, further confirming their strength.

The R^2 values for the text features and surprise series hover around 5%, indicating that the instruments explain a moderate amount of variation in the oil price residuals, which is typical in settings with external instruments.

The comparable performance of the proposed instrument (text features) with the traditional instrument (surprises) is a significant validation of the method used in this paper. The high F-statistics and R^2 values for the text features show that the sentiment and narratives extracted from OPEC-related news articles are just as effective in identifying oil price shocks as the more established instruments in the literature.

5.4.2 Impulse Response Analysis - Text Features as External Instrument

Figure 9: Impulse Response Functions: A shock to oil price expectations driven by OPEC news.



First stage regression: F: 29.24, robust F: 18.66, R^2: 5.05%, Adjusted R^2: 4.88%

Notes: (Impulse Response) A shock to oil price expectations driven by OPEC news. The solid line is the point estimate and the dark and light shaded areas are 68% and 90% confidence bands respectively, based on 10,000 bootstrap replications.

Figure 9 presents the IRFs for the model using text features as external instrument.

A shock to oil price expectations driven by OPEC news leads to a significant and immediate increase in the real price of oil. The price jumps by 10% following the shock, with the peak impact occurring within the first month (12.64%). Over time, the price effect diminishes, returning to pre-shock levels after approximately 50 months. This indicates that OPEC news shocks cause a temporary but substantial spike in oil prices, which eventually normalize as the market adjusts.

World oil production shows a modest initial decline (which is not statistically significant) in response to the shock, becoming more pronounced after about 10 months. The decline stabilizes after approximately 20 months, remaining slightly negative thereafter. This gradual production response suggests that producers adjust output levels over time, possibly in reaction to both supply signals and changing demand conditions conveyed by the OPEC news.

World oil inventories experience a slight initial decrease following the shock but begin to rise steadily after around 10 months, increasing by about 0.75% by the end of the 50-month horizon. The initial decline in inventories implies that immediate consumption demand draws down existing stocks. The subsequent accumulation of inventories may reflect delayed adjustments in storage behavior as market participants reassess future supply and demand expectations.

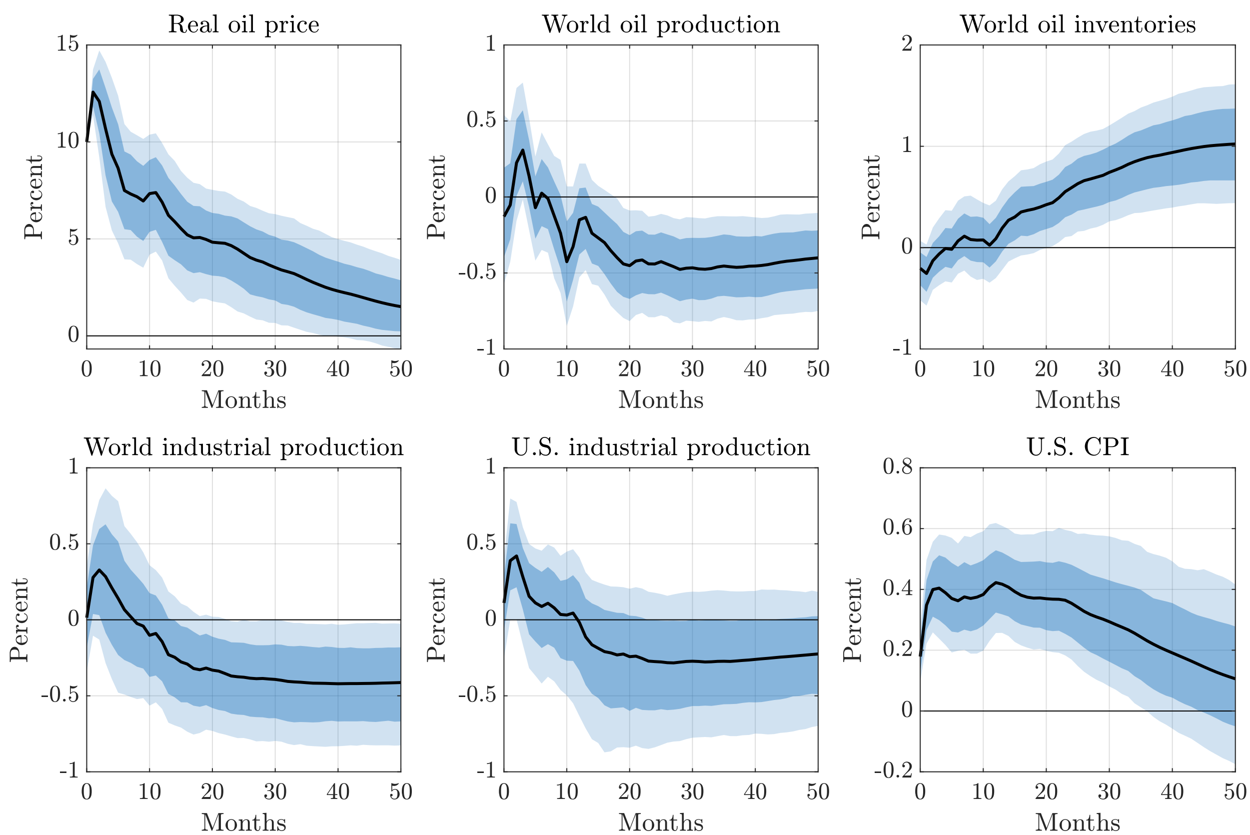
World industrial production does not significantly respond to the shock on impact. However, it begins to decline after two months and continues to fall persistently beyond the 50-month horizon. This behavior presents a puzzle. If the shock identified is a supply shock, as suggested by Känzig (2021), economic theory would expect world industrial production to fall immediately following the shock, and world oil inventories to increase on impact. Here, we observe a decline in world oil inventories on impact, suggesting that the shock may be a flow demand shock. If this were the case, one would expect a significant increase in world industrial production, but this does not occur. This presents a puzzle, as echoed in Degasperi (2023) and Känzig (2021).

Similarly, **U.S. industrial production** exhibits a comparable behavior to world industrial production. It does not react significantly on impact but falls persistently thereafter.

The **U.S. Consumer Price Index (CPI)** increases by approximately 0.2% following the shock, with this inflationary effect persisting for about 23 months before gradually returning to pre-shock levels. The rise in CPI reflects both the direct impact of higher oil prices on consumer energy costs and potential demand-driven price pressures. As the economy adjusts and oil prices normalize, inflationary pressures subside.

5.4.3 Impulse Response Analysis - Surprises as External Instrument

Figure 10: Impulse Response Functions: A shock to oil price expectations driven by OPEC news.



First stage regression: F: 30.61, robust F: 19.54, R^2: 5.27%, Adjusted R^2: 5.10%

Notes: (Impulse Response) A shock to oil price expectations driven by OPEC news. The solid line is the point estimate and the dark and light shaded areas are 68% and 90% confidence bands respectively, based on 10,000 bootstrap replications.

Figure 10 shows the IRFs using the surprise series, which is the observed change in oil prices around OPEC announcement days, as the external instrument. This approach is analogous to that of Känzig (2021), where the deviation from expected price serves as a proxy for OPEC-related oil price shocks.

The IRFs from using the surprise series as the instrument exhibit nearly identical patterns to those obtained using the text features as the external instrument (Figure 9). Both methods show consistent responses in terms of real oil prices, production, inventories, and macroeconomic variables like U.S. and world industrial production and CPI.

The similarity in IRFs strongly suggests that the model is robust in capturing the dynamics of oil price shocks around OPEC announcement days. The expected price changes, based on the textual analysis of OPEC-related news, seem to align closely with the actual observed changes in oil prices and the broader sentiment captured through news narratives.

5.4.4 Mixed Nature of Shocks in Impulse Response Analysis

The IRFs obtained using the text features as the external instrument closely resemble those from using the surprise series. The puzzle present in the IRFs of both instruments suggest that the instruments (text features and surprises) are likely capturing more than one type of shock, potentially leading to contradictory or mixed interpretations.

This issue, observed in both Degasperi (2023) and Kilian (2024), points to the possibility that these instruments are not isolating a single, clean shock but rather reflect a mixture of *supply* and *demand shocks* related to OPEC announcements. The high and robust F-statistics reported in the first-stage regressions (Table 5) suggest that the instruments are strong and significant predictors of oil price changes. However, these statistics also indicate that the instruments may be picking up multiple, overlapping shocks.

The results from the IRFs provide further support for this hypothesis. Both methods (using text features or surprises) yield nearly identical patterns, with the IRFs suggesting responses that could be associated with both *flow demand* and *flow supply* shocks. For instance, the increase in oil prices is consistent with both a positive demand shock and a negative supply shock, while the mixed responses in world industrial production and inventories imply that more than one shock type is influencing the dynamics.

This overlap in shocks calls for the need to separate OPEC-related shocks more carefully to measure their distinct effects. Without such a separation, the instruments may conflate the demand-side effects driven by sentiment with supply-side effects driven by OPEC's production decisions, thereby obscuring the true nature of the shocks.

In the next section, I disentangle these shocks by distinguishing between *OPEC-demand news shocks* (related to market expectations) and *OPEC-supply news shocks* (related to actual production or supply adjustments). This approach will help isolate the effects of OPEC's announcements on oil prices more precisely and provide a clearer picture of how these shocks propagate through the economy.

5.5 Disentangling OPEC News Shocks

In this section, I disentangle the news surrounding OPEC announcements, between two key drivers: **sentiment-driven news**, which reflects market expectations about oil demand, and **objective news**, which pertains to actual OPEC production decisions and supply adjustments.

OPEC's **supply announcements** are typically straightforward and objective, detailing production cuts or increases that directly impact the physical supply of oil. For instance, a headline such as *"OPEC announces a 1 million barrel per day production cut"* conveys an objective change in supply.

In contrast, **sentiment-driven news** often reflects the market's expectations and opinions about future demand. Analysts and journalists may speculate on how OPEC's actions will affect global oil demand, such as *"Analysts expect oil prices to rise following OPEC's production cut amidst strong demand"*.

OPEC-related news can be broken down into two components:

$$\text{OPEC News} = \text{OPEC Demand News} + \text{OPEC Supply News}$$

In other words, the total news surrounding OPEC announcements includes both sentiment-driven demand expectations and factual supply-side adjustments.

Computing the News Sentiment To quantify market sentiment from OPEC-related news, I compute sentiment scores using the textual content of articles published on OPEC meeting days. The sentiment index is constructed by counting the occurrences of *positive* and *negative* words, based on the oil-specific dictionary from Loughran et al. (2019):

$$\text{News Sentiment}(\mathbf{S}_t) = \text{Positive Count} - \text{Negative Count}$$

Here, *positive count* and *negative count* represent the frequency of positive and negative words found in the dictionary, respectively. These words reflect the tone of news coverage related to OPEC announcements. To ensure comparability across different OPEC announcements, the sentiment scores are normalized to fall within the range $[-3, 3]$.[28]

To estimate the non-sentiment portion of OPEC news, which represents OPEC supply news, I regress the sentiment index against the surprise component attributed to OPEC (estimated in Table 2). The unexplained portion (the residual) represents the change in oil prices that occurs due to OPEC supply news. The regression equation is expressed as follows:

$$\text{Surprise}_{d,t}^{\text{OPEC}} = \beta_0 + \mathbf{S}_t\gamma + \mu_t$$

where $\text{Surprise}_{d,t}^{\text{OPEC}}$ is the surprise in oil prices attributed to OPEC on announcement day d at time t, β_0 is the constant term, $\mathbf{S}_t$ is the sentiment index at time t, which captures market sentiment, γ is the coefficient measuring the impact of sentiment on the OPEC surprise, μ_t is the residual, representing the unexplained portion of the price change, which is interpreted as the OPEC supply news effect.

Table 6 presents the results of the regression model, where the dependent variable is the surprise change in oil price attributable to OPEC on announcement days. The independent variable is the sentiment index ($\mathbf{S}_t$), which captures the demand-side news related to OPEC announcements.
The coefficient for the sentiment index ($\mathbf{S}_t$) is positive and statistically significant, indicating that an increase in positive sentiment regarding market demand and expectations is associated with

[28] I chose the scale from -3 to 3 to align with the natural standardized scale of the text features ($\mathbf{F}$), ensuring consistency across the broader analysis.

Table 6: Regression of Sentiment on OPEC Price Surprise

	$\text{Surprise}^{\text{OPEC}}$
Constant	-0.2204
	(0.074)
Sentiment Index	0.1969***
	(0.077)
N	110
R-squared	7.0%
R-squared Adj.	**6.2%**

Notes:
** $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*
N = Number of observations.
Standard Error in parenthesis.

an increase in oil price movements. Despite its statistical significance, the model exhibits low explanatory power, with an R-squared value of 7.0% and an adjusted R-squared of 6.2%. This suggests that OPEC supply news accounts for the majority of the variability in oil price changes during OPEC announcement periods.

Thus, the results can be summarized as follows:

> *"Of the 91.3% of the surprise change in oil prices attributable to OPEC on announcement days, 85.6% is due to OPEC supply news, while 5.7% can be attributed to OPEC demand news."*

$$\underbrace{\text{Surprise}^{\text{OPEC}}}_{91.3\%} = \underbrace{\text{OPEC Demand News}}_{5.7\%} + \underbrace{\text{OPEC Supply News}}_{85.6\%}$$

The results suggest that, unlike in Känzig (2021), where 100% of the surprise (change in oil price) is attributed to OPEC supply news, this paper finds that approximately 86% of the surprise is driven by pure OPEC supply news shocks. To assess the macroeconomic implications of OPEC demand news and OPEC supply news, I use the estimated OPEC demand (captured by the sentiment index) and supply news as instruments in an IV-SVAR framework. In the following section, I discuss the results.

5.6 Results

5.6.1 Strength of External Instruments

I assess the strength of the instruments using the first-stage F-statistics and R^2 values from the regressions of the oil price residuals on the sentiment index (OPEC demand news features) and OPEC supply news features.

Table 7: Instrument Strength: First-Stage Regression Results

	OPEC Demand News (**S**)	OPEC Supply News
F-stat	12.84	20.84
F-stat (robust)	5.54	11.49
R^2	2.28	3.65
R^2 (adjusted)	2.10	3.48
Observations	552	552

Notes: The table shows the results of the first-stage regressions of the oil price residual $\hat{v}_{1,t}$ *on the OPEC demand news (sentiment index (**S**)) and OPEC supply news. F-statistics above 10 indicate strong instruments.*

As shown in Table 7, the F-statistics for both OPEC demand and supply news features are well above the conventional threshold of 10, indicating that these instruments are strong predictors of the oil price residuals. However, the robust F-statistic for OPEC demand news is notably lower, suggesting that while OPEC supply news remains a strong instrument under heteroscedasticity conditions, the OPEC demand news features may not.

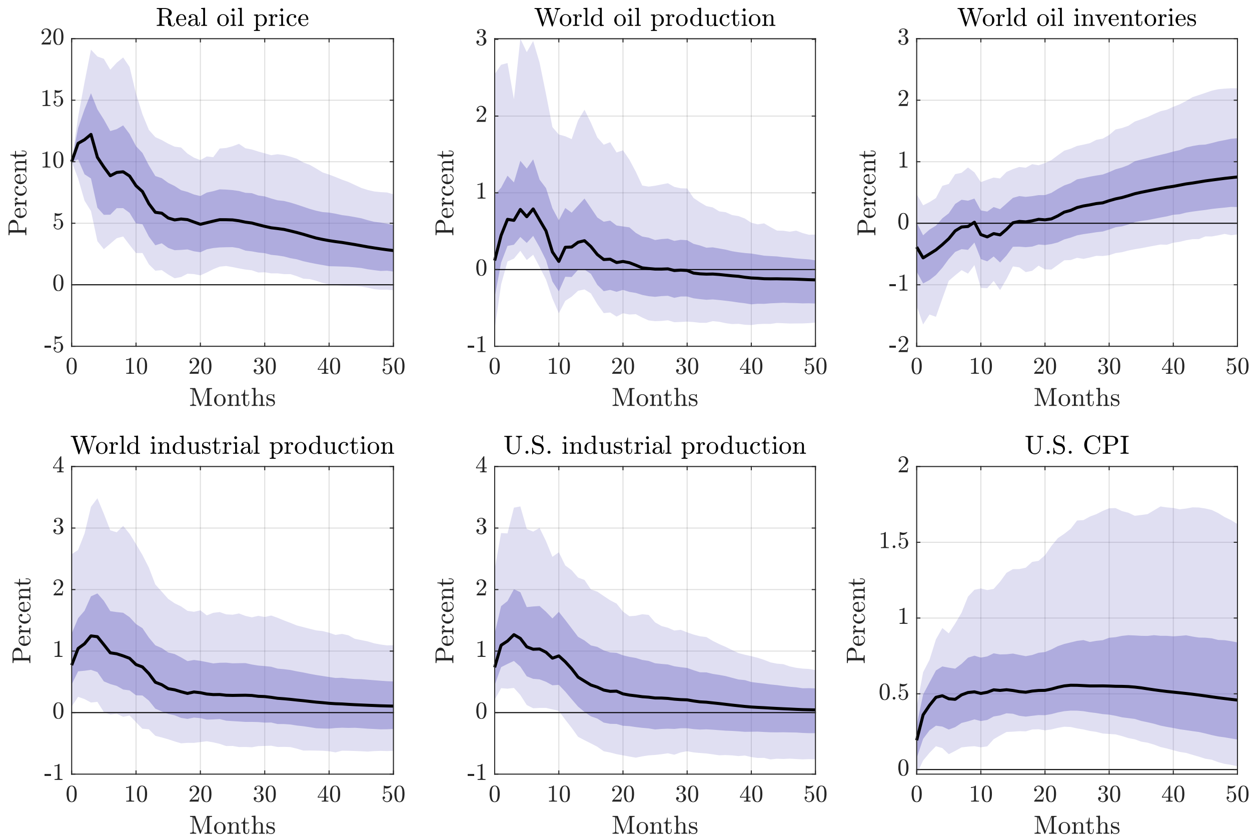
The R^2 values for OPEC demand and supply news features hover around 2% and 3%, respectively, indicating that these instruments explain a moderate amount of variation in the oil price residuals. This is typical in settings with external instruments.

5.6.2 Impulse Response Analysis - OPEC Demand News

Figure 11 presents the IRFs for the model using sentiment index as external instrument.

A shock to oil demand expectations driven by OPEC demand news leads to a significant and immediate increase in the real price of oil. The price rises by 10% following the shock, reaching its peak in the third month (12.97%). Over time, the effect gradually diminishes, with the price

Figure 11: Impulse Response Functions: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 12.84, robust F: 5.54, R^2: 2.28%, Adjusted R^2: 2.10%

Notes: The figure shows the impulse responses to a shock in oil demand expectations. The solid line represents the point estimate, while the dark and light shaded areas indicate the 68% and 90% confidence bands, respectively. Confidence intervals are based on 10,000 bootstrap replications.

returning to pre-shock levels after approximately 50 months. This suggests that OPEC demand-related news causes a substantial but temporary spike in oil prices, which eventually stabilize as the market adjusts to new information.

World oil production shows a modest initial increase in response to the shock, though this increase is not statistically significant. After two months, the production response becomes more pronounced, rising by 0.64%. This increase stabilizes for a brief period before beginning to decline after six months, eventually returning to pre-shock levels by the 20th month. This pattern suggests that producers initially increase output in response to positive demand expectations, but adjust their production levels downward as market conditions evolve.

World oil inventories exhibit a slight initial decline, falling by approximately 0.35% following the shock. However, inventories begin to rise steadily after the first month, increasing by around 0.05% and returning to pre-shock levels by the sixth month. The initial drawdown in inventories likely reflects immediate consumption needs, as market participants reduce stockpiles to meet increased demand. The subsequent rise in inventories suggests that, as demand expectations adjust, market participants begin replenishing their stocks.

World industrial production responds positively and significantly to the shock, increasing by 0.75% on impact and reaching a peak of 1.22% in the fourth month. However, this increase is short-lived, as industrial production begins to decline after the fourth month and gradually returns to pre-shock levels over the 50-month horizon. This behavior contrasts sharply with the results obtained using the surprise series and helps resolve the earlier contradiction. The initial decline in world oil inventories, combined with the increase in world oil production, supports the interpretation that the shock is driven by **demand expectations**, rather than a supply shock as proposed by Känzig (2021). These results indicate that the **news sentiment index** more accurately identifies the nature of the shock, aligning with economic theory's expectation of increased industrial production in response to demand-driven shocks.

Similarly, **U.S. industrial production** exhibits a comparable response to world industrial production. It increases significantly by 0.73% on impact, reaching a peak of 1.22% in the third month, before beginning a persistent decline, returning to pre-shock levels by the 30th month.

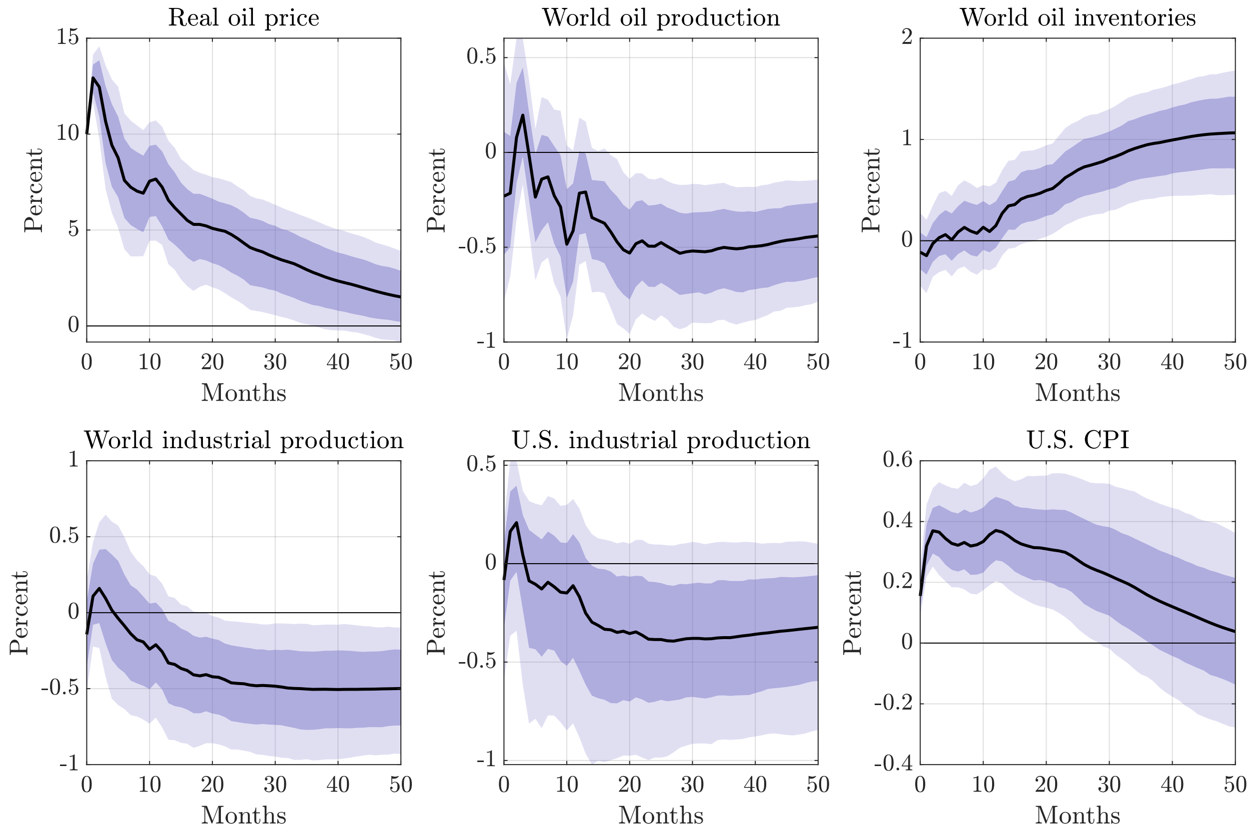
The **U.S. Consumer Price Index (CPI)** also rises following the shock, increasing by approximately 0.2%. This inflationary effect persists for about 23 months before gradually subsiding as the economy adjusts and oil prices normalize. The increase in CPI reflects both the direct impact of higher oil prices on consumer energy costs and the indirect price pressures stemming from increased demand expectations. As oil prices stabilize, inflationary pressures recede, bringing CPI back to pre-shock levels.

5.6.3 Impulse Response Analysis - OPEC Supply News

Figure 12 presents the IRFs for the model using OPEC supply news as the external instrument.

A shock to oil supply expectations driven by OPEC supply news leads to a significant and

Figure 12: Impulse Response Functions: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 20.84, robust F: 11.49, R^2: 3.65%, Adjusted R^2: 3.48%

Notes: The figure shows the impulse responses to a shock in oil supply expectations. The solid line represents the point estimate, while the dark and light shaded areas indicate the 68% and 90% confidence bands, respectively. Confidence intervals are based on 10,000 bootstrap replications.

immediate increase in the real price of oil. The price rises by 10% following the shock, peaking at 12.93% in the first month. Over time, the effect gradually diminishes, with the price returning to pre-shock levels after approximately 50 months. This suggests that OPEC supply-related news causes a substantial but temporary spike in oil prices, which eventually stabilize as the market adjusts to new information.

World oil production shows an initial decline of 0.23% in response to the shock, though this decline is not statistically significant. Part of the initial drop reverses within three months. Afterward, the production decline becomes more pronounced, reaching its lowest point in the 10th month and continuing to decline persistently through the 50th month. This pattern suggests that producers decrease output in response to negative supply news, causing the spike in the real price of oil.

World oil inventories do not respond significantly on impact. Initially, there is a slight decline of approximately 0.35% following the shock. However, inventories begin to rise steadily after the first month. The initial drawdown likely reflects that, in response to a supply shock, oil inventories decrease as stored oil is used to smooth the production of refined products. The subsequent rise in inventories suggests that, as supply expectations adjust, market participants begin replenishing their stocks.

World and U.S. industrial production also do not respond significantly on impact, showing initial declines of 0.14% and 0.08%, respectively. Both begin to fall significantly and persistently after the second month. This pattern suggests that the increase in real oil prices from the supply shock leads to a decrease in real economic activity. This behavior is consistent with the effect of a flow supply shock described in Kilian and Murphy (2014).

The **U.S. Consumer Price Index (CPI)** rises following the shock, increasing by approximately 0.2%. This inflationary effect persists for about 23 months before gradually subsiding as the economy adjusts and oil prices normalize. The increase in CPI reflects both the direct impact of higher oil prices on consumer energy costs and the indirect price pressures stemming from higher oil prices. As oil prices stabilize, inflationary pressures recede, bringing CPI back to pre-shock levels.

The results of this analysis highlight the role that textual data plays in understanding the true nature of oil price shocks around OPEC announcements. While traditional approaches, such as those by Känzig (2021), rely primarily on price data to capture supply shocks, this paper demonstrates that a significant portion of the oil price surprise is also driven by demand-side expectations, as reflected in news sentiment.

5.6.4 Historical Decomposition and Forecast Error Variance Decomposition

The historical decomposition (HD) provides a breakdown of the contributions from different structural shocks over time, helping to understand how each shock has influenced the real price of oil during key historical periods. The forecast error variance decomposition (FEVD), on the other hand, allows us to quantify the relative importance of each shock in explaining the variations in oil prices and other macroeconomic variables over different horizons.

In this section, I present the HD and FEVD results, which offer a clearer perspective on the role of OPEC announcements in driving fluctuations in crude oil prices and broader economic indicators.

Figures 13 and 14 present the historical decomposition of the real oil price, capturing the contributions from both OPEC supply and demand news shocks over the period from 1975 to 2020. These figures illustrate how major events influenced oil prices through the lens of both supply and demand dynamics.

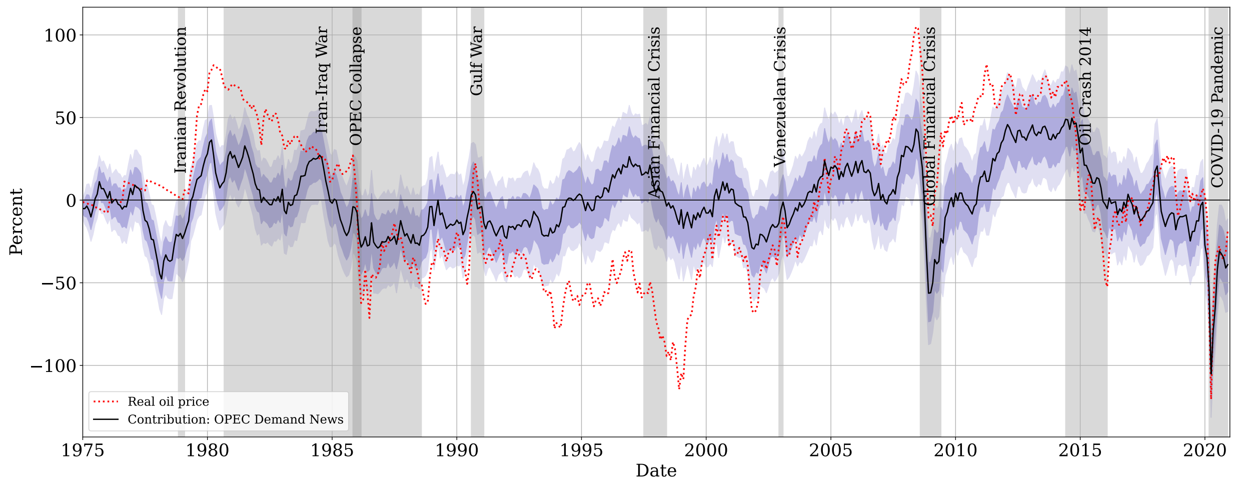
During significant geopolitical disruptions, such as the **Iranian Revolution** and the **Iran-Iraq War**, **supply news** played a **larger role** in driving oil price changes compared to **demand news**. The supply-side contributions (represented by the black line in Figure 14) were marked by substantial positive deviations during these periods, indicating that expectations around constrained supply had a strong effect on the market. **Demand news**, on the other hand, showed more modest contributions during these crises, suggesting that the market's primary focus was on disruptions to supply rather than changes in demand.

In contrast, during economic events like the **Global Financial Crisis** and the **COVID-19 Pandemic**, **demand news** took on a more **dominant role**. During these periods, expectations around a significant drop in economic activity and oil demand were reflected in the decomposition (Figure 13). The contribution from demand shocks to oil price changes was higher compared to supply news, highlighting that economic slowdowns shifted market sentiment more toward concerns about demand.

One interesting period of comparison is the OPEC Collapse in the mid-1980s, where both supply and demand news contributed significantly, but in different ways. Supply news explained much of the initial price drop as OPEC struggled to maintain cohesive production cuts, while demand news became more influential as the market adjusted to a period of lower prices and shifting consumption patterns.

Another key observation can be made during the Gulf War, where both supply and demand news spikes are observed, but the supply shock clearly had a larger immediate impact, likely due to fears of production disruptions in the region. In contrast, demand news played a smaller, though still important, role in explaining the subsequent stabilization of oil prices after the initial shock.

Figure 13: Historical Decomposition of Real Oil Price: Contribution from OPEC Demand News Shock



Notes: The figure shows the historical decomposition of the real oil price attributed to OPEC demand news shocks. The black line represents the contribution of the demand shock, while the shaded areas indicate the confidence bands. Key economic and geopolitical events are highlighted to contextualize the impact of these shocks on oil prices.

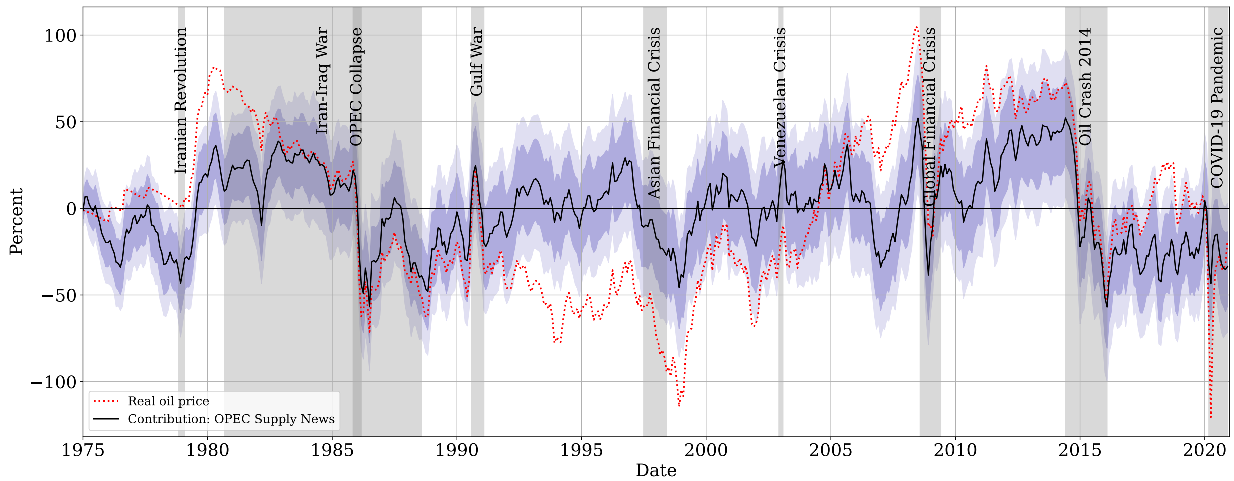
The shaded areas in both figures represent confidence intervals, reflecting the uncertainty in these contributions. During events with high uncertainty, like the Asian Financial Crisis or Venezuelan Crisis, the confidence bands widen, suggesting greater variability in how news was processed by the market.

The comparison between OPEC supply and demand news contributions shows that the nature of the underlying shock plays a critical role in determining the impact on oil prices. Supply shocks tend to dominate during geopolitical crises, while demand shocks are more influential during broader economic downturns that affect consumption expectations.

Figure 15 presents the forecast error variance decomposition (FEVD) for the six key macroeconomic variables. The FEVD provides insights into how much of the variation in these variables can be attributed to shocks driven by OPEC demand and supply news over a 50-month horizon.

For the real oil price, the contributions from both OPEC demand and supply news are noticeable, especially in the short term. The initial impact is dominated by supply shocks, contributing around 0.8% to the variance, which slowly decreases over time, while the demand contribution remains lower but persistent. This indicates that supply shocks have a stronger immediate influence

Figure 14: Historical Decomposition of Real Oil Price: Contribution from OPEC Supply News Shock



Notes: The figure shows the historical decomposition of the real oil price attributed to OPEC supply news shocks. The black line represents the contribution of the supply shock, while the shaded areas indicate the confidence bands. Key economic and geopolitical events are highlighted to contextualize the impact of these shocks on oil prices.

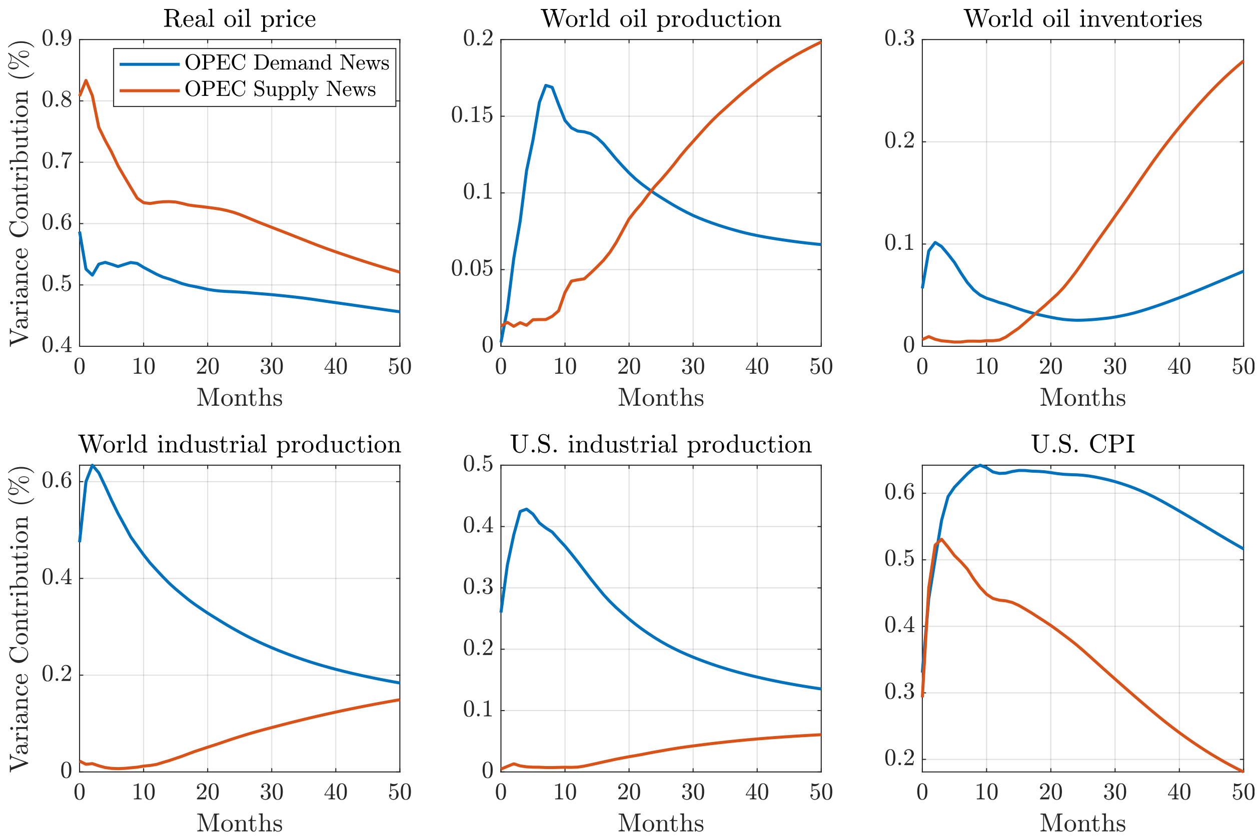
on oil prices, while demand shocks play a more consistent but smaller role.

World oil production shows an interesting dynamic where both types of news influence production, but in contrasting ways. Initially, supply news accounts for a small but growing portion of the variance, which becomes more pronounced after the 20th month, suggesting that supply-side news has a delayed but increasingly significant impact on production decisions. In contrast, demand news starts strong but declines steadily, highlighting a quick adjustment by producers to demand signals.

When looking at world oil inventories, the pattern is clear: both demand and supply shocks contribute to changes, but the influence of supply news becomes more dominant over time. Initially, the variance is quite low for both, but as the months progress, supply news begins to play a larger role. This suggests that, as supply expectations adjust, market participants increasingly rely on inventories to manage the evolving situation.

For world industrial production and U.S. industrial production, the demand shock contribution is more significant, especially in the early months. This makes sense since positive news about oil demand can signal broader economic growth, leading to a rise in industrial activity. Over time,

Figure 15: Forecast Error Variance Decomposition



Notes: The figure shows the forecast error variance decomposition. The decomposition highlights the contributions of OPEC demand and supply news shocks to the forecast error variance.

however, the impact of demand shocks wanes, and supply news, though less impactful initially, begins to contribute slightly, possibly as industries adjust to changing cost structures driven by supply dynamics.

Finally, for the U.S. Consumer Price Index (CPI), we observe that both OPEC demand and supply news affect the variance, but the demand shock appears to have a more consistent and dominant influence. This is understandable, as demand-driven increases in oil prices can lead directly to higher consumer costs, contributing to inflation. Supply shocks also have an initial impact, but their influence decreases gradually as the market stabilizes.

This analysis suggests that the sentimental demand-driven news primarily affects economic variables quickly but fades as markets adjust, while the factual supply-driven news leads to a more gradual adjustment, particularly in oil production and inventory management.

5.7 Different Narrative Lead to Different Outcomes

To ascertain how different news tone and narrative affect oil prices, I perform a novel experiment using synthetic news. In this section, I use large language models, specifically GPT-4, to generate synthetic news related to OPEC announcements. The experiment aims to vary the **tone** and **sentiment** of the news to observe if different narratives lead to different economic outcomes. By doing so, we can assess whether the way news is framed and the sentiment it carries have measurable impacts on market reactions. The synthetic news was created by issuing the following prompt to GPT-4:

> *"Create 110 article-style crude oil-related news pieces on OPEC meeting announcements. Make them lengthy, including discussions, implications, and detailed analysis."*

1. *"The recent OPEC meeting in Vienna saw key producers advocating for an increase in oil production, citing rising global demand and the need to stabilize market supply. The meeting, which lasted several days, featured heated debates over how much production should be increased, with some members arguing for a significant boost to meet growing global energy needs. Ultimately, OPEC agreed on a modest increase, a decision that was met with approval from the markets, as oil prices rose slightly in response. The increase is seen as a cautious approach to prevent any supply shortages while ensuring that prices do not skyrocket, which could harm global economic growth. As the global economy recovers from recent downturns, the role of OPEC in maintaining price stability will continue to be a key focus for both market participants and policymakers".*

2. *"In an unexpected move, OPEC opted to leave production levels unchanged despite pressures from some member nations to increase output. This decision, which surprised many market participants, was seen as a cautious approach aimed at maintaining price stability in the face of uncertain global demand. OPEC ministers emphasized the need for a careful balancing act, taking into account both short-term market conditions and long-term economic trends. Several key players in the oil market had anticipated a production hike, but the decision to hold steady reflects OPEC's ongoing efforts to avoid drastic price fluctuations. The meeting also included discussions on potential collaborations with non-OPEC producers and the growing influence of renewable energy sources on global oil demand".*

Figure 16: Examples of Synthetic News Generated by GPT-4

The motivation behind this experiment is twofold. First, I aim to determine if the tone and sentiment of OPEC-related news significantly affect economic indicators like oil prices and production. Second, I intend to test if the real news data contains genuine signals that influence market behavior, or if these signals are merely statistical artifacts that can be replicated using synthetic

Figure 17: Word Cloud of the Top 200 Most Frequent Bigrams from Synthetic News Data

data generated by a language model. Examples of these synthetic news articles are shown in Figure 16.

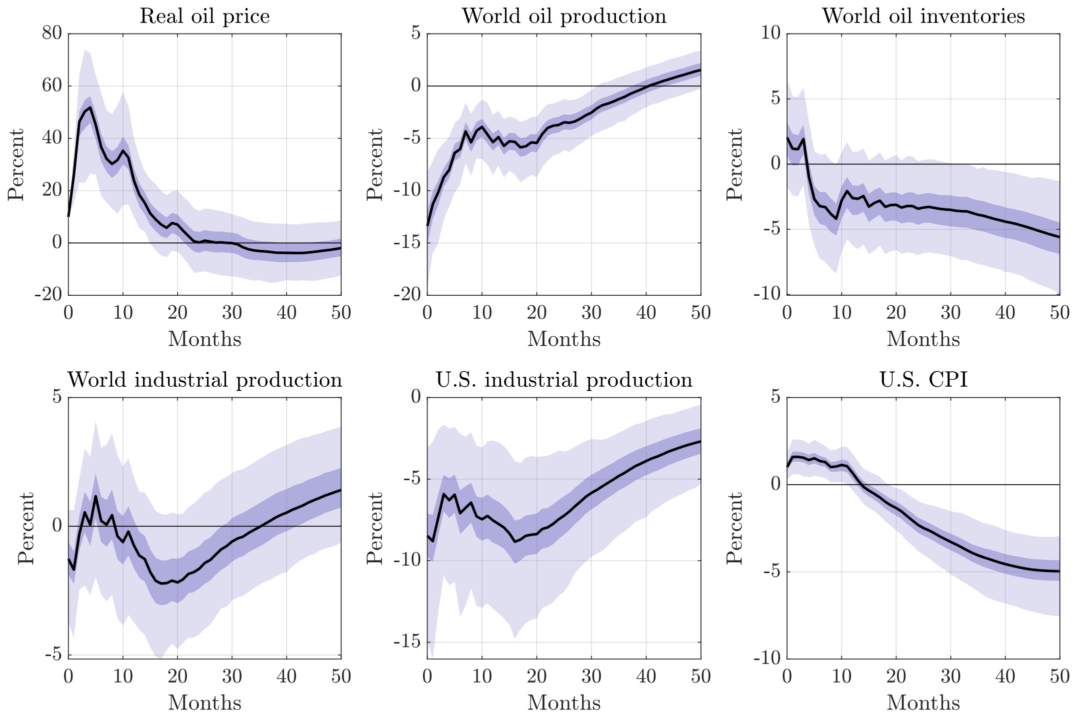
The synthetic news is taken through the same pre-processing stage of cleaning as the real news. Figure 17 is a word cloud of the top 200 most frequent bigrams, obtained from the synthetic news.

Next, I extract the sentimental demand-driven news as well as the factual supply news, as I did in the real news data case and use them in my IV-SVAR model. The IRFs are presented in figure 18.

The results from the first-stage regression show that both instruments have an F-statistic value below the threshold of 10, indicating that the instruments are weak and not valid. When compared with the real news data, this result shows the effectiveness of real news in extracting valid instruments that contain actual market signals. It suggests that real news contains genuine information that impacts the market.

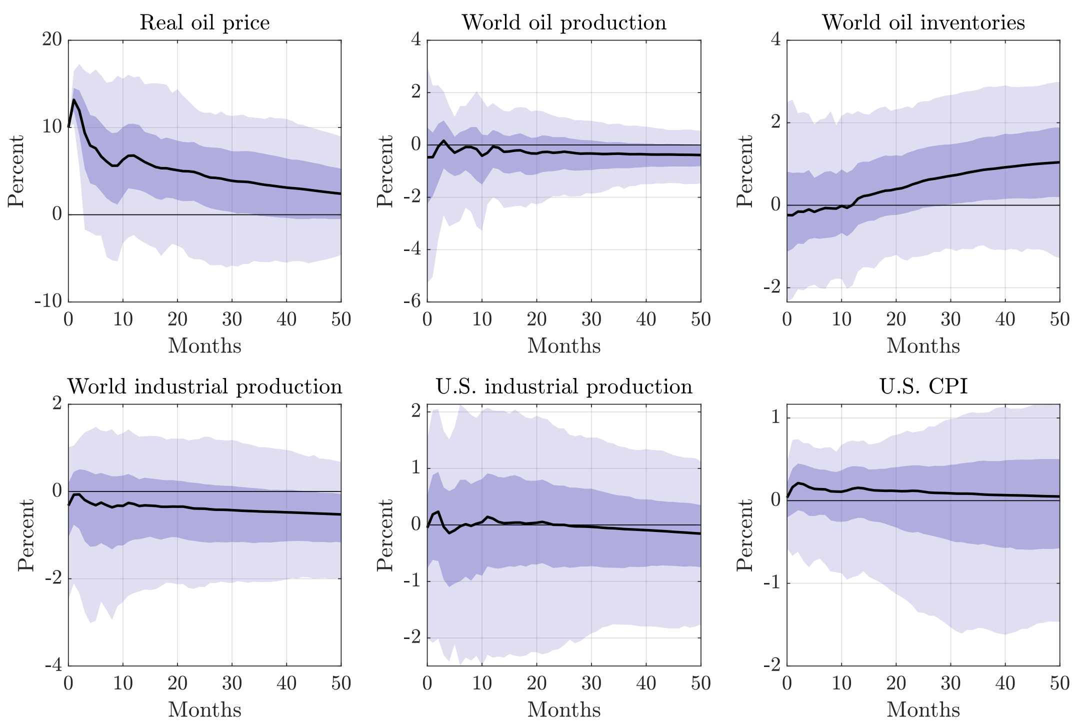
Despite the weakness of the synthetic instruments, I still interpret the IRFs for demonstration purposes. The IRFs for oil demand expectations using synthetic news (Figure 18a) differ significantly from those obtained using real news (Figure 11). For example, while the real news data shows an increase in world oil production, the synthetic news results in a significant fall in world oil production by over 12%. Similar discrepancies are observed in world and U.S. industrial production. For supply expectations, apart from real oil prices, there are substantial differences between the IRFs generated using real news (Figure 12) and synthetic news (Figure 18b).

(a) IRFs: A shock to oil demand expectations driven by synthetic OPEC demand news.



First stage regression: F: 0.02, robust F: 0.01, R^2: 0.00%, Adjusted R^2: -0.18%

(b) IRFs: A shock to oil supply expectations driven by synthetic OPEC supply news.



First stage regression: F: 1.83, robust F: 1.93, R^2: 0.33%, Adjusted R^2: 0.15%

Figure 18: *Impulse Response Functions (IRFs) for Synthetic News: Demand vs. Supply News*

These results suggest that real news data contains unique, meaningful signals that synthetic news struggles to replicate, emphasizing the importance of genuine, context-specific information in driving market outcomes. Additionally, they demonstrate that different narratives lead to different outcomes, underscoring how news framing significantly impacts economic reactions.

6 Robustness Check

In this section, I use the extracted structural shocks from the IV-SVAR models to estimate impulse responses via local projections (LP). Local projections provide an alternative way to estimate impulse responses, offering greater flexibility compared to VAR models, but often at the cost of precision.

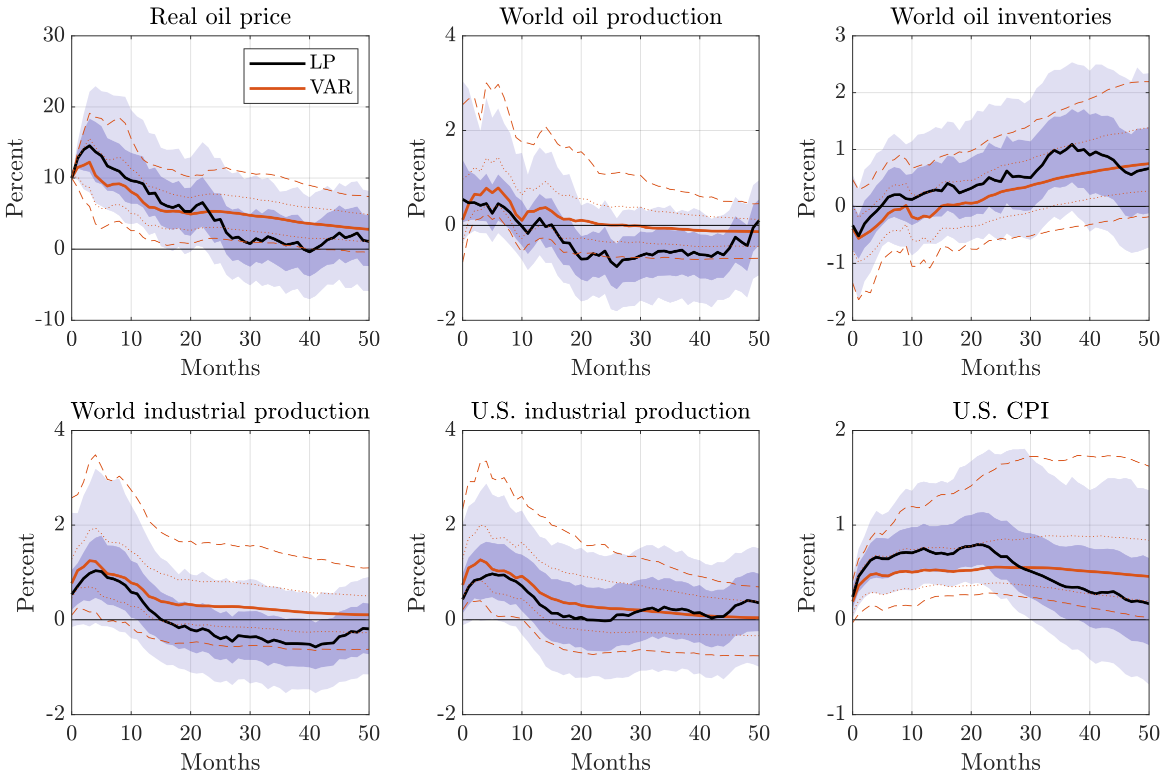
Comparison of VAR and LP Results: The results from Figure 19 shows that the VAR-based and LP-based impulse responses are quite similar in terms of the general direction and timing of the shocks' impact on key macroeconomic variables. However, as is often the case with local projections, the LP-based IRFs are more erratic and less precisely estimated, especially at longer horizons. This aligns with previous findings in the literature regarding the external instruments approach, where local projections tend to have wider confidence intervals compared to VAR estimates. Despite this, the key takeaway is that the overall patterns remain consistent across both methods.

7 Conclusion

This paper presents a new approach to understanding the impact of OPEC announcements on oil prices by incorporating textual analysis of news articles. The findings highlight the importance of considering both supply decisions and market sentiment when analyzing oil price shocks. By using textual features as external instruments in a proxy SVAR model, I provide a more nuanced understanding of the factors driving oil price movements around OPEC announcements.

The results suggest that future research should continue to explore the role of news and sentiment in shaping market behavior, particularly in the context of major policy announcements. The

(a) IRFs: A shock to oil demand expectations driven by OPEC demand news.



(b) IRFs: A shock to oil supply expectations driven by OPEC supply news.

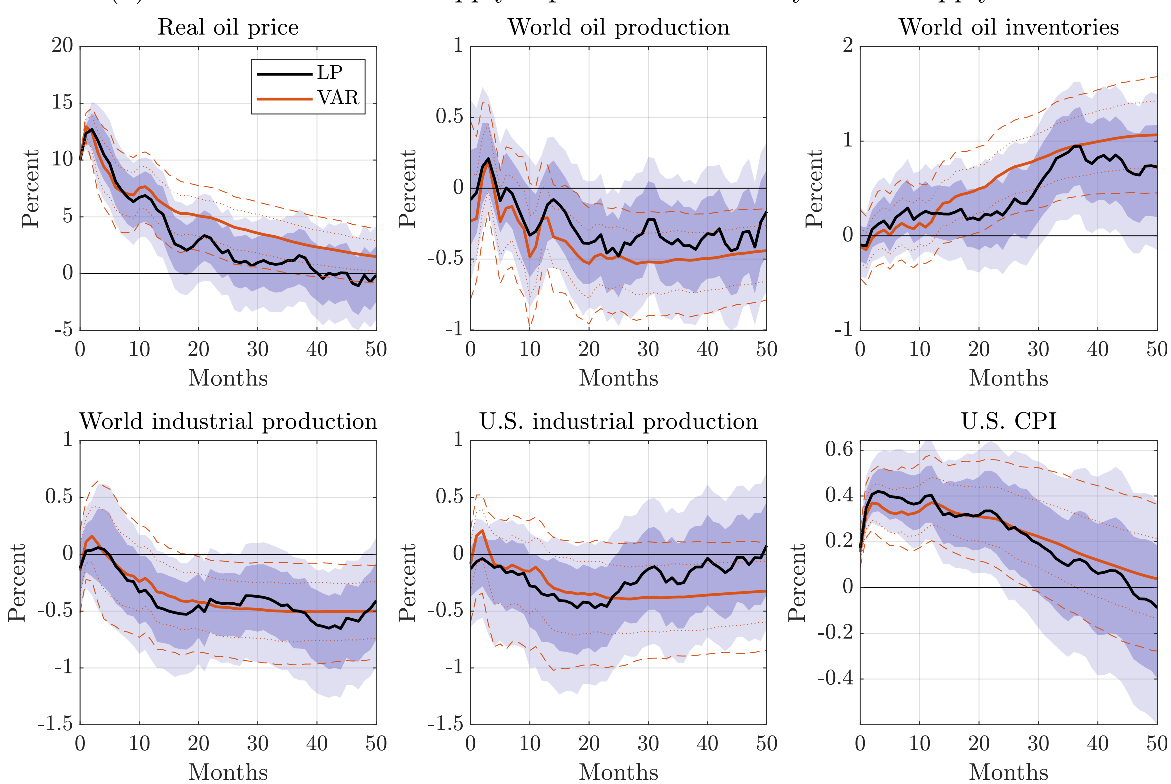


Figure 19: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News.*

use of textual data offers a valuable complement to traditional price-based analyses and provides deeper insights into the dynamics of crude oil markets.

References

Abeyie, Gabriel (2024) "Beyond News Headlines and TF-IDF: Enhancing Text-Based Forecasting Models with Validated Collocations and Improved Attention," *Available at SSRN 4822767.*

Alquist, Ron and Lutz Kilian (2010) "What do we learn from the price of crude oil futures?" *Journal of Applied econometrics*, 25 (4), 539–573.

Antolín-Díaz, Juan and Juan F Rubio-Ramírez (2018) "Narrative sign restrictions for SVARs," *American Economic Review*, 108 (10), 2802–2829.

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

Bai, Yun, Xixi Li, Hao Yu, and Suling Jia (2022) "Crude oil price forecasting incorporating news text," *International Journal of Forecasting*, 38 (1), 367–383.

Baumeister, Christiane and James D Hamilton (2019) "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks," *American Economic Review*, 109 (5), 1873–1910.

Blanchard, Olivier Jean and Danny Quah (1993) "The dynamic effects of aggregate demand and supply disturbances: Reply," *The American Economic Review*, 83 (3), 653–658.

Blinder, Alan S, Robert M Solow et al. (1973) "Does fiscal policy matter?" *Journal of public economics*, 2 (4), 319–337.

Bruns, Martin (2021) "Proxy vector autoregressions in a data-rich environment," *Journal of Economic Dynamics and Control*, 123, 104046.

Bruns, Martin and Helmut Lütkepohl (2023) "Have the effects of shocks to oil price expectations changed?: Evidence from heteroskedastic proxy vector autoregressions," *Economics Letters*, 233, 111416.

Caldara, Dario, Michele Cavallo, and Matteo Iacoviello (2019) "Oil price elasticities and oil price fluctuations," *Journal of Monetary Economics*, 103, 1–20.

Caravello, Tomás and Pedro Martinez-Bruera (2024) "Disentangling Sign and Size Non-linearities," *Available at SSRN 4704050*.

Christiano, Lawrence J (1999) "Monetary Policy Shocks: What Have We Learned and to What End," *Handbook of Macroeconomics*, 1.

Clarida, Richard, Jordi Gali, and Mark Gertler (1999) "The science of monetary policy: a new Keynesian perspective," *Journal of economic literature*, 37 (4), 1661–1707.

Datta, Deepa D and Daniel A Dias (2019) "Oil shocks: A textual analysis approach," *manuscript, Federal Reserve Board*.

Degasperi, Riccardo (2023) "Identification of expectational shocks in the oil market using OPEC announcements."

Demirer, Rıza and Ali M Kutan (2010) "The behavior of crude oil spot and futures prices around OPEC and SPR announcements: An event study perspective," *Energy Economics*, 32 (6), 1467–1476.

Gagliardone, Luca and Mark Gertler (2023) "Oil prices, monetary policy and inflation surges."

Handlan, Amy (2020) "Text shocks and monetary surprises: Text analysis of fomc statements with machine learning," *Published Manuscript*.

Hansen, Lars Peter and Kenneth J Singleton (1982) "Generalized instrumental variables estimation of nonlinear rational expectations models," *Econometrica: Journal of the Econometric Society*, 1269–1286.

Hoerl, Arthur E and Robert W Kennard (1970) "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, 12 (1), 55–67.

Inoue, Atsushi and Lutz Kilian (2013) "Inference on impulse response functions in structural VAR models," *Journal of Econometrics*, 177 (1), 1–13.

Jiao, Xingrui, Yuping Song, Yang Kong, and Xiaolong Tang (2022) "Volatility forecasting for crude oil based on text information and deep learning PSO-LSTM model," *Journal of Forecasting*, 41 (5), 933–944.

Känzig, Diego R (2021) "The macroeconomic effects of oil supply news: Evidence from OPEC announcements," *American Economic Review*, 111 (4), 1092–1125.

Kilian, Lutz (2008) "The economic effects of energy price shocks," *Journal of economic literature*, 46 (4), 871–909.

——— (2009) "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American economic review*, 99 (3), 1053–1069.

——— (2024) "How to Construct Monthly VAR Proxies Based on Daily Surprises in Futures Markets," *Journal of Economic Dynamics and Control*, 104966.

Kilian, Lutz and Daniel P Murphy (2012) "Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models," *Journal of the European Economic Association*, 10 (5), 1166–1188.

——— (2014) "The role of inventories and speculative trading in the global market for crude oil," *Journal of Applied econometrics*, 29 (3), 454–478.

Kilian, Lutz and Xiaoqing Zhou (2020) "Does drawing down the US Strategic Petroleum Reserve help stabilize oil prices?" *Journal of Applied Econometrics*, 35 (6), 673–691.

——— (2023) "The econometrics of oil market VAR models," in *Essays in honor of joon y. park: Econometric methodology in empirical applications*, 65–95: Emerald Publishing Limited.

Li, Xuerong, Wei Shang, and Shouyang Wang (2019) "Text-based crude oil price forecasting: A deep learning approach," *International Journal of Forecasting*, 35 (4), 1548–1560.

Lippi, Francesco and Andrea Nobili (2012) "Oil and the macroeconomy: a quantitative structural analysis," *Journal of the European Economic Association*, 10 (5), 1059–1083.

Loughran, Tim, Bill McDonald, and Ioannis Pragidis (2019) "Assimilation of oil news into prices," *International Review of Financial Analysis*, 63, 105–118.

Lütkepohl, Helmut and Anton Velinov (2016) "structural vector autoregressions: checking identifying long-run restrictions via heteroskedasticity," *Journal of Economic Surveys*, 30 (2), 377–392.

Mertens, Karel and Morten O Ravn (2013) "The dynamic effects of personal and corporate income tax changes in the United States," *American economic review*, 103 (4), 1212–1247.

Ochs, Adrian CR (2021) "A New Monetary Policy Shock with Text Analysis."

Patzelt, Paula and Ricardo Reis (2024) *Estimating the rise in expected inflation from higher energy prices*: London School of Economics and Political Science.

Peersman, Gert and Ine Van Robays (2009) "Oil and the Euro area economy," *Economic Policy*, 24 (60), 603–651.

Plagborg-Møller, Mikkel and Christian K Wolf (2022) "Instrumental variable identification of dynamic variance decompositions," *Journal of Political Economy*, 130 (8), 2164–2202.

Plante, Michael (2019) "OPEC in the news," *Energy Economics*, 80, 163–172.

Ramey, Valerie A (2011) "Identifying government spending shocks: It's all in the timing," *The Quarterly Journal of Economics*, 126 (1), 1–50.

Salton, Gerard, Anita Wong, and Chung-Shu Yang (1975) "A vector space model for automatic indexing," *Communications of the ACM*, 18 (11), 613–620.

Sims, Christopher A (1980) "Macroeconomics and reality," *Econometrica: journal of the Econometric Society*, 1–48.

Stock, James H and Mark W Watson (2012) "Disentangling the Channels of the 2007-2009 Recession," Technical report, National Bureau of Economic Research.

Sugihara, George, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch (2012) "Detecting causality in complex ecosystems," *science*, 338 (6106), 496–500.

Tibshirani, Robert (1996) "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58 (1), 267–288.

Uhlig, Harald (2005) "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52 (2), 381–419.

Wold, Svante, Michael Sjöström, and Lennart Eriksson (2001) "PLS-regression: a basic tool of chemometrics," *Chemometrics and intelligent laboratory systems*, 58 (2), 109–130.

Wu, Mr Tao and Mr Michele Cavallo (2012) *Measuring oil-price shocks using market-based information*: International Monetary Fund.

Zou, Hui and Trevor Hastie (2005) "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67 (2), 301–320.

Appendix

A Example of Text Preprocessing

This section provides an example illustrating how the text changes after each preprocessing stage in subsection 3.3. The original text is transformed step by step to demonstrate the effect of different preprocessing methods applied in the analysis.

1. **Original Text:**

   "OPEC and other oil-producing nations led by Russia, trying to gauge the strength of the global economy as the coronavirus continues to rage but with vaccines on the horizon, reached a compromise on Thursday to modestly increase production in January. But the talks revealed strains in the unwieldy group, known as OPEC Plus, which has tried to manage the oil market since 2016. These tensions could make it more difficult for the producers to stay in line with production targets as the global economy recovers in the coming months."

2. **Removal of Noise:**

   Text after Noise Removal:

   "OPEC oil-producing nations led Russia trying gauge strength global economy coronavirus continues rage vaccines horizon reached compromise Thursday modestly increase production January talks revealed strains unwieldy group known OPEC Plus tried manage oil market since tensions make difficult producers stay line production targets global economy recovers months"

   Explanation: HTML tags, URLs, email addresses, websites, numbers, and words consisting of one or two characters have been removed to eliminate irrelevant information.

3. **Normalization:**

   Text after Normalization:

   "opec oil producing nations led russia trying gauge strength global economy coronavirus continues rage vaccines horizon reached compromise thursday modestly increase production january talks revealed strains unwieldy group known opec plus tried manage oil market since ten-

sions make difficult producers stay line production targets global economy recovers months"

Explanation: Accented characters, hyphens, apostrophes, and common stopwords that do not contribute significant meaning to the analysis have been removed.

4. **Stemming:**

   Text after Stemming:

   "opec oil produc nation led russia try gaug strength global economi coronaviru continu rage vaccin horizon reach compromis thursday modestli increas product januari talk reveal strain unwieldi group known opec plu tri manag oil market sinc tension make difficult produc stay line product target global economi recov month"

   Explanation: Words have been stemmed to reduce them to their base forms, ensuring that similar words like "producing" and "production" are treated equivalently.

5. **Tokenization into Bigrams:**

   Text after Tokenization into Bigrams:

   "opec oil", "oil produc", "produc nation", "nation led", "led russia", "russia try", "try gaug", "gaug strength", "strength global", "global economi", "economi coronaviru", "coronaviru continu", "continu rage", "rage vaccin", "vaccin horizon", "horizon reach", "reach compromis", "compromis thursday", "thursday modestli", "modestli increas", "increas product", "product januari", "januari talk", "talk reveal", "reveal strain", "strain unwieldi", "unwieldi group", "group known", "known opec", "opec plu", "plu tri", "tri manag", "manag oil", "oil market", "market sinc", "sinc tension", "tension make", "make difficult", "difficult produc", "produc stay", "stay line", "line product", "product target", "target global", "global economi", "economi recov", "recov month"

   Explanation: Words have been tokenized into bigrams (pairs of words) to capture important phrases and contextual information, providing more insight into the relationships between terms.

B Algorithm for Extracting Information from Text Data

This section outlines the step-by-step process used for extracting and analyzing information from text data surrounding OPEC announcement days.

1. **Input Data**
   - **Data Source**: Daily aggregated news articles related to OPEC announcements.
   - **Target Variable**: Log change in oil prices, $\Delta P_{d,t}^{h}$.
2. **Document-Term Matrix Construction**
   - Convert each daily aggregated article into a vector using the vector space model, where each bigram (pair of words) is treated as a unique term.
   - Build a **document-term matrix** $\mathbf{X}_t(n, m)$, where:
     - n: number of articles (aggregated on OPEC days).
     - m: number of unique bigrams.
   - **Result**: A matrix with bigram frequencies for each article.
3. **Bigram Selection Using TF-IDF**
   - Apply **Term Frequency-Inverse Document Frequency (TF-IDF)** to evaluate the importance of each bigram.
   - Set a minimum TF-IDF threshold of 0.05 to filter out less relevant bigrams.
   - **Result**: The document-term matrix is reduced to 2,057 key bigrams.
4. **Model Incorporation**
   - Use the document-term matrix in a regression model:

$$\Delta P_{d,t}^{h} = \alpha + \mathbf{X}_t \beta + \epsilon_t$$

   - Where $\mathbf{X}_t$ is the matrix of selected bigrams, β represents the coefficients, and ϵ_t captures unexplained variations (Non-OPEC surprise component).

5. **Elastic Net Regularization to Combat High Dimensionality**

   - **Challenge**: The number of bigrams ($m = 2,057$) is much larger than the number of observations ($n = 110$), which is known as the "curse of dimensionality."
   - Apply **Elastic Net Regularization**, which blends Lasso (L_1) and Ridge (L_2) penalties to handle high-dimensional data and multicollinearity:

$$\min_{\beta} \left(\frac{1}{2n} \sum_{i=1}^{n} \left(\Delta P_{d,t}^{h} - \mathbf{X}_t^{\top} \beta \right)^2 + \lambda \left(\alpha ||\beta||_1 + (1-\alpha)||\beta||_2^2 \right) \right)$$

   - Use 5-fold time series cross-validation to find the optimal λ and α.
   - **Result**: Retain 405 non-zero features after Elastic Net regularization.

6. **Further Dimensionality Reduction with Partial Least Squares (PLS)**

   - **Forward Selection**: Iteratively add the most relevant features by evaluating the adjusted R^2 at each step.
   - Apply **Partial Least Squares (PLS)** to handle multicollinearity and reduce dimensionality by extracting latent variables that explain the maximum variance.
   - **Result**: A final set of 34 features is identified, and the first PLS component $\mathbf{F}_t$ is extracted.

7. **Final Regression Model**

   - Fit an **Ordinary Least Squares (OLS)** regression model using the first PLS component $\mathbf{F}_t$ as a predictor:

$$\Delta P_{d,t}^{h} = \alpha + \mathbf{F}_t \beta + \epsilon_t$$

B.1 Partial Least Squares (PLS) Regression

Partial Least Squares (PLS) regression is a dimensionality reduction technique that finds latent variables (or components) that maximize the covariance between the predictors and the response variable. It is particularly useful in contexts where the number of predictors exceeds the number of observations or when multicollinearity is present, as in our analysis with the document-term matrix of bigrams.

1. Problem Setup. In our context, we have a response variable $\Delta P^h_{d,t}$ (the log change in oil prices around OPEC announcement days) and a set of predictors $\mathbf{X}_t \in \mathbb{R}^{n\times m}$, where n is the number of observations (OPEC announcement days) and m is the number of bigrams (features). The predictors and the response are centered, and PLS finds components of $\mathbf{X}_t$ that explain the most variance in both $\mathbf{X}_t$ and the response $\Delta P^h_{d,t}$.

2. Goal of PLS. The goal of PLS is to find latent variables (also called *components*), $\mathbf{F}_k$, from $\mathbf{X}_t$ that have maximum covariance with the response variable $\Delta P^h_{d,t}$. These components are linear combinations of the original predictors and are used to reduce the dimensionality of the problem while retaining the most relevant information for predicting the response.

3. Latent Variables and Projections. PLS works by projecting both the predictor matrix $\mathbf{X}_t$ and the response variable $\Delta P^h_{d,t}$ onto a new set of components $\mathbf{F}_k$, where $k = 1, 2, \ldots$, and each component is chosen to maximize the covariance between $\mathbf{X}_t$ and $\Delta P^h_{d,t}$.

Step 1: Define the Components. Let $\mathbf{w}_k$ be the weight vector that defines the linear combination of the columns of $\mathbf{X}_t$ for the k-th component. Then, the k-th PLS component $\mathbf{F}_k$ is:

$$\mathbf{F}_k = \mathbf{X}_t \mathbf{w}_k$$

The weight vector $\mathbf{w}_k$ is chosen to maximize the covariance between $\mathbf{F}_k$ and the response variable $\Delta P^h_{d,t}$.

Step 2: Compute the Weight Vector. The weight vector $\mathbf{w}_k$ is computed by maximizing the covariance between the predictor component $\mathbf{F}_k$ and the response $\Delta P^h_{d,t}$:

$$\mathbf{w}_k = \arg\max_{\mathbf{w}} \mathrm{Cov}(\mathbf{F}_k, \Delta P^h_{d,t}) = \arg\max_{\mathbf{w}} \mathbf{w}^\top \mathbf{X}_t^\top \Delta P^h_{d,t}$$

This weight vector $\mathbf{w}_k$ is used to project $\mathbf{X}_t$ onto the new component $\mathbf{F}_k$.

Step 3: Deflate $\mathbf{X}_t$ and $\Delta P^h_{d,t}$. After each component $\mathbf{F}_k$ is calculated, both the predictor matrix $\mathbf{X}_t$ and the response variable $\Delta P^h_{d,t}$ are *deflated* to remove the variance explained by the component $\mathbf{F}_k$. This ensures that subsequent components explain new variance not captured by the previous components.

The deflation step is:

$$\mathbf{X}_t^{(k+1)} = \mathbf{X}_t^{(k)} - \mathbf{F}_k \mathbf{p}_k^\top$$

where $\mathbf{p}_k$ is the loadings vector for the predictors, which describes how much of each predictor is explained by the component $\mathbf{F}_k$.

Similarly, the response variable is deflated as:

$$(\Delta P^h_{d,t})^{(k+1)} = (\Delta P^h_{d,t})^{(k)} - \mathbf{F}_k \mathbf{q}_k$$

where $\mathbf{q}_k$ is the loading coefficient for the response variable in the k-th component.

Step 4: Repeat for Subsequent Components. Repeat the process to extract additional components until the desired number of components K is reached, or until the additional components no longer provide significant explanatory power.

4. Final PLS Regression Model. Once the components $\mathbf{F}_k$ have been extracted, we perform regression on the response variable $\Delta P^h_{d,t}$ using the latent components:

$$\Delta P^h_{d,t} = \alpha + \sum_{k=1}^{K} \mathbf{F}_k \mathbf{b}_k + \epsilon_t$$

where $\mathbf{F}_k$ are the PLS components, $\mathbf{b}_k$ are the corresponding regression coefficients, and ϵ_t is the residual term.

Graphical Representation of PLS Loadings. Figure 20 provides a graphical view of the loadings for the PLS components. Loadings indicate the weight of each predictor (bigrams in this case) in constructing the PLS components. This visualization helps to identify which predictors contribute the most to each component.

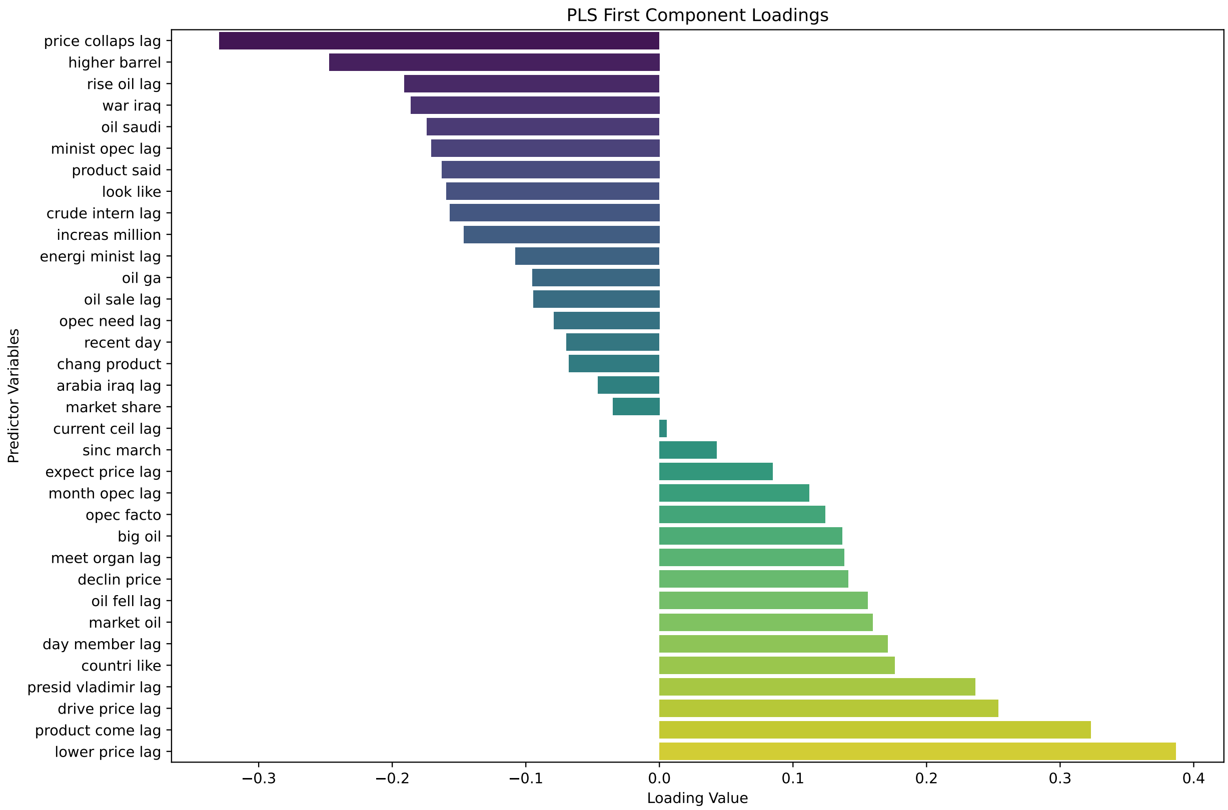


Figure 20: Loadings of the first PLS component. The loadings represent the contribution of each predictor (bigram) in the construction of the first PLS component, showing how different textual features contribute to explaining the variation in oil prices.

Sample News Extracts Containing Bigrams from the PLS Loadings. The following table presents sample news extracts from the newpaper articles in the corpus, with specific bigrams highlighted. These bigrams were identified as important predictors in the PLS model and played a significant role in constructing the latent components used to explain oil price variations. The highlighted text showcases how these bigrams appeared within the original news articles.

Bigram	**Extract**
price collaps	*1. OPEC raises the oil ceiling so Iraq can come back without causing a price collapse.* *2. It's an awkward modus operandi, said Nadine. Hides Lousine, manager of Naokosa, an oil consulting company in Geneva. "I think only when, or should I say if, prices collapse will OPEC be strongly motivated to pay attention to the situation.*
higher barrel	*The Iranian proposal seemed most likely to be accepted, although some delegates said the price might be set a little higher at $21 or $22 a barrel to satisfy Iraq. Libya, Ecuador, and Nigeria are all backing higher prices.*
minist opec	*We will have a new minimum price," Nordine Ait-Laoussine, Algeria's oil minister and the OPEC conference president, told reporters today.*
war iraq	*Oil are shot up as high as $41.00 a barrel since then, but it tumbled back to around $25 last week as efforts intensified to avert war with Iraq. The price of a barrel of crude in New York close to today at $25.35, still above Opec's target price of $21.00 a barrel.*
oil saudi	*The United Arab Emirates, Venezuela and Iran, lifted their outputs by slightly more than a million barrels a day to compensate for the loss of 4 million barrels a day of oil. Resulting from the embargo on Iraqi and Kuwaiti oil, Saudi Arabia has raised its output by three million barrels a day to about 8.3 million barrels.*
product said	*Officials of Saudi Arabia, which spent some $50 billion in the war against Iraq and invested huge sums to expand oil production. Said at the tensed session that they did not want to be pressured into cutting outputs.*
look like	*Oil stocks suddenly look like higher scores now that the UN has put Iraq in the game.*
big oil	*But the case for big oil goes beyond the Iraqi deal. Crude oil and natural gas prices are higher than years ago levels. The chemicals business is recovering after a brief slump. And refining and marketing are looking better than have in years, driven partly by the record high gasoline prices of recent month. All the pistons are finally firing at the same time. Says analyst Paul Ting of Solomon Brothers.*
increas million	*While Opec's action was widely expected, observers were surprised at the size of the increase, 2.5 million barrels a day above Opec's current production ceiling of 25 million barrels a day. The increase is a signal that the elimination group believes what demand will continue to grow despite a raft of bearish short term fundamentals and unexpected economic slowdown in Asia.*
countri like	*In those talks, some countries like Algeria and Iran pressed for reallocation of the group's output quarters, and others, like Saudi Arabia, Kuwait and the United Arab Emirates proposed that OPEC simply increase Iraq's allotment to take into. The oil exports allowed by the UN.*
market oil	*On the other commodity markets, oil futures fell and livestock and meat futures were mixed.*

Bigram	Extract
declin price	*OPEC signals it may curb output. Attempts to tighten spigot may be effort by cartel to avoid decline in price.*
current ceiling	*A majority of the OPEC ministers are supporting a new total production ceiling of 22.5 million barrels a day for the remainder of the year, whilst 400,000 barrels a day above Opec's current ceiling. That ceiling would be below, the organization's actual output most of this year because the old ceiling was routinely exceeded.*
oil ga	*Yesterday, US Federal Reserve Chairman Alan Greenspan warned a congressional panel that high natural gas prices are likely to endure for sometime, comments that spurred oil and gas features even higher, some energy experts said.*
energi minist	*OPEC President Chakib Khelil, Algeria's energy minister, said recent price declines supported the group's prevailing view that the cartel supply restraints weren't to blame for soaring prices.*
recent day	*In an unexpected decision, made after a six hour meeting. That lasted war into the night. The OPEC oil cartel said it will reduce its oil production by about half a million barrels a day in a bid to stem a rapid decline in oil prices in recent days.*
crude intern	*On Friday, immediately after the OPEC decision to keep its production targets, the price of Brent crude, the international benchmark, rose slightly to $62.50 a barrel, then fell slightly in the afternoon trading.*
chang product	*This year, the group hasn't made any changes to production targets as oil prices have returned to comfortable levels, though compliance among members with the targets appears to have waned.*
market share	*The problem for OPEC is that it can't, for the fear of flooding the market, set a ceiling on its total output high enough to fit the sales ambitions of Kuwait and the Emirates without others sacrificing their market share.*
opec facto	*Under the leadership of Ali Naimi, Saudi Arabia's oil minister and OPEC de facto leader, the cartel in recent years has prided itself on its ability to manage oil markets and avoid politics.*
sinc march	*July corn rose 6.50 cents, to $2.83 bushel, its highest price since March 30, 1994, close at $2.84.*
month opec	*Continued overproduction by Kuwait and the United Arab Emirates, in addition to the 1 million barrels a day increase that OPEC approved, is expected to increase substantial oil supplies over the next six months. OPEC officials announced that the group agreed to raise its total production ceiling to 19.5 million barrels a day, saying the move was justified by rising demand for oil in the summer.*
expect price	*Since late April, wholesale unleaded gasoline prices are down about 15 cents a gallon, Mr. Boutle said. We expect prices to continue on a downward bias at least through July.*

Bigram	Extract
opec need	*A surprise at the meeting was the sudden cordiality between Gholam Reza Aghazadeh, Iran's oil minister, and Saudi Arabia's Mr. Nazar, who are often critical of each other. Mr. Aghazadeh complimented Mr. Nazar on Saudi's efforts to strengthen oil prices and told him Iran would do the same. " OPEC needs very strong support by the big producers", Mr. Aghazadeh said.*
oil sale	*Iraq's chief OPEC delegate, Osama Al Hiti, said he is expected to resume oil sales within a matter of weeks.*
arabia iraq	*While the group held together on its output reduction, choosing a leader for the organization proved more troublesome. Ministers were unable to select a new secretary general after nearly 8 hours of debate. With Iran, Saudi Arabia , and Iraq vying for the job, the ministers decided to delay a decision until OPEC's next meeting in March.*
oil fell	*Crude oil fell on Monday after OPEC decided to raise output and on a weakening technical picture.*
presid vladimir	*President Vladimir Putin of Russia, in an interview with an American journalist on Saturday, indicated that his country will not offer further cut. Moreover, Russia's oil companies have been privatized, unlike those of the other producers, making it difficult for the government to control production.*
lower price	*Oil prices reached a two-year low in November after OPEC told rivals such as Russia and Norway to make cutbacks or face lower prices . Those countries, along with Mexico, Oman, and Angola, agreed to curtail supplies by a total of 462,500 barrels a day.*
drive price	*OPEC's move underlined its determination to increase supplies if necessary to prevent shortages in the international oil market, where there is considerable worry that a war in Iraq could drive prices sharply higher.*
meet organ	*Bond prices got a big boost later yesterday on news of lower oil futures prices, fueled by reports of discord at the Vienna meeting of the Organization of Petroleum Exporting Countries.*
month opec	*Most of the world's big oil producers decided today not to reduce oil production in the coming months, gambling that demand would increase this winter and allow crude oil prices to recover somewhat from the sharp drop of the last few months . OPEC Secretary General Subroto made the announcement after two days of talks here on whether to restrain supplies in the hope of pushing up prices, now around a three-year low.*
rise oil	*Yesterday's move by the Organization of Petroleum Exporting Countries to raise output limits marks the group's attempt to cast itself as a white knight to a world worried about the impact of rising oil prices.*
day member	*In a quick meeting at their headquarters here, Ministers of the Organization of the Petroleum Exporting Countries yesterday decided to stick with their current output target of 25.8 million barrels a day for the 10 members in its quota system.*

Bigram	Extract

Table 8: Sample news extracts containing the bigrams identified in the PLS loadings. These bigrams, highlighted in yellow, were among the key predictors in the PLS model for explaining oil price variations.

C Additional Robustness Checks

In this section, I perform a series of additional robustness checks. Specifically, I re-estimate the baseline model under different specifications: using VAR(7), VAR(24), and by reducing the sample size to start from April 1984.

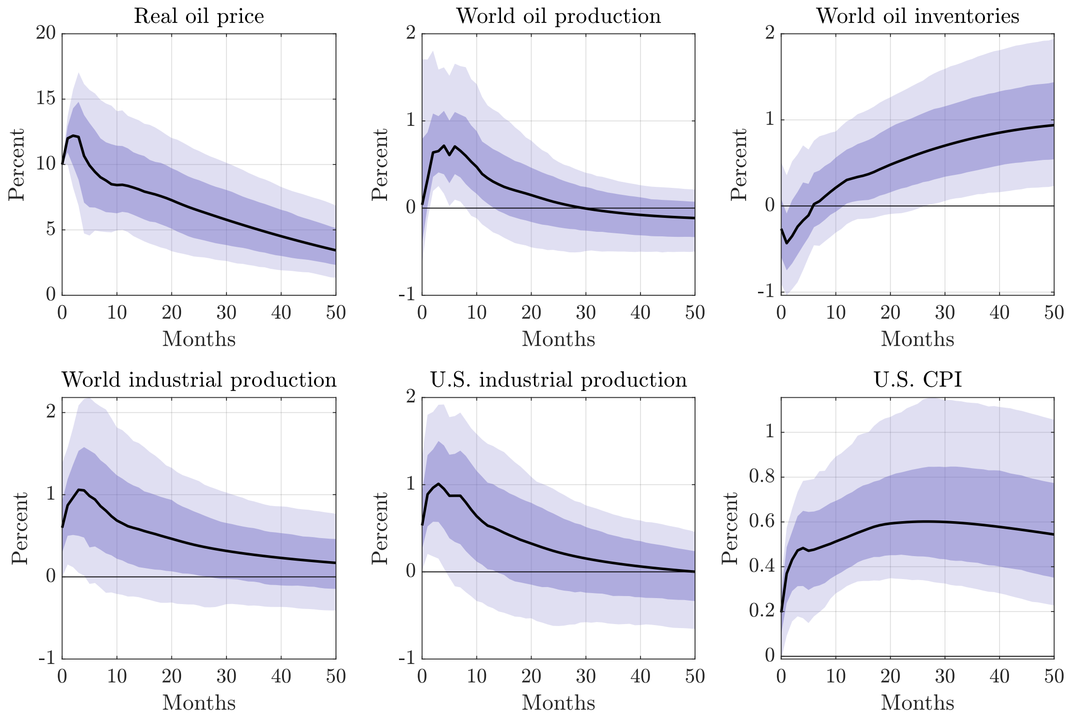
The VAR(7) specification is employed because the optimal lag length determined by the Akaike Information Criterion (AIC) was 7. Additionally, I estimate a VAR(24) following the approach of Kilian and Murphy (2014). I also reduce the sample size to begin from April 1984, as trading in oil futures markets before this date, particularly for longer maturities, was limited.

The results from these robustness checks are consistent with the findings presented in the main text (see Section 5.6).

D Additional Results

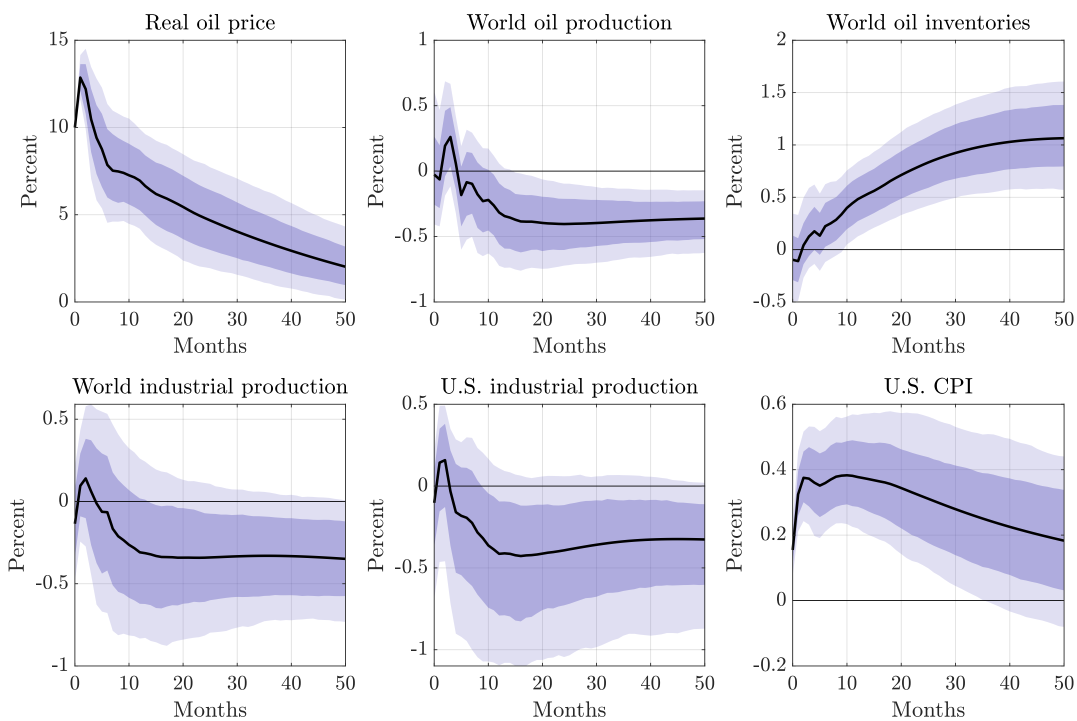
In this section, I present a series of additional impulse response functions (IRFs) using the VAR(12) specification. The analysis is organized into four groups: macroeconomic indicators, CPI components, PPI components, industrial production categories, and financial market indicators. Notable differences are observed between the demand-driven and supply-driven news shocks. For instance, the S&P 500 increases on impact in the case of demand shocks but decreases on impact for supply shocks. Similarly, CPI medical expenses exhibit an increase on impact for demand shocks, while they decrease for supply shocks. All components of industrial production show a consistent pattern, increasing on impact for demand shocks and decreasing on impact for supply shocks. These results provide further insight into the differential effects of demand and supply shocks on various economic sectors.

(a) VAR(7) Specification IRFs: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 16.86, robust F: 6.99, R^2: 2.95%, Adjusted R^2: 2.77%

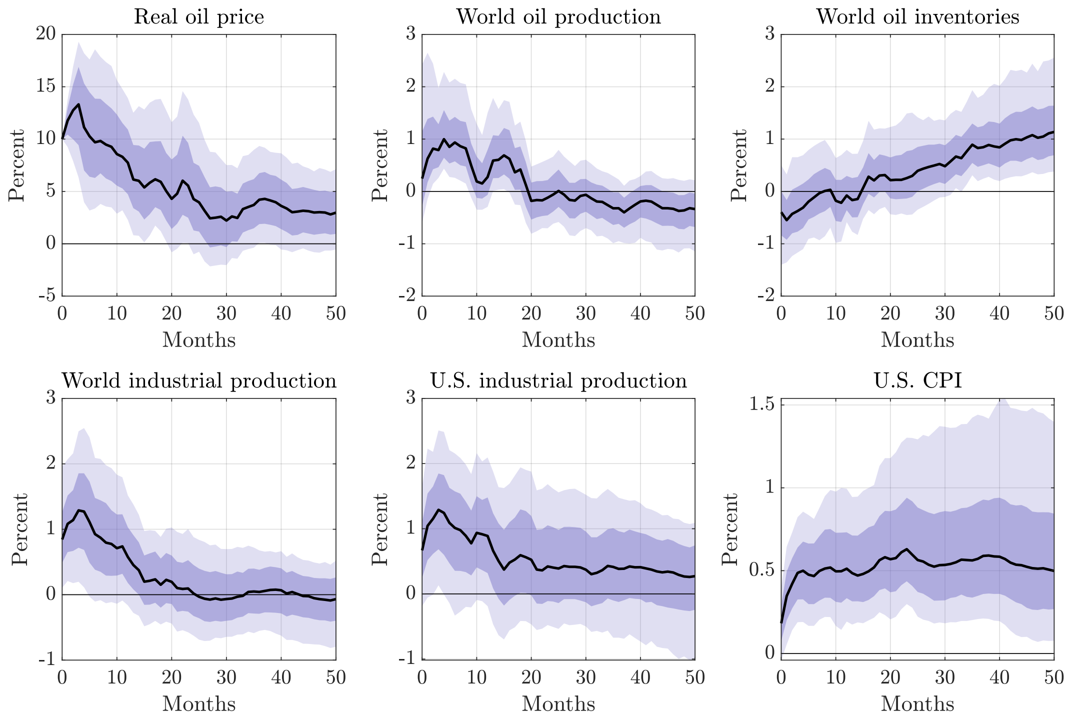
(b) VAR(7) Specification IRFs: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 20.26, robust F: 9.80, R^2: 3.52%, Adjusted R^2: 3.35%

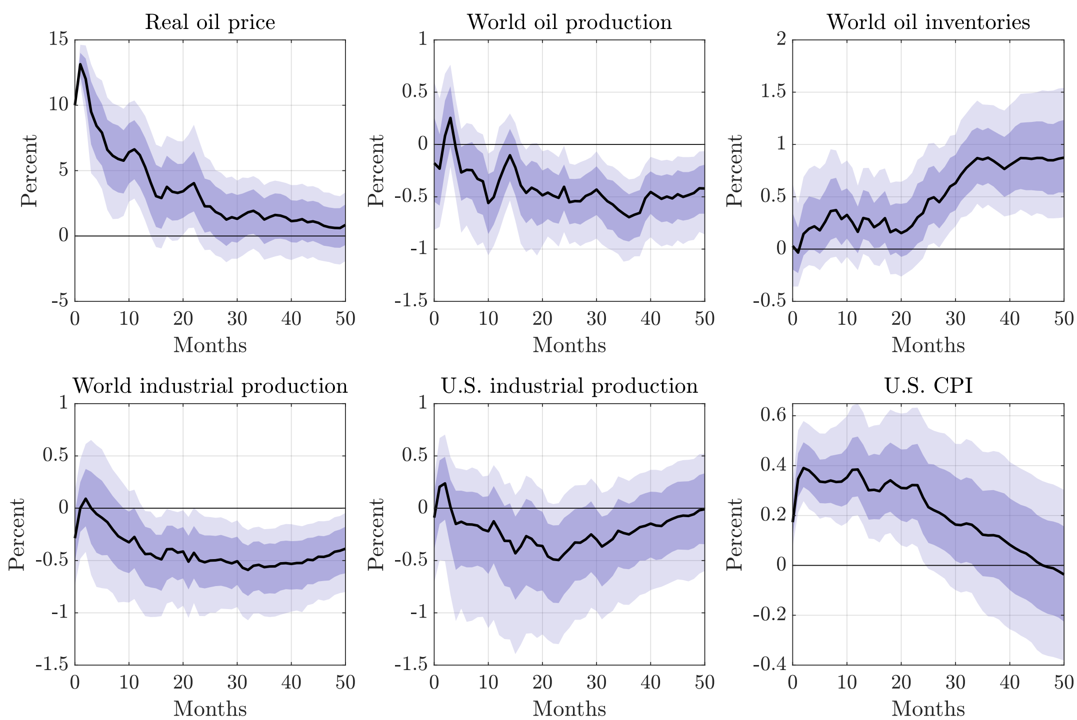
Figure 21: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

(a) VAR(24) Specification IRFs: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 11.71, robust F: 5.47, R^2: 2.13%, Adjusted R^2: 1.95%

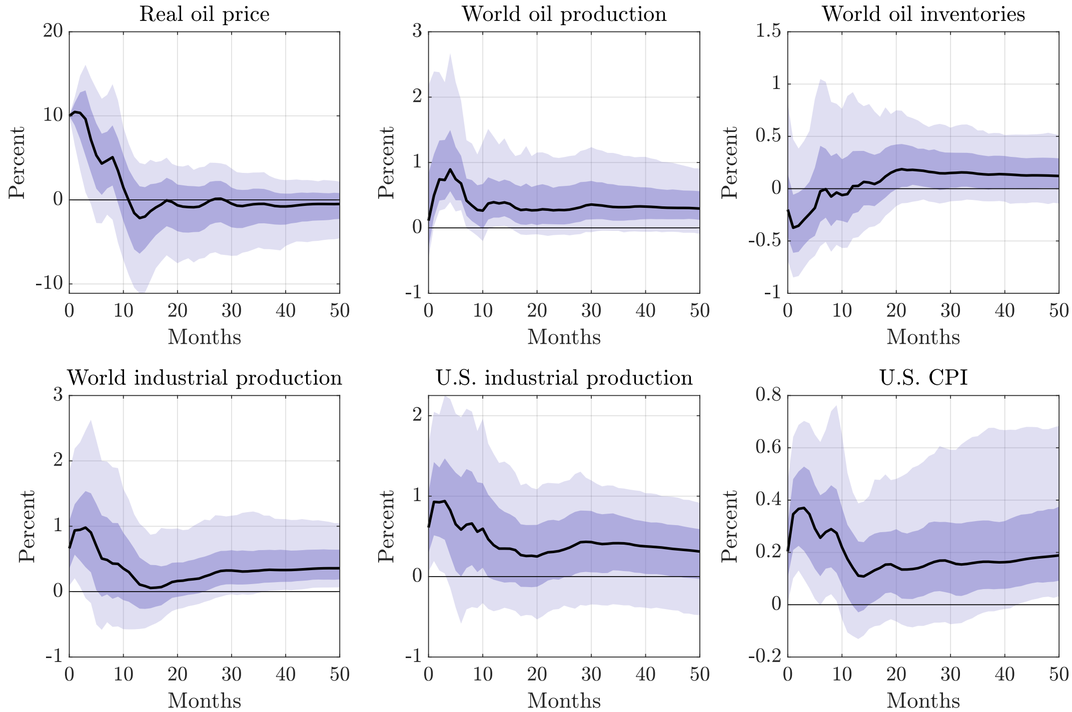
(b) VAR(24) Specification IRFs: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 17.09, robust F: 9.88, R^2: 3.08%, Adjusted R^2: 2.90%

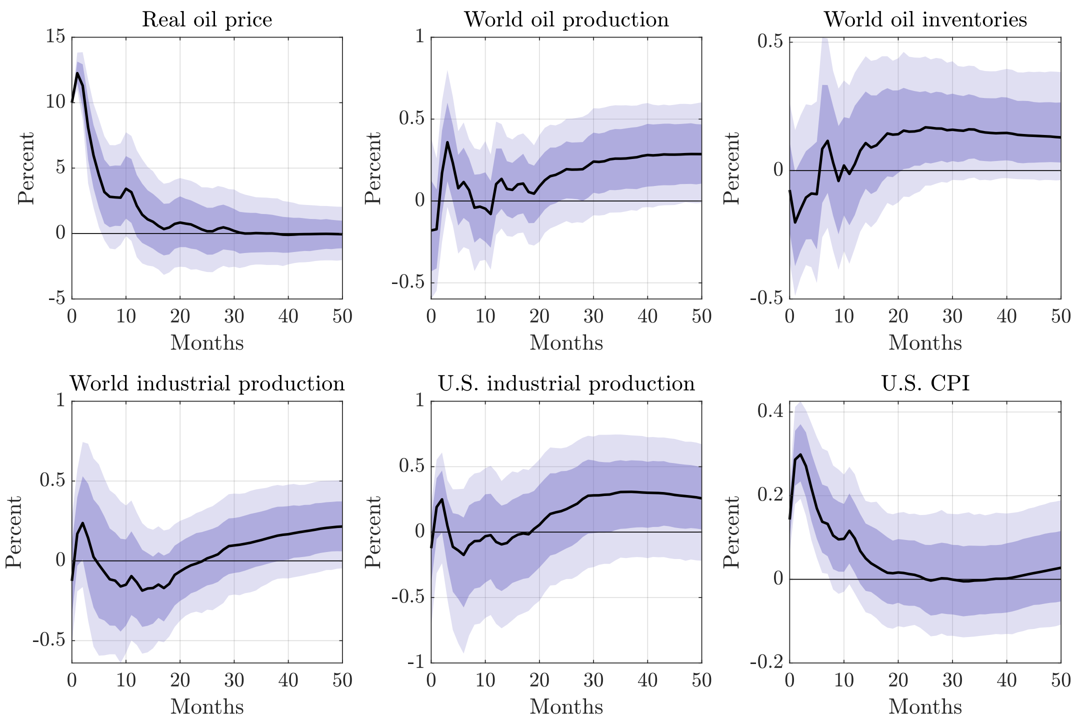
Figure 22: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

(a) IRFs: A shock to oil demand expectations driven by OPEC demand news, 1989.04-2020.12



First stage regression: F: 12.97, robust F: 6.75, R^2: 3.31%, Adjusted R^2: 3.05%

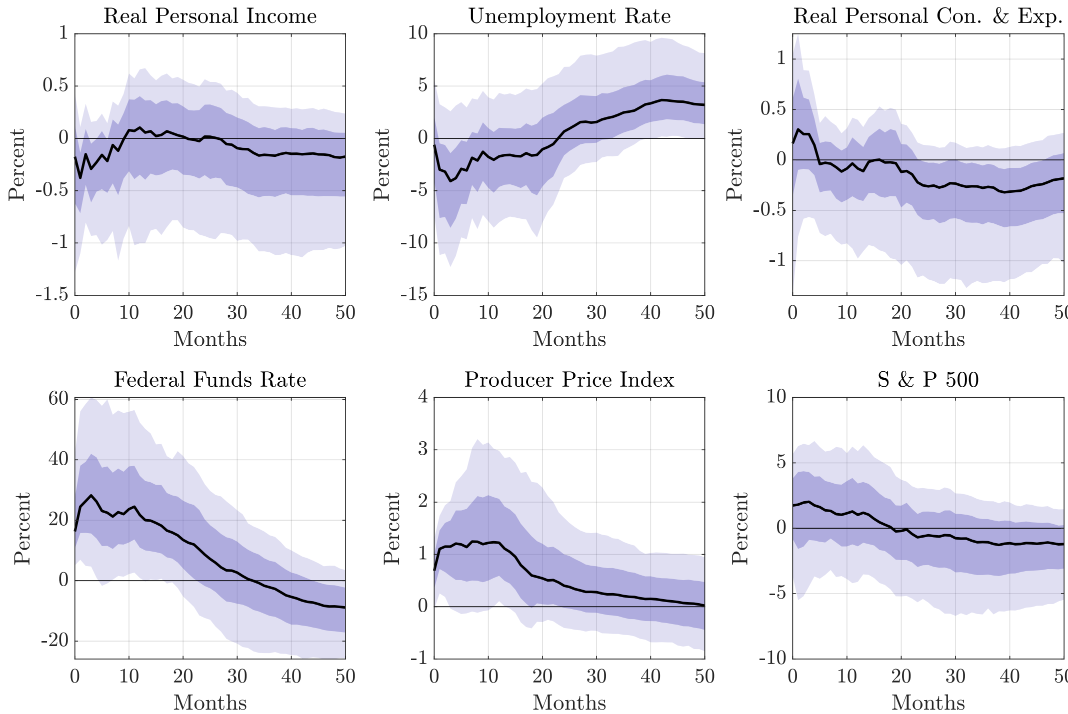
(b) IRFs: A shock to oil supply expectations driven by OPEC supply news, 1989.04-2020.12



First stage regression: F: 22.07, robust F: 15.34, R^2: 5.50%, Adjusted R^2: 5.25%

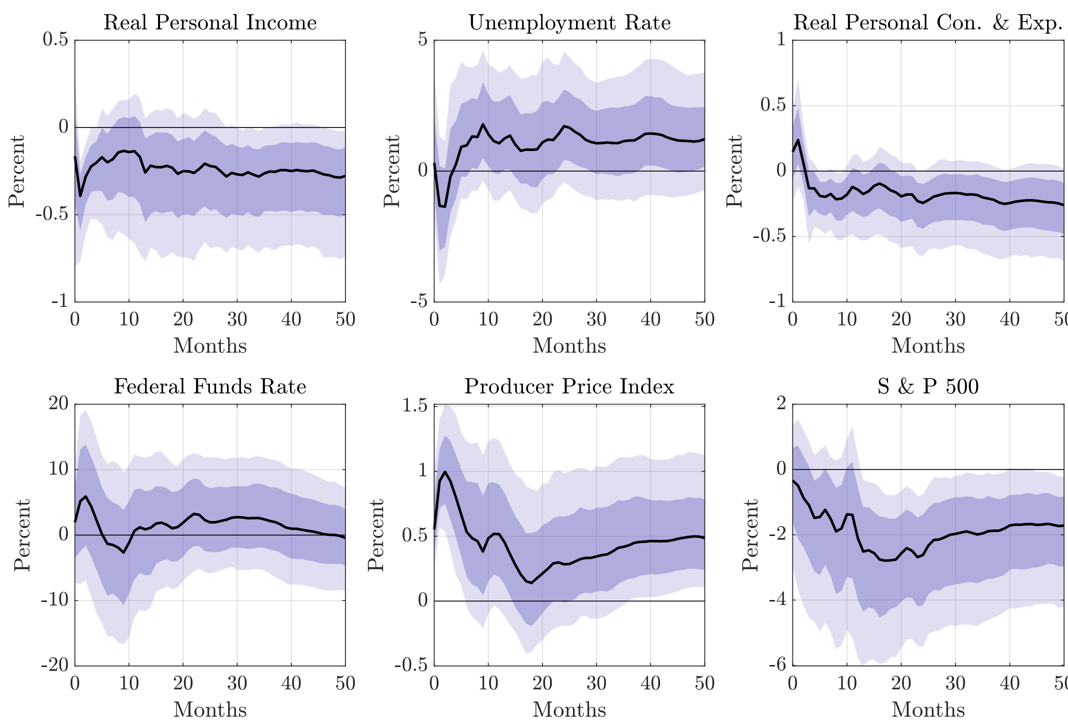
Figure 23: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

(a) IRFs: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 10.97, robust F: 7.09, R^2: 1.96%, Adjusted R^2: 1.78%

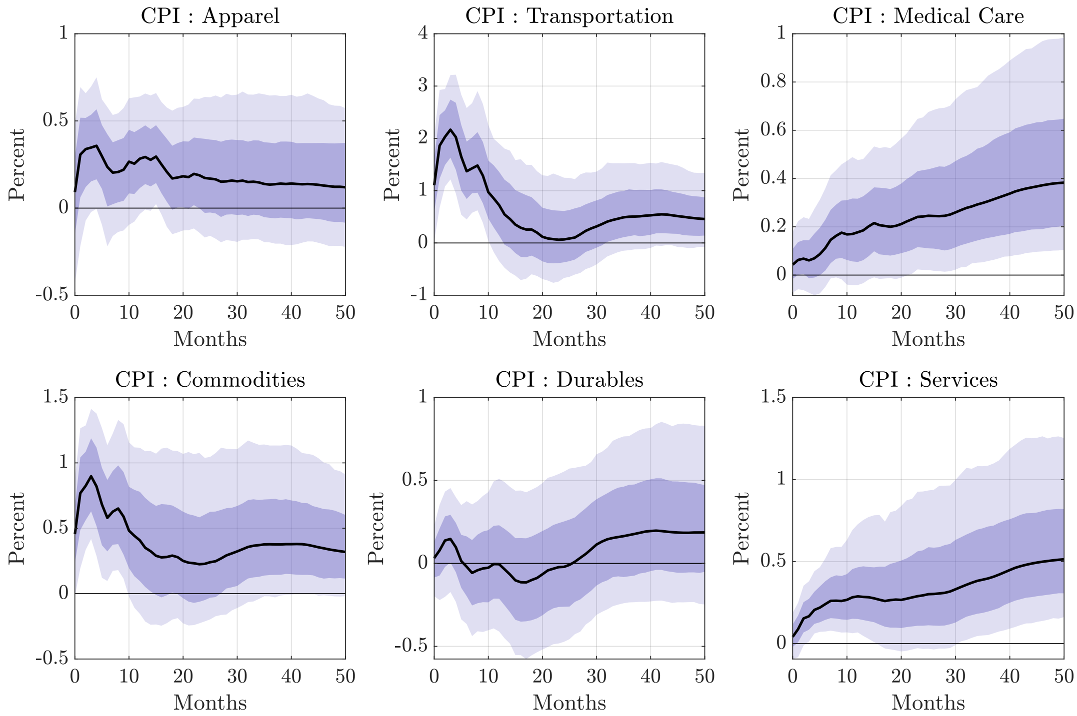
(b) IRFs: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 23.20, robust F: 13.99, R^2: 4.05%, Adjusted R^2: 3.87%

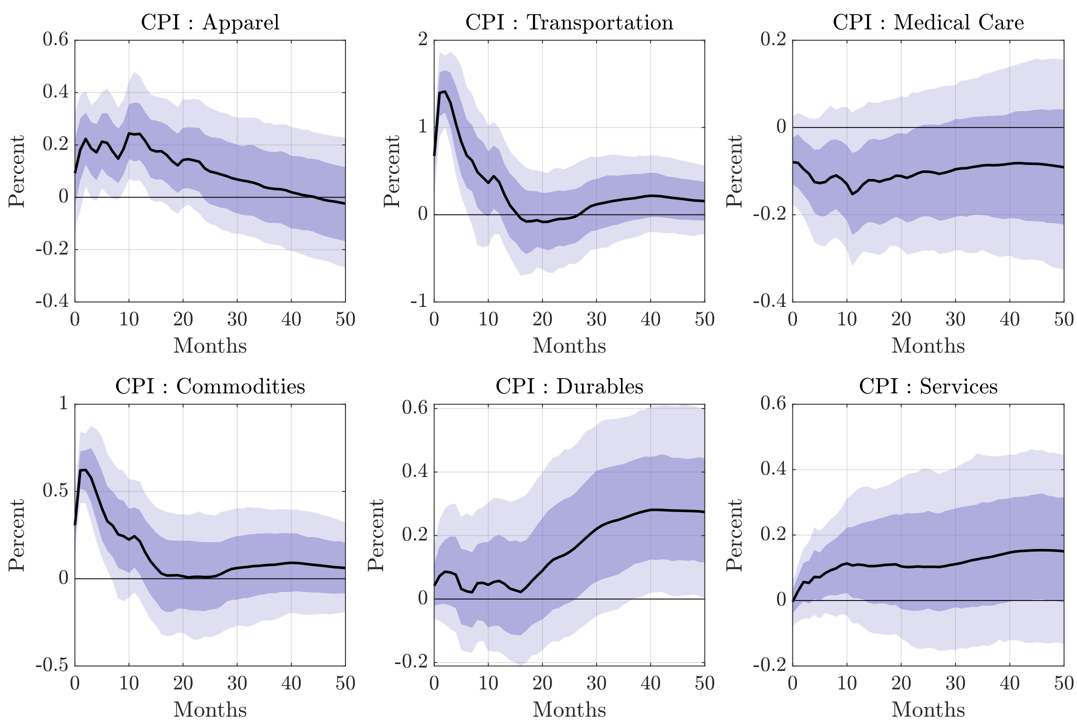
Figure 24: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

(a) IRFs: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 12.76, robust F: 6.16, R^2: 2.27%, Adjusted R^2: 2.09%

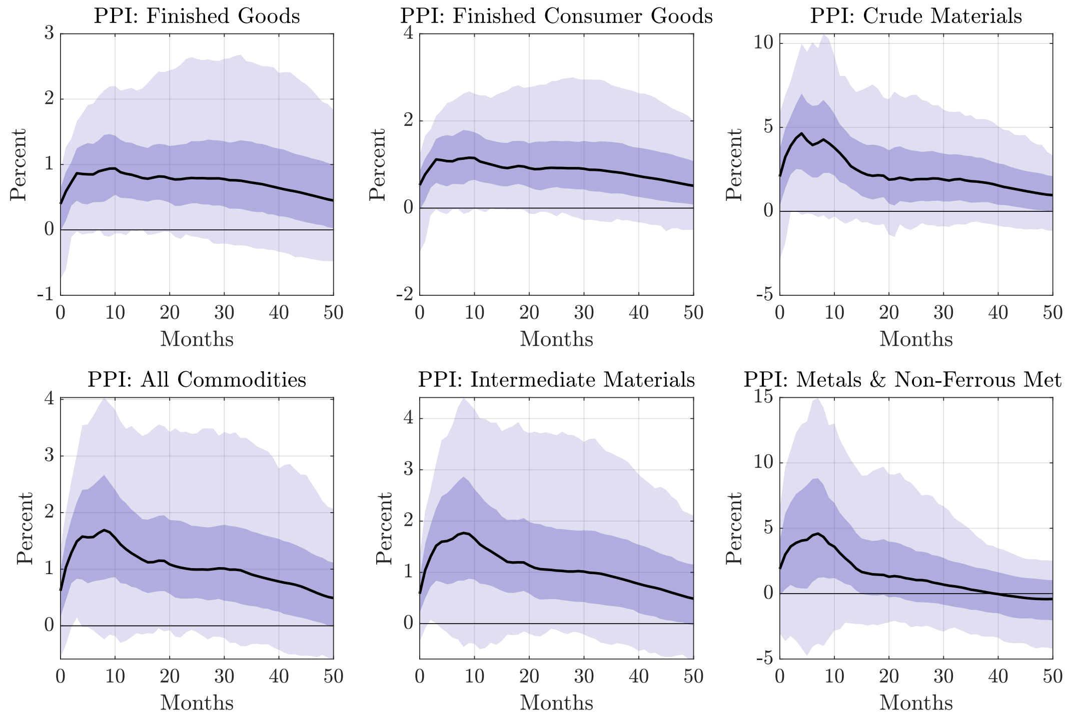
(b) IRFs: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 20.63, robust F: 9.25, R^2: 3.62%, Adjusted R^2: 3.44%

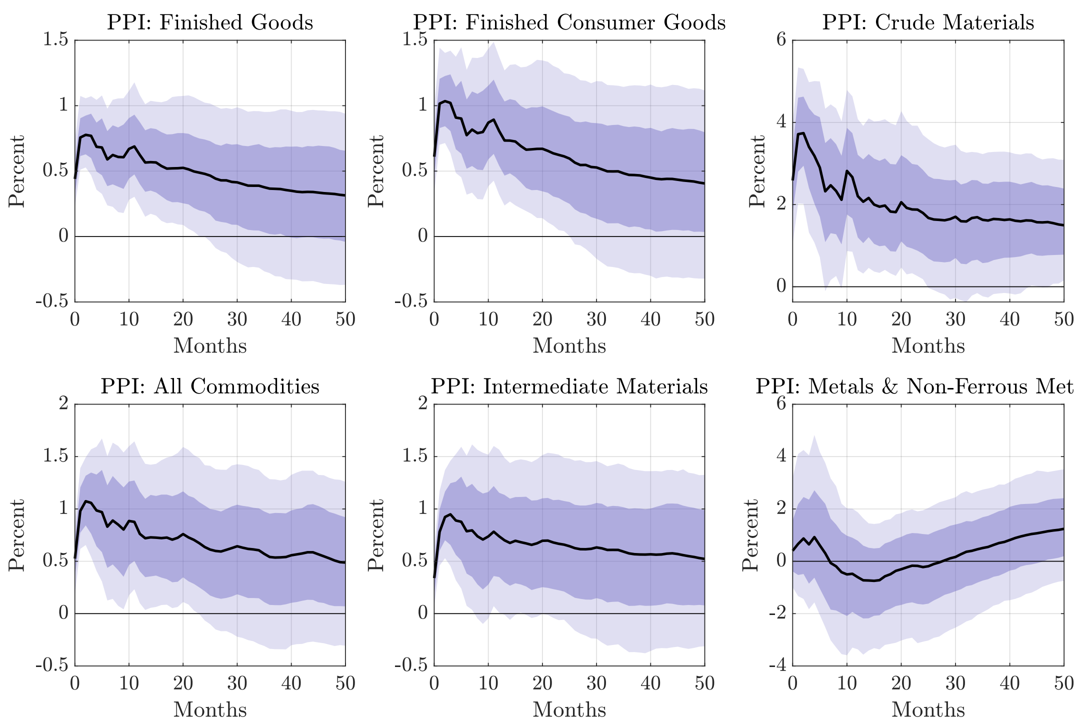
Figure 25: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

(a) IRFs: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 11.05, robust F: 5.82, R^2: 1.97%, Adjusted R^2: 1.79%

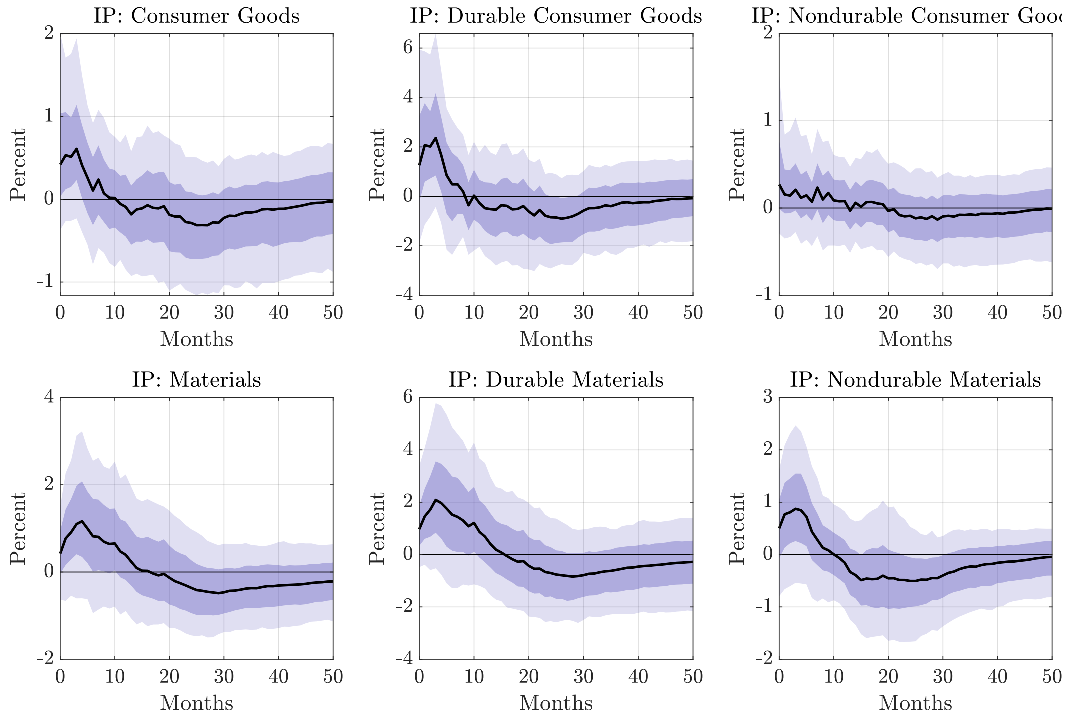
(b) IRFs: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 16.41, robust F: 9.16, R^2: 2.90%, Adjusted R^2: 2.72%

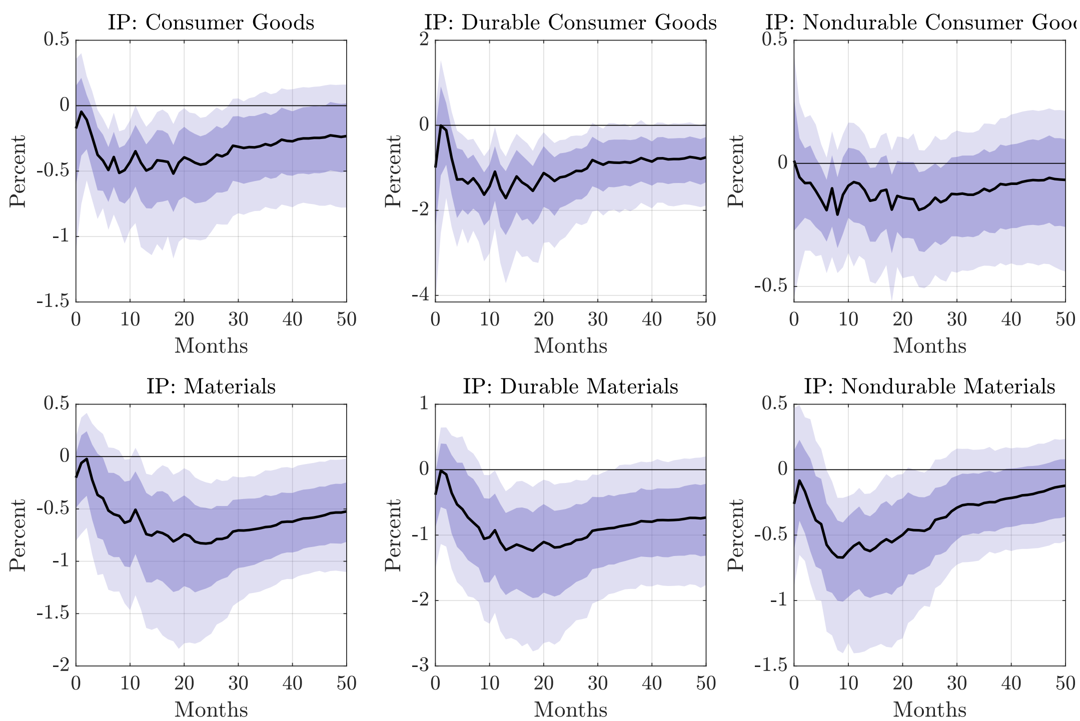
Figure 26: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

(a) IRFs: A shock to oil demand expectations driven by OPEC demand news.



First stage regression: F: 10.83, robust F: 5.13, R^2: 1.93%, Adjusted R^2: 1.75%

(b) IRFs: A shock to oil supply expectations driven by OPEC supply news.



First stage regression: F: 14.37, robust F: 11.79, R^2: 2.55%, Adjusted R^2: 2.37%

Figure 27: *Impulse Response Functions (IRFs) for OPEC News: Demand vs. Supply News*

E Data Sources

Variable	Description	Source	Code	Trans.
Real Oil Price	WTI spot crude oil price (deflated by CPI)	FRED	WTISPLC, CPIAUCSL	100 * log
World Oil Production	Crude Oil Production, World, Mbbl/day	EIA	EIA1955	100 * log
World Oil Inventories	Computed using Kilian and Murphy (2014)	EIA	EIA1976, EIA1533, EIA1541	100 * log
World Industrial Production	IP of OECD + 6 countries	Baumeister and Hamilton (2019)		100 * log
US Industrial Production	USIP, Index 2017=100, Seasonally Adjusted	FRED	INDPRO	100 * log
US Consumer Price Index	CPI for All Urban Consumers: All Items	FRED	CPIAUCSL	100 * log
Real Personal Income	Real Personal Income	FRED	RPI	100 * log
Unemployment Rate	Unemployment Rate	FRED	UNRATE	100 * log
Real Personal Consumption	Real Personal Consumption	FRED	DPCERA3M086SBEA	100 * log
Federal Funds Rate	Federal Funds Rate	FRED	FEDFUNDS	100 * log
Producer Price Index	Producer Price Index	FRED	PPIACO	100 * log
S&P 500	S&P 500 Index	FRED	SP500	100 * log
CPI: Apparel	CPI for Apparel	FRED	CPIAPPSL	100 * log
CPI: Transportation	CPI for Transportation	FRED	CPITRNSL	100 * log
CPI: Medical Care	CPI for Medical Care	FRED	CPIMEDSL	100 * log
CPI: Commodities	CPI for Commodities	FRED	CUSR0000SAC	100 * log
CPI: Durables	CPI for Durables	FRED	CUSR0000SAD	100 * log
CPI: Services	CPI for Services	FRED	CUSR0000SAS	100 * log
PPI: Finished Goods	PPI for Finished Goods	FRED	WPSFD49207	100 * log
PPI: Finished Consumer Goods	PPI for Finished Consumer Goods	FRED	WPSFD49502	100 * log
PPI: Crude Materials	PPI for Crude Materials	FRED	WPSID62	100 * log
PPI: All Commodities	PPI for All Commodities	FRED	PPIACO	100 * log
PPI: Intermediate Materials	PPI for Intermediate Materials	FRED	WPSID61	100 * log
PPI: Metals & Non-Ferrous Metals	PPI for Metals and Non-Ferrous Metals	FRED	PPICMM	100 * log
IP: Consumer Goods	Industrial Production: Consumer Goods	FRED	IPCONGD	100 * log
IP: Durable Consumer Goods	Industrial Production: Durable Consumer Goods	FRED	IPDCONGD	100 * log
IP: Nondurable Consumer Goods	Industrial Production: Nondurable Consumer Goods	FRED	IPNCONGD	100 * log
IP: Materials	Industrial Production: Materials	FRED	IPMAT	100 * log
IP: Durable Materials	Industrial Production: Durable Materials	FRED	IPDMAT	100 * log
IP: Nondurable Materials	Industrial Production: Nondurable Materials	FRED	IPNMAT	100 * log
3-Month Treasury Bill	3-Month Treasury Bill Rate	FRED	TB3MS	100 * log
6-Month Treasury Bill	6-Month Treasury Bill Rate	FRED	TB6MS	100 * log
Switzerland/U.S. Foreign Exchange Rate	Exchange Rate: Swiss Franc per USD	FRED	EXSZUSx	100 * log
Japan/U.S. Foreign Exchange Rate	Exchange Rate: Japanese Yen per USD	FRED	EXJPUSx	100 * log
U.S./U.K. Foreign Exchange Rate	Exchange Rate: British Pound per USD	FRED	EXUSUKx	100 * log
Canada/U.S. Foreign Exchange Rate	Exchange Rate: Canadian Dollar per USD	FRED	EXCAUSx	100 * log

Table 9: Data sources, transformations, and codes for the variables used in the analysis.