

Reading OPEC: Text-Based Identification of Oil Supply and Demand News

Gabriel A. Abeyie* Andrew Hanson†

October 16, 2026

Abstract

We use newspaper coverage to identify the supply and demand content of OPEC announcements. From 7,941 articles around 206 OPEC and OPEC+ announcements from 1971 to 2025, we construct separate text-based measures of supply and demand news without using futures prices. The supply measure produces macroeconomic responses similar to futures-based identification, while demand news generates distinct effects: both raise oil prices, but supply news contracts industrial production whereas demand news initially expands it.

Keywords: oil supply news; narrative identification; proxy SVAR; text as data; OPEC announcements; large language model

JEL Classification: C36, C55, E32, Q41, Q43

*Department of Economics and Finance, Cameron School of Business, University of North Carolina Wilmington; Email: abeyieg@uncw.edu.

†Corresponding author. Department of Economics, Haslam College of Business, University of Tennessee, Knoxville; Email: ahanson8@utk.edu.

1 Introduction

News about crude oil supply, demand, inventories, and oil market expectations affects crude oil prices. A major source of this information is OPEC’s meetings.¹ OPEC meetings are typically scheduled in advance, with extraordinary meetings convened when needed. This feature has made OPEC announcements useful for identifying oil-market shocks. A common approach is to use the change in oil futures prices around announcement dates as measures of unexpected OPEC supply news (Demirer and Kutan, 2010; Känzig, 2021).

This approach comes with an implied assumption that oil futures price movements around OPEC meeting announcement dates can be attributed solely to OPEC’s production decisions. Although OPEC’s production decisions are the central theme of the announcements, they may also contain information about global demand, spare capacity, compliance, or geopolitical conditions. Therefore, an increase in oil futures price around an OPEC meeting announcement can reflect an unexpected tightening of supply, stronger demand prospects, or both. All these components cannot be separated using only the price response. Degasper (2023) shows that oil futures price surprises around OPEC meetings can combine revisions in supply and demand expectations and separates the two using the high-frequency co-movement of oil futures and stock prices. A related concern arises from the findings of Kilian (2024), who shows that the structure and time aggregation of OPEC futures surprises affect their magnitude and economic interpretation, and that the resulting proxies contain substantial demand-related news. These results make the economic content of an OPEC announcement, rather than the price response alone, central to identification. Another limitation is that, historically, oil futures data begin only in 1983. For this reason, they cannot be used to study the major OPEC decisions of the 1970s.

Due to these concerns and limitations, we propose a different approach. We measure the content of an OPEC announcement directly from contemporaneous newspaper cover-

¹OPEC is an intergovernmental organization of major oil exporters and accounts for more than 35% of global crude oil production during our sample period.

agerather than inferring the content from the price response. We collect 7,941 articles surrounding 206 OPEC and OPEC+ announcements between 1971 and 2025 from the *New York Times*, *Wall Street Journal*, and *Washington Post*, including their historical editions. Prior to structural estimation, we classify each article by the OPEC decision announced, the oil market expectations prior to the decision, whether the decision was considered a surprise, and the direction and intensity of the supply and demand news.

The classified articles produce two separate measures of OPEC news: supply and demand measures. The supply measure is based on descriptions of the announced OPEC decision relative to oil market expectations. Thus, it reveals whether the OPEC decision was more or less restrictive than the markets expected. The demand measure is constructed independently from the demand outlook reported in the same OPEC announcement window and orthogonalized to predictable content.² Supply and demand distinction is important because two measures capture different information content in the same OPEC announcement. For instance, an OPEC announcement might be unexpectedly restrictive in terms of supply, while simultaneously there are reports of weakening global demand. Changes in oil futures price on OPEC meeting date combine all these components, while newspaper coverage allows capturing them separately.

We use the supply and demand measures as external instruments in the separate proxy-SVAR specifications (Stock and Watson, 2012; Mertens and Ravn, 2013). Our baseline VAR model is the same as in Känzig (2021) and contains real oil price, world oil production, world oil inventories, world industrial production, U.S. industrial production, and U.S. consumer price index. We use a common 1983:7-2017:12 sample, allowing comparison of the futures-based supply instrument, the text-based supply instrument, and the text-based demand instrument with Känzig (2021).

We present three main findings. First, OPEC announcements carry distinct supply and demand information with different macroeconomic consequences. Text measures are con-

²We report the raw and residualized measures separately and give the full diagnostic of predictability and autocorrelation in Appendix B.

structed independently and are only weakly correlated. A supply news shock which raises the real oil price decreases world and U.S. industrial production and raises consumer prices. The demand news shock generates another pattern of reaction, including expansion of industrial activity.

Second, the text-based supply measure is strongly associated with (but not subsumed by) the futures-based surprise. Over the common sample, the two measures are nonzero jointly in 91 months with correlation of 0.40 and sign agreement in 65% of observations. Also, both instruments are independently relevant for the innovation in the real oil price. When included together both coefficients are positive and significantly different from zero. Furthermore, the first-stage R^2 increases relatively to using either measure alone. Impulse responses identified by the two supply measures are similar. We interpret this similarity as validation of the newspaper measure capturing economically meaningful OPEC supply news, while also containing information not fully captured by the futures-price measure.

Third, the newspaper record extends announcement-based analysis before oil futures trading and into the OPEC+ period. The longer-sample specifications retain the contrasting supply- and demand-news responses, with first-stage relevance varying across sample windows (Table 5).

Our paper contributes to three related literatures. First, it contributes to the literature identifying structural shocks in oil markets. Existing approaches use timing restrictions, exclusion restrictions, sign restrictions, narrative information, and external instruments (Kilian, 2008, 2009; Alquist and Kilian, 2010; Kilian and Murphy, 2012; Inoue and Kilian, 2013; Caldara et al., 2019). The contribution of our paper is the measurement of the supply and demand information contained in the OPEC meeting announcements, rather than the composition of the news inferred from the subsequent price response. This approach is close to Degasperi (2023), who separates supply and demand using the high-frequency co-movement of oil future price and stock prices, and to Kilian (2024), who shows that OPEC future surprises might contain the demand information and are sensitive to the temporal aggregation.

We address the same identification problem from the other perspective; that is, the news type is measured from the content of the announcement itself, rather than inferred from additional financial market response.

Second, our paper is also connected with the literature that employs the movements in oil future prices due to OPEC announcement as the external instrument ([Demirer and Kutan, 2010](#); [Känzig, 2021](#)). We employ [Känzig \(2021\)](#) in order to benchmark the text-based supply measure. This allows us to achieve two goals. First, it lends external validation to the text measure over the period when the well-established oil price futures proxy exists. Secondly, it allows us to ask whether the contemporary newspaper coverage contains information beyond the oil price futures response itself. Unlike the oil price futures measure, text allows us to distinguish supply from demand news content and to study OPEC announcements before the availability of oil price futures data.

Third, we contribute to the expanding literature employing text for measuring economic news and shocks identification ([Aruoba and Drechsel, 2021](#); [Ochs, 2021](#); [Handlan, 2020](#)). The text-based literature for oil markets has mostly focused on forecasting ([Li et al., 2019](#); [Bai et al., 2022](#); [Jiao et al., 2022](#); [Abeyie, 2026](#)), construction of news indexes ([Plante, 2019](#)), narrative event coding ([Caldara et al., 2019](#)) and keyword-based measures ([Datta and Dias, 2020](#)). Instead of that, we classify the contemporary newspaper coverage in relation to the fixed calendar of OPEC announcements with the classification scheme designed prior to structural estimation. This allows measuring expectations, surprises and the supply and demand content of the same event while considering the measurement issues mentioned in [Kilian and Zhou \(2020, 2023\)](#).

The rest of the paper is organized as follows. In section [2](#), we describe the OPEC meeting calendar, newspaper corpus, classification scheme and construction of the text-based news measures. In section [3](#), we ask what the text adds validating the measures against historical episodes and the futures-based surprise. In section [4](#), we describe the econometric framework and macroeconomic data. In section [5](#), we estimate the macroeconomic effects

of OPEC supply and demand news and compare the two channels. In section 6, we study the changes in announcement-based identification as the benchmark sample is expanded over different OPEC eras. In section 7, we perform robustness exercises and additional instrument diagnostic. In section 8, we conclude.

2 Measuring OPEC News

2.1 OPEC announcements and institutional eras

The definition of an "OPEC announcement" has evolved, which is crucial for identification. During the early 1970s and until the start of the 1980s, OPEC announcements were mainly about adjustment of posted crude-oil prices, while production levels were less important. Following the mid-1980s until 2016, production quotas became the dominant policy instrument. Decisions were announced at scheduled ministerial meetings and sometimes during extraordinary meetings. This period is the standard for event studies using futures data.

The mechanism was altered since the creation of OPEC+ in late 2016. Meetings and announcements increased in frequency. Forward guidance about future production plans became a policy lever. Beginning in 2023 several OPEC+ members also announced voluntary production cuts over weekends, before oil futures markets reopened. These announcements complicate the measurement of policy surprises, because the resulting futures-price adjustments may incorporate other information released while markets were closed. We thus do not consider all OPEC announcements between 1971 and 2025 to be of the same kind of event. In this paper, our event calendar captures the institutional context in which each of the announcements was made. As shown in Section 6, the distinction is important because the informational content is regime-specific.

2.2 The OPEC announcement calendar, 1971-2025

Our event calendar contains 206 OPEC and OPEC+ production policy announcements. We begin with the 119 dates of announcements found in [Känzig \(2021\)](#). By using the common corpus, we can directly compare the text-based and futures-based measures of the same set of events. Then, we extend the calendar through December 2025 by including additional OPEC+ ministerial meetings and voluntary cut announcements, many of which occur on Sundays. Finally, we extend the calendar backwards to February 1971, including 37 decision events that occur prior to the existence of oil futures.

Our construction of the pre-1983 calendar proceeds in two stages. First, we use the newspaper corpus to detect periods of anomalously high coverage of OPEC relative to a local baseline of coverage. This process identifies the major OPEC conferences from 1974 to 1983 and ends with the July 1983 meeting that is the beginning of the sample in [Känzig \(2021\)](#). Second, we check the candidate dates identified in the coverage-based procedure for appearance in the records of the OPEC Secretariat’s production agreement documents and US State Department archives. Events that are not detected in the coverage-based procedure are included based on documentation provided in these records.

We label four of these events as *masked coverage* events: the Tripoli agreement, the second Geneva currency adjustment, the October 1973 Gulf committee price increase, and the December 1973 Tehran price increase. The latter two occurred during times of significant war and embargo coverage. We mask them so that they do not produce anomalous OPEC coverage in our procedure.

Two exclusion criteria are applied. Decisions by Organization of Arab Petroleum Exporting Countries (OAPEC) are recorded but excluded from the baseline calendar because OAPEC is a distinct entity. For example, the October 1973 embargo was an OAPEC decision, while the posted price increase that happened at that time was an OPEC decision. Additionally, we record the unilateral pricing decisions of the individual member countries from 1979-1981 as a separate event category. They are not included in the baseline event

calendar; only in our robustness checks.

Each of the pre-1983 events is either externally verified, coverage dated, or included in the event calendar from an official source. We also record the institutional era of each event and, for OPEC+ announcements, whether the meeting was a scheduled rollover. A scheduled rollover reaffirms an existing production agreement or confirms the continuation of a previously announced production path.³

Table 1 summarizes the calendar by institutional era. Two features stand out. First, the pre-futures era contributes 37 announcement events that cannot be studied using oil-futures surprises. Second, the OPEC+ calendar includes frequent meetings that confirm existing production plans rather than revise them. These rollovers are still policy decisions and can contain unexpected news when markets anticipate a change. Meeting counts therefore should not be interpreted as counts of distinct policy changes or announcement surprises.

The event calendar is fixed before we classify the articles or use any price data. Event inclusion follows the rules described above and reported in Appendix A. The article-classification scheme is also fixed before estimation (Section 2.4). Therefore, event calendar and article-level classification are thus not chosen based on the price changes or structural effects that the instruments aim to explain.⁴ Oil price futures are used much later, when we construct and estimate the instruments (Section 2.6).

2.3 Newspaper corpus

Articles in this study are collected using a ProQuest TDM Studio dataset that includes *The New York Times*, *The Wall Street Journal*, and *The Washington Post*. We employ a full-text search for articles containing OPEC or the Organization of the Petroleum Exporting Countries in their titles, which excludes advertisements, classified ads, and stock tables. There are around 47,000 articles in total for the 1960-2025 sample period. News about OPEC

³Continuing a previously announced production path need not mean holding production constant. It can involve implementing scheduled increases or reductions already agreed at an earlier meeting.

⁴The subsequent mapping from the fields that have been classified and the instruments themselves was changed once after the first stage diagnostic tests, as explained in Section 2.6 and Appendix B.

Table 1: The OPEC announcement calendar by institutional era, 1971-2025

	Posted-price	Leapfrog	Quota era		OPEC+
	1971–78	1979–83	1983–99	2000–16	2016–25
Decision events	24	13	49	67	53
Conferences	14	11	49	67	0
Committee decisions	10	2	0	0	0
ONOMM / OPEC+ meetings	0	0	0	0	53
of which scheduled rollovers	0	0	0	0	35
Events in Känzig (2021)	0	0	49	67	3
Tier: externally verified	8	0	–	–	–
Tier: coverage-dated	14	13	–	–	–
Tier: official-list added	2	0	–	–	–
of which masked-coverage	4	0	–	–	–
Event-window articles	611	997	2,096	2,621	1,616
Articles per event (median)	22	79	34	35	24

Notes: There are 206 OPEC and OPEC+ production policy decision-making events. We employ the 119 announcement dates from Känzig (2021) as is; since the quota era columns begin with the July 1983 conference, which marks the starting point of that sample, the first two columns contain exactly 37 pre-futures events. Committee events relate to price decisions made by ministerial committees instead of the larger conferences; ONOMM stands for OPEC and Non-OPEC Ministerial Meetings in the OPEC+ era. The pre-1983 events are classified according to their source tier (external verification from OPEC Secretariat and U.S. State Department records; date coding based on abnormal-coverage clusters; officially listed addition). Coverages with masked events have been officially added as new cases due to their absence of abnormal coverages, since they are announced during the period of news environment dominated by the war and embargo coverages. Unilateral price decisions of the members in the leapfrogging period are kept separate as a distinct category from the baseline calendar. News stories between two days prior and three days following an event make up the event window articles; each of the 7,941 corpus news articles is matched to its nearest event in the calendar. The quota era split is used for exposition purposes only, since the estimate samples do not include this split point.

prior to 1971 can be considered virtually non-existent (OPEC had essentially no presence in U.S. newspapers before the Tehran conference), thereby determining the starting point for our dataset. Our estimation corpus consists of the 7,941 articles from a two- to three-day window around the calendar event.

Our coverage audit justifies this design choice. Newspaper print coverage arrives the day *after* an OPEC announcement. Hence, articles published on the OPEC announcement day itself preview the meeting. Whereas in the online era same-day news articles are frequently first-instance decision reports. Fixed date windows therefore misclassify article roles differ-

entially across eras. We instead assign each news article’s role (preview, decision report, retrospective analysis) from its content, as described in Section 2, retaining date-window assignment only as a robustness check.

2.4 Classification scheme

The classification scheme was fixed prior to merging newspaper text and oil futures price data. Changes made afterwards are documented in the dated changelog available in Appendix A. This is important, because a narrative instrument should not be constructed in light of the outcomes it tries to explain. Kilian and Zhou (2023) pointed to this problem in judgment-based narrative series.

For each news article, the classification scheme includes the answer to whether the article is about an OPEC production-policy meeting or a pricing decision. It includes the role of the news article in relation to the decision.⁵ The classification is done according to the content of the news article, not its publication date. Information is included about whether the decision pertains to posted oil futures prices or production amounts, its direction, and its magnitude, if any. Coding of the direction of surprise takes place only if the article contains a comparison between the decision and the expectations, for instance “*larger than expected by the analysts*” or “*in line with forecasts.*”

In addition, two separate supply and demand intensities are coded from 0 to 3 for each news article. The intensities can take large values on both axes. The direction of demand outlook and whether there is a description of a geopolitical situation in the article are also coded.

Two aspects of the design matter. First, the supply and demand news are measured independently. Second, the surprise variable is based entirely on the text. Since the surprise measure uses the language of expectations used in contemporaneous coverage, it can be constructed for Sunday announcements when futures markets are closed, announcements

⁵Preview, first news of the decision, retrospection, or background information.

whose outcome was anticipated, and OPEC decisions preceding the introduction of the oil futures market.

2.5 Classification and Validation

We perform the classification exercise using a large language model (LLM) (`claude-sonnet-4-6`, temperature zero, strict JSON output validated field-by-field against the schema), applied to the 7,941 news articles falling within a window of two days before to three days after each calendar event. The classifier receives only the news article’s title, text, date, and source; it is never shown the event calendar, meeting identities, oil futures prices, or any market data.

We assess the classifications using an independent human audit drawn from the newspaper corpus. A human coder observes the same news article text but not the machine labels, event identities, oil futures prices, or subsequent market outcomes. Agreement is high for the fields entering the empirical measures: human and machine classifications agree on the demand-news inclusion threshold in 98.9 percent of the validation sample ($\kappa = 0.974$), while weighted agreement for demand-content intensity is $\kappa = 0.931$. Detailed field-level results, including validation for historical editions, are reported in [Appendix A.2](#).

2.6 Constructing supply and demand news

We set the rules for converting news article classifications into instruments in advance. We revised them once, after examining the first-stage results and before structural estimation. [Appendix A](#) documents the revision, and [Section 6](#) discusses what it reveals about OPEC decisions.

For event e , let $\mathcal{A}_e$ contain the decision reports and retrospective news articles that report the direction of the surprise. We set the values $s_a = +1$ if news article a describes the policy as tighter than expected, -1 if looser than expected, and also 0 if it matches expectations.

Supply news for the event is

$$z_e^{\text{sup}} = \frac{1}{\sqrt{n_e}} \sum_{a \in \mathcal{A}_e} s_a, \quad n_e = |\mathcal{A}_e|, \quad (1)$$

so agreement across more news articles receives more weight. Also, positive and negative news article assessments offset one another. The measure is zero if they cancel.

Equation (1) uses the decision relative to expectations, not the announced policy change alone. As Section 6 shows, oil production cuts often occur when oil markets are weak. The decision therefore responds to market conditions and can have the wrong sign in first-stage regressions when used without the comparison with expectations.

We construct demand news without using the supply measure. Let $\mathcal{D}_e$ contain the event's news articles with a demand-content score of at least two. Let w_a be that score and $d_a \in \{+1, 0, -1\}$ indicate a stronger, mixed or absent, or weaker demand outlook. The raw demand measure is

$$z_e^{\text{dem,raw}} = \frac{\sum_{a \in \mathcal{D}_e} w_a d_a}{\sum_{a \in \mathcal{D}_e} w_a}, \quad (2)$$

a weighted average between -1 and 1 . This records the oil demand outlook reported around the OPEC announcement. It does not compare that outlook with expectations before the OPEC announcement. Hence, it may include information already available before the meeting.

Thus we subtract the component fitted from this past information. Let X_{e-1} contain oil-market and macroeconomic information available before event e . We estimate

$$z_e^{\text{dem,raw}} = \alpha + \Gamma' X_{e-1} + \nu_e, \quad (3)$$

using events from 1975:1-2017:12. The predictors are one lag of each of the six variables in the baseline oil-market VAR, transformed to stationary form. We use the residual as the

demand news instrument:

$$z_e^{\text{dem}} = \hat{\nu}_e. \quad (4)$$

Exclusion of the supply news is intentional; the adjustment does not require the supply and demand news measures to be uncorrelated. We retain $z_e^{\text{dem,raw}}$ for describing the series and comparing it with historical events. Section 5.1 reports first-stage results for both the raw and adjusted demand measures. Appendix B.2 reports the predictability and autocorrelation tests.

We convert supply news to monthly values using the timing rule of Kilian (2024). The real oil price in the VAR is a monthly average, so an announcement near the end of a month affects only part of that month’s average. For an announcement on day d of a month with D days, we assign the share $(D - d + 1)/D$ to that month and the remainder to the following month:

$$z_m^{\text{sup}} = \sum_{e: m(e)=m} \frac{D_e - d_e + 1}{D_e} z_e^{\text{sup}} + \sum_{e: m(e)=m-1} \frac{d_e - 1}{D_e} z_e^{\text{sup}}. \quad (5)$$

Months receiving no contribution from OPEC announcements in the current or preceding month are assigned zero, following the zero-coding convention in Känzig (2021). Also, the timing adjustment matters because OPEC conferences often occur near the month-end. This places much of their weight in the following month.

We then assign each oil demand measure to the announcement month without splitting it across months. These measures describe the demand outlook reported around the announcement, rather than a price response dated within the month. If several events occur in a month, we sum their values:

$$z_m^{\text{dem,raw}} = \sum_{e: m(e)=m} z_e^{\text{dem,raw}}, \quad z_m^{\text{dem}} = \sum_{e: m(e)=m} z_e^{\text{dem}}. \quad (6)$$

Months receiving no event contribution are assigned zero. As robustness checks, we replace equation (1) with the news article mean or the unscaled sum. We also check an alternative

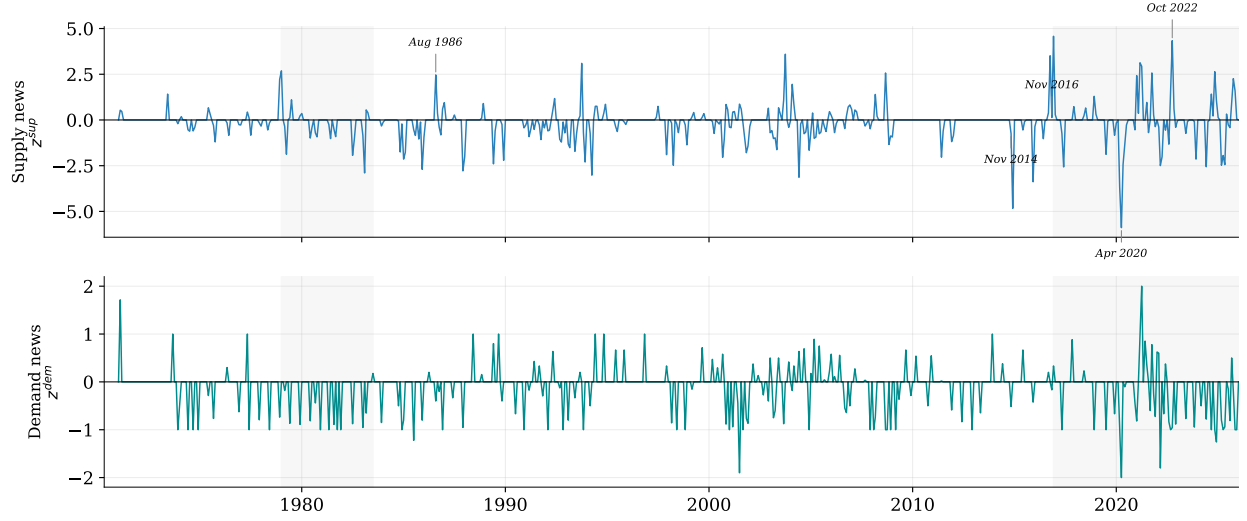


Figure 1: Text-based OPEC news measures, 1971-2025. The figure plots monthly oil supply news (z^{sup}) and the raw oil demand-outlook measure ($z^{\text{dem,raw}}$) from the classified newspaper articles. Shaded bands mark institutional periods. The labelled episodes are listed in Table 2.

that assigns all supply news to the announcement month.

The raw text measures cover all periods in the sample (Figure 1), including the years before oil futures trading began and Sunday announcements during the OPEC+ period. The structural analysis compares the futures, text-supply, and adjusted demand instruments over the common 1983:7-2017:12 sample. Section 6 uses the longer text record to examine how identification from OPEC announcements changes outside this period.

3 Informational Content of Text

In this section, we examine what the news text-based measures capture. We first compare the measures with well-known OPEC episodes and then with the oil futures-based surprise over their common sample period. Next, we ask whether the text contains information beyond the oil futures surprise and whether the separately constructed supply and demand news measures capture different types of news.

3.1 Historical validation

Table 2 shows the news measurements associated with events for which one might expect the most frequent checking by readers. The November 2014 decision to produce through the glut is the largest dovish realization observed in the sample. Thirty-five news stories carrying a surprise, almost unanimously believing that the surprise was more dovish than expected (-5.58). Also, the October 2022 two-million-barrel cut ($+4.67$) and the November 2016 Algiers agreement ($+4.00$) are the strongest hawkish measures in the sample, with the August 1986 conference that ended the price war by restoring quotas at $+2.83$. The April 2020 agreement case is illuminating. The largest cut in OPEC’s history is seen not as a hawkish but as a very dovish event (-4.01) because of its contemporaneous media coverage uniformly believing it to be insufficient to compensate for the demand collapse due to the pandemic. The oil price responded in kind. A measure that would use the level of announced decisions would see this event as the sample’s most hawkish realization while the surprise measure sees it exactly as media did, predicting the result of the endogeneity test in Section 5. The demand column tells a complementary story. The October 2022 announcement, seen as the sample’s most hawkish realization in the supply news measure, arrives along with a unanimously weak demand coverage ($z^{\text{dem, raw}} = -0.96$). This motivates constructing two separate measures. As an external reference, the largest realizations of Känzig (2021)’s futures surprise (November 2014, August 1986, March 2020, November 2016) have the same sign as the text measure. While being constructed from entirely disjoint data.

3.2 Text news and the futures surprise

Over the common 1983:7-2017:12 sample period, the text-based supply measure and the futures surprise are jointly nonzero in 91 months. Their correlation is 0.40, signs agree in 65 percent of those months, and the squared correlation is 0.16. The two measures contain related but far from identical information. Over the extended sample (through 2025), the measures are jointly nonzero in 129 months, their correlation is 0.39, and sign agreement

Table 2: Face validity: News measures at major episodes

Event	Decision	z^{sup}	n	$z^{\text{dem, raw}}$
August 1986	Re-establishment of quota system, end of price war	+2.83	18	-0.40
November 2014	Maintenance of production in glut	-5.58	35	-0.52
November 2016	Implementation of Algiers/Vienna cuts	+4.00	16	-0.17
March 2020	Failure of OPEC+ negotiations	-4.38	23	-1.00
April 2020	Record 9.7 mb/d cut into demand collapse	-4.01	18	-1.00
October 2022	Two-million-barrel cut	+4.67	31	-0.96

Notes: z^{sup} is the value of supply news for an event included in the baseline instrumental variable. Sum of directions of surprises at the level of articles (+1 = more hawkish, 0 = as expected, -1 = more dovish) in n articles containing surprises within the event, weighted by $\sqrt{n}$. $z^{\text{dem, raw}}$ is the mean demand outlook direction in the articles containing demand surprises (in $[-1, 1]$).

remains 65 percent (Figure 2).

The agreement is also visible in the major OPEC episodes. The largest realizations of the futures surprise have the same signs in the text measure, including August 1986, November 2014, November 2016, and March 2020 (Table 2). Beyond these large episodes the two measures differ more often, because they measure OPEC news in different ways. The text measure uses the expectations reported in contemporaneous newspaper coverage, whereas the futures surprise measures the price change around the announcement.

Table 3 tests whether the text supply measure adds information beyond the futures surprise over the common sample. Both coefficients remain positive and statistically significant when the proxies enter jointly, and first-stage R-squared increases from 4.7 percent with futures alone to 6.7 percent. The text measure therefore contains incremental information about the reduced-form oil-price innovation rather than simply reproducing the futures response.

3.3 Supply and demand news

The supply and demand news measures are constructed from separate fields in the classification scheme. Supply news is based on surprise language surrounding OPEC decisions, while demand news is based on the direction of the demand outlook reported in articles with

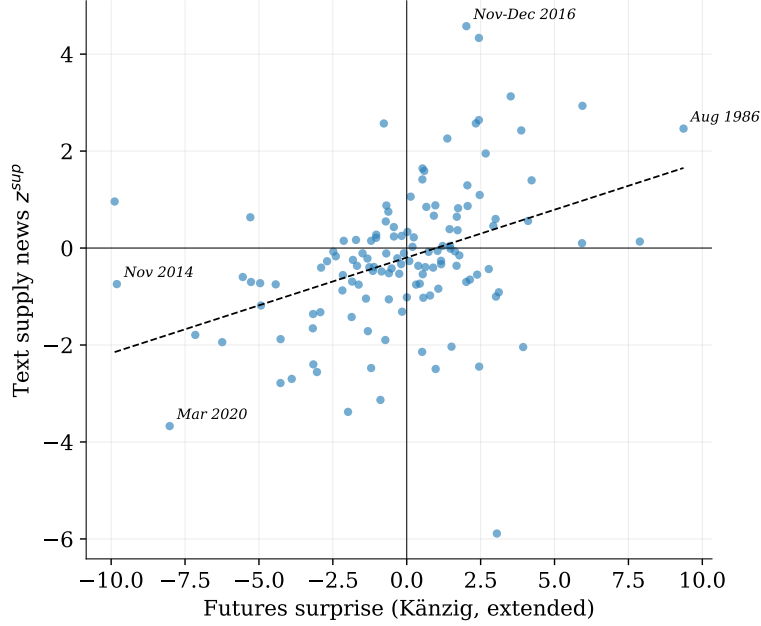


Figure 2: Text supply news and the extended futures-based OPEC surprise. The figure plots the 129 months between 1983 and 2025 in which both measures are nonzero. The dashed line shows the least-squares regression fit. Across the extended sample, the correlation is 0.39 and sign agreement is 65 percent. Over the benchmark 1983:7-2017:12 sample period, the measures are jointly nonzero in 91 months, their correlation is 0.40, and sign agreement is also 65 percent. Labeled observations correspond to the episodes reported in Table 2.

Table 3: Information in text beyond the futures surprise

	Futures only	Text only	Both
Futures surprise	1.11 (3.44)	-	0.89 (2.72)
Text Supply news	-	1.81 (3.83)	1.37 (2.84)
Robust F	11.84	14.68	12.33
R^2 (%)	4.7	4.0	6.7
N	414	414	414

Notes: The dependent variable is the reduced-form innovation in the real oil price from the baseline VAR. All the columns use the common 1983:7-2017:12 sample period. The heteroskedasticity-robust t -statistics are reported in parentheses. The robust F follows Montiel Olea et al. (2021). The oil futures surprise is from Känzig (2021) and the text measure is defined in Section 2.6.

substantial demand content.

Despite the two measures being constructed independently, the raw supply and demand

news measures are only weakly correlated. These correlations are calculated before the predictability adjustment to the demand measure. They therefore reflect the information produced by the classification itself rather than the subsequent construction of the structural demand instrument. Their correlation is 0.15 in the monthly series over 1971-2025 and 0.19 across the 192 events for which both measures are defined. The low correlation therefore arises from the classified news rather than from an econometric restriction.

Nevertheless, individual announcements can contain both types of news. For example, the October 2022 announcement combines the strongest supply-tightening reading in Table 2 with a strongly negative demand outlook. This illustrates why we measure the two dimensions separately. An OPEC decision may convey news about supply while contemporaneous coverage also contains information about the state of oil demand.

4 Econometric Approach

The text-based series reflect information about unexpected changes in OPEC meetings but are not structural shocks themselves. They can represent only a subset of the actual news and contain measurement errors. Therefore, we employ the text measures as proxies for latent shocks in a vector autoregression. This approach lets us identify the news shock based on information included in the announcement without interpreting the measure of news as the shock itself (Stock and Watson, 2012; Mertens and Ravn, 2013).

4.1 External-instrument identification

Let x_t be an $n \times 1$ vector of monthly variables. We consider the reduced-form VAR

$$x_t = c + A_1 x_{t-1} + \cdots + A_p x_{t-p} + v_t, \quad (7)$$

where c is a vector of constants, $A_1, \dots, A_p$ are coefficient matrices, and v_t is the vector of reduced-form innovations with covariance matrix Σ_v . The reduced-form innovations are

related to the structural shocks by

$$v_t = H\eta_t, \quad (8)$$

where η_t contains mutually uncorrelated structural shocks and H is the matrix of contemporaneous effects. We are interested in a structural OPEC supply-news shock, denoted by η_t^s , and the corresponding column of H , denoted by h_s .

Let z_t^s be the monthly text-based supply-news instrument constructed in Section 2.6. The identification assumes that the instrument is relevant to the supply news shock and orthogonal to the rest of structural shocks:

$$E[z_t^s \eta_t^s] \neq 0, \quad (9)$$

$$E[z_t^s \eta_t^j] = 0, \quad j \neq s. \quad (10)$$

The first condition means that the text measure should contain information about the structural supply news shock. The second one that it is not systematically related to other structural shocks that affect the oil market and the macroeconomy. Under these conditions,

$$E[z_t^s v_t] = h_s E[z_t^s \eta_t^s], \quad (11)$$

and the covariance of the instrument and the reduced-form innovations identifies the contemporaneous effect of the supply news shock up to sign and scale. We chose the sign and scale in such a way that a positive supply news shock increases the real price of oil by 10 percent on impact. All the impulse response functions considered below are normalized in such a way.

4.2 Macroeconomic data and baseline specification

To make the results comparable with those obtained via the futures-based identification, we follow [Känzig \(2021\)](#) and construct the baseline VAR with the same system of variables. There are six monthly macro variables: real price of oil, world oil production, world oil inventories, world industrial production, U.S. industrial production, and U.S. consumer price index.

The real price of oil is defined by the U.S. refiners’ acquisition cost of imported crude oil adjusted by the U.S. CPI. World oil inventories are based on OECD petroleum stocks and converted to crude-oil equivalents ([Kilian and Murphy, 2014](#)). All six variables are included in the VAR in log levels.

The baseline VAR is estimated with twelve lags and a constant over 1974:1-2017:12. The external instruments are included in the VAR over 1983:7-2017:12, the period of common overlap with futures-based measures. Section 6 exploits the longer text-based series for comparing the pre-futures and post-2017 environment of announcements separately.

4.3 Estimation and inference

We apply the external-instrument framework separately to supply and demand news in the baseline analysis. The supply specification uses z_t^{sup} to identify an OPEC supply-news shock, while the demand specification uses the residualized measure z_t^{dem} to identify a demand-news shock. Both specifications use the same baseline VAR and the common 1983:7–2017:12 identification window. Responses are normalized to a 10 percent increase in the real oil price on impact. The underlying text measures are constructed separately, and the demand predictability adjustment does not use the supply measure. No orthogonalization between the two proxies is imposed in their baseline construction.⁶

⁶The separate baseline specifications do not provide a jointly identified orthogonal decomposition of announcement news. Separate construction and low proxy correlation do not establish the exclusion restrictions: each proxy is assumed to be orthogonal to structural shocks other than the shock it identifies. As a supplementary exercise, Appendix C.6 reports exploratory joint supply–demand responses obtained by symmetrically orthogonalizing the two proxy covariance directions. This additional orthogonalization is not

We assess instrument relevance by regressing the reduced-form innovation in the real price of oil on a constant and the relevant proxy. We report conventional and heteroskedasticity-robust first-stage F -statistics, together with R^2 and the number of observations. Section 5 presents the baseline first stages before the impulse responses, and Appendix C.1 provides the expanded diagnostics.

For the text-based supply specification and the futures-based supply benchmark, we report 68 and 90 percent confidence bands from a moving-block bootstrap with 10,000 replications. Baseline demand inference instead uses the Montiel Olea–Stock–Watson (MSW) weak-identification-robust confidence sets of Montiel Olea et al. (2021), reflecting the more limited relevance of the demand instrument, whose robust first-stage F -statistic is 9.06.⁷ Appendix C.3 documents the construction and implementation checks.

The baseline proxy-SVAR also relies on the shock of interest being invertible with respect to the current and past information in the VAR. Appendix C.4 reports degree-of-invertibility diagnostics following Plagborg-Møller and Wolf (2022); the associated Granger-causality pre-test rejects invertibility for demand news. We therefore also estimate an internal-instrument VAR following Plagborg-Møller and Wolf (2021), augmenting the system with the external instrument and ordering it first. Under the maintained instrument-validity conditions, this approach identifies relative impulse responses without requiring invertibility with respect to the original six-variable information set (Appendix D.2).

Separately, we assess sensitivity to the VAR’s dynamic propagation using local projections on the recovered structural shocks, with a constant and one lag of the dependent variable (Appendix D.3). These horizon-by-horizon point-estimate comparisons retain the baseline shock identification and therefore do not address invertibility or provide an independent test of the exclusion restrictions.

imposed in the baseline and does not constitute a test of the exclusion restrictions.

⁷We do not treat a particular value of the robust F -statistic as a universal validity threshold. The MSW procedure changes the confidence sets, not the underlying proxy-SVAR point estimates.

5 Macroeconomic Consequences of OPEC News

In this section, we estimate the macroeconomic consequences of OPEC news using the text-based measures. We use the common sample period and the six-variable VAR model for all our baseline specifications, which were described in Section 4. First, we report the instrument relevance and the diagnostics that justify the demand-news adjustment. Second, we estimate the effects of supply news and compare the text-based identification with the futures benchmark. Third, we consider the effects of demand news. Finally, the last subsection compares the effects of supply and demand news.

5.1 Instrument Relevance and Diagnostics

Table 4 reports baseline first-stage diagnostics. The text supply instrument has relevance comparable to the futures benchmark. The raw demand outlook is partly predictable from pre-announcement conditions; the adjustment in Section 2.6 reduces this predictability and modestly improves relevance (Appendix B.2). The residualized demand instrument nevertheless remains comparatively weak (robust $F = 9.06$), motivating the weak-identification-robust confidence sets described in Section 4.3 rather than a claim of strong identification. Additional residualization using twelve own lags yields similar response patterns (Appendices C.2 and D.1). We retain the event-level adjustment as the baseline because it targets information available before the announcement.

5.2 Supply news

Figure 3 reports the responses to an OPEC supply-news shock normalized to raise the real price of oil by 10 percent on impact. The solid line identifies the shock using the text-based supply measure; the dashed line reports the corresponding responses using the futures-based instrument of Känzig (2021). Both specifications use the same 1983:7-2017:12 instrument sample and the same reduced-form VAR.

Table 4: First-stage relevance of the OPEC news measures

	Futures supply	Text supply	Raw demand	Purged demand
Instrument coefficient	1.11	1.81	2.62	3.32
Robust t	3.44	3.83	2.55	3.01
Conventional F	20.14	16.96	7.19	8.80
Robust F	11.84	14.68	6.52	9.06
R^2 (%)	4.7	4.0	1.7	2.1
N	414	414	414	414

Notes: The dependent variable is the innovation in the real price of oil in the reduced form of the baseline VAR. All columns employ the standard 1983:7-2017:12 sample for the instruments. “Raw demand” is the demand outlook variable defined in equation (2). “Purged demand” purges the part that is forecastable based on predetermined conditions in the oil market and economy as shown in equation (3). The robust standard errors are heteroskedasticity-robust.

The text shock induces the standard contractionary response to a sudden decrease in oil supplies. The real price of oil jumps more than 10 percent initially and gradually falls. The world oil production decreases with a small lag and world oil stocks increase eventually. Economic activity declines in both world and U.S. economies. World industrial production declines gradually, whereas U.S. industrial production decreases more and more persistently. The consumer prices in the U.S. rise due to the demand news and stay high for several years afterwards.

The responses based on text and futures instruments are alike in terms of sign and dynamics. They are very similar for the real price of oil, world oil production, world industrial production and consumer prices. The most significant difference can be found in the case of U.S. industrial production where futures-based instrument indicates a larger contraction. It should be emphasized that the similarity is not incorporated into the text measure construction. Therefore, we regard the futures comparison as an external validation of the text-based supply shock rather than the primary result. The text measure identifies the OPEC supply shock, and its macroeconomic effects are quite similar to those identified using the futures instrument based on announcement day changes.

Text vs. futures identification, common era 1983--2017

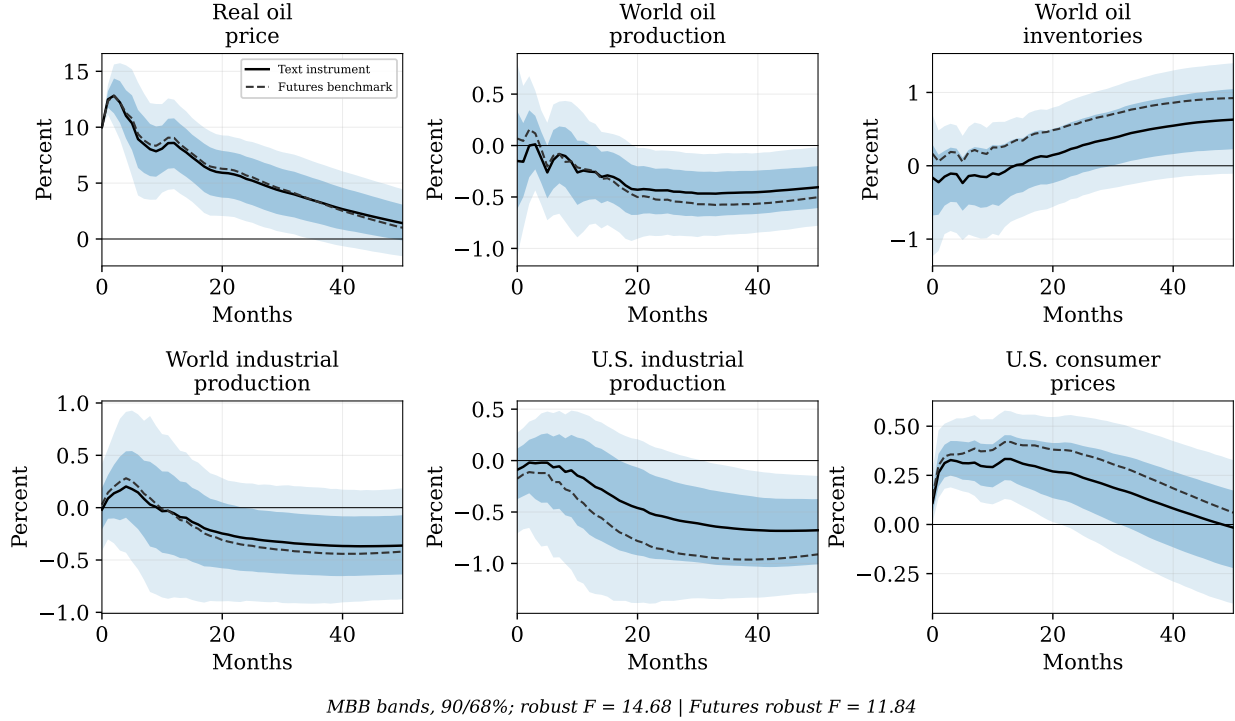


Figure 3: Macroeconomic effects of OPEC supply news. The solid line reports impulse responses identified with the text-based supply news instrument and the dashed line reports responses identified with the futures-based instrument of [Känzig \(2021\)](#). Both specifications use the common 1983:7-2017:12 instrument sample. The shock is normalized to raise the real price of oil by 10 percent on impact. Shaded areas are 68 and 90 percent moving-block-bootstrap confidence bands for the text-based specification. The heteroskedasticity-robust first-stage F -statistics are 14.68 for the text instrument and 11.84 for the futures instrument.

5.3 Demand news

Figure 4 displays the responses to an OPEC demand news shock with the same normalizing assumption that the real price of oil jumps by 10 percent on impact. The figure plots MSW confidence sets robust to weak identification rather than the standard confidence bands based on bootstrapping.

The responses are distinct from those arising from supply news. A positive shock to demand news leads to a similar real oil price movement as with the supply shock. There is a small increase in world oil production after several months, followed by a decline back

to zero. Inventories also decline after the shock, and there is an eventual gradual recovery. This response pattern is consistent with demand shocks putting pressure on the stock of oil and not a reduction in supply shocks.

The responses for real activity are different. World industrial production increases following the shock, peaks in the first year after the shock and declines thereafter. U.S. industrial production shows even stronger expansion after the shock, but declines back to zero over time. Consumer prices also rise. In other words, the increase in the oil price following good news about the economy takes place amid economic expansion and not economic contraction following a shock to supply news.

This is relevant in assessing OPEC announcement day oil price changes. A rise in oil prices following an OPEC meeting is not necessarily due to a bad supply shock. If contemporaneous newspaper coverage suggests good news about demand, the same rise in oil prices can reflect a positive shock to output and not economic contraction. Thus, classification of newspapers allows one to separately measure the effects of demand and supply shocks on oil prices.

Despite being economically distinct from the supply responses, inference based on demand news is much more uncertain due to the weakness of the first stage. Confidence sets in Figure 4 show the uncertainty arising from this. Appendix B.2 shows that additional own-lag residualization in the first stage strengthens the first stage and delivers similar qualitative responses.

5.4 Comparing supply and demand effects

Figure 5 summarizes the central economic distinction at the common 10-percent oil-price normalization. Supply news is followed by weaker industrial production at medium horizons; demand news produces an initial expansion. Oil production falls after supply news but initially rises after demand news, while inventories eventually accumulate after supply news and initially decline after demand news. Consumer prices rise following both shocks. Thus, similarly signed oil-price responses need not imply similar real effects.

Oil demand news shock: text identification, 1983--2017

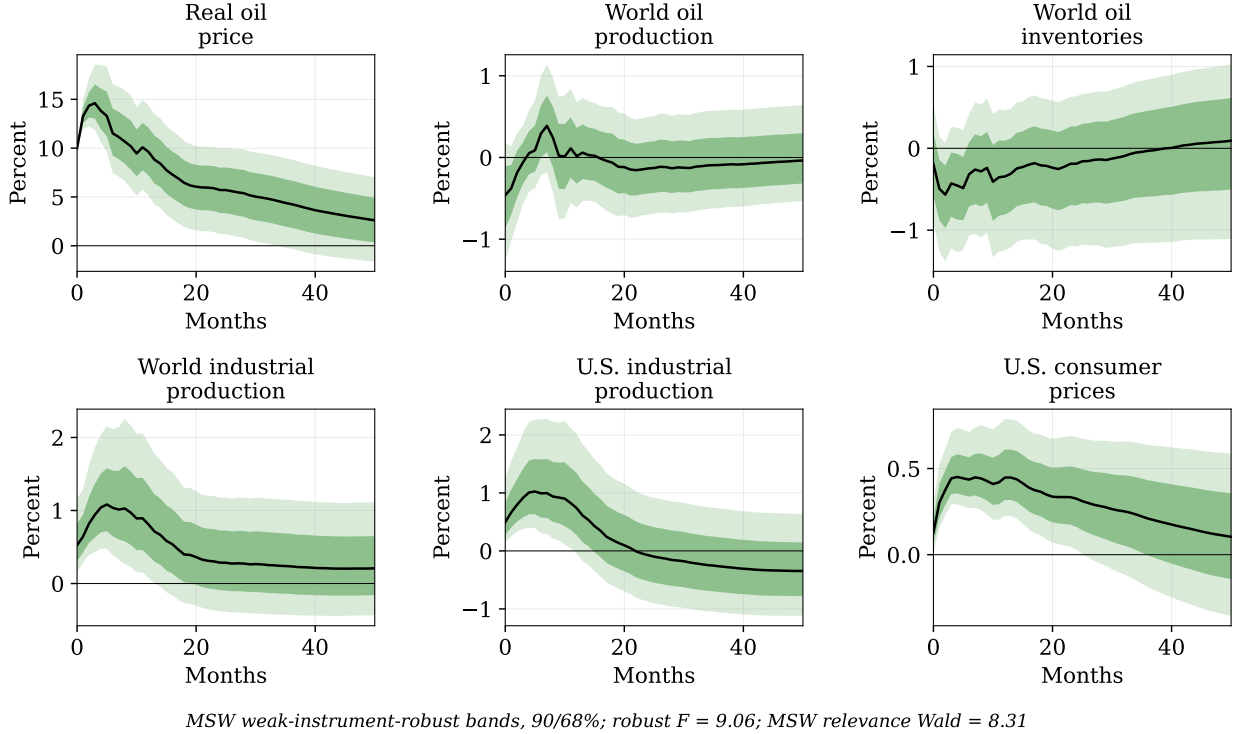
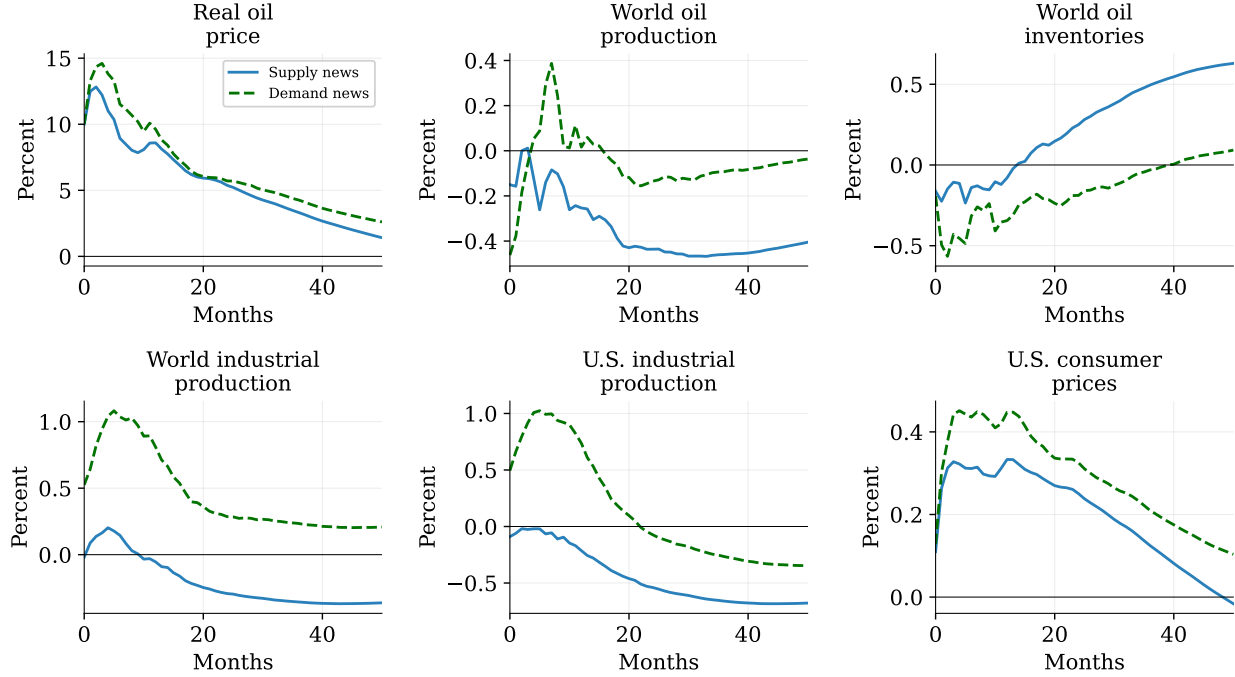


Figure 4: Macroeconomic effects of OPEC demand news. The figure reports impulse responses identified with the residualized text-based demand-news instrument over the 1983:7-2017:12 sample. The shock is normalized to raise the real price of oil by 10 percent on impact. Shaded areas are 68 and 90 percent MSW weak-identification-robust confidence sets. The heteroskedasticity-robust first-stage F -statistic is 9.06; the MSW relevance Wald statistic is 8.31.

We supplement this comparison with joint smooth local projections following [Barnichon and Brownlees \(2019\)](#). Each regression includes the recovered futures-based shock, the recovered text-supply shock residualized against the futures-based shock, and the recovered demand-news shock, together with a constant and one lag of the dependent variable. Figure 6 reports the conditional responses per one standard deviation of each regressor. At twelve months, world industrial production is approximately 0.35 percent lower for the residual text-supply component and 0.63 percent higher for the demand component; the corresponding pointwise 90 percent analytical intervals exclude zero. Demand is also associated with an expansion in U.S. industrial production, whereas the U.S. response to residual text supply

Supply and demand news compared, 1983--2017



Both shocks normalized to raise the real oil price by 10%; separate proxy-SVAR specifications

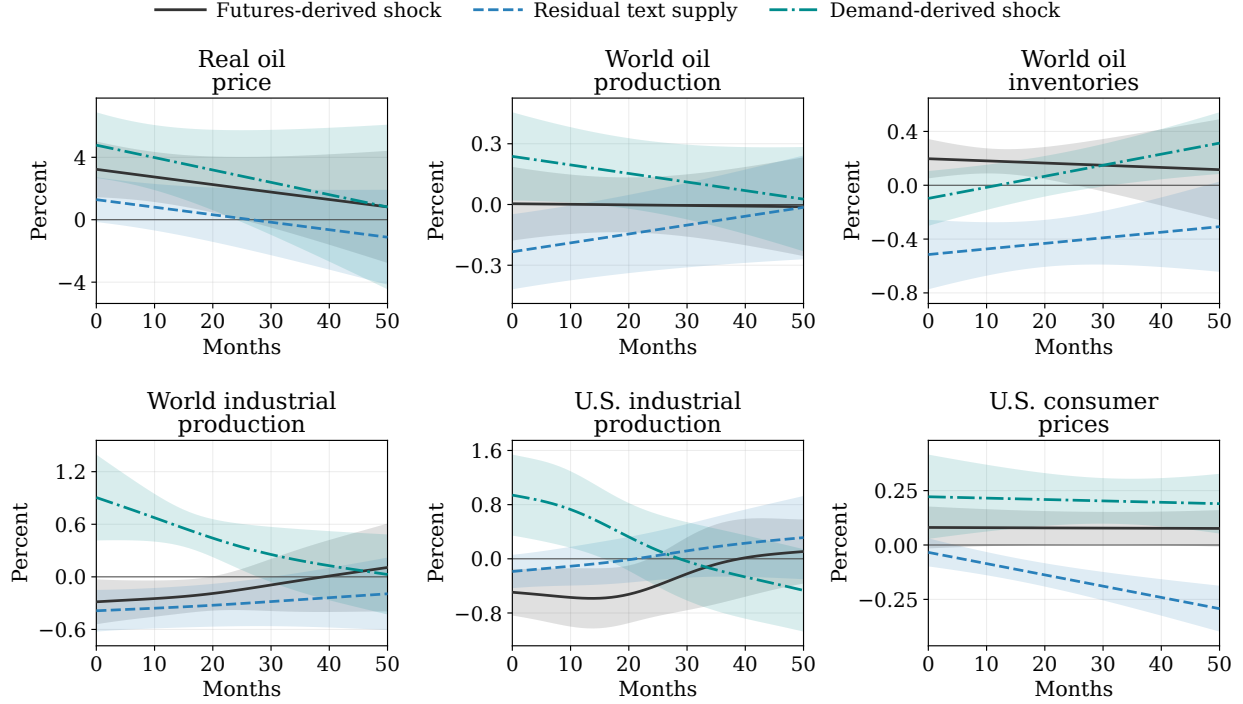
Figure 5: Macroeconomic effects of supply and demand OPEC news. The solid line reports responses identified with the text-based supply-news instrument and the dashed line reports responses identified with the residualized demand-news instrument. Both specifications use the common 1983:7-2017:12 sample, and both shocks are normalized to raise the real price of oil by 10 percent on impact. The shocks are estimated in separate proxy-SVAR specifications. Confidence sets are reported in Figures 3 and 4.

is imprecisely estimated.

6 Extending the Sample

Given the long history of the newspaper announcements, we can explore how the announcement-based identification behaves when we expand our sample beyond the benchmark 1983:7-2017:12 sample. In particular, the 1975:1-2017:12 sample covers the pre-futures period but excludes the OPEC+ regime; 1983:7-2025:12 sample keeps the benchmark starting point but includes the post-2017 OPEC+ period; finally, 1975:1-2025:12 sample uses all text data

Joint smooth LP: conditional 90% analytical bands



*One-standard-deviation responses; color-matched pointwise 90% HAC intervals.
Conditional on recovered shocks and tuning; not weak-identification robust.*

Figure 6: Joint comparison using smooth local projections. The curves report conditional responses associated with the recovered futures-based shock, the recovered text-supply shock residualized against the futures-based shock, and the recovered demand-news shock. All three series enter simultaneously, together with a constant and one lag of the dependent variable. Responses are expressed per one standard deviation of each regressor, not per identical oil-price impact. Color-matched shading reports pointwise 90 percent analytical HAC confidence intervals using a 50-month Bartlett bandwidth, conditional on the recovered shocks and smoothing choices. The regressions use 364 common shock dates, 1983:7–2013:10, across horizons zero through 50 months, with outcomes observed through December 2017.

available.

Table 5 extends the benchmark identification window backward to 1975, forward through 2025, and in both directions. The pre-futures extension exploits announcements unavailable to futures-based measurement; the later extension includes the more frequent meetings and scheduled rollovers of OPEC+. For specifications ending in 2025, we include separate March–June 2020 pulse dummies in every VAR equation and re-estimate the demand-news predictability adjustment over the extended event sample using the same predictors. Adding

Table 5: Instrument relevance across alternative sample windows

	Benchmark 1983:7-2017:12	+ Pre-futures 1975:1-2017:12	+ OPEC+ 1983:7-2025:12	Full sample 1975:1-2025:12
Text supply news	14.68	15.23	11.95	12.88
Residualized demand news	9.06	10.59	13.46	13.43
N	414	516	510	612

Notes: First two rows entries are heteroskedasticity-robust first-stage F -statistics from regressions of the reduced-form real-oil-price innovation on the designated instrument. First column entries are the identification sample used as benchmark in Section 5. Second column entries extend the sample backwardly to cover the pre-futures era but end in 2017. Third column entries extend the benchmark sample forwardly to the OPEC+ era but start in 1983. Fourth column entries consider the full 1975-2025 text record sample. For samples that end in 2017, residualized demand uses event-level adjustment based on the 1975-2017 event sample. For samples that end in 2025, the same event-level adjustment is estimated on the 1975-2025 event sample. Specifications ending in 2025 include separate dummies for March, April, May, and June 2020 events in every reduced-form VAR equation. N represents the number of monthly observations in each first-stage regression.

the pre-futures observations modestly increases both first-stage statistics. In the COVID-adjusted specifications extending through 2025, supply relevance is somewhat lower and demand relevance higher than in the benchmark. Appendix H.3 separately reports weak first stages for the short tail windows under its specification without COVID pulse controls. Figure 7 shows that the full-sample point estimates retain the central real-activity distinction. Supply news is followed by lower world and U.S. industrial production at medium horizons, whereas demand news produces an initial expansion. Consumer prices rise after both shocks. Appendix F reports confidence bands, and Appendix G examines sensitivity to acute-pandemic months and individual post-2017 observations. The longer text record thus extends the period over which the benchmark response patterns can be studied.

7 Robustness and Additional Results

The robustness exercises address distinct threats to interpretation. Removing twelve-own-lag predictability leaves the main response patterns largely unchanged (Appendices C.2 and D.1). The internal-instrument VAR preserves the central supply–demand contrast without requiring invertibility with respect to the original six-variable information set, under

Supply and demand news: full text sample, 1975--2025

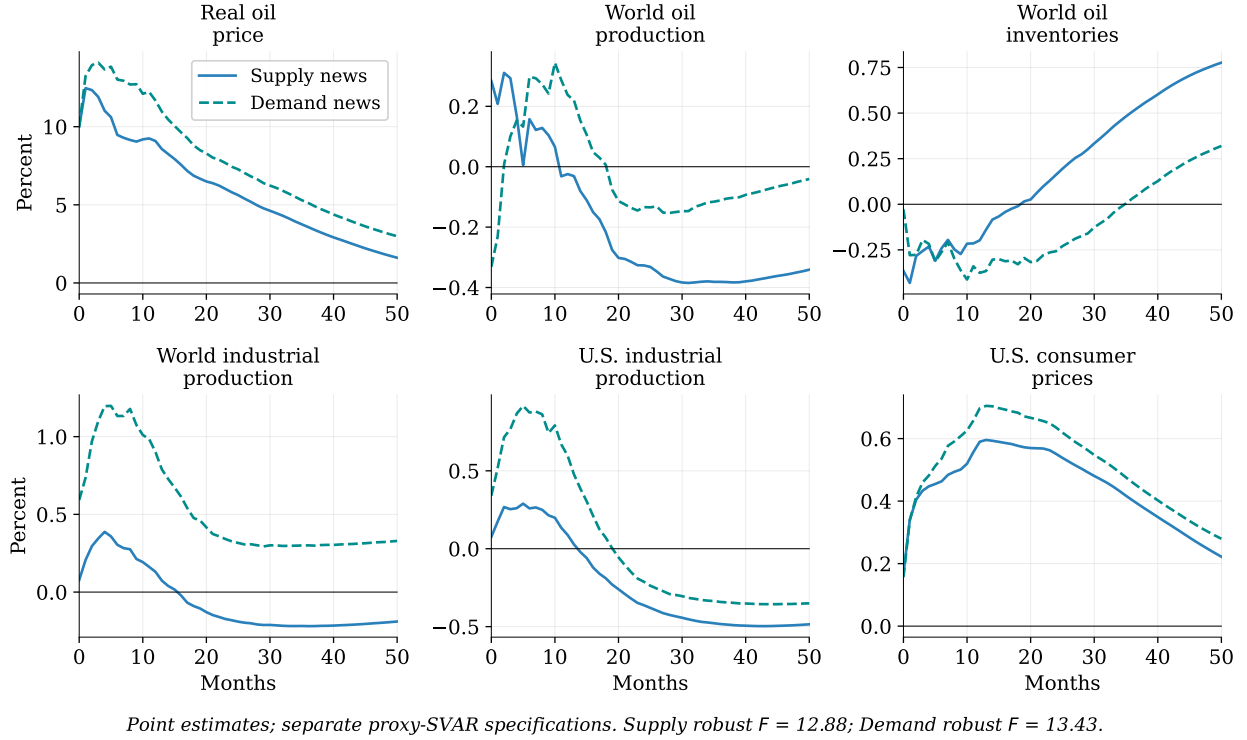


Figure 7: Macroeconomic consequences of supply and demand OPEC news, 1975-2025. The solid line represents the response to supply news identified in the text and the dashed line represents the response to residual demand news. Both shocks have been scaled to increase the real oil price by 10 percent and are estimated using different proxy-SVAR models. The VAR reduces form equation uses separate dummy variables for March, April, May, and June 2020. Confidence intervals of 68 and 90 percent are provided in the appendix.

the maintained instrument-validity conditions (Appendix D.2). Recovered-shock local projections also preserve the main real-activity contrast, although some oil-market responses and longer-horizon persistence differ. Because these projections use the baseline recovered shocks, they assess dynamic propagation rather than independently validating identification (Appendix D.3). Using the futures and text measures jointly yields supply responses similar to the text-only specification (Appendix E.1). The text supply measure remains significantly associated with the [Mori and Peersman \(2024\)](#) supply-news shock after controlling for the futures proxy, whereas no significant association is detected for the demand measure (Appendix C.5). Alternative scaling and monthly allocation preserve first-stage relevance

(Appendix E.2). Pandemic and recent-event sensitivity for the extended sample are reported in Appendix G.

8 Conclusion

In this paper, we measure information embedded in OPEC announcements directly using text of newspaper articles. In particular, we classify 7,941 articles around 206 OPEC and OPEC+ announcements from 1971 to 2025 into two different groups according to the content and form the corresponding supply and demand news measures.

The text-based supply measure yields the same macroeconomic responses as obtained by futures-based identification. The supply news increasing the real price of oil reduces industrial production and increases consumer price level. The demand news increases the price of oil, but it is followed by increased industrial production and reduced inventories. These two distinct responses make clear why the price reaction around an OPEC announcement does not contain the information about the type of news causing it.

The extended text data also allows using OPEC announcement news for structural analysis for a longer period. The addition of the pre-futures sample to the benchmark one does not reduce the relevance of the instruments, and an extension of the sample to the OPEC+ period leads to somewhat lower supply identification and higher demand identification. In the full 1975-2025 sample, the instruments stay relevant, and the difference between the macroeconomic impact of the news on supply and demand is also maintained. Shorter pre-futures and post-2017 samples have lower identification power independently, but their presence in the longer sample does not reduce identification power. Thus, text not only makes it possible to measure the content of OPEC announcements but also extends their usage as an instrument of identification.

References

- Abeyie, G. A. (2026). Beyond news headlines and tf-idf: Enhancing text-based forecasting models with validated collocations and improved attention. *International Journal of Forecasting*.
- Alquist, R. and Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, 25(4):539–573.
- Aruoba, S. B. and Drechsel, T. (Forthcoming). Identifying monetary policy shocks: A natural language approach. *American Economic Journal: Macroeconomics*.
- Bai, Y., Li, X., Yu, H., and Jia, S. (2022). Crude oil price forecasting incorporating news text. *International Journal of Forecasting*, 38(1):367–383.
- Barnichon, R. and Brownlees, C. (2019). Impulse response estimation by smooth local projections. *Review of Economics and Statistics*, 101(3):522–530.
- Caldara, D., Cavallo, M., and Iacoviello, M. (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*, 103:1–20.
- Datta, D. D. and Dias, D. A. (2020). Oil shocks: A textual analysis approach. *Working Paper*.
- Degasperi, R. (2023). Identification of expectational shocks in the oil market using opec announcements. *The Warwick Economics Research Series*.
- Demirer, R. and Kutan, A. M. (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.
- Handlan, A. (2020). Text shocks and monetary surprises: Text analysis of fomc statements with machine learning. *Working Paper*.

- Inoue, A. and Kilian, L. (2013). Inference on impulse response functions in structural var models. *Journal of Econometrics*, 177(1):1–13.
- Jiao, X., Song, Y., Kong, Y., and Tang, X. (2022). Volatility forecasting for crude oil based on text information and deep learning pso-lstm model. *Journal of Forecasting*, 41(5):933–944.
- Känzig, D. R. (2021). The macroeconomic effects of oil supply news: Evidence from opec announcements. *American Economic Review*, 111(4):1092–1125.
- Kilian, L. (2008). The economic effects of energy price shocks. *Journal of Economic Literature*, 46(4):871–909.
- Kilian, L. (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.
- Kilian, L. (2024). How to construct monthly var proxies based on daily surprises in futures markets. *Journal of Economic Dynamics and Control*, 168:104966.
- Kilian, L. and Murphy, D. P. (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.
- Kilian, L. and Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29(3):454–478.
- Kilian, L. and Zhou, X. (2020). Does drawing down the us strategic petroleum reserve help stabilize oil prices? *Journal of Applied Econometrics*, 35(6):673–691.
- Kilian, L. and Zhou, X. (2023). The econometrics of oil market var models. In *Essays in honor of joon y. park: Econometric methodology in empirical applications*, pages 65–95. Emerald Publishing Limited.
- Li, X., Shang, W., and Wang, S. (2019). Text-based crude oil price forecasting: A deep learning approach. *International Journal of Forecasting*, 35(4):1548–1560.

- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the united states. *American Economic Review*, 103(4):1212–1247.
- Montiel Olea, J. L., Stock, J., and Watson, M. (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, 225:74–87.
- Mori, L. and Peersman, G. (2024). Estimating the macroeconomic effects of oil supply news. Technical report, CESifo Working Paper.
- Ochs, A. C. (2021). A new monetary policy shock with text analysis. *Cambridge Working Papers in Economics*.
- Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2):955–980.
- Plagborg-Møller, M. and Wolf, C. K. (2022). Instrumental variable identification of dynamic variance decompositions. *Journal of Political Economy*, 130(8):2164–2202.
- Plante, M. (2019). Opec in the news. *Energy Economics*, 80:163–172.
- Stock, J. H. and Watson, M. W. (2012). Disentangling the channels of the 2007-2009 recession. Technical report, National Bureau of Economic Research.

Online Appendix

Reading OPEC: Text-Based Identification of Oil Supply and Demand News

Gabriel Abeyie

Andrew Hanson

Contents

A	Text Classification and Human Validation	2
A.1	Article selection and classification procedure	3
A.2	Human and cross-model validation	7
A.3	Illustrative classified OPEC events	9
B	Construction and Validation of the Instruments	11
B.1	Supply-news construction	11
B.2	Demand-news construction and purification	13
B.3	Text supply news and the futures-based surprise	16
C	Instrument Relevance and Identification Diagnostics	17
C.1	Baseline first stages	18
C.2	Predictability and own-lag diagnostics	19
C.3	Weak-identification-robust inference	20
C.4	Invertibility diagnostics	21
C.5	External validation with an alternative oil-supply-news shock	22
C.6	Exploratory joint supply–demand diagnostic	23

D	Alternative Dynamic Specifications	24
D.1	Own-lag residualized instruments	25
D.2	Invertibility-robust internal-instrument VAR	26
D.3	Local projections using recovered structural shocks	28
E	Alternative Supply Identification	30
E.1	Futures and text as two proxies for the same shock	31
E.2	Alternative construction of the text supply measure	34
F	Extended-Sample Results, 1975-2025	35
F.1	Full-sample supply responses	36
F.2	Full-sample demand responses	36
G	Pandemic and Recent-Period Sensitivity	38
G.1	The acute COVID period	38
G.2	Influence of recent OPEC observations	39
H	Additional Historical and Regime Evidence	39
H.1	Pre-futures OPEC events	39
H.2	OPEC+ event calendar, 2018-2025	40
H.3	Full regime first-stage diagnostics	43
I	Data Sources and Replication Details	44

A Text Classification and Human Validation

This appendix contains further information about the creation of the text measures at the article level included in the paper. First, we describe the article selection and classification process, specifying the inputs for the classifier, the output format, and ambiguous cases.

Next, in Appendix [A.2](#), we present the human validation of the process.

A.1 Article selection and classification procedure

Newspapers and event windows. The article corpus is drawn from the *New York Times*, *Wall Street Journal*, and *Washington Post* using ProQuest TDM Studio, including both current and historical editions where available. As described in Section 2.2, the OPEC calendar contains 206 OPEC and OPEC+ announcements between 1971 and 2025. For each event, we gather articles published from two days before to three days after the announcement. Articles that appear within an overlapping period of events are mapped to the closest event so that there is no duplicate article in the corpus. Deduplication leads to a total of 7,941 articles in the event window corpus.

Classifier inputs. Each article is classified using `claude-sonnet-4-6` at temperature zero. The model receives only the article title, article text, publication date, and newspaper source. Article text is truncated at 9,000 characters; the same truncation rule is used in the human validation exercise. The classifier is not provided with the OPEC event identifier, the event calendar, futures prices, oil-price movements, or any other market data. These restrictions are intended to ensure that the article labels are obtained from the reported content of the news coverage rather than from the subsequent financial-market response.

The classifier returns a strict JSON object. Each output is checked field-by-field against a fixed schema, including the permitted labels and the conditional relationships among fields. Invalid outputs are retried up to three times and are otherwise retained in an audit log. Of the 7,941 articles, 130 failed the initial classification pass because of transient interface or formatting failures. Of these, 125 were recovered during the final run and the remaining five were processed in a separate backfill batch using the identical frozen prompt and validation rules. The final dataset therefore contains a validated classification for every article in the corpus.

Decision, expectation, and surprise fields. The classification schema records several distinct dimensions of OPEC coverage. First, `meeting_relevance` indicates whether the OPEC event is central, peripheral, or unrelated to the substantive content of the article. Second, `coverage_role` classifies the article as a `preview`, `decision_report`, `post_analysis`, or `background` article. This field is assigned from the content of the article rather than mechanically from its publication date.

For articles discussing an OPEC decision, the classifier records the policy domain, the type of decision, and the announced magnitude when it is explicitly stated. Preview articles additionally record the outcome expected before the meeting. Decision reports record the direction of the announcement surprise relative to prior expectations. The surprise field is not used for preview, retrospective, or background coverage. Surprise direction is coded only when the article contains sufficient language comparing the announced outcome with what had been anticipated; it is never inferred from the direction of the oil price or from information that became available after the announcement. Conditional fields receive their designated neutral value when the relevant condition does not apply.

Table 6 summarizes the principal fields and their distribution in the 7,941-article corpus.

Supply and demand content. Supply and demand information are recorded separately rather than treated as mutually exclusive article types. The classifier assigns an independent supply-content score and demand-content score from zero to three, so that an article may contain substantial information about both sides of the oil market. A separate `demand_direction` field records whether the demand outlook described in the article is stronger, weaker, mixed, or absent.

The two external instruments use different classifier fields. The supply-news measure uses the surprise-direction field from coverage of the OPEC decision. The demand-news measure instead uses the demand-content and demand-direction fields and is restricted to articles with a demand-content score of at least two. Thus, the two measures do not share

Table 6: Article classification schema and label distributions

Field	Label	<i>N</i>
Meeting relevance	central	2,517
	peripheral	4,728
	none	696
Coverage role	preview	1,765
	decision_report	1,831
	post_analysis	638
	background	3,707
Decision domain	price	1,046
	quantity	4,204
	both	475
	none	2,216
Decision type	cut	1,872
	hold	1,580
	raise	1,522
	no_agreement	480
	framework_change	95
	none	2,392
Surprise direction	more_hawkish	455
	more_dovish	673
	as_expected	320
	unclear	6,493
Expected outcome	cut	413
	hold	458
	raise	341
	uncertain	512
	none	6,217
Supply content (0-3)	0 / 1 / 2 / 3	931 / 3,597 / 1,214 / 2,199
Demand content (0-3)	0 / 1 / 2 / 3	3,002 / 2,571 / 2,113 / 255
Demand direction	stronger	1,137
	weaker	3,103
	mixed	575
	none	3,126
Geopolitical flag	0 / 1	5,355 / 2,586

Notes: Counts are over the 7,941 articles in the event-window corpus. The field **decision_magnitude** additionally records the announced magnitude as stated in the article (e.g., “−2000 kbd”) and is used only descriptively. The machine label **post_analysis** corresponds to retrospective analysis in the text. Surprise direction refers to the supply-price implication of the announcement relative to expectations under the corrected classification rule. **more_hawkish** denotes less supply, or an oil-price implication that is higher than expected.

an underlying classifier field, although a single newspaper article can contain both supply-related and demand-related information.

Weighting and aggregation. Article-level classifications are aggregated to the event level using the procedures described in Section 2. The supply measure in equation (1) sums the directional surprises contained in surprise-bearing decision coverage and scales the sum by the square root of the number of contributing articles. This scaling allows broader independent coverage of an unexpected decision to increase the event-level signal without allowing the measure to grow proportionally with the number of articles.

The raw demand measure in equation (2) is a content-weighted average of the demand outlook reported across qualifying articles. Unlike the supply measure, demand news is not defined as the residual component of supply news and the two measures are not orthogonalized against one another.

Monthly aggregation follows equation (5). For the supply instrument, an event occurring on day d of a month with D days is allocated between the announcement month and the following month using the timing weights described in Section 2. This adjustment reflects the fact that the real oil price in the VAR is measured as a monthly average. The demand measure is assigned to the announcement month without this timing allocation. Months containing no qualifying OPEC event receive a value of zero.

Treatment of ambiguous classifications. The classification procedure records ambiguity rather than forcing a directional label. In particular, decision coverage without sufficient language about the relationship between the realized decision and prior expectations receives `unclear` in the surprise-direction field and contributes no directional information to the supply instrument. Articles stating that the outcome was broadly anticipated receive `as_expected`; these observations contribute a zero directional surprise while remaining part of the underlying decision coverage.

The same principle applies to demand coverage. A demand-direction label is used only when the article provides sufficient information about the direction of the demand outlook. Articles without such content receive the neutral category rather than having a direction

inferred from oil prices, production decisions, or other classifier fields.

Correction to the hawkish/dovish definition. The definition of the surprise-direction field was amended once during the validation stage. An audit of the initial classifier output revealed that the first version of the prompt sometimes imported the conventional monetary-policy meaning of the term “hawkish.” In an OPEC setting, the economically relevant definition is instead based on the implication for oil supply and oil prices: a more hawkish-than-expected decision implies less supply, or a higher oil price, than had been anticipated.

The wording of the field definition was therefore corrected, and all articles potentially affected by the revised definition were reclassified under the corrected rule. Among the 374 articles initially classified as `more_hawkish`, 255 retained that classification, 101 were reclassified as `more_dovish`, and 18 became `unclear`. Previously dovish classifications were 97 percent stable under the revised definition.

This amendment was made once, after the initial classification output had been inspected as part of the validation and diagnostic process and before the final instruments and structural estimates reported in the paper were produced. The revised definition was subsequently held fixed. The amendment, the affected field definition, and the reclassification step are recorded in the dated classification changelog. This documentation is retained so that the final text measures can be reconstructed from the frozen classification rules rather than inferred from the structural results.

A.2 Human and cross-model validation

We evaluate the article classifications using an independent human audit conducted by the first author. The completed validation sample contains 95 articles spanning 1971-2025, including 35 articles from the *Wall Street Journal*, 32 from the *Washington Post*, and 28 from the *New York Times*. Sixty articles are drawn from historical editions and 35 from modern print or online editions.

The human audit focuses on the fields that determine the external instruments, together with article relevance and coverage role. The human coder classifies OPEC meeting relevance, coverage role, surprise direction for decision reports, demand-content intensity, and demand direction. The coder observes the same article title and truncated text provided to the production classifier but is not shown the machine labels, OPEC event identities, futures prices, or subsequent oil-price movements.

Table 7 reports agreement between the human coding and the production classifier. Agreement is high for the classification fields used in the empirical analysis. OPEC meeting relevance agrees in 88.4 percent of articles, with Cohen’s $\kappa = 0.794$, while coverage role agrees in 89.5 percent, with $\kappa = 0.736$. Demand-content intensity agrees exactly in 89.5 percent of articles. Because this field is ordinal, we report weighted Cohen’s κ , which equals 0.931.

The economically relevant demand threshold is even more stable. The demand-news instrument uses articles with a demand-content score of at least two. Human and machine classifications agree on this inclusion indicator for 94 of the 95 articles, an agreement rate of 98.9 percent and a Cohen’s κ of 0.974. Thus, most disagreements over the 0-3 intensity scale concern adjacent categories that do not alter whether an article enters the demand measure.

For surprise direction, the comparison is necessarily based on a much smaller set because the field is defined only for first-instance decision reports containing information about prior expectations. The human coder identifies five decision reports in the validation sample, and the human and machine surprise classifications agree in all five cases. We therefore report the underlying count rather than placing substantial weight on a kappa statistic based on so few informative observations. Conditional on articles containing demand information, human and machine classifications of demand direction also agree in all 56 cases.

A particular concern for the historical extension is whether OCR noise reduces classification reliability. We find little evidence of such a deterioration. Among the 60 human-coded articles from historical editions, coverage-role agreement is 91.7 percent ($\kappa = 0.772$),

Table 7: Human validation of article classifications

Field	N	Agreement (%)	Cohen’s κ
OPEC meeting relevance	95	88.4	0.794
Coverage role	95	89.5	0.736
Demand content	95	89.5	0.931
Demand content ≥ 2	95	98.9	0.974
Demand direction	56	100.0	1.000
Surprise direction	5	100.0	-

Notes: The table compares independent human coding with the production language-model classifications. Cohen’s κ for demand content uses ordinal weights. Demand direction is evaluated only for articles with positive human-coded demand content. Surprise direction is evaluated only for human-coded decision reports; because there are only five such observations in this validation subset, we report agreement but do not emphasize the corresponding kappa statistic. The indicator “Demand content ≥ 2 ” corresponds to the inclusion threshold used to construct the demand-news measure.

compared with 85.7 percent ($\kappa = 0.674$) among the 35 modern-edition articles. Weighted agreement for demand-content intensity is similarly high in both groups: $\kappa = 0.930$ for historical articles and 0.921 for modern articles. Most importantly, human and machine classifications agree on the demand-content inclusion threshold in all 60 historical articles. The corresponding agreement rate among modern articles is 97.1 percent. The historical results therefore provide no indication that OCR-era articles are disproportionately responsible for the classifications that enter the demand instrument.

As an additional robustness exercise, we independently reclassify the validation articles using alternative language models under the same schema and information restrictions. These comparisons are used only to assess whether the production labels are model-specific; they do not replace or modify the classifications used to construct the instruments.

A.3 Illustrative classified OPEC events

As a complement to the statistical validation, Table 8 reports the article-level aggregation for six economically salient OPEC episodes. The examples span both positive and negative supply-news realizations and illustrate why the announced policy action need not have the

Table 8: Illustrative OPEC events and text-based news measures

Episode	Decision	Dominant surprise	z^{sup}	n	$z^{\text{dem,raw}}$
August 1986	Quota system restored; price war ends	Tighter than expected	+2.83	18	-0.40
November 2014	Production maintained into oil glut	Looser than expected	-5.58	35	-0.52
November 2016	Algiers/Vienna cuts implemented	Tighter than expected	+4.00	16	-0.17
March 2020	OPEC+ negotiations collapse	Looser than expected	-4.38	23	-1.00
April 2020	Record 9.7 mb/d production cut	Looser than expected	-4.01	18	-1.00
October 2022	Two-million-barrel production cut	Tighter than expected	+4.67	31	-0.96

Notes: z^{sup} is the event-level supply-news measure used to construct the monthly instrument. Positive values indicate that contemporaneous decision coverage characterized the announcement as tighter than expected, while negative values indicate a looser-than-expected outcome. The column n is the number of surprise-bearing decision-report articles contributing to the event-level supply measure, not the total number of articles in the event window. $z^{\text{dem,raw}}$ is the content-weighted mean direction of the demand outlook among articles with substantial demand content and is bounded between -1 and 1 . Negative values indicate weaker demand news. The April 2020 row refers to the April 12 agreement.

same sign as the surprise contained in the announcement. They also show that supply and demand information can move independently within the same event window.

The November 2014 meeting illustrates the distinction between a policy decision and a policy surprise particularly clearly. OPEC maintained production, but contemporaneous coverage overwhelmingly characterized the outcome as looser than expected. Thirty-five surprise-bearing articles produce an event-level supply score of -5.58 , the largest dovish realization among the episodes shown in the table. The measure therefore does not mechanically classify a production hold as neutral; it records how the announced outcome compared with what had been anticipated.

The April 2020 agreement provides an even sharper example. OPEC+ agreed to a historically large production cut, yet the text measure assigns the announcement a strongly negative supply-news value of -4.01 . Coverage judged the cut insufficient relative to the scale of the pandemic-driven collapse in oil demand. A measure based only on the announced production change would classify this event as an unusually large tightening, whereas the text-based surprise measure captures the substantially looser outcome relative to the information environment surrounding the announcement.

The table also illustrates why supply and demand news are measured separately. The October 2022 two-million-barrel production cut produces a strong positive supply-news re-

alization of +4.67, while the raw demand score is -0.96 . Thus, the same announcement window contains news of a tighter-than-expected supply decision and an unusually weak demand outlook. Similarly, both March and April 2020 combine strongly dovish supply-news readings with the minimum raw demand score of -1 . These episodes provide direct examples of the information bundling that the separate supply and demand measures are designed to distinguish.

B Construction and Validation of the Instruments

This appendix provides additional details on the construction of the text-based supply- and demand-news instruments. We first describe the mapping from article-level classifications to event- and monthly-level supply news. We then document the event-level adjustment applied to the demand measure to remove variation predictable from lagged oil-market and macroeconomic conditions. Finally, we compare the text-based supply measure with the futures-based OPEC surprise over their common sample.

B.1 Supply-news construction

The supply-news measure is based exclusively on explicit descriptions of the announced OPEC decision relative to prior expectations. For event e , let $\mathcal{A}_e$ denote the set of decision-report articles with a usable surprise classification. Define

$$s_a = \begin{cases} +1, & \text{if article } a \text{ describes the outcome as more hawkish than expected,} \\ 0, & \text{if the outcome is described as expected,} \\ -1, & \text{if the outcome is described as more dovish than expected.} \end{cases}$$

The event-level supply-news measure is

$$z_e^{\text{sup}} = \frac{1}{\sqrt{n_e}} \sum_{a \in \mathcal{A}_e} s_a, \quad n_e = |\mathcal{A}_e|. \quad (12)$$

The normalization allows broader agreement across independent newspaper coverage to increase the strength of the event-level signal without making the measure proportional to article volume. For example, ten articles that unanimously characterize a decision as tighter than expected generate a larger signal than two articles with the same direction, whereas additional articles coded as expected do not create directional news. Announced production or price changes themselves do not enter equation (12). The measure therefore captures the reported surprise relative to expectations rather than the level of the OPEC policy action.

Monthly timing adjustment. The structural model uses the monthly average real price of oil. Treating an announcement on the final day of a month as if it affected the entire month’s average price would therefore misalign the instrument with the reduced-form innovation. Following the temporal-aggregation logic of [Kilian \(2024\)](#), an event occurring on day d_e of a month containing D_e days is divided between the announcement month and the following month:

$$z_m^{\text{sup}} = \sum_{e:m(e)=m} \frac{D_e - d_e + 1}{D_e} z_e^{\text{sup}} + \sum_{e:m(e)=m-1} \frac{d_e - 1}{D_e} z_e^{\text{sup}}. \quad (13)$$

Months receiving no event contribution are assigned zero. The two weights sum to one, so equation (13) changes the timing of the event-level signal without changing its total magnitude.

The November 30, 2016 OPEC announcement provides a simple illustration. Its event-level supply-news value is +4.00. Because November has 30 days, only 1/30 of this event enters the November instrument, while 29/30 enters December. Thus, the November 30 event contributes approximately 0.13 to November and 3.87 to December. This adjustment is quantitatively relevant because OPEC announcements frequently occur near the end of

the month.

As robustness exercises, we also consider an unscaled article-sum measure, a simple-mean measure, and a specification that assigns the supply signal only to the announcement month. These alternatives are reported in Appendix E.

B.2 Demand-news construction and purification

Demand news is constructed independently of the supply measure. Let $\mathcal{D}_e$ denote articles around event e with a demand-content score of at least two. Let $q_a \in \{2, 3\}$ denote the article’s demand-content intensity and let $d_a \in \{-1, 0, +1\}$ indicate a weaker, mixed, or stronger demand outlook. The raw event-level demand measure is

$$z_e^{\text{dem,raw}} = \frac{\sum_{a \in \mathcal{D}_e} q_a d_a}{\sum_{a \in \mathcal{D}_e} q_a}, \quad (14)$$

and is therefore bounded between -1 and 1 . Unlike the supply instrument, demand news is not prorated across adjacent months. The event-level value enters the announcement month, and values from multiple events in the same month are summed. Months without qualifying OPEC events receive zero.

The raw demand measure may nevertheless reflect information already predictable from the state of the oil market or the macroeconomy. We therefore remove this predictable component at the OPEC-event level before using demand news as an external instrument. For each event with a defined demand measure, we estimate

$$z_e^{\text{dem,raw}} = \alpha + \Gamma' X_{e-1} + \nu_e, \quad (15)$$

where X_{e-1} contains one lag of the six transformed variables used in the Känzig oil-market system: the real price of oil (POIL), world oil production (OILPROD), world oil inventories (OILSTOCKS), world industrial production (WORLDIP), U.S. industrial production (IP), and the U.S. consumer price index (CPI). Only information dated prior to the announcement

month enters the regression. The residual $\hat{\nu}_e$ defines the purified event-level demand-news measure.

The baseline adjustment is estimated over 1975:1-2017:12 and contains 142 usable OPEC events. The six lagged state variables explain 20.45 percent of the variation in the raw event-level demand measure. The correlation between the raw measure and the resulting residual is 0.892. The purification therefore removes a nontrivial predictable component without fundamentally changing the ordering of demand-news events.

The event residuals are subsequently summed within calendar month:

$$z_m^{\text{dem}} = \sum_{e:m(e)=m} \hat{\nu}_e, \quad (16)$$

with non-event months again set to zero. Importantly, equation (15) does not contain the supply-news measure. Demand news is therefore not constructed as the residual of supply news, and no supply-demand orthogonalization is imposed.

Extended-sample construction. For the 1975-2025 structural extension, we apply the same procedure to the longer event sample. Through December 2017, the predictors are the same transformed series used in the baseline construction. From 2018 through 2025, we extend the six-variable information set using the same variables and transformations from the updated monthly data. The extended regression contains 192 usable events. Its R^2 is 0.127, and the correlation between raw and residualized event-level demand news is 0.934. Thus, the extended measure preserves the same construction as the baseline rather than introducing a different definition of demand news.

Predictability diagnostics. Table 9 reports Granger-style forecastability tests for the raw and purified monthly demand series over 1975-2017. Each row reports the p -value for the null that twelve lags of the indicated variable do not predict the demand-news measure, conditional on twelve own lags of the instrument and a constant. The first six variables are

Table 9: Forecastability of raw and purified demand news

Predictor	Raw demand <i>p</i> -value	Purified demand <i>p</i> -value
Real oil price (POIL)	0.472	0.815
World oil production (OILPROD)	0.323	0.557
World oil inventories (OILSTOCKS)	0.120	0.639
World industrial production (WORLDIP)	0.057	0.352
U.S. industrial production (IP)	0.021	0.432
U.S. CPI (CPI)	0.460	0.902
Policy rate (PRATE)	0.078	0.781
Stock prices (PSTOCK)	0.084	0.323
Nominal effective exchange rate (NEER)	0.633	0.442
Geopolitical risk (GPR)	0.077	0.646
Joint macro-state test	0.088	0.891

Notes: Entries are *p*-values from regressions containing a constant, twelve own lags of the demand-news series, and twelve lags of the indicated predictor. The joint row tests the lagged macroeconomic and financial predictors jointly. The first six variables enter the event-level purification in equation (15); the remaining four are held out from that regression. The tests are diagnostics of forecastability and should not be interpreted as tests of the external-instrument exclusion restriction. Sample: 1975:1-2017:12.

those entering equation (15); the remaining variables-the policy rate, stock prices, the nominal effective exchange rate, and geopolitical risk-are held out of the event-level purification.

The comparison shows that residualization substantially reduces forecastability from lagged macroeconomic conditions. For example, U.S. industrial production predicts the raw demand measure at conventional levels ($p = 0.021$), but not the purified measure ($p = 0.432$). The joint *p*-value rises from 0.088 to 0.891. The same pattern extends to the variables deliberately excluded from the purification regression. These results provide evidence that the residualization removes systematic dependence on the lagged macroeconomic state. They do not establish the proxy-SVAR exclusion restriction, which remains an identifying assumption.

Because equation (15) removes macroeconomic predictability rather than serial dependence in the instrument itself, we examine own-lag predictability separately in Appendix C.2. As an additional robustness exercise, we residualize the monthly demand instrument with

respect to twelve of its own lags before structural estimation.

B.3 Text supply news and the futures-based surprise

The futures-based OPEC surprise provides a useful external benchmark for the text supply measure because the two series are constructed from different information. The futures measure records the financial-market price response around an OPEC announcement, whereas the text measure records contemporaneous descriptions of the announced decision relative to expectations. The text classifier never observes futures prices or other market data.

Over the common 1983:7-2017:12 sample, the two measures are jointly nonzero in 91 months. Their correlation is 0.40, their signs agree in 65 percent of these months, and the squared correlation is 0.16. The association is therefore economically meaningful but far from mechanical. Extending the comparison through 2025 yields nearly the same relationship: the measures are jointly nonzero in 129 months, with a correlation of 0.39 and sign agreement of 65 percent.

The difference between the two measures is also visible in a conditional first-stage regression. Table 10 regresses the reduced-form real-oil-price innovation on the futures surprise, the text supply measure, and both instruments jointly.

When entered separately, both measures are positively related to the oil-price innovation. More importantly, both remain informative when entered together. The text coefficient is 1.37 with a robust t -statistic of 2.84, while the futures coefficient is 0.89 with a robust t -statistic of 2.72. The regression R^2 rises from 4.7 percent with the futures surprise alone and 4.0 percent with text alone to 6.7 percent when both are included.

The conditional regression therefore supports two conclusions. First, the text measure captures the same broad class of OPEC supply news as the futures-based proxy, consistent with their positive correlation and agreement at major historical episodes. Second, the text measure is not simply a textual reconstruction of the futures surprise: it retains incremental explanatory content for the oil-price innovation after the futures measure is included. This

Table 10: Information in text beyond the futures-based surprise

	Futures only	Text only	Both
Futures surprise	1.11 (3.44)	-	0.89 (2.72)
Text supply news	-	1.81 (3.83)	1.37 (2.84)
Robust F	11.84	14.68	12.33
R^2 (%)	4.7	4.0	6.7
N	414	414	414

Notes: The dependent variable is the reduced-form innovation in the real price of oil from the baseline six-variable VAR. All specifications use the common 1983:7-2017:12 sample. Heteroskedasticity-robust t -statistics are reported in parentheses. The futures-based surprise is the OPEC announcement proxy of [Känzig \(2021\)](#). The text measure is defined in Appendix [B.1](#).

distinction is important because the text series is constructed without using market prices and remains observable for announcements outside the conventional futures-event window.

C Instrument Relevance and Identification Diagnostics

This appendix collects additional diagnostics for the external instruments used in the structural analysis. We first compare first-stage relevance across the alternative proxies on the common 1983:7-2017:12 sample. We then examine own-lag predictability and the effect of residualizing the monthly text instruments with respect to their own histories. Finally, we describe the weak-identification-robust inference used for the baseline demand-news specification.

Table 11: First-stage relevance on the common sample

Instrument	F	Robust F	R^2 (%)	N
Futures supply	20.14	11.84	4.66	414
Text supply	16.96	14.68	3.95	414
Raw demand	7.19	6.52	1.72	414
Purified demand	8.80	9.06	2.09	414
Text supply, AR(12)-residualized	19.46	17.62	4.51	414
Purified demand, AR(12)-residualized	10.84	11.39	2.56	414
Futures + text supply	14.87	12.33	6.75	414

Notes: The dependent variable is the reduced-form innovation in the real price of oil from the baseline VAR. The identification sample is 1983:7-2017:12 in every row. For single-instrument specifications, the reported statistics correspond to the coefficient on the indicated proxy. In the final row, the conventional and robust F -statistics test the two supply proxies jointly. “Purified demand” denotes the event-level demand measure residualized with respect to the six predetermined oil-market and macroeconomic variables described in Appendix B.2. AR(12) specifications are robustness exercises and do not replace the baseline instruments.

C.1 Baseline first stages

Instrument relevance is measured by the relationship between each proxy and the reduced-form innovation in the real price of oil from the baseline six-variable VAR. Table 11 reports both the conventional and heteroskedasticity-robust first-stage F -statistics. All specifications use the same VAR and the same 1983:7-2017:12 identification sample.

The text-based supply instrument is comparable in strength to the futures benchmark. Its robust F -statistic is 14.68, compared with 11.84 for the futures proxy. The raw demand measure is weaker, with a robust F of 6.52. Removing the component predictable from the lagged oil-market and macroeconomic state raises the corresponding statistic to 9.06.

Residualizing the monthly proxies with respect to twelve of their own lags strengthens both text-based instruments further. The robust F -statistic rises from 14.68 to 17.62 for supply news and from 9.06 to 11.39 for demand news. These specifications provide a useful check that the structural results are not driven by predictable persistence in the monthly instruments.

The final row evaluates the futures and text supply measures jointly. The robust joint

Table 12: Own-lag predictability and first-stage relevance

Instrument	Own-lag p -value	Robust F
Text supply	0.002	14.68
Text supply, AR(12)-residualized	> 0.999	17.62
Purified demand	0.005	9.06
Purified demand, AR(12)-residualized	> 0.999	11.39

Notes: The second column reports the p -value for a joint test that twelve own lags of the monthly instrument have zero coefficients. Tests use the common 1983:7-2017:12 sample. The final column reproduces the heteroskedasticity-robust first-stage F -statistic from Table 11. The AR(12) instruments are residuals from regressions of each baseline monthly proxy on twelve of its own lags.

F -statistic is 12.33 and the regression explains 6.75 percent of the oil-price innovation, compared with 4.66 percent for futures alone and 3.95 percent for text alone. As shown in Appendix B.3, both coefficients remain positive and individually informative when entered together.

C.2 Predictability and own-lag diagnostics

Appendix B.2 examines whether lagged oil-market and macroeconomic variables forecast the demand-news proxy. Here we focus on a separate issue: predictability from the instrument's own history. Following the diagnostic used for the futures-based OPEC proxy, we regress each monthly instrument on twelve of its own lags and test their joint significance.

The baseline supply measure exhibits statistically detectable serial predictability. This is not unexpected given the temporal-aggregation procedure in equation (13): a late-month announcement contributes mechanically to both the announcement month and the following month. The purified demand series also retains own-lag predictability. The event-level purification in equation (15) is designed to remove information predictable from lagged macroeconomic conditions, not to whiten the resulting monthly series.

As a robustness exercise, we therefore estimate

$$z_t = c + \sum_{j=1}^{12} \rho_j z_{t-j} + \eta_t \quad (17)$$

separately for the supply and purified-demand proxies and use $\hat{\eta}_t$ as the external instrument. The resulting series show essentially no remaining own-lag predictability. At the same time, first-stage relevance improves rather than deteriorates: the robust F rises to 17.62 for supply and 11.39 for demand. The corresponding structural responses are reported in Appendix D and remain close to the baseline estimates.

C.3 Weak-identification-robust inference

First-stage relevance is more limited for the baseline demand-news instrument than for the two supply proxies. Its heteroskedasticity-robust first-stage F -statistic is 9.06. We therefore do not rely exclusively on moving-block-bootstrap inference for this specification. Instead, the baseline demand figures report weak-identification-robust confidence sets following the MSW procedure of Montiel Olea et al. (2021).⁸ Our use of weak-identification-robust inference should not be interpreted as treating any particular value of the robust F -statistic as a universal validity threshold.

For the baseline demand specification, the MSW relevance Wald statistic is 8.31. The MSW procedure changes the confidence sets, not the reported proxy-SVAR point estimate. As a numerical implementation check, the maximum absolute difference between the MSW point response and the stored baseline proxy-SVAR impulse response across all variables and horizons is 2.3×10^{-8} . We therefore plot the same structural point estimate throughout, using the MSW 68 and 90 percent confidence sets for the baseline demand specification.

The AR(12)-residualized demand instrument provides a complementary relevance check. Its robust first-stage F rises to 11.39, and the resulting impulse responses remain similar when conventional moving-block-bootstrap inference is used. Thus, the interpretation of the demand-news responses does not depend on applying conventional inference to the weaker baseline first stage.

⁸MSW refers to the Montiel Olea-Stock-Watson weak-identification-robust construction.

Table 13: Degree-of-invertibility diagnostics

	Lower bound	Upper bound	Pre-test p -value
Känzig futures surprise	0.060	0.345	0.372
Text supply news	0.045	0.258	0.082
Residualized demand news	0.038	0.179	< 0.001

Notes: The first two columns report estimated bounds on the degree of invertibility following [Plagborg-Møller and Wolf \(2022\)](#). The final column reports the p -value from the corresponding Granger-causality pre-test of invertibility. All specifications use the 1983:7-2017:12 benchmark sample, the six-variable VAR, and 12 lags.

C.4 Invertibility diagnostics

Identification based on external instruments in SVAR typically uses the property that structural shocks can be recovered by looking at the information available in the VAR both now and in the past. This condition is tested through the bounds of degree of invertibility of [Plagborg-Møller and Wolf \(2022\)](#).

Table 13 provides the identified set estimates for the degree of invertibility for our benchmark period from 1983:7 to 2017:12. The upper bound is 0.35 for the Känzig futures instrument, 0.26 for text-based supply news, and 0.18 for residualized demand news. The associated Granger-causality pre-test rejects invertibility for residualized demand news. For text-based supply news, the p -value is 0.082, implying rejection at the 10 percent level but not at the 5 percent level. The test does not reject invertibility for the [Känzig \(2021\)](#) futures instrument.

This implies that the six-variable VAR might not encompass all the information contained in OPEC announcement shocks, especially for demand news. As such, we do not rely only on the invertible proxy-SVAR representation. Appendix D.2 provides response functions using the internal instruments VAR, which identifies relative impulse responses without invertibility.

Table 14: External validation using the Mori-Peersman oil-supply-news shock

	(1) Baseline	(2) + Futures proxy	(3) AR(12) text measures
Text supply news	0.195*** (3.76)	0.148*** (2.95)	0.210*** (4.50)
Residualized demand news	0.060 (1.05)	0.016 (0.26)	0.069 (1.24)
Känzig futures surprise		0.172** (2.16)	
N	414	414	414
R^2	0.042	0.068	0.050

Notes: The dependent variable is the Mori-Peersman oil-supply-news shock. All variables are standardized. Parentheses report HAC t -statistics using 12 lags. Column (2) additionally controls for the Känzig futures-based OPEC surprise. Column (3) uses text measures residualized with respect to twelve of their own lags. The sample is 1983:7-2017:12. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

C.5 External validation with an alternative oil-supply-news shock

As an external validation of the structural interpretation of the text-based supply measure, we compare it with the oil-supply-news shock estimated by [Mori and Peersman \(2024\)](#). Their shock is recovered from a financially augmented structural VAR designed to separate oil-supply news from information effects that can contaminate announcement-based futures surprises.

Over the benchmark 1983:7-2017:12 sample, we regress the Mori-Peersman supply-news shock on the text-based supply and residualized demand measures. All variables are standardized before estimation. The text-based supply measure is positively and significantly associated with the alternative supply-news shock, whereas the text-based demand measure is not. This pattern remains when the Känzig futures surprise is included simultaneously and when the text instruments are residualized with respect to twelve of their own lags.

Table 14 reports the results. The exercise should be interpreted as external validation rather than a direct test of the proxy-SVAR exclusion restriction, since the Mori-Peersman shock is itself obtained from a structural model.

C.6 Exploratory joint supply–demand diagnostic

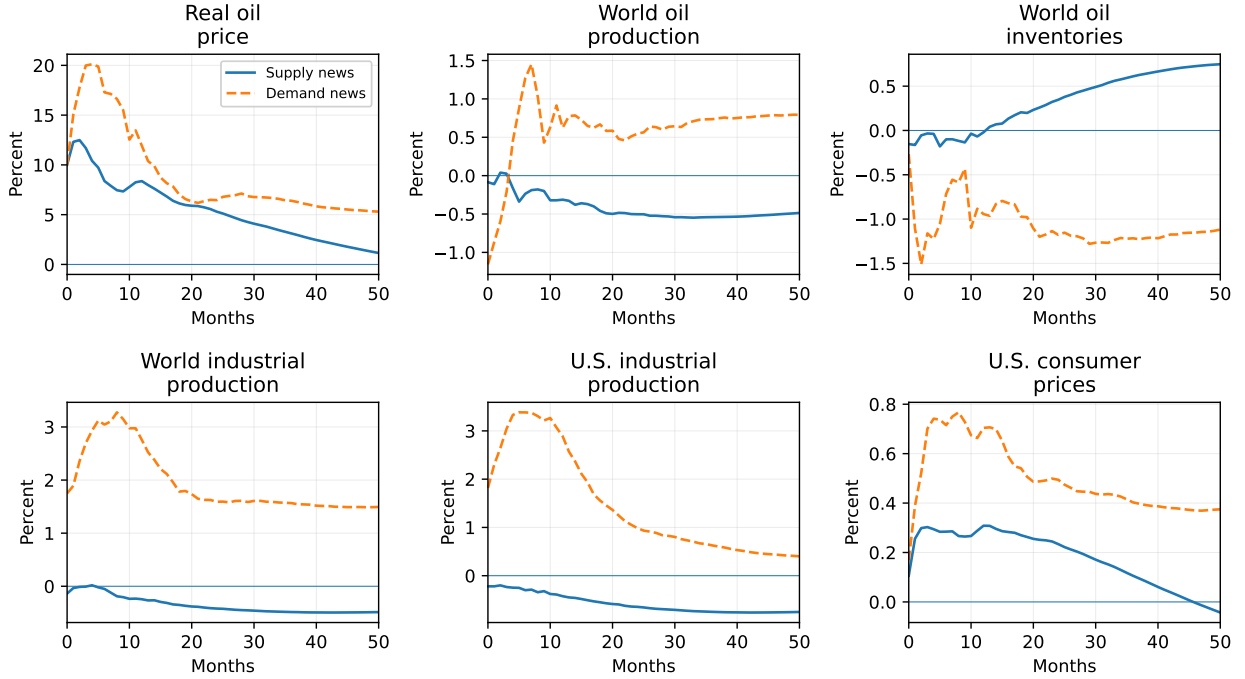
The baseline analysis identifies supply and demand news separately. As a supplementary diagnostic, we examine the responses obtained when the two proxy covariance directions are orthogonalized within the same VAR. Using the benchmark six-variable VAR with twelve lags and the common 1983:7–2017:12 proxy window, we compute the sample covariance between each text instrument and the reduced-form innovations. We apply a symmetric orthogonalization to these two covariance vectors, using the inverse reduced-form residual covariance matrix as the metric, and propagate the resulting impact vectors through the estimated VAR. Each response is normalized to raise the real oil price by 10 percent on impact.

Figure 8 reports the resulting point estimates. The supply-associated response features declining oil production and weaker world and U.S. industrial production, together with an eventual increase in inventories. The demand-associated response instead features a pronounced expansion in industrial production, an increase in oil production after the initial months, and a persistent inventory decline.

The diagnostic also illustrates why low correlation between the proxy series does not establish orthogonality of the associated structural directions. Although the sample correlation between the two proxies is 0.021, the cosine between their covariance directions in the inverse-residual-covariance metric is 0.676, corresponding to an angle of approximately 47.5 degrees rather than 90 degrees. These are descriptive geometric diagnostics, not formal tests of the exclusion restrictions.

The magnitudes also depend on the orthogonalization rule. For example, the impact response of world industrial production to the demand-associated direction is approximately 1.75 percent under symmetric orthogonalization, compared with 4.58 percent under an ordered construction that holds the supply covariance direction fixed. We therefore interpret this exercise as an exploratory comparison under an additional orthogonalization convention, not as a uniquely established joint structural decomposition. The figure reports no

Exploratory joint two-shock identification



Symmetric orthogonalization of the two proxy covariance directions; diagnostic only

Figure 8: Exploratory joint supply–demand responses. The solid line reports the supply-associated response and the dashed line reports the demand-associated response obtained by symmetrically orthogonalizing the two text-proxy covariance directions in the inverse reduced-form residual covariance metric. Both responses are propagated through the benchmark VAR and normalized to raise the real oil price by 10 percent on impact. The common proxy window is 1983:7–2017:12. Point estimates only; no confidence bands are reported. This construction is a supplementary diagnostic rather than the baseline identification strategy or a test of the proxy exclusion restrictions.

confidence bands and does not replace the separate baseline estimates.

D Alternative Dynamic Specifications

This appendix examines whether the estimated responses are sensitive to predictability in the monthly instruments, the invertibility assumption underlying the baseline proxy-SVAR, or the estimation of dynamic propagation. First, we use supply- and demand-news instruments residualized with respect to twelve of their own lags. Second, we estimate an internal-

instrument VAR that identifies relative impulse responses without requiring invertibility with respect to the original six-variable information set, under the maintained instrument-validity conditions. Finally, we estimate local projections using the structural shocks recovered from the baseline proxy-SVAR. These horizon-by-horizon comparisons assess sensitivity to the VAR’s recursive propagation while retaining the baseline identifying assumptions.

D.1 Own-lag residualized instruments

Section C.2 shows that both baseline text instruments exhibit some predictability from their own histories. As a robustness exercise, we therefore replace each monthly instrument with the residual from

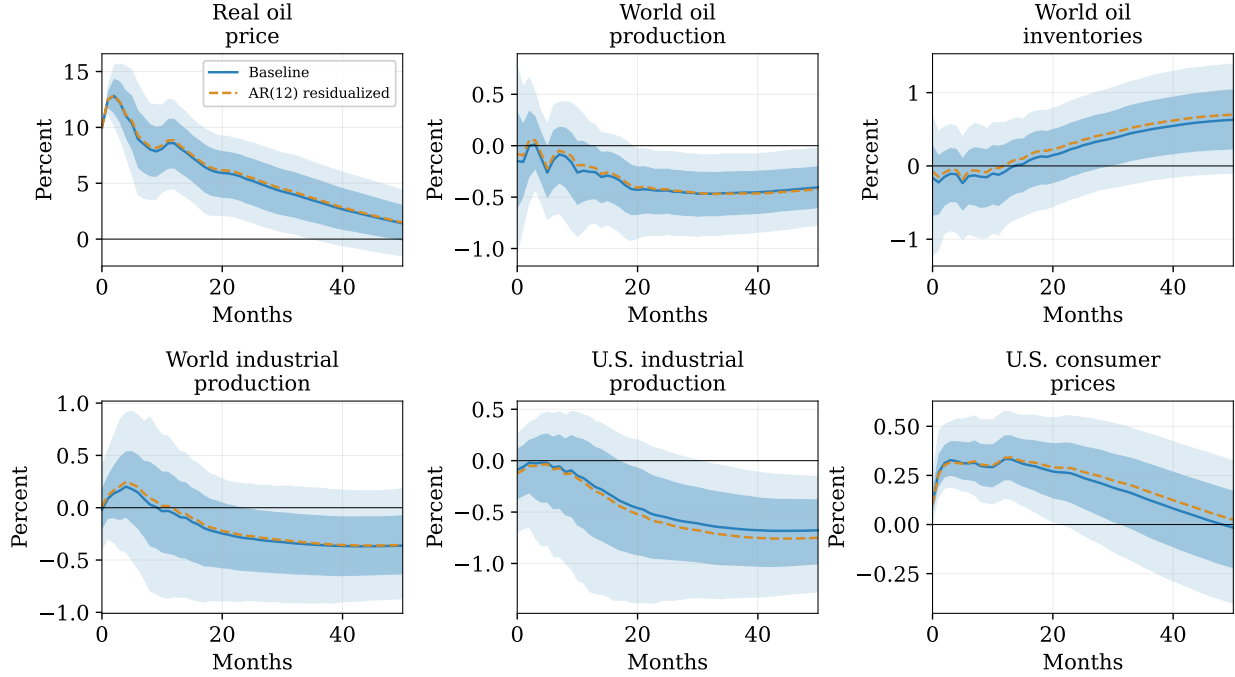
$$z_t = c + \sum_{j=1}^{12} \rho_j z_{t-j} + \eta_t, \quad (18)$$

and use $\hat{\eta}_t$ as the external instrument. This adjustment is performed separately for supply and demand news and does not change the underlying article classifications or event-level construction.

Figures 9 and 10 compare the resulting impulse responses with the baseline estimates. The responses remain close across the two instrument constructions. For supply news, the decline in world and U.S. industrial production and the increase in consumer prices are preserved after removing own-lag predictability. The demand specification is similarly stable: the oil price rises on impact, real activity initially expands, and inventories decline.

The residualized instruments are also somewhat more relevant in the first stage. As reported in Appendix C.1, the heteroskedasticity-robust F -statistic rises from 14.68 to 17.62 for supply news and from 9.06 to 11.39 for demand news. The similarity of the impulse responses therefore indicates that the baseline results are not generated by the predictable component of the monthly proxy series.

Supply news: baseline and own-lag residualization



MBB 90/68% bands; robust $F = 14.68$; AR(12) robust $F = 17.62$

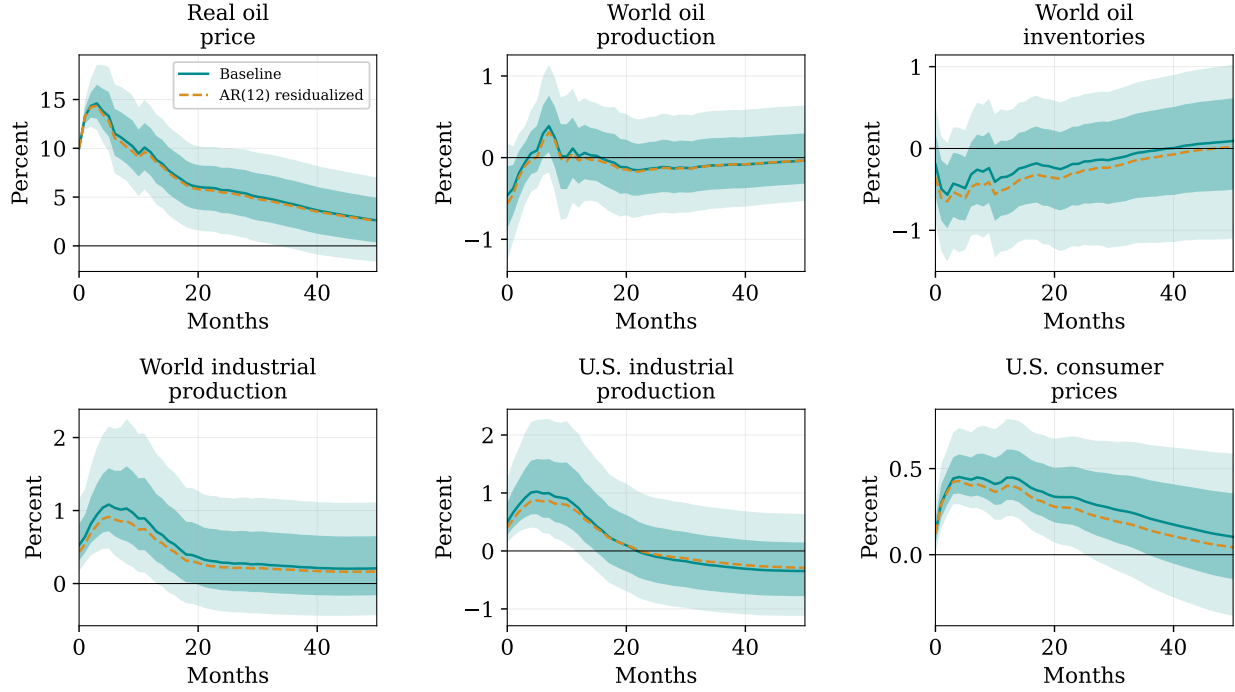
Figure 9: Supply-news responses using the baseline and AR(12)-residualized instruments. The solid line reports the baseline text-supply proxy-SVAR and the dashed line uses the residual from a regression of the monthly supply instrument on twelve of its own lags. Both shocks are normalized to raise the real oil price by 10 percent on impact. Shaded regions report the 68 and 90 percent baseline moving-block-bootstrap confidence bands. The identification sample is 1983:7-2017:12.

D.2 Invertibility-robust internal-instrument VAR

The estimates from the proxy-SVAR in the main text use the standard external-instrument specification. The invertibility diagnostics in Appendix C.4 indicate that the six-variable VAR may not fully span the information contained in the OPEC announcement shocks, particularly for demand news. We therefore estimate the responses using the internal-instrument approach of Plagborg-Møller and Wolf (2021).

We augment the benchmark VAR with the external instrument, order the instrument first, and identify its innovation recursively. Relative structural impulse responses obtained in this way remain identified even when the structural shock is noninvertible. As in the

Demand news: baseline and own-lag residualization



MSW weak-ID-robust 90/68% bands; robust $F = 9.06$; relevance Wald = 8.31; AR(12) robust $F = 11.39$

Figure 10: Demand-news responses using the baseline and AR(12)-residualized instruments. The solid line reports the baseline demand-news proxy-SVAR and the dashed line uses the residual from a regression of the monthly purified-demand instrument on twelve of its own lags. Both shocks are normalized to raise the real oil price by 10 percent on impact. Shaded regions report the 68 and 90 percent weak-identification-robust confidence sets for the baseline specification. The identification sample is 1983:7-2017:12.

baseline specification, the responses are normalized to a 10 percent increase in the real oil price on impact.

Figures 11 and 12 compare the baseline proxy-SVAR responses with the corresponding responses from the internal-instrument VAR. The shaded regions report the confidence sets for the baseline proxy-SVAR, while the internal-instrument VAR responses are shown as dashed lines.

The central supply-demand distinction is preserved. Supply news is followed by lower oil production and generally weaker real activity. Demand news, in contrast, produces a pronounced expansion in world and U.S. industrial production and a short-run increase

Supply news: Proxy-SVAR and internal-instrument VAR

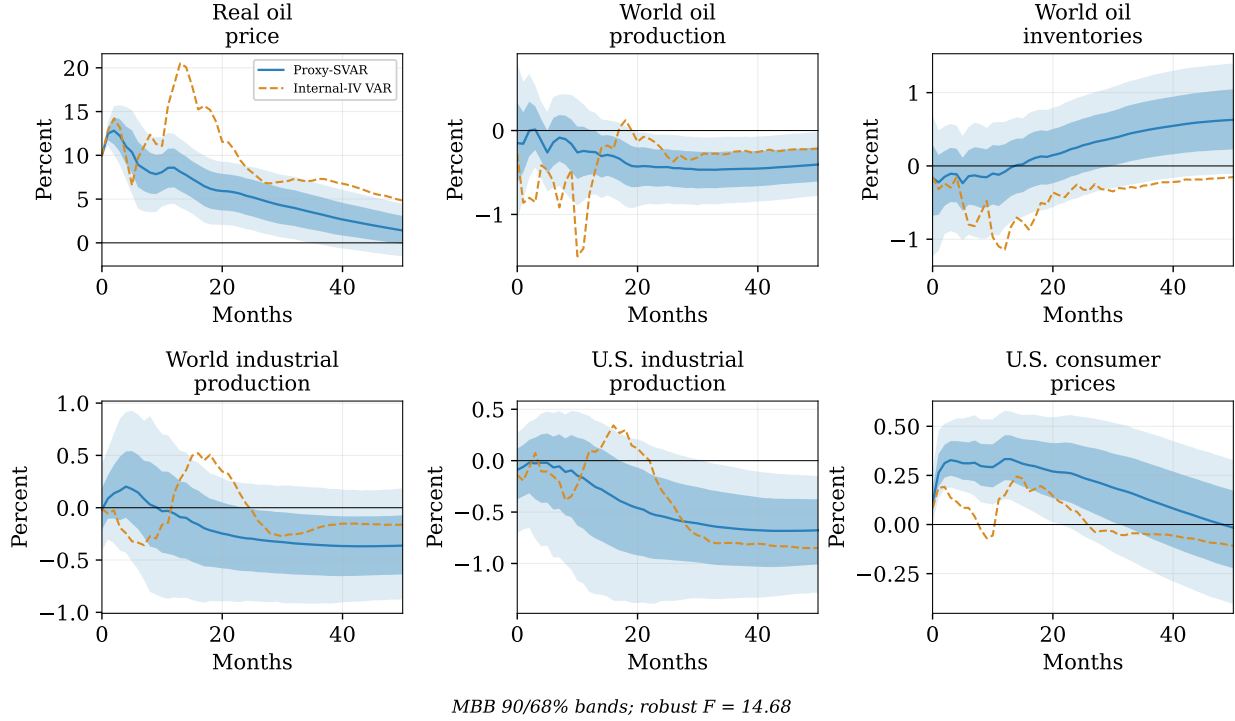


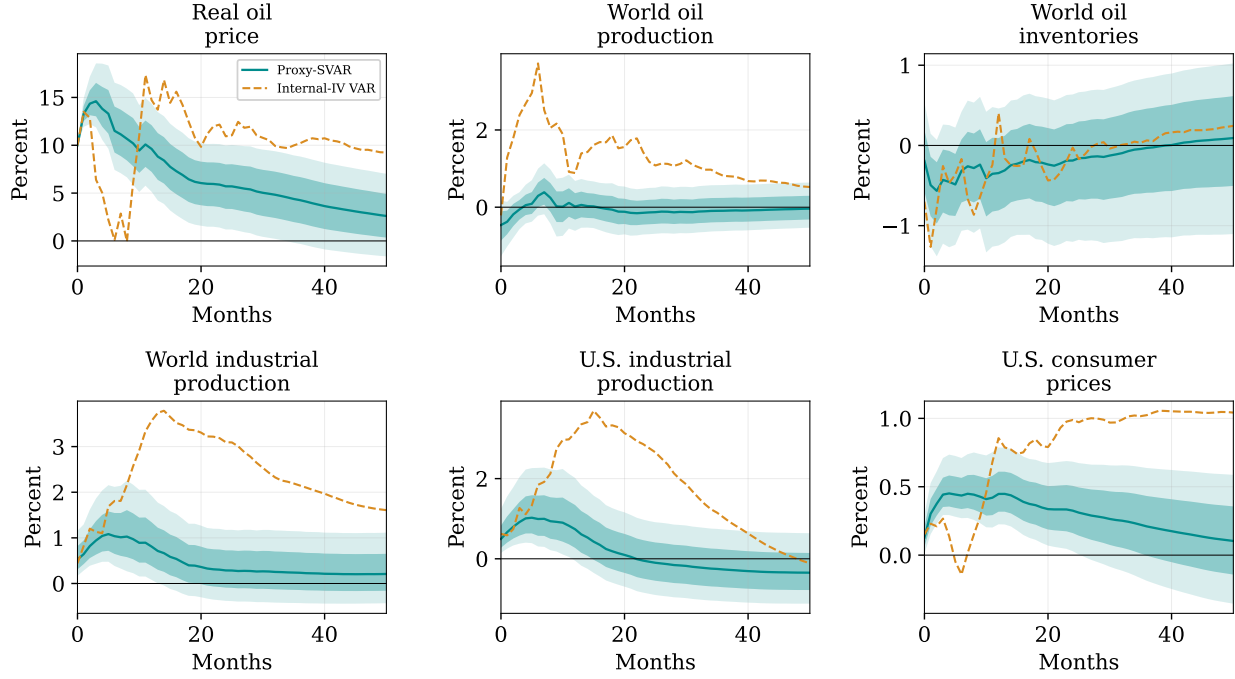
Figure 11: Supply-news responses from the proxy-SVAR and the internal-instrument VAR. The solid line reports the baseline proxy-SVAR estimate and the dashed line reports the internal-instrument VAR point estimate using the same text-based supply-news measure. Both responses are normalized to raise the real oil price by 10 percent on impact. Shaded regions report the 68 and 90 percent moving-block-bootstrap confidence bands for the baseline proxy-SVAR. The identification sample is 1983:7-2017:12.

in world oil production. In the internal-instrument specification, the expansion in world industrial production remains positive throughout the reported horizon. The qualitative demand result therefore does not depend on the invertibility assumption underlying the benchmark proxy-SVAR representation.

D.3 Local projections using recovered structural shocks

Following the recovered-shock local-projection approach in [Känzig \(2021\)](#), we assess the sensitivity of the estimated responses to replacing the VAR's recursive propagation with horizon-by-horizon estimation. For each baseline proxy-SVAR specification, we recover the

Demand news: Proxy-SVAR and internal-instrument VAR



MSW weak-ID-robust 90/68% bands; robust $F = 9.06$; relevance Wald = 8.31

Figure 12: Demand-news responses from the proxy-SVAR and the internal-instrument VAR. The solid line reports the baseline proxy-SVAR estimate and the dashed line reports the internal-instrument VAR point estimate using the same residualized demand-news measure. Both responses are normalized to raise the real oil price by 10 percent on impact. Shaded regions report the 68 and 90 percent MSW weak-identification-robust confidence sets for the baseline proxy-SVAR. The identification sample is 1983:7–2017:12.

corresponding structural shock series and estimate, for each variable i and horizon h ,

$$y_{i,t+h} = \alpha_{i,h} + \beta_{i,h}\widehat{\varepsilon}_t + \gamma_{i,h}y_{i,t-1} + u_{i,t+h}, \quad (19)$$

where $\widehat{\varepsilon}_t$ denotes the recovered supply- or demand-news shock. Each regression is estimated by ordinary least squares and includes a constant and one lag of the dependent variable. The structural shocks are recovered from the baseline six-variable VAR with twelve lags.

The proxy identification window is 1983:7–2017:12. The local projections use a common set of shock dates across horizons $h = 0, \dots, 50$, restricting these dates to 1983:7–2013:10 so

that outcomes through the maximum horizon are observed by December 2017. Each set of local-projection responses is rescaled by a common factor so that the real oil price rises by 10 percent on impact, matching the normalization of the baseline proxy-SVAR.

Figures 13 and 14 compare the local-projection point estimates with the baseline proxy-SVAR responses. The shaded regions report confidence sets for the baseline proxy-SVAR only; no local-projection confidence bands are displayed. The overlays therefore provide a comparison of estimated response patterns rather than a formal test of equality between the two estimators.

The local-projection estimates preserve the initial expansion in world and U.S. industrial production following demand news, together with a temporary increase in world oil production. Supply news is followed by weaker real activity at intermediate horizons, particularly in the United States. Agreement is not uniform across variables and horizons, however. The local-projection response of world oil production to supply news remains closer to zero than the baseline response, and several longer-horizon responses are less persistent. The comparison therefore supports the qualitative distinction in real-activity responses while indicating sensitivity in the magnitude and persistence of some oil-market responses.

This exercise is distinct from the internal-instrument VAR in Appendix D.2. The local projections retain the structural shocks recovered from the baseline proxy-SVAR and therefore inherit its identifying assumptions, including those required for shock recovery. Their purpose is to assess sensitivity to the estimation of dynamic propagation, not to provide an independent validation of identification. The internal-instrument VAR addresses the separate concern about invertibility.

E Alternative Supply Identification

This appendix considers two additional approaches to identifying OPEC supply news. First, we use the futures-based and text-based surprises jointly as two proxies for the same struc-

Supply news: Proxy-SVAR and recovered-shock local projections

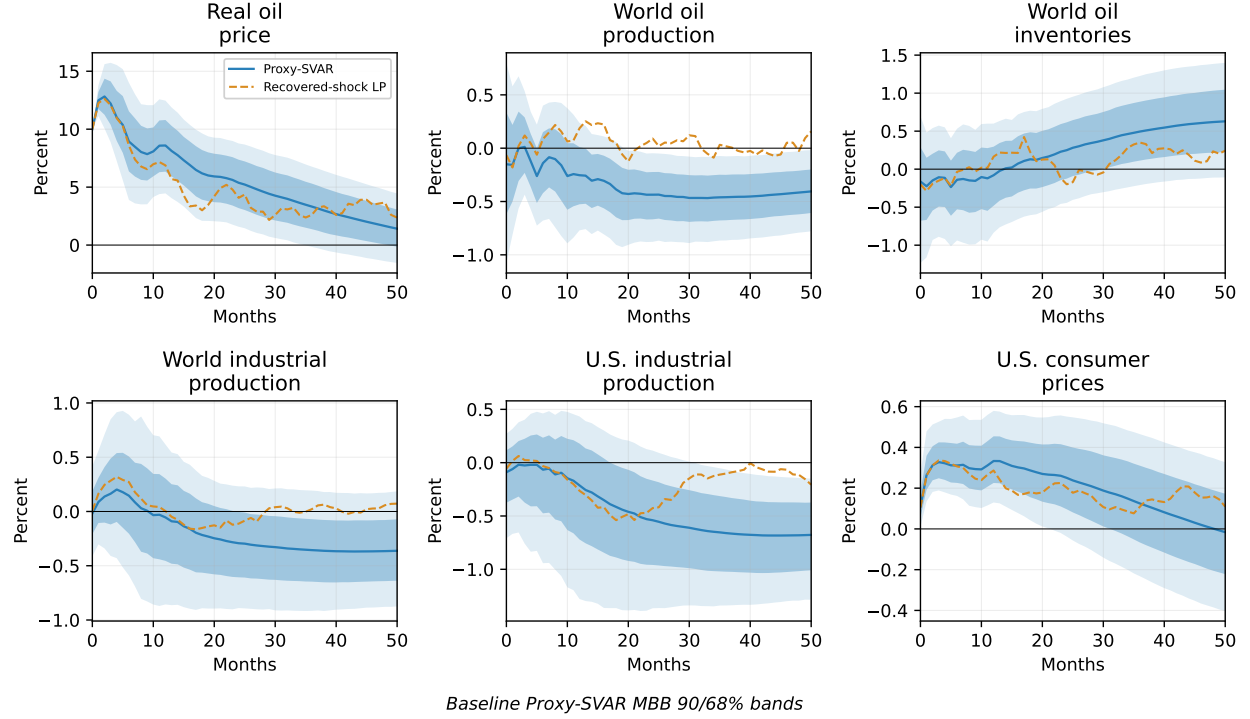


Figure 13: Supply-news responses from the proxy-SVAR and local projections using the recovered structural shock. The solid line reports the baseline proxy-SVAR response, and the dashed line reports the local-projection estimate using the recovered supply-news shock, a constant, and one lag of the dependent variable. Both sets of responses are normalized to raise the real oil price by 10 percent on impact. Shaded regions report the 68 and 90 percent moving-block-bootstrap confidence bands for the baseline proxy-SVAR only; no local-projection confidence bands are shown. The proxy identification window is 1983:7–2017:12. The local projections use the fixed shock-date sample 1983:7–2013:10, with outcomes observed through December 2017.

tural supply-news shock. Second, we examine whether first-stage relevance depends on the scaling and temporal aggregation used to construct the text instrument.

E.1 Futures and text as two proxies for the same shock

The baseline analysis uses the futures and text measures separately in order to compare their implied structural responses. An alternative is to treat them as two imperfect proxies for the same underlying OPEC supply-news shock. We therefore estimate an overidentified

Demand news: Proxy-SVAR and recovered-shock local projections

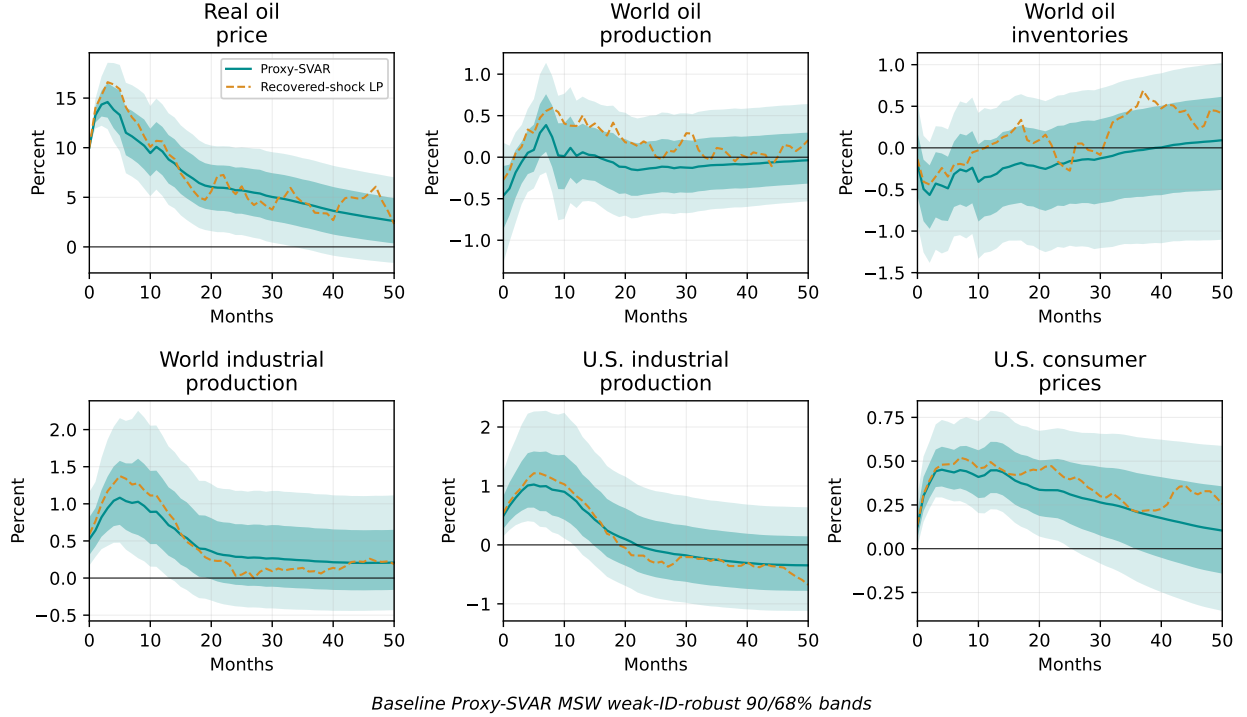


Figure 14: Demand-news responses from the proxy-SVAR and local projections using the recovered structural shock. The solid line reports the baseline proxy-SVAR response, and the dashed line reports the local-projection estimate using the recovered demand-news shock, a constant, and one lag of the dependent variable. Both sets of responses are normalized to raise the real oil price by 10 percent on impact. Shaded regions report the 68 and 90 percent MSW weak-identification-robust confidence sets for the baseline proxy-SVAR only; no local-projection confidence bands are shown. The proxy identification window is 1983:7–2017:12. The local projections use the fixed shock-date sample 1983:7–2013:10, with outcomes observed through December 2017.

external-instrument specification that uses both measures simultaneously.

The joint first stage remains informative. Over the common 1983:7-2017:12 sample, the heteroskedasticity-robust joint F -statistic is 12.33 and the two proxies jointly explain 6.75 percent of the reduced-form oil-price innovation. Moreover, both proxies retain conditional explanatory power: the futures coefficient has a robust t -statistic of 2.72 and the text coefficient a robust t -statistic of 2.84. Thus, the text measure contains information about the oil-price innovation that is not subsumed by the futures surprise, and vice versa.

Supply news: futures and text as two proxies for the same shock

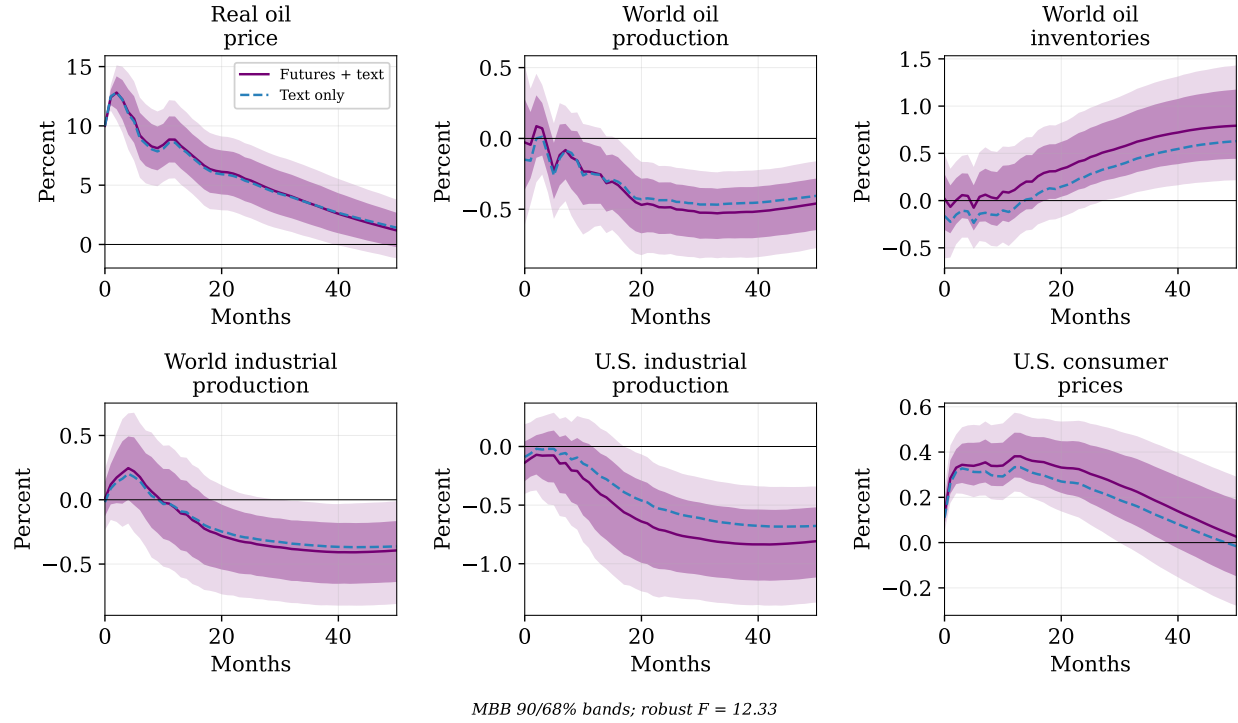


Figure 15: Supply-news responses using futures and text as two proxies for the same structural shock. The solid line reports the specification using both the futures-based OPEC surprise and the text-based supply-news measure as external instruments. The dashed line reports the baseline text-only estimate. The shock is normalized to raise the real oil price by 10 percent on impact. Shaded regions report 68 and 90 percent moving-block-bootstrap confidence bands for the dual-proxy specification. The heteroskedasticity-robust joint first-stage F -statistic is 12.33. The identification sample is 1983:7-2017:12.

Figure 15 reports the resulting structural responses. The dual-proxy estimates remain close to those obtained using the text instrument alone. In particular, the central supply-news pattern is preserved: an OPEC supply shock that raises the real oil price is followed by weaker real activity and higher consumer prices. Combining the two sources of identifying variation therefore reinforces the supply interpretation without materially changing the macroeconomic responses.

Table 15: Alternative constructions of the text supply instrument

Supply-instrument construction	F	Robust F	R^2 (%)	N
Baseline timing-weighted measure	16.96	14.68	3.95	414
Simple mean	13.01	11.34	3.06	414
Unscaled sum	19.19	19.82	4.45	414
Announcement month only	11.31	10.12	2.67	414

Notes: Each row reports the first-stage relationship between the indicated text-based supply measure and the reduced-form innovation in the real price of oil from the baseline six-variable VAR. The baseline measure scales the sum of article-level surprise directions by the square root of the number of surprise-bearing articles and applies the monthly timing adjustment described in Appendix B.1. The simple-mean measure replaces the event-level scaling with the average article-level surprise direction. The unscaled-sum measure uses the raw sum of article-level surprise directions. The announcement-month-only specification removes the next-month spillover from the baseline timing allocation. All specifications use the common 1983:7-2017:12 identification sample.

E.2 Alternative construction of the text supply measure

The baseline text instrument makes two construction choices. First, the event-level surprise is the sum of article-level surprise directions scaled by the square root of the number of surprise-bearing articles. This gives more weight to an interpretation supported by broad independent coverage without making the instrument proportional to raw article volume. Second, the event-level measure is allocated between the announcement month and the following month according to the timing rule described in Appendix B.1.

Table 15 examines the sensitivity of first-stage relevance to both choices. The simple-mean specification gives each event the average surprise direction across contributing articles. The unscaled-sum specification instead sums the article-level surprise directions without the square-root normalization. Finally, the announcement-month-only specification retains the baseline event-level measure but removes the next-month allocation generated by the temporal- aggregation adjustment.

First-stage relevance is not specific to the baseline scaling choice. Both the simple-mean and unscaled-sum measures remain informative, with robust F -statistics of 11.34 and 19.82, respectively. In particular, the stronger first stage for the unscaled sum shows that the

baseline result is not created by the square-root article-count normalization. We nevertheless prefer the baseline normalization because it allows agreement across a broader set of articles to strengthen an event without allowing article volume itself to determine the magnitude of the instrument.

Removing the next-month timing allocation weakens the first stage more noticeably, reducing the robust F from 14.68 to 10.12. The announcement-month-only measure nevertheless remains positively relevant. This pattern is consistent with the temporal-aggregation motivation for the baseline construction: assigning a late-month OPEC announcement entirely to its calendar month discards information that is more naturally associated with the following month’s average oil price. Taken together, the alternative constructions show that the relevance of the text-based supply measure is not driven by a particular choice of scaling or aggregation.

F Extended-Sample Results, 1975-2025

The baseline analysis focuses on 1983:7-2017:12 so that the futures, text-supply, and demand-news specifications can be compared over a common identification sample. The longer text record allows us to extend the structural analysis back to 1975 and forward through 2025. The extended sample, however, contains the extraordinary macroeconomic and oil-market movements associated with the COVID-19 pandemic. We therefore treat the COVID-adjusted specification as the preferred full-sample extension.

Specifically, the reduced-form VAR includes separate pulse dummies for March, April, May, and June 2020. Each dummy enters every reduced-form VAR equation as an exogenous control. The external instruments themselves are left unchanged. We otherwise retain the same six-variable VAR, twelve lags, and 10-percent oil-price normalization used in the baseline analysis. For demand news, the predetermined-information adjustment is re-estimated over the extended event sample using the same six lagged oil-market and macroeconomic

variables described in Appendix B.2.

Both text instruments remain relevant in this specification. The heteroskedasticity-robust first-stage F -statistics are 12.88 for supply news and 13.43 for purified demand news. We therefore report 68 and 90 percent moving-block-bootstrap confidence bands based on 10,000 replications.

F.1 Full-sample supply responses

Figure 16 reports the responses to text-based OPEC supply news over 1975-2025. The full-sample estimates preserve the main supply-news pattern from the benchmark sample. Following a shock normalized to raise the real oil price by 10 percent on impact, world and U.S. industrial production initially rise modestly but subsequently decline. World oil production also falls at medium horizons, while U.S. consumer prices increase persistently.

The COVID adjustment matters most for the propagation of real activity. Relative to the untreated full-sample estimate shown for reference in the figure, the preferred specification produces a clearer medium-run contraction in world and U.S. industrial production. The central interpretation is nevertheless unchanged: OPEC news associated with tighter-than-expected supply raises oil prices and is subsequently followed by weaker real activity.

F.2 Full-sample demand responses

Figure 17 reports the corresponding results for purified demand news. The distinction from supply news remains clear in the extended sample. Demand news that raises the real oil price is followed by a pronounced short-run expansion in both world and U.S. industrial production. World oil production also rises initially, while world oil inventories decline. Consumer prices increase following the shock.

This pattern is economically different from the supply response in Figure 16. Both shocks raise the oil price on impact, but supply news is followed by weaker real activity at medium horizons whereas demand news generates a substantial initial expansion. The longer sample

Supply news: full sample with COVID pulse dummies

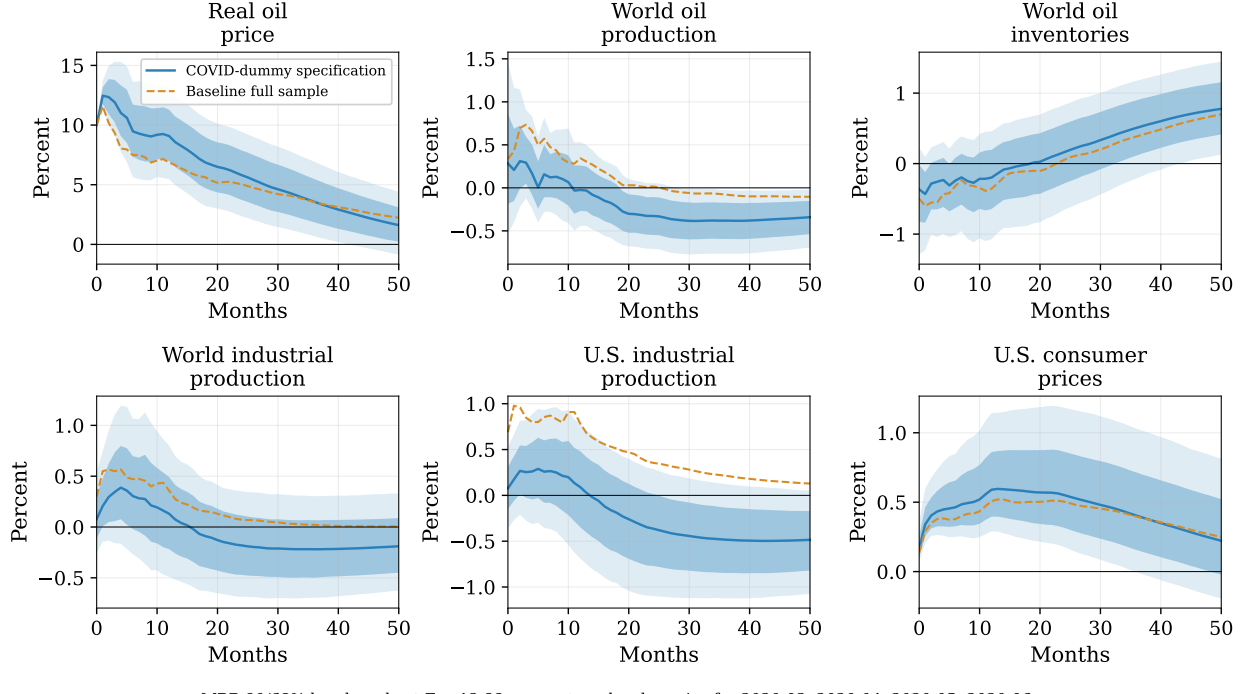


Figure 16: Full-sample responses to OPEC supply news, 1975-2025. The solid line reports the preferred specification in which the reduced-form VAR includes separate pulse dummies for March, April, May, and June 2020; the dashed line reports the corresponding untreated full-sample point estimate for comparison. The external supply-news instrument is otherwise unchanged. The shock is normalized to raise the real oil price by 10 percent on impact. Shaded regions report 68 and 90 percent moving-block-bootstrap confidence bands for the COVID-adjusted specification based on 10,000 replications. The heteroskedasticity-robust first-stage F -statistic is 12.88.

therefore preserves the central result from the benchmark period: the sign of the oil-price response alone does not reveal whether an OPEC announcement contains supply or demand information.

The first stage is also stronger than in the common-sample baseline demand specification. The heteroskedasticity-robust F -statistic is 13.43, compared with 9.06 over 1983:7-2017:12. Accordingly, the full-sample demand responses are reported with moving-block-bootstrap rather than weak-identification-robust MSW confidence sets.

Demand news: full sample with COVID pulse dummies

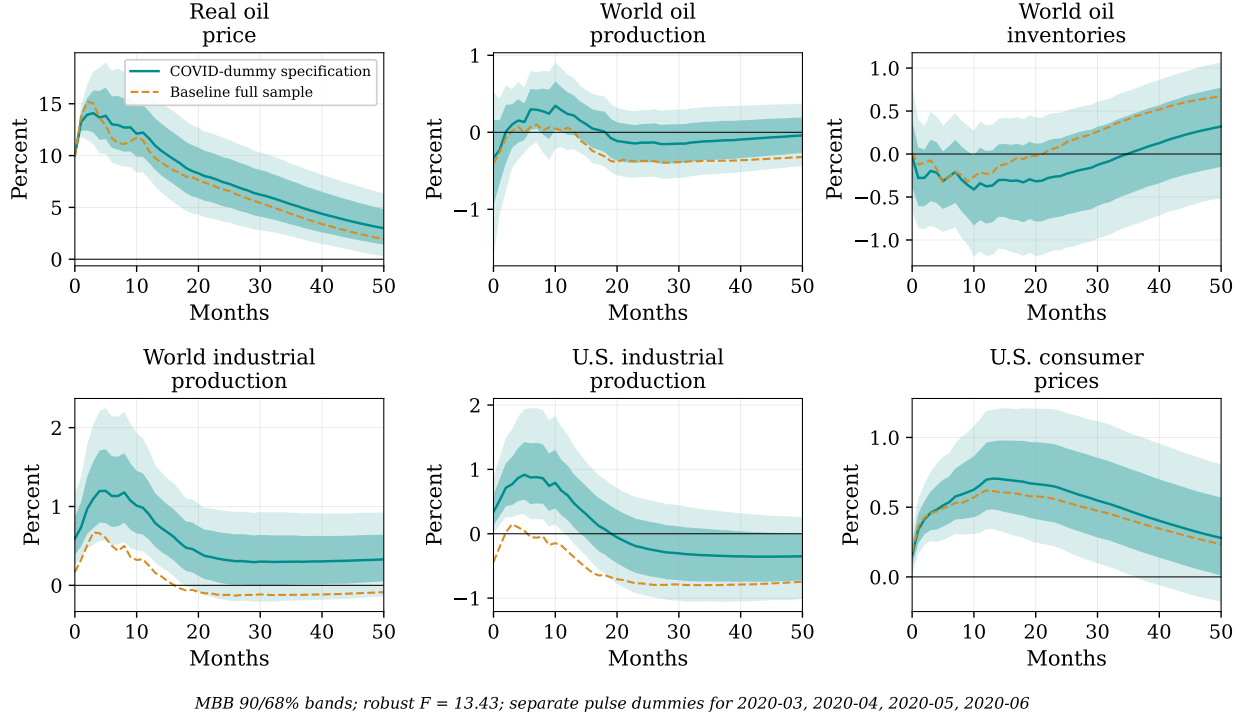


Figure 17: Full-sample responses to OPEC demand news, 1975-2025. The solid line reports the preferred specification in which the reduced-form VAR includes separate pulse dummies for March, April, May, and June 2020; the dashed line reports the corresponding untreated full-sample point estimate for comparison. The demand-news measure uses the extended-sample predetermined-information adjustment described in Appendix B.2. The shock is normalized to raise the real oil price by 10 percent on impact. Shaded regions report 68 and 90 percent moving-block-bootstrap confidence bands for the COVID-adjusted specification based on 10,000 replications. The heteroskedasticity-robust first-stage F -statistic is 13.43.

G Pandemic and Recent-Period Sensitivity

G.1 The acute COVID period

The preferred 1975-2025 specification controls for the exceptional macroeconomic and oil-market observations surrounding the onset of the COVID-19 pandemic using separate pulse dummies for March through June 2020. As an additional check, we instead set the external instruments to zero in these four months while leaving the reduced-form VAR otherwise unchanged. The resulting heteroskedasticity-robust first-stage F -statistics are 13.89 for sup-

ply news and 12.65 for demand news, and the central supply-demand distinction in the impulse responses is preserved. Thus, the full-sample results do not depend on using the acute-COVID announcement observations for identification.

G.2 Influence of recent OPEC observations

We also examine whether the post-2017 results are driven by a small number of individual OPEC or OPEC+ announcements. For each nonzero instrument month after 2017, we set that month’s instrument value to zero, re-estimate the model, and record the largest absolute change in the impulse-response functions.

The most influential post-2017 observation for both measures is April 2020. Censoring that month changes the largest response by approximately 1.06 percentage points for supply news and 1.43 percentage points for demand news. In contrast, the observations from 2023-2025 are substantially less influential. Across these recent years, the largest change in any response is approximately 0.08 percentage points for supply news and 0.26 percentage points for demand news. The extended results therefore are not driven by a particular recent OPEC announcement during the period of heightened geopolitical tension in the Middle East.

H Additional Historical and Regime Evidence

This appendix provides the underlying event calendars and additional first-stage diagnostics for the institutional-regime results discussed in Section 6. The purpose is documentary. The structural results for the extended sample are reported separately in Appendix F.

H.1 Pre-futures OPEC events

The text corpus identifies 37 OPEC decision events before the beginning of the futures-based identification sample in July 1983. Table 16 lists the complete calendar. These events are available to the text-based measurement procedure even though no corresponding oil-futures

Table 16: Pre-futures OPEC event calendar

Year	Announcement dates
1971	Feb. 3, Feb. 14, Apr. 2
1972	Jan. 20
1973	Jun. 1, Sep. 16, Oct. 16, Dec. 23
1974	Jan. 9, Jun. 17, Sep. 13, Dec. 13
1975	Jun. 11, Sep. 25
1976	Apr. 21, May 27, Dec. 17
1977	May 13, Jun. 26, Jun. 30, Dec. 22
1978	May 8, Jun. 18, Dec. 18
1979	Mar. 27, Jun. 28, Dec. 20
1980	Jun. 10, Sep. 18, Dec. 16
1981	May 26, Aug. 21, Oct. 29, Dec. 10
1982	Jul. 12
1983	Jan. 25, Mar. 14

Notes: The table lists the 37 OPEC decision events in the text calendar prior to the July 1983 beginning of the benchmark futures-based identification sample. Inclusion in the event calendar does not imply that a usable supply surprise is observed for the event; a supply-news score requires contemporaneous decision-report coverage that explicitly characterizes the outcome relative to prior expectations.

surprise can be constructed.

The calendar begins in 1971, whereas the pre-futures first-stage diagnostics in Section 6 begin in 1975:1 because they require reduced-form innovations from the monthly macroeconomic VAR. Thus, the 37-event historical calendar should not be confused with the number of identifying observations entering the pre-futures monthly first stage.

H.2 OPEC+ event calendar, 2018-2025

Table 17 reports the 50 OPEC and OPEC+ announcement events in the post-2017 portion of the text sample. We report the event-level supply-news score and the raw demand-news score so that the identifying variation can be inspected directly. The demand column is the text-based measure before the predetermined- information adjustment described in Appendix B.2; the structural demand instrument uses the residualized version.

A missing supply score indicates that no article for the event provides a usable decision-

surprise classification. This differs from a supply score of zero, which indicates usable decision coverage whose surprise directions aggregate to zero.

Table 17: OPEC and OPEC+ event calendar, 2018-2025

Announcement date	z_e^{sup}	$z_e^{\text{dem,raw}}$
2018-06-23	+0.89	+0.23
2018-12-07	+1.60	−1.00
2019-07-02	−1.94	−1.00
2019-12-06	−1.00	−0.67
2020-03-06	−4.38	−1.00
2020-04-09	−3.61	−1.00
2020-04-12	−4.01	−1.00
2020-06-06	−1.63	−0.10
2020-12-03	−0.58	−0.44
2021-01-05	+2.83	−0.82
2021-03-04	+3.46	+0.93
2021-04-01	+2.60	+1.00
2021-04-27	-	+1.00
2021-06-01	0.00	+0.85
2021-07-05	+2.00	+0.59
2021-07-18	−1.73	−0.23
2021-09-01	0.00	−0.60
2021-10-04	+2.85	+0.78
2021-11-04	0.00	−0.25
2021-12-02	−0.38	−0.72
2022-01-04	0.00	+0.62
2022-02-02	−0.41	+0.60

Announcement date	z_e^{sup}	$z_e^{\text{dem,raw}}$
2022-03-02	-2.50	-1.00
2022-03-31	-2.00	-0.80
2022-05-05	0.00	-0.67
2022-06-02	-0.58	-0.12
2022-06-30	0.00	+0.50
2022-08-03	-1.41	-0.83
2022-09-05	+2.00	-1.00
2022-10-05	+4.67	-0.96
2022-12-04	-0.45	-0.88
2023-06-04	-0.82	-0.77
2023-11-30	-2.21	-0.94
2024-03-03	0.00	-0.50
2024-06-02	-2.65	-0.78
2024-09-05	+1.63	-1.00
2024-11-03	+2.83	-1.00
2024-12-05	+1.00	-0.78
2024-12-10	-	-0.48
2025-03-03	-2.65	-0.80
2025-04-03	-1.90	-1.00
2025-05-03	-2.50	-0.94
2025-05-28	0.00	0.00
2025-05-31	+0.50	0.00
2025-07-05	-0.30	-0.14
2025-08-03	-0.41	-0.81
2025-09-07	+1.13	+0.50

Announcement date	z_e^{sup}	$z_e^{\text{dem,raw}}$
2025-10-05	+2.33	−0.20
2025-11-02	+1.33	−1.00
2025-12-01	-	−1.00

Notes: z_e^{sup} is the event-level text supply-news measure defined in Appendix B.1. $z_e^{\text{dem,raw}}$ is the raw content-weighted demand-outlook measure defined in Appendix B.2. A dash denotes an event for which no usable supply-surprise classification is available. A value of zero denotes a defined measure whose directional components net to zero. The monthly supply instrument subsequently applies the timing allocation described in equation (13).

H.3 Full regime first-stage diagnostics

Table 18 expands the first-stage evidence summarized in Section 6. In addition to the heteroskedasticity-robust F -statistic reported in the main text, the table gives the conventional F -statistic, first-stage R^2 , sample size, and the number of nonzero monthly proxy observations.

The standalone regime diagnostics should be distinguished from the overlapping sample-extension exercises in Section 6. When estimated in isolation, the pre-futures and post-2017 periods provide weak first stages. These windows contain only 102 and 96 monthly observations, respectively, compared with 414 observations in the benchmark period.

This does not imply that adding these observations weakens identification in the pooled sample. As shown in Section 6, extending the benchmark backward to 1975 raises the robust first-stage F -statistics from 14.68 to 15.23 for supply news and from 9.06 to 10.59 for demand news. Extending the benchmark forward through 2025 produces robust F -statistics of 11.95 and 13.46, respectively. The complete 1975-2025 specification yields corresponding statistics of 12.88 and 13.43.

Table 18: First-stage diagnostics across OPEC regimes

Instrument (Era)	F	Robust F	R^2 (%)	N	Nonzero months
<i>Panel A: Pre-futures, 1975:1-1983:6</i>					
Text supply	1.06	1.22	1.05	102	37
Raw demand	1.72	2.76	1.69	102	22
Purified demand	1.73	3.41	1.70	102	23
<i>Panel B: Benchmark, 1983:7-2017:12</i>					
Futures supply	20.14	11.84	4.66	414	117
Text supply	16.96	14.68	3.95	414	167
Raw demand	7.19	6.52	1.72	414	111
Purified demand	8.80	9.06	2.09	414	116
<i>Panel C: Post-2017, 2018:1-2025:12</i>					
Futures supply	0.86	0.37	0.91	96	42
Text supply	5.06	1.52	5.10	96	56
Raw demand	7.81	3.63	7.67	96	42
Purified demand	0.53	0.34	0.56	96	42

Notes: Each row regresses the reduced-form innovation in the real price of oil on a constant and the indicated monthly proxy. The reduced-form VAR contains the six baseline variables and twelve lags. The baseline purified-demand measure is used through 2017. The post-2017 row labeled “Extended purified demand” uses the 1975-2025 event-level purification described in Appendix B.2. The pre-futures period contains no futures-based proxy. “Nonzero months” counts monthly observations of the external instrument rather than OPEC events; because the supply instrument allocates late-month announcements partly to the following month, the number of nonzero supply months need not equal the number of announcement dates. These regime diagnostics are based on the baseline reduced-form specification without the March–June 2020 pulse dummies and therefore differ from the COVID-adjusted post-2017 and full-sample diagnostics reported in Table 5.

The regime results therefore make a narrower point: neither short tail period provides strong standalone identification. The longer text record can nevertheless contribute useful identifying variation when those observations are incorporated into the broader sample.

I Data Sources and Replication Details

The monthly macroeconomic data employ the six-variable framework of Känzig (2021) that consists of the real price of oil, world oil production, world oil inventories, world industrial production, U.S. industrial production, and U.S. consumer price index. The real oil price is the refiners’ acquisition cost of imported crude oil adjusted by the U.S. CPI. All six variables

are used in the VAR model in level logarithms. The structural VAR model employs twelve-month-lagged variables and the constant term. The baseline VAR is estimated on the sample of months from 1974:1 to 2017:12, using external instruments from 1983:7 to 2017:12. The extended specification is applied until 2025:12 and employs pulse dummies for March, April, May, and June 2020 in all the reduced-form equations.

The benchmark based on futures is the surprise in OPEC announcements from [Känzig \(2021\)](#). The text measures were constructed using 7,941 articles containing 206 OPEC and OPEC+ announcements from 1971 to 2025 in the *New York Times*, *Wall Street Journal*, and *Washington Post*, including their historical editions. Appendices [A](#) and [B](#) summarize the article classification, event aggregation, month correction, and news filtering procedure.

Estimation is conducted using Python and the included `ivsvar` package and replication code. The impulse responses are presented up to 50 months and normalized such that the shock increases the real oil price by 10 percent. The confidence intervals for moving-block bootstrap are based on 10,000 replications. For specifications with low first stage relevance, we also report weak identification robust MSW confidence sets as in Appendix [C.3](#). The replication files include the code to create the instruments and perform all first stage tests, estimation of the structural specifications, and all tables and figures in the main paper and online appendix.