Oil Shocks: A Textual Analysis Approach*

Deepa D. Datta† Daniel A. Dias‡

August 30, 2020

Abstract

We use textual analysis to gather information on oil market developments contained in news articles from energy-market specialized publications. Relative to existing measures of oil market developments, the indexes of oil market developments we construct contain information on current and prospective oil market developments and can be updated in real time. We show that our indexes correlate with existing measures of oil market developments in the expected way, and that they contain information that can predict future oil price movements, thereby changing the views on what are expected and unexpected oil price movements. In an application, we use our indexes to estimate a monthly and a weekly structural VAR models of the oil market. All the results from these models are in line with economic theory.

JEL classification codes: Q40, Q41, Q43, C19.

Keywords: Textual analysis; Oil supply and demand; Structural VAR; Oil price shocks.

---

*We thank Alice Abboud, Mariya Pominova, Ashley Schwanebeck, and Madeleine Yi for excellent research assistance, and Jessica Liu for her contributions to the initial empirical investigations. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. All errors are our own. Please do not cite or circulate without authors' permission.

†Board of Governors of the Federal Reserve System. Email: deepa.d.datta@frb.gov.

‡Board of Governors of the Federal Reserve System and CEMAPRE. Email: daniel.dias@frb.gov.

1 Introduction

Understanding whether oil prices changes are driven by supply or demand developments is important because it has been shown that other macroeconomic variables respond differently to oil price movements depending on the source of the shock (see Kilian (2009)). For policy making in particular, it is of paramount importance to be able to distinguish the sources of the movement in oil prices in real time, as policy decisions are made. Unfortunately, however, most direct measures of oil supply and demand conditions are produced with a significant lag. Additionally, while price movements are often driven by news about future market conditions, existing oil market measures tend to provide information only on current conditions. To address this gap, we make use of a specialized daily publications whose single purpose is to provide timely news pertaining to the oil market. By applying textual analysis techniques to the newly digitized archive, we show how these publications provide a unique source of information in real time about current and prospective oil market developments.

In this paper, we develop a systematic and automated process to gather information on oil market developments from specialized industry publications – the *Oil Daily* and the *International Oil Daily*, both published by the Energy Intelligence Group (EIG). We categorize words into supply- and demand-related phrases, and we use the publication archive to construct a historical time series of supply and demand indexes containing information about oil market developments.

Our paper contributes to several strands of the literature. We contribute the oil literature by constructing new variables containing information about oil supply and oil demand developments, which, among other things, can be used to study the sources of oil price movements. We also contribute to the broader macroeconomics literature by providing further evidence that important macroeconomic information can be derived from text, and that this information can be used to better understand and measure expected and unexpected movements in macroeconomic variables.

Our work improves upon existing approaches first, by allowing us to provide information on oil price movements in real time, as the news develops. Using this method, there is less need to wait for data on oil production or measures of economic activity to be compiled and published. Secondly, our work quantifies the supply and demand signals available directly in the oil industry press, rather than inferring the relative importance of supply and demand from external markets, such as those for metals or equities. Thirdly, our indexes contain information on current and prospective developments in the oil market, and therefore will allow for a much better understanding of movements in oil prices.

We show that our indexes are highly correlated with existing measures of oil production and global economic activity, both contemporaneously and at a long time horizons. Similarly, we find strong correlation between our indexes and contemporaneous and longer time horizons movements in oil prices. Importantly, these correlations are in line with economic theory. After constructing and validating our indexes, we use them to estimate a structural model of the oil market, which we then use to decompose well-known episodes of large movements of oil prices into supply- and demand-driven.

Relative to the existing literature using textual analysis to study economic phenomena, we expand the use of text-based analysis to the study of oil market developments. Moreover, and unlike many existing papers in this field, our work moves towards extracting quantitative signals from the text by counting phrases, instead of articles (Baker et al. (2016), Caldara and Iacoviello (2018), Husted et al. (2017)). That is, we search the full text of articles for supply and demand phrases, which is more computationally intensive than counting the number of articles mentioning a given topic across a variety of publications.

There are other papers using textual analysis to study oil market developments, but, in most cases, the purpose of these papers is different from ours. Loughran et al. (2018) uses textual analysis to construct an oil market sentiment index. This paper uses the index to study how information gets incorporated into prices. Our paper does some of this at lower frequencies and is more concerned with the effect of oil prices and less with the effect of news on oil prices. Brandt and Gao (2019) use textual analysis to study how news about macroeconomic fundamentals and about geopolitical events affect oil prices differently. In our approach we don't distinguish between purely macroeconomic news and geopolitical events in the construction of our oil demand indexes because we are less interested in this nuance and also because major geopolitical events are relatively rare.

Among the existing literature applying textual analysis to study the oil market, Cavallo and Wu (2012) is the paper that is closest to ours. These authors combine narrative and quantitative approaches to construct measures of oil price shocks. Their narrative approach involves human auditing of *Oil Daily*, *Oil & Gas Journal*, and the *Monthly Energy Chronology.* Using this approach, they attribute daily changes in oil prices to 22 types of oil-related events, e.g. weather changes, oil field discoveries, political and military actions, and changes in actual or expected inventories. Next, they aggregate select event types to generate exogenous oil shocks series, and show substantial effects of these shocks. While this work is very promising, it depends on human auditing and therefore is extremely difficult to replicate and update. By accomplishing similar analysis using automated processing instead of individual judgment to classify articles, we increase replicability and produce a data set that is easily updated over time. Additionally, whereas the Cavallo and Wu classification only provides an indicator variable for whether each type of event is discussed in a given article, our full text searches for the number of phrases in each article allow us to quantify the relative importance of various factors within the news narrative and to provide partial attribution of daily price changes based on these counts.

Our paper is obviously also related to the narrative approach to identifying macroeconomic shocks literature see, for example, Romer and Romer (2004), Romer and Romer (2010), or Cloyne (2013). Two important differences to this literature are the use of textual analysis techniques that allow us to easily replicate and extend our results and the fact that we are not interested in creating a measure of economic shocks. More generally, there have been recent papers showing that certain developments take time to be reflected in macroeconomic variables, but that not accounting for them muddles our view of how certain economic variables develop. For example, Cascaldi-Garcia and Vukoti (0) show that patent-based news shocks have little effect in the short-run but induce strong permanent effects on total factor productivity after five years. Using our indexes, we find strong evidence that the indexes contain important information about future oil supply, demand, and, consequently, price movements, which, if not taken into account in empirical analysis will give rise to a distorted view on expected and unexpected oil price movements.

The rest of the paper is organized as follows: in section 2, we describe the data and how we construct the oil supply and demand indexes; in section 3, we show empirical evidence on the relationship between our oil price shocks indexes and movements in oil production, economic activity, and oil prices; in section 4, we use our supply and demand indexes to construct a structural model of the oil market that can be used to learn about oil price movements; in section 5 we provide some concluding remarks; and, finally, in the appendices we provide additional information on how we constructed our oil supply and demand indexes as well as some additional results.

2 Measuring Oil Supply and Demand Developments

The main contribution of this paper is to use word and phrase counts in oil-related news articles to construct indexes reflecting oil supply and demand developments. In this section, we describe the data source and methodology used to construct the indexes and provide some summary statistics of the indexes themselves.

2.1 *The Oil Daily*

Our corpus includes articles published in two daily publications, the *Oil Daily* (OD) and the *International Oil Daily* (IOD), from the well-regarded Energy Intelligence Group (EIG). EIG is so well regarded for its high-quality coverage of oil market developments that their oil market statistics are used by OPEC as an official secondary source, e.g. for monitoring adherence to production quotas.[1] These publications, which specialize in the oil market, provide us with deeper and broader coverage would be available from the general financial press, including the *Wall Street Journal*, *Financial Times*, Bloomberg, or Thomson Reuters.

As shown in Table 1, aside from the earlier availability of the OD, the two publications are relatively similar along several dimensions.[2] Each publication has about 20 articles per daily issue averaging 300 words per article. As shown in Figure 1, the IOD tends to have, on average, a few more articles than the IOD each day. Importantly, both publications always have one article summarizing the previous day's oil price movements as well as other articles about current and prospective oil market developments. The key difference between the two publications is that since the debut of the IOD, the OD has been more focused on the U.S. oil market, while the IOD has been weighted towards the international (global) oil market. That said, because the oil market is global in nature, there is some overlap between the two publications, which results in some articles being published in both the OD and IOD on the same day. Our supply and demand indexes do not remove the duplicates, as in our view, articles that are published in both publications are likely to be more important and informative than articles that are only published in one of the two publications.

Before implementing our index construction methodology, we clean the data by filtering out non-oil-related articles and reducing typographical and digitization errors. In particular, we remove articles that focus on natural gas or gasoline and correct errors related to punctuation

[1] Although EIG's published oil market statistics are also of high quality, we use only the article text, or narratives, to create the supply and demand indexes.

[2] Although the OD has been published since 1951, our electronic archive begins only in 1992.

as further explained in Appendix B. To give an idea how the cleaned data compare with the original data, we provide in Table 1 the summary statistics for the cleaned data. In the cleaned sample, the average number of articles per day is reduced by about 30 percent, from approximately 20 to 14 articles per issue. The dropped articles include a number of blurbs that are unrelated to oil supply and demand market developments, including announcements of personnel changes at major oil companies, blurbs that very briefly report statistical releases without additional commentary, and brief corrections to earlier articles. In line with dropping this type of brief article, our cleaned sample has a slightly higher average word count of about 350 words per article. Finally, we can see in the figure that the number of articles per issue declines somewhat in the later years of the sample. In the later period, each daily issue tends to have fewer, longer articles, and a total word count that is little changed from the earlier period. Notably, we also find (as discussed below) that the share of words that are related to oil supply and demand remains stable over time.

2.2 Index Construction

Our indexes are constructed to capture the fundamental movements in oil markets: increases and decreases in supply and demand. To accomplish this, we generate broad vocabulary lists of supply and demand, increase and decrease words. Next, we count the number of times supply and demand words appear in proximity to increase and decrease words. These counts are used to generate our directional supply and demand indexes.

Appendix C provides the full vocabulary lists as well as details on how they are constructed. In sum, we have 49 supply words, 34 demand words, 243 increase words, and 298 decrease words. We list the top five supply, demand, increase, and decrease words and their frequencies in Table 2. Notably, there is a large concentration of supply and demand words in just a few expressions – the top five words from our supply and demand word lists account for about 60 and 70 percent of all the supply and demand words. Similarly, the top five words from our increase and decrease words account for about 40 and 30 percent of all the increase and decrease words, respectively, that are used in the text of the articles. Importantly, the five most frequently used supply, demand, increase, and decrease words, shown in Table 2, all seem reasonable and are expected to reflect changes in oil supply and demand conditions.

In Table 3, we provide an example from our article corpus to illustrate this process. This article from the midst of the Global Financial Crisis discusses how weak oil demand conditions are weight on prices despite possible OPEC supply cuts in response. It also discusses potential supply cuts that may boost prices imminently or at some point in the future, pointing to the forward-looking nature of our supply and demand indexes. Using our methodology, this example yields a count of 4 supply words and a count of 3 demand words, and it yields 0 supply increase, 3 supply decrease, 0 demand increase, and 2 demand decrease counts. These counts are well in line with the gist of the article, which is to discuss current weak demand and prospective supply cuts.

Of course, our methodology may capture a fair amount of noise in addition to the information we are able to process. However, the results we obtain for these two examples illustrate how our methodology might capture the main signals from the articles. In the next section, we provide statistical evidence that our methodology does provide useful information.

2.3 Index Properties

Having described the data we use and how we construct our oil supply and demand indexes, we now show the results of our textual analysis algorithms. In Figure 2, the top panel shows the raw counts of supply and demand words at the monthly frequency, while the bottom panel shows the monthly average of the individual article ratios of supply and demand words to total words. Notably, there is significant time variation in the monthly counts, and there are approximately twice as many supply words as demand words in the articles. After accounting for the fact that IOD articles are slightly longer than OD articles, IOD articles have similar shares of supply and demand words as OD articles. On average, the shares of supply and demand words per article hover around 1.1 and 0.5 percent, respectively.

The word counts give some indication of the relative importance of supply relative to demand for the reporting on oil market developments. Ultimately, however, our goal is to create separate indexes for supply and demand developments that can be associated with movements in oil prices. To construct these, we count supply increase, supply decrease, demand increase, and demand decrease phrases for each article. To create each monthly directional index, as shown in Figure 3, we take the ratio of the weekly sum of e.g. supply increase phrases to the weekly sum of total words in the articles.

This figure shows that, in general, there are more increase phrases than decrease phrases for both supply and demand. This may be due to the fact that both oil supply and demand grew considerably over our sample period, or it may be due to the higher salience of increases in the news. As before, we find more supply than demand phrases.

Next, we construct net supply and net demand ratios, such that net supply is the ratio of supply increase minus supply decrease phrases divided by the total word count, net demand is the ratio of demand increase minus demand decrease phrases divided by the total word count. Before using our indexes in statistical models, we transform them by standardization. That is, for each of our four directional indexes and for the net indexes as well, we subtract the full sample mean and divide by the standard deviation of the series. The standardized versions of the net supply and demand indexes are shown in Figure 4. Notably, both measures fall sharply during U.S. recessions (the shaded periods).

3 Evaluating Our Oil Supply and Demand Indexes

In this section we show how our oil supply and demand indexes correlate with existing measures of oil supply and demand and with prices. First, we relate our measures to monthly global oil production and industrial production. Next, we relate our measures to spot and farther-dated futures contracts. For prices, we consider both monthly and weekly oil prices.

3.1 Empirical Approach

To conduct this analysis we estimate the following equations:

$$
\begin{aligned}
y_t - y_{t-1} = \beta_0 &+ \beta_1 SupplyIncrease_t + \beta_2 SupplyDecrease_t \\
&+ \beta_3 DemandIncrease_t + \beta_4 DemandDecrease_t \\
&+ Controls(1, L) + \epsilon_t
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
y_t - y_{t-1} = \beta_0 &+ \beta_1 NetSupply_t + \beta_2 NetDemand_t \\
&+ Controls(1, L) + \epsilon_t
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
y_{t+h} - y_t = \beta_0 &+ \beta_1 SupplyIncrease_t + \beta_2 SupplyDecrease_t \\
&+ \beta_3 DemandIncrease_t + \beta_4 DemandDecrease_t \\
&+ Controls(0, L) + \epsilon_t
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
y_{t+h} - y_t = \beta_0 + \beta_1 NetSupply_t + \beta_2 NetDemand_t \\
+ Controls(0, L) + \epsilon_t
\end{aligned}
\tag{4}
$$

For oil supply, we use the log of total oil production for the world, as reported by the International Energy Agency. For real economic activity, which is an indirect measure of oil demand, we use the log of global industrial production as in Hamilton (2008).

In these equations, the left-hand-side variables are either the logarithm of oil production for the world, as reported by the International Energy Agency, the logarithm of global industrial production, which Hamilton (2008) proposes as a measure of real economic activity (REA), or REA), or the logarithm of WTI oil futures prices (the front month and the 12-month ahead contracts, considered separately). We consider contemporaneous changes in the left-hand-side variable (equations 1 and 2) and h-periods ahead (equations 3 and 4). We estimate these equations at monthly and weekly frequencies – at the weekly frequency we can only use oil prices as left-hand-side variable because oil production and global industrial production are not available at a weekly frequency.

As explanatory variables, we include the directional indexes (equations 1 and 3) or the net supply and demand indexes (equations 2 and 4). In addition, we also include lagged values of all the left-hand-side variables as controls. The exact composition of these controls and the number of lags depends on the time horizon of the regression (contemporaneous difference and h-periods-ahead difference) and the frequency of the data (monthly and weekly). To be more specific, the *Controls(j,l)* function is defined as:

$$
\begin{aligned}
Controls(j,l) = &\sum_{i=j}^{l} \rho_i^{Prod} ln(OilProd_{t-i}) + \sum_{i=j}^{l} \rho_i^{REA} ln(REA_{t-i}) \\
&+ \sum_{i=j}^{l} \rho_i^{WTI^{Sport}} ln(WTI_{t-i}^{Spot}) + \sum_{i=j}^{l} \rho_i^{WTI^{12-month}} ln(WTI_{t-i}^{12-month})
\end{aligned} \tag{5}
$$

For the weekly regressions, we can only control for oil prices because oil production and global industrial production are only available at a monthly frequency. In the monthly and weekly regressions we use 1-year equivalent lags ($L = 12$ in the monthly regressions and $L = 52$ in the weekly regressions).

Equations 1 and 2 and equations 3 and 4 serves a different purpose. With equations 1 and 2 we want to see whether our indexes contain any information on oil production, REA, and oil prices besides what could be predicted by the past values of these variables (oil production, REA, and oil prices). With equations 3 and 4 we want to see whether our indexes contain any additional information about future movements of the four left-hand-side variables in addition to what could be predicted by the current and previous values of these variables (oil production, REA, and oil prices). Additionally, we want to see whether our indexes correlate with the left-hand side variables in the expected way. The estimation results are shown in Tables 4 (monthly data) and 4 (weekly data). Note that, for ease of comparison, we have annualized all changes before using them in the regressions. This way, the results are comparable across equations and tables.

In the next subsections we discuss the results separately for oil production and oil demand and for oil prices.

3.2 Our Indexes and Other Measures of Oil Supply and Demand

Table 4 shows that our indexes capture directional information about movements in supply and demand. As our indexes are standardized, the results in the first column of the table show that a one standard deviation increase in the supply increase index is associated with an increase in oil production at about a 3.1 percentage points annual rate. Similarly, a one standard deviation increase in the supply decrease index is associated with a decrease in oil production that month, at a 3.4 percentage points annual rate. In both cases, the effect of on oil production is statistically significant. In contrast, neither the demand increase or demand decrease indexes have a statistically significant effect on oil production. In the next column we show that our net supply index is also strongly positively associated with contemporaneous oil production, but the net demand index is not. A one standard deviation increase in the net supply index is associated with an increase in oil production that month, at a 2.8 percent annual rate.

We also find a strong and statistically significant contemporaneous relationship with real economic activity, in that a one standard deviation increase in the demand increase index is associated with a 1.2 percentage points higher growth rate in industrial production, and a one standard deviation increase in the demand decrease index is associated with a 1.8 percentage points lower growth rate in industrial production. These results also hold for the net demand index, with the effect of a one standard deviation increase in the net demand increase index increasing industrial production growth rate by 1 percentage point. For REA, we find that the supply decrease index has statistically significant effect of negative 1.4 percentage points on the growth rate of industrial production. One possible explanation is that oil production is itself part of global industrial production, and therefore a decline in oil supply would negatively affect global industrial production. The effect of the supply increase index on REA is also positive, but not statistically significant. In the case of the net-supply index, the effect is also positive, but, as for the supply increase index, it is not statistically significant.

These relationships also hold when relating the indexes to the 12-month-ahead change in oil production and economic activity, as shown in the bottom part of of the table. For oil production, we find that the 12-month-ahead coefficients are smaller in magnitude and somewhat less precisely estimated than for the contemporaneous coefficients in the earlier regressions. By contrast, the estimated relationships between the indexes and real economic activity remain fairly strong. A one standard deviation increase in the demand increase or decrease index in a given month is associated with a 1 percentage point increase and a 1.3 percentage points decrease in real economic activity over the following 12 months, respectively.

We also included in Table 4 an estimate of the contribution of the indexes for the goodness-of-fit (measured by the R-square) of the equation (see row "Indexes R^2 share" on the table).[3] For oil production, in the contemporaneous regression, the four directional indexes and the two net indexes account for 16% and 14%, respectively of each equation's R-square. In the 12-month ahead equations, the four directional indexes account for 17% of the regression's R-square and the two net indexes account for 9%. For REA, the four directional indexes account for 21% and 23% of the R-square of the contemporaneous and 12-month ahead regressions, respectively. The two net indexes account for 14% and 16% of the R-square of the contemporaneous and 12-month ahead regressions, respectively.

The various results from Table 4 show that the indexes we constructed contain information about the developments in the oil market, both contemporaneously and in the future, and that the information is in general consistent with each of the indexes (that is, the supply indexes affect oil production as expected and the demand indexes affect REA as expected).

[3] To estimate the effect of the indexes on the R-square of each equation, we use the Shapley-Owen's R-square decomposition proposed by Huettner et al. (2012). This approach considers all possible combinations of variables or group of variables in the regression to estimate the contribution of each variable or groups of variables to the total R-square of the equation.

3.3 Our Indexes and Oil Prices

Having documented the close relationship between our indexes and oil production and real economic activity, we now turn to establish the relationship between our indexes of supply and demand and oil spot and 12-month futures prices. We first consider regressions at the monthly frequency before moving on to consider the weekly frequency. As with the oil production and economic activity regressions, we consider the 1-month contemporaneous change in the dependent variable, as well as the 12-month-ahead change. As before, we have annualized all dependent variables before using them in the regressions for ease of comparison. The results are shown in Tables 4, monthly regressions, and 5, weekly regressions.

In Table 4, the results in the fifth column of the table show that a one standard deviation increase in the supply increase index is associated with a contemporaneous decline in the spot oil price of 14.5 percent at an annual rate. Similarly, a one standard deviation increase in the supply decrease index is associated with a decrease in the spot oil price that month of 7.2 percent at an annual rate. The effect of the supply increase index is statistically significant but the effect of the supply decrease index is not. However, the signs of the estimated coefficients are in line with what would be expected, oil prices decrease after an increase in supply while they decrease after a decrease in supply. In the next column we show that our net supply index is also strongly negatively associated with contemporaneous oil production. A one standard deviation increase in the net supply index is associated with a decrease in oil prices that month, at a 10.7 percent annual rate.

The demand indexes also show strong relationship with the 1-month oil price changes. For example, a one standard deviation increase in the demand increase is associated with a contemporaneous 24.9 percent increase of the spot price of oil. For the demand decrease index, we estimate that a one standard deviation increase in the index is associated with a 19.4 percent decrease of the spot price of oil. The results for the net supply and net demand indexes are similar to those observed for the four directional indexes, with oil prices increasing by 22.4 percent at annual rate in response to a one standard deviation of the net demand index. All demand-related coefficients are statistically significant.

Turning to the results for the 12-month futures price, we find similar results with somewhat smaller magnitudes, in line with the lower volatility observed for 12-month futures prices. While all demand-related coefficients are statistically significant, none of the supply-related coefficients are. However, once again, both supply- and demand-related coefficients have the expected signs.

These relationships also hold when relating the indexes to the 12-month-ahead change in oil spot and futures prices, as shown in the last 4 columns of the table. In general, we find that a higher supply decrease and demand increase indexes are associated with prices rising over the next 12 months. Similarly, prices fall in the 12 months following an increase in the supply increase and supply decrease indexes. For the demand indexes, we find that the 12-month-ahead coefficients are around 7 to 10 percent as compared to 15 to 25 percent for the contemporaneous effects. By contrast, the estimated relationships between the supply indexes and 12-month-ahead spot and futures prices remain large in magnitude, and now all these relationships are statistically significant. A one standard deviation increase in the net supply index is associated with a 10 percent (a.r.) decline in the contemporaneous futures price, and a nearly 15 percent decline in the futures price over the following 12 months.

In comparison to the effects of our indexes on oil production and REA, the effects of the indexes on oil prices are much larger. One possible reason is that our indexes contain not only information about current oil production and current REA but also about future oil production and future REA, and that additional information translates into larger effects on oil prices.

Once again, the indexes contribute significantly to the goodness-of-fit of the regressions, especially in the case of the 12-month-ahead regressions. For the contemporaneous regressions, the indexes contribution to the R-square varies between 8.5 and 16.5 percent. For the 12-

month-ahead the contribution of the indexes to the R-square of the regressions is estimated to be close to 50 percent. This is an astonishing result, which suggests that the information contained in our indexes will have very important for determining what are expected and unexpected price movements. The very large increase in the R-square of the 12-month-ahead regression due to the indexes indicates that oil price movements are more predictable than what movements in oil prices, production, and real economic activity would suggest.

Table 5 provides the results for the regressions at the weekly frequency. Here, we consider the contemporaneous change in the spot or futures price in the first two columns. Subsequent columns examine the relationship between the indexes and the price change observed over the following 12, 26, and 52 weeks. In general, we find strong associations between the indexes and the contemporaneous spot and futures price movements. For example, a one standard deviation increase in the net supply index is associated with a 22 percent decline in spot prices and a 12 percent decline in 12-month futures prices. When looking at results for the following weeks, we find weaker results at the 12- and 26-week horizon, and somewhat stronger results again at the 52-week horizon.

While for the contemporaneous, 12-week-ahead, and 26-week-ahead regressions the contribution of the indexes is more modest than in the case of the monthly regressions, for the 52-week-ahead regressions the contribution of the indexes is even higher than in the case of the monthly regressions.

Overall, the results in this section show that our indexes contain important information about developments in the oil market and that this information may be used to analyze developments in the oil market.

4 Using the Supply and Demand Indexes to Learn About Oil Price Movements

In this section we propose a structural model of the oil market that utilizes the high information content of the oil supply and demand indexes by comparing the reduced form shocks that are implied by the VAR model underlying the structural VAR model we propose to a standard three-variable VAR model of the oil market. Next, we present estimates of the short-run elasticities, the impulse-response functions, and the forecast error variance decomposition of the SVAR model introduced. And last, we use the SVAR model to decompose oil price movements into demand and supply driven at monthly and weekly frequencies.[4]

4.1 A Structural Model of the Oil Market using Supply and Demand Indexes

The workhorse model of the oil market has three main components: 1) information on supply; 2) information on demand; and 3) information on the price of oil. These three components are often represented by just three variables, including monthly global oil production, a measure of global real economic activity, and an oil price, as in Kilian (2009) and subsequent research. In other cases, these three components are represented by more than three variables. For example, Kilian and Murphy (2014) use oil inventories as an additional source of information on oil supply and demand conditions, while Caldara et al. (2018) use metals prices and separate measures of real economic activity for advanced and emerging economies to better model the demand side of the oil market.

Our approach follows a similar structure, in that we use the supply increase and supply decrease indexes as the sources of information for oil supply conditions and the demand

[4]Note that, there is nothing but our preference preventing us from estimating a daily SVAR model because the indexes can be constructed at a daily frequency.

increase and demand decrease indexes as the sources of information for oil demand conditions. Notably, we use the individual directional indexes in this model, based on our earlier results showing they have higher informational content than the net supply and net demand indexes. As in the rest of the literature, we combine these measures of supply and demand with observed oil prices for a more complete picture of the oil market.

One key advantage of using our indexes to study the oil market is that, unlike the typical measures of monthly oil production and real economic activity, our supply and demand measures contain both current and prospective information. That is, we don't just have information on flow oil supply and flow oil demand, we also have information on *expected future* oil supply and demand. This prospective information can be especially valuable as oil prices are determined by both current and expected supply and demand conditions.

We present a model with just three variables as an example to help guide our discussion, and abstracting from the model dynamics, we have that each variable is a function of three structural shocks:

$$\begin{cases} Q_t = h_1(\epsilon_t^S, \epsilon_t^D, \epsilon_t^{OD}) \\ REA_t = h_2(\epsilon_t^S, \epsilon_t^D, \epsilon_t^{OD}) \\ P_t = h_3(\epsilon_t^S, \epsilon_t^D, \epsilon_t^{OD}) \end{cases} \tag{6}$$

In this model, our variables for monthly global oil production (Q_t), global real economic activity (REA_t), and the price of oil (P_t) are each affected by the same three structural shocks: a flow supply shock (ϵ_t^S), a flow demand shock (ϵ_t^D), and precautionary demand shock (ϵ_t^{PD}).

In this setting, the precautionary demand shock corresponds to the oil price movements that cannot be explained by shocks to *current* or *flow* supply or demand. In particular, many developments have little bearing on current oil supply or oil demand for immediate use, but they can still affect oil prices immediately via expectations of future oil market conditions.

For example, oil prices often move in response to new economic projections published by organizations including the International Monetary Fund or the International Energy Agency. These price movements reflect a change in expected future oil demand instead of a change in flow oil demand. Similarly, oil prices often move sharply after OPEC announcements about changes in future supply targets. Again, these oil price movements reflect changing expected future supply conditions, and not changes in the current flow supply or flow demand for oil. In these two examples, as long as oil prices are moving in response to developments not captured by our measures of supply and global real economic activity, the oil model described in equation (6) would attribute the price movements to precautionary demand shocks. Notably, price movements could be driven by changes in expected future supply or expected future demand. Yet, the model in equation (6) would bundle them together into the precautionary demand shock.

To fully decompose oil price changes into supply- and demand-driven components, we must unpack the oil-specific demand shock. The existing literature shows this decomposition is important because GDP and inflation are affected differently by supply-driven and demand-driven oil price changes, and a better understanding of the decomposition can help determine the optimal monetary policy response.

Our model incorporates 5 structural shocks: positive supply developments shock, $\epsilon_t^{S^+}$, negative supply developments shock, $\epsilon_t^{S^-}$, positive demand developments shock, $\epsilon_t^{D^+}$, negative demand developments shock, $\epsilon_t^{D^-}$, and a non-fundamentals price shock, ϵ_t^{NFP}. Separating the positive and negative shocks instead of using a measure that nets them out has several advantages. In particular, the positive and negative supply shocks may have different dynamic effects, and the additional degrees of freedom can allow for better identification of these effects. For example, suppose there is news of both an increase in OPEC production targets and a pipeline disruption on the same day. While that day's net price effect may be zero, the dynamic effects of these two shocks may be very different than that of a day with no shocks. More generally, it is useful to think of the oil market as a combination of many players and participants, which are constantly being hit by both oil supply and demand shocks.

The fifth shock, which we call the non-fundamentals price shock, captures oil price movements that are not explained by oil supply and demand developments, as measured by our supply and demand indexes. If our indexes capture all relevant oil supply and demand developments, then any price movements that are not explained by the indexes must be due to non-fundamental factors, such as risk aversion.

In the model, we assume that, contemporaneously, our indexes only interact with each other through the price of oil. However, the price of oil responds to all four indexes contemporaneously. Abstracting from the dynamic effects, the model we propose can be written as follows:

$$\begin{cases} SI_t = \alpha_P^{SI} P_t + \epsilon_t^{S^+} \\ SD_t = \alpha_P^{SD} P_t + \epsilon_t^{S^-} \\ DI_t = \alpha_P^{DI} P_t + \epsilon_t^{D^+} \\ DD_t = \alpha_P^{DD} P_t + \epsilon_t^{D^-} \\ P_t = \gamma_{SI} SI_t + \gamma_{SD} SD_t + \gamma_{DI} DI_t + \gamma_{DD} DD_t + \epsilon_t^{NFP} \end{cases} \tag{7}$$

Note that if we assume the supply increase and supply decrease indexes have a symmetric effect on oil prices ($\gamma_{SI} = -\gamma_{SD} = \gamma_S$), and the demand increase and demand decrease indexes also have a symmetric effect on oil prices ($\gamma_{DI} = -\gamma_{DD} = \gamma_D$), we can rearrange the

5-variable model from equation 7 into a 3-variable one:

$$
\begin{cases}
SI_t - SD_t = (\alpha_P^{SI} - \alpha_P^{SD})P_t + (\epsilon_t^{S^+} - \epsilon_t^{S^-}) \\
DI_t - DD_t = (\alpha_P^{DI} - \alpha_P^{DD})P_t + (\epsilon_t^{D^+} - \epsilon_t^{D^-}) \\
P_t = \gamma_S(SI_t - SD_t) + \gamma_D(DI_t - DD_t) + \epsilon_t^{NFP}
\end{cases} \tag{8}
$$

While this model looks more like the one in equation 6, a key difference is that in this model, the supply and demand variables contain information about current and prospective supply and demand conditions, respectively, rather than just incorporating data on flow-supply and flow-demand.

4.2 Reduced Form Oil Price Shocks

In this section, we examine the reduced form shocks that are obtained using the oil price equation in models similar to those presented in the previous section. In particular, we estimate the following models and then compare the residuals that are obtained from each one of them:

$$
Baseline \; : \; \Delta P_t = \alpha + \sum_{i=1}^{L} (\rho_i^P \Delta P_{t-i} + \rho_i^S \Delta S_{t-i} + \rho_i^{REA} \Delta REA_{t-i}) + \omega_t^P \tag{9}
$$

$$
\begin{aligned}
Augmented : \Delta P_t = \alpha + \sum_{i=1}^{L} (\rho_i^P \Delta P_{t-i} + \rho_i^S \Delta S_{t-i} + \rho_i^{REA} \Delta REA_{t-i} + \rho_i^{SI} SI_{t-i} + \\
\rho_i^{SD} SD_{t-i} + \rho_i^{DI} DI_{t-i} + \rho_i^{DD} DD_{t-i}) + \omega_t^P
\end{aligned} \tag{10}
$$

$$Indexes : \Delta P_t = \alpha + \sum_{i=1}^{L} (\rho_i^P \Delta P_{t-i} + \rho_i^{SI} SI_{t-i} + \rho_i^{SD} SD_{t-i} + \rho_i^{DI} DI_{t-i} + \rho_i^{DD} DD_{t-i}) + \omega_t^P \tag{11}$$

The first equation corresponds to the price equation in a standard 3-variable VAR model of the oil market (as in 6); the second equation adds our indexes to the first equation; and the third equation corresponds to the price equation of the 5-variable VAR model of equation (7). To compare across models, we estimate the the three equation to obtain an estimate of the residuals from each of the these equations. Next, we compute the standard deviation of the residuals as well as the root mean square error (RMSE) of the regression. To compare across models, we compute the deviation of the results for the Augmented and Indexes models relative to those of the Baseline model. The results of this exercise are shown in Table 6.

In looking at the results, we first note that for the Augmented model, all the entries are negative, suggesting that the inclusion of the indexes reduces the pricing errors relative to the Baseline model. Of course, because adding more variables to a regression model always improves the fit, we also provide the RMSEs for additional evidence. These statistics are adjusted for the the number of variables in the regression model, and still show a reduction in the errors for the Augmented model, relative to the Baseline. Next, looking at the Indexes model, we also see negative values for both the standard error and the RMSEs. This comparison tells us that, even if there is some loss information from using indexes to measure flow oil supply and flow oil demand, there must be also some informational gain from the indexes containing information about prospective oil supply and prospective economic activity.

The results in Table 6 provide additional evidence that the oil market indexes we construct in this paper have relevant oil market information, which goes beyond the information contained in measures of flow oil supply and current economic activity.

4.3 Structural VAR Results

Following the discussion of the SVAR model in the previous sub-section, we now discuss the estimation results of the short-run dynamics in the model, forecast error-variance decomposition of the structural shocks, the impulse response functions, and historical decomposition of oil price movement during well-defined economic episodes.

4.3.1 Short-Run Dynamics

In Table (7) we show the estimates of the short-run relationships between each of the five variables included in the monthly (Panel A) and weekly (Panel B) models. For both models, the estimated relationships are in line with economic theory. While the supply increase (column 1) and the demand decrease (column 4) indexes respond positively to oil price increases, the supply decrease (column 2) and the demand increase (column 3) indexes respond negatively. In both models, the oil price (column 5) responds negatively to the supply increase and demand decrease indexes, and responds positively to the supply decrease and demand increase indexes. With the exception of the response of the supply decrease index in the weekly model, all estimates are statistically significant.

While the results for the monthly and weekly models are qualitatively very similar, there are some noteworthy quantitative differences. In particular, while prices respond similarly to supply and demand indexes in the monthly model, we find that in the weekly model, the response of prices to demand indexes is about double of that for the supply indexes. Similarly, while in the monthly model the responsiveness of supply and demand indexes to oil price changes is about the same magnitude for all four indexes, in the weekly model the response of demand indexes to oil price changes is much larger than that of supply indexes.

These differences suggest that at higher frequencies of data observation, oil price changes have more impact on demand than on supply, and demand shocks are more important than supply shocks for oil price fluctuations. This result is consistent with the idea that oil supply is constrained in the short run, but not as much at longer time horizons. At the same time, oil demand appears to be very responsive to oil price changes. Regarding the differences in the response of the oil price, we conjecture that supply shocks at a monthly frequency may be more persistent than at a weekly frequency, which could explain why prices react more to supply shocks at lower frequencies (monthly model) than at higher frequencies (weekly model).

4.3.2 Oil Price Forecast Error Variance Decomposition

Table (8) shows the forecast error variance decomposition of oil prices in the short run ($t = 1$) and in the long run ($t = \infty$) in response to the five structural shocks in the model. To help with the discussion we added one column showing the sum of the importance of supply increase and supply decrease shocks (S) and another column showing the sum of the importance of demand increase and demand decrease shocks (D). For the monthly model, the supply, demand, and non-fundamentals price shocks are approximately equally important, though the demand shock is the most important by a small margin. Interestingly, in the monthly model, especially in the long run, supply increase and supply decrease have about the same importance and the same is true for demand increase and demand decrease shocks.

In the weekly model, as expected given the results in Table 7, supply shocks are much less important than demand shocks in both the short and long run. In the short run, supply shocks account for only about 5 percent of oil price fluctuations while demand shocks account for close to 60 percent of oil price fluctuations. In the long run these differences abate, but supply shocks still only account for less than 20 percent of oil price fluctuations while demand shocks account for nearly 50 percent of oil price fluctuations.

In both models, the importance of non-fundamentals price shocks for oil price fluctuations is around 35 percent in the short run, and around 30 percent in the long run. The main difference between the two models relates to the importance of supply and demand shocks at the two different frequencies. We note that the relative importance of supply and demand shocks for oil price fluctuations is still an open question (see, for example, Kilian (2009), Caldara et al. (2018), and Baumeister and Hamilton (2019)). Our results suggest that the time horizon of the oil price fluctuations may also play an important role in this discussion.

While high-frequency oil price movements, such as weekly oil price movements, are likely to be more driven by demand shocks, lower frequency oil price movements are more likely to be equally driven by supply and demand shocks. These results are consistent with the idea that supply shocks are more likely to be temporary than demand shocks. For example, at a weekly frequency we may observe temporary production disruptions, which may be netted out by offsetting production resumptions in later weeks but within the same month. By contrast, the demand shocks observed at the weekly frequency may be considered to be more permanent and may therefore have a larger effect. At the monthly frequency, temporary supply shocks may net out, and the surviving monthly shocks therefore have a larger effect.

4.3.3 Structural Impulse Response Functions

Figure 5 shows the impulse response function of oil prices to each of the five structural shocks based on the monthly (panel a) and weekly (panel b) models.[5] Despite some small differences between the two sets of impulse response functions, the response of oil price to the supply and demand shocks is very similar in both models. The price of oil declines in response to either a supply increase or a demand increase shock, while the price of oil increases in response to either a supply decrease or a demand decrease shock. With respect to the non-fundamentals price shock, the price of oil price first increases, then decreases, and then converges to near zero for both models. The fact that the the non-fundamentals price shock has a non-permanent effect on the price of oil is, in our view, consistent with

[5]Figures D1 and D2 in the Appendix show all the structural impulse response functions associated with the monthly and weekly SVAR models, respectively.

our interpretation of this shock. After controlling for the most relevant supply and demand developments, oil prices may move in response to some non-fundamental developments, but these movements should be short-lived.

Turning now to the impulse responses of the supply and demand indexes to a non-fundamentals price shock, shown in Figure 6. Once again, the responses are very similar in both the monthly and weekly models, despite some quantitative differences. Both models have that the supply increase and the demand decrease indexes increase in response to a positive non-fundamentals price shock, while the supply decrease and the demand increase indexes decrease in response to the same shock.

All in all, the impulse response functions based on the SVAR model presented in equation (7) accord with standard economic theory. These results provide further evidence in favor of the oil supply and demand indexes that we construct.

4.3.4 Historical Decomposition of Oil Price Movements

Using our structural model of the oil market represented by equation (7), we now turn to historical decompositions that separate oil price movements into their supply- and demand-driven components. This is particularly useful for guiding monetary policy responses, as monetary authorities are more likely to take a signal and adjust the monetary policy stance in reaction to rising oil prices when the oil price increases are demand-driven rather than supply-driven. The rationale is that the demand pressures that are pushing oil prices higher are also likely to push prices of other goods higher over time. In contrast, if oil prices are rising due to supply-side developments, there is much less concern about overall inflation, as other prices don't have the same reasons to move up.

Relative to other structural models of the oil market that have been proposed in the literature, our model has the advantage of also allowing us to distinguish oil price movements that are due to oil market forces from those that are not due to non-oil market forces. That is, if oil prices are not moving due to supply or demand developments, then it must be due to what we call non-fundamental reasons. For example, as we already discussed, oil prices may be moving due to changes in investors' risk aversion or due to speculative factors. This information is also useful, because, as shown in Figure 5, these shocks are short lived and therefore it is just a matter of time until they are reversed.

To demonstrate the value of the historical decomposition of oil price changes obtained from our model, we next analyze two episodes with very large oil price changes: 1) the 2007-2009 Great Recession and 2) the 2014 oil price collapse.

The 2007-2009 Great Recession As the Great Recession was the most severe global recession since the 1929 Great Depression, it is perhaps unsurprising that it was also a time of an extraordinary oil price collapse. On net, oil prices declined nearly 25 percent, from about $90 per barrel at the start of the recession in December 2007 to $70 per barrel at the end of the recession in June 2009. Even more astounding is that this net price decline masks even greater oil price volatility. Prices rose sharply through the start of the recession before beginning their decline in June 2008. Over the next six months, oil prices collapsed by almost 70%, from around $140 per barrel at the end of June 2008 to close to $45 per barrel at the end of December 2008.

Following Burbidge and Harrison (1985), we use the SVAR model to decompose the the observed oil price change into the contribution of each of the five structural shocks. The main results can be seen in the top left panel of Figure 7. Additionally, to summarize the net effects of the positive and negative shocks, the panels on the right combine the contributions of the two supply and two demand shocks.

Notably, as shown by our impulse response analysis, shocks at time t can have an effect long afterwards. Also following Burbidge and Harrison (1985), we decompose price changes into those driven by shocks before the end of June 2008 and those driven by shocks in July 2008 and later. That is, when considering the drop in oil prices between June 2008 and December 2008, we can decompose the price change into the effects of shocks that occurred prior to the start of the episode – in this case all the shocks that occurred before the end of June 2008 – and all the shocks that occurred after the start of the episode – in this case, all the shocks that occurred in July 2008 and after. We consider price changes driven by shocks from before July 2008 to be expected price changes, and provide a separate decomposition for these shocks in the middle panels of Figure 7. Price changes that are due to shocks that occurred during or after July 2008 are unexpected, and the decomposition of these changes are shown in the bottom panels of Figure 7.

According to the results of the monthly and the weekly models, the collapse of oil prices between June and December 2008 was mostly due to demand and non-fundamentals price shocks. A little more than half of the price collapse was unexpected, with demand shocks driving the prices down after June 2008. For the expected price declines after June 2008 shown in the middle left panel, we find that non-fundamentals price shocks are the main driver.

To further investigate these results, we show in Figure 8 the price decomposition for the period November 2007 to June 2008. We find that oil prices increased about 40%, driven by both demand shocks and non-fundamentals price shocks. In particular, we find a large role for non-fundamentals price shocks pushing prices higher before July 2008. The impulse response function of oil prices shows that, in response to a non-fundamentals price shock, oil prices increases at first and decline later (see Figure 5), with the long-term effect on prices being basically neutral. This is exactly what we see happening in this episode: between November 2007 and June 2008, oil prices increased mostly due non-fundamental reasons (that is, the supply and demand developments did not justify this large increase in oil prices), but this price increase reversed and contributed to later price declines between June and December 2008.

The 2014 Oil Price Collapse Starting in mid–2014, oil prices plunged by more than 70 percent, from about $100 per barrel at the end of June 2014 to less than $30 per barrel in February 2016. In Figure 9 we show the historical decomposition of this oil price change based on the monthly and weekly SVAR models.

As before, the left hand-side charts show the results for the five structural shocks while the right hand-side charts show the results after combining the positive and negative shocks to supply and demand. Our model shows that supply was a major reason for the price decline, with a significant but slightly smaller role for demand. Interestingly, both models show that supply decrease shocks pushed oil prices up a bit during this period. That is, without some supply decrease shocks, oil prices would have declined even more.

Separating between expected and unexpected shocks (the middle and bottom panels, respectively), shows that a significant part of the oil price decline was expected prior to July 2014. In the middle panels, for both models, it can be seen that supply shocks that occurred before July 2014 were still playing out and pushing oil prices down after July 2014. The middle panels also show that the non-fundamentals price shocks that hit the oil market before July 2014 were now contributing to the oil price decline. These negative effects of the non-fundamentals price shock are due to the unravelling of positive non-fundamentals price shocks that helped maintain the oil price higher prior to July 2014. Figure 10 shows the historical decomposition of oil price changes in the year before the start of the oil price collapse in mid-2014. The top left and top right panels show that oil prices were mostly unchanged between June 2013 and June 2014, but that is the result of supply shocks pushing oil prices down and non-fundamentals price shocks pushing oil prices up.

Finally, the bottom panels of Figure 9 shows that a mix of unexpected supply and demand shocks contributed to pushing oil prices down. The monthly model attributes less importance to demand shocks and more to non-fundamentals price shocks, while the weekly model attributes more importance to demand shocks but less to non-fundamentals price shocks. Both models attribute about equal importance to supply shocks. Once again, supply negative shocks contributed to increasing oil prices, but the effect of supply increase shocks dominated.

To summarize, the mid-2014 oil price collapse was due to a combination of supply, demand, and non-fundamentals price shocks. Most demand shocks occurred after June 2014, most non-fundamentals price shocks occurred on or before June 2014, and supply shocks occurred before, on, and after June 2014.

5 Concluding Remarks

In this paper, we construct several indexes tracking developments in the oil market based on textual analysis of two highly regarded daily energy-related publications. We show that our indexes, which separate developments that are expected to increase oil supply or demand from those that are expected to decrease oil supply or demand, are correlated in the expected way with observed oil production changes, changes in real economic activity conditions, and oil price changes. We also show that our indexes contain information regarding current and prospective oil market developments, a feature that is unique relative to most of the existing measures oil market developments. Moreover, our findings suggest that important information about the economy can be obtained from specialized industry publications. Indeed, we found that our indexes had significant predictive power of oil prices changes at longer horizons. By not accounting for this information the economist may consider certain developments in the oil market, or in other markets, as unexpected while these developments could be predicted given past developments. Other key characteristics of our indexes are that they can be updated on a daily basis and that these measures can be constructed for different time frequencies.

In an application, we use our indexes to estimate monthly and weekly structural VAR models of the oil market. For both models, the estimated relationships between all the variables included in the model and the structural shocks are in line with economic theory. When using these structural models to decompose large oil price movements around the 2008-09 Great Recession and the 2014 oil price collapse, we find that a large portion of these oil price movements was expected prior to the start of the episodes. For both episodes, the resulting historical decomposition of oil price movements during these episodes are plausibly in line with what one would expect to happen to oil around price around an economic recession and after a large increase in oil production.

In future work we plan to use our oil supply and demand indexes to estimate oil production and real economic activity in real time. We also would like to use our oil supply and demand indexes to forecast oil prices and compare those projections against other alternative forecasts, such as those based on oil futures.

References

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4):1593–1636.

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

Brandt, M. W. and Gao, L. (2019). Macro fundamentals or geopolitical events? a textual analysis of news events for crude oil. *Journal of Empirical Finance*, 51:64 – 94.

Burbidge, J. and Harrison, A. (1985). An historical decomposition of the great depression to determine the role of money. *Journal of Monetary Economics*, 16(1):45–54.

Caldara, D., Cavallo, M., and Iacoviello, M. (2018). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*.

Caldara, D. and Iacoviello, M. (2018). *Measuring geopolitical risk*. International Finance Discussion Papers 1222.

Cascaldi-Garcia, D. and Vukoti, M. (0). Patent-based news shocks. *The Review of Economics and Statistics*, 0(ja):1–45.

Cavallo, M. and Wu, T. (2012). *Measuring oil-price shocks using market-based information*. International Monetary Fund Working Paper No. 12/19.

Cloyne, J. (2013). Discretionary tax changes and the macroeconomy: new narrative evidence from the united kingdom. *American Economic Review*, 103(4):1507–28.

Hamilton, J. D. (2008). Daily monetary policy shocks and new home sales. *Journal of Monetary Economics*, 55(7):1171–1190.

Huettner, F., Sunder, M., et al. (2012). Axiomatic arguments for decomposing goodness of fit according to shapley and owen values. *Electronic Journal of Statistics*, 6:1239–1250.

Husted, L., Rogers, J. H., and Sun, B. (2017). *Monetary policy uncertainty*. FRB International Finance Discussion Paper 1215.

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):1053–69.

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.

Loughran, T., McDonald, B., and Pragidis, I. (2018). Assimilation of oil news into prices. *Available at SSRN 3074808*.

Romer, C. D. and Romer, D. H. (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review*, 94(4):1055–1084.

Romer, C. D. and Romer, D. H. (2010). The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. *American Economic Review*, 100(3):763–801.

A Figures and Tables

Figure 1: Monthly average of articles per issue in the "Oil Daily" and "International Oil Daily" publications, before and after removing articles unrelated to the oil market.

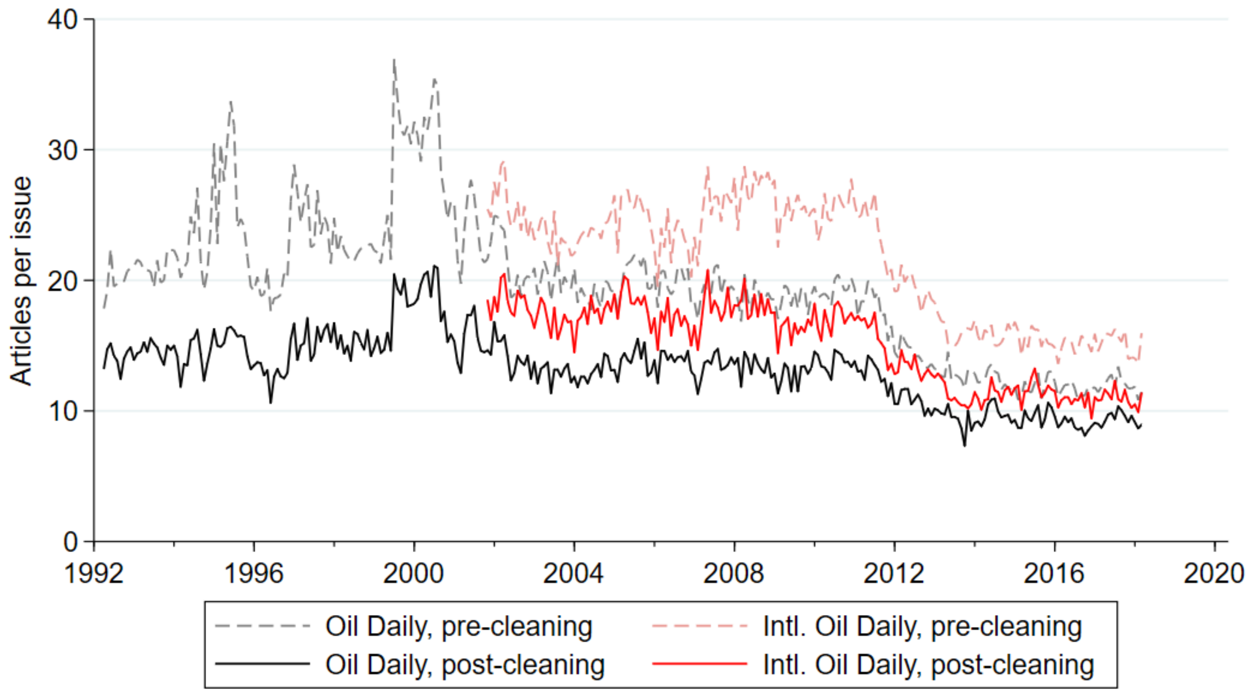


Figure 2: Supply and demand counts per month in the "Oil Daily" and "International Oil Daily" publications - raw word counts and as share of total number of words

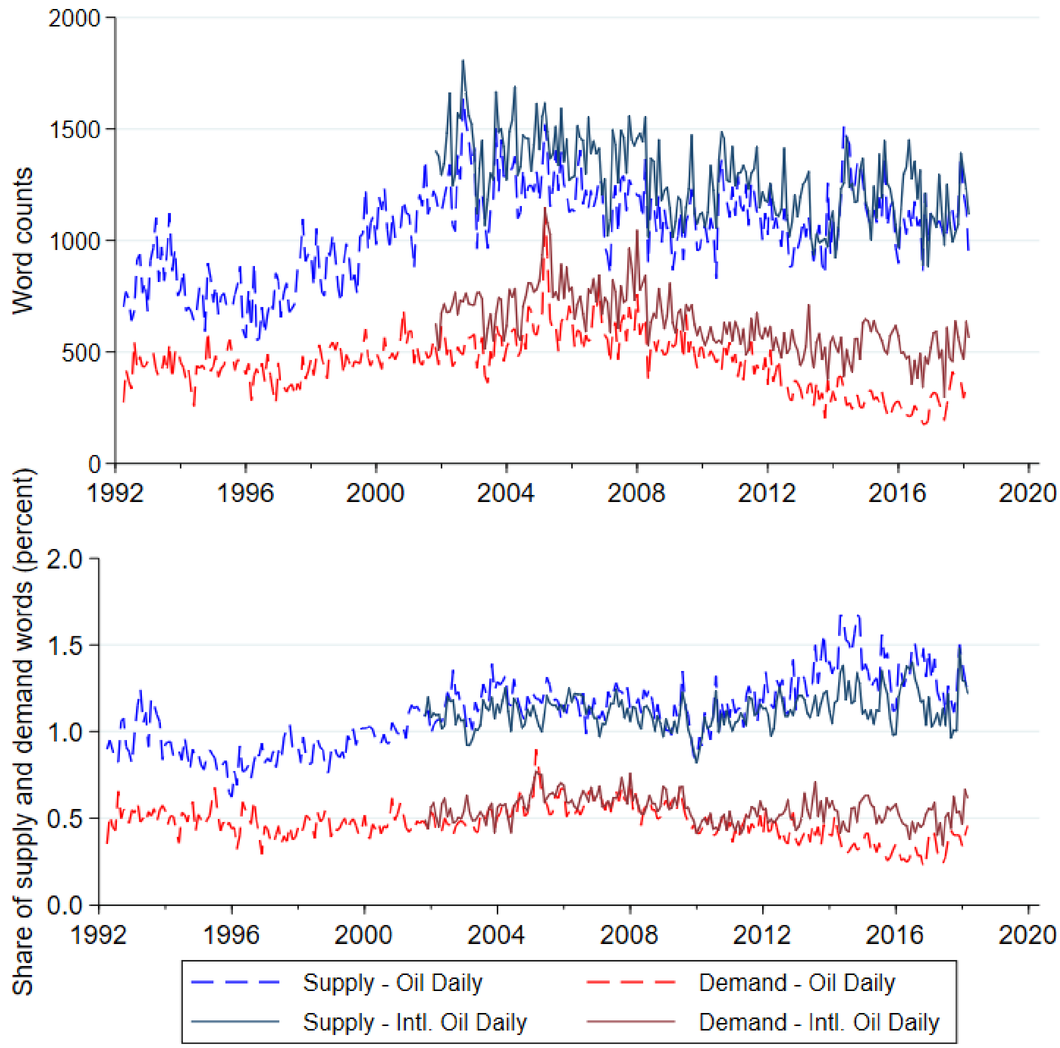


Figure 3: Share of phrase counts per month

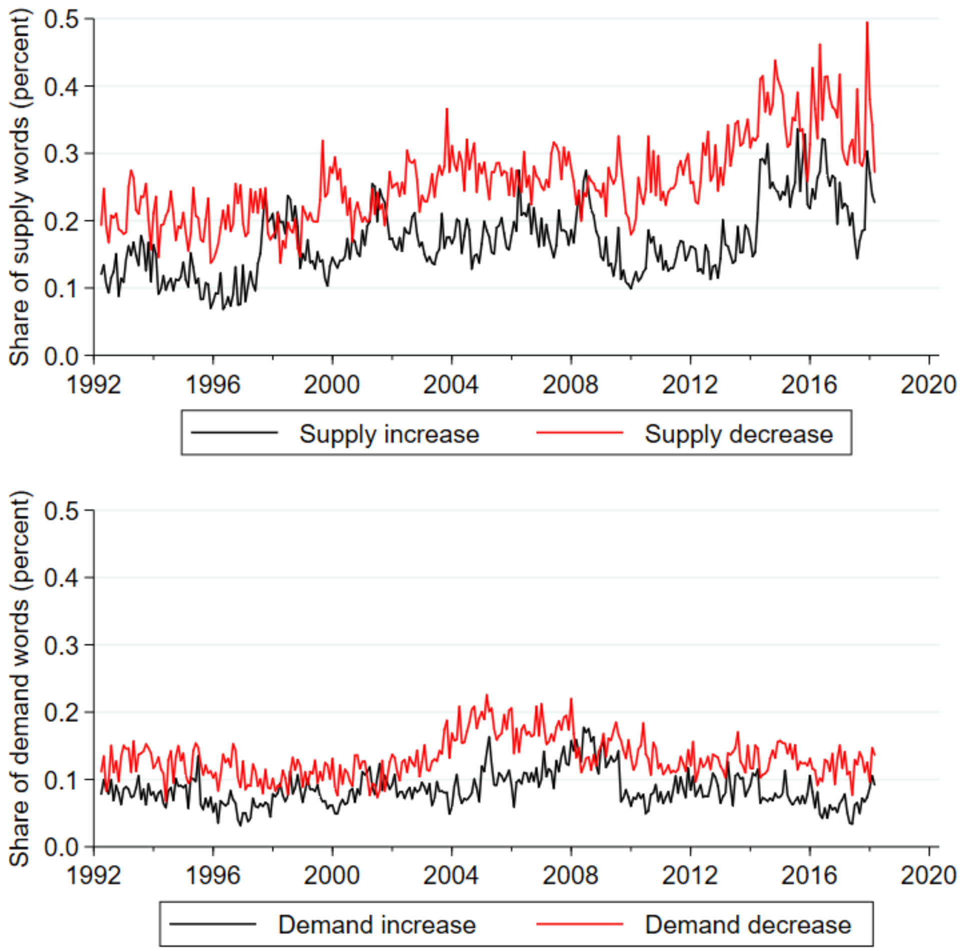


Figure 4: Net supply and net demand word counts per month in the "Oil Daily" and "International Oil Daily" publications combined - standardized.

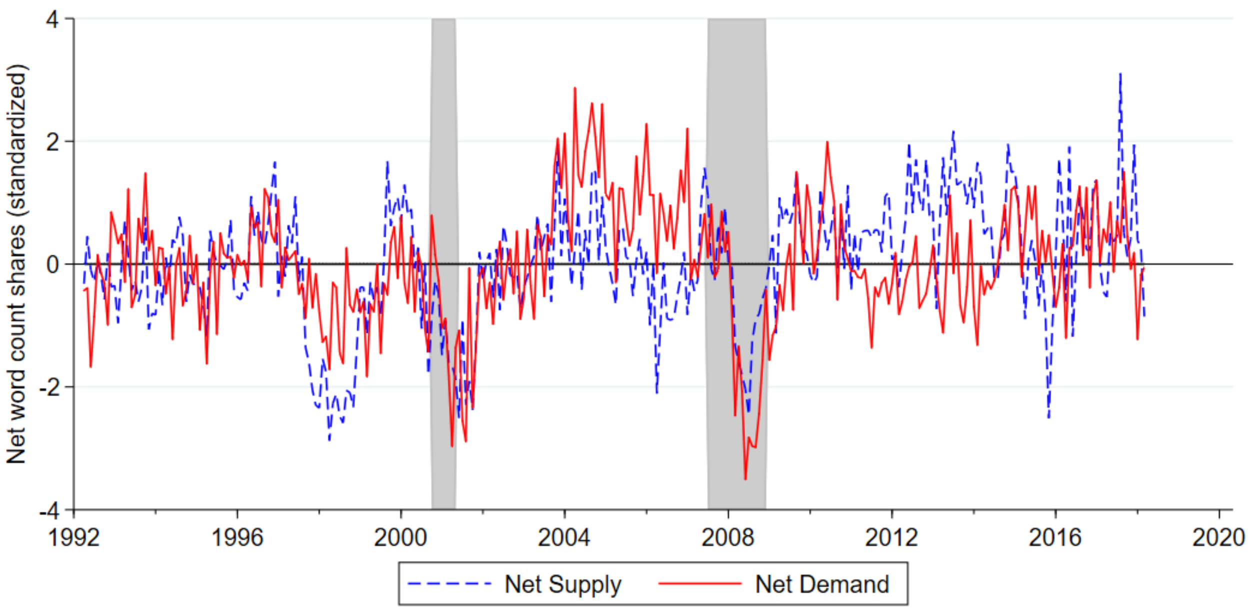


Figure 5: Oil price impulse response to supply demand, and non-fundamentals price shocks - monthly and weekly models

(a) Monthly model

(b) Weekly model

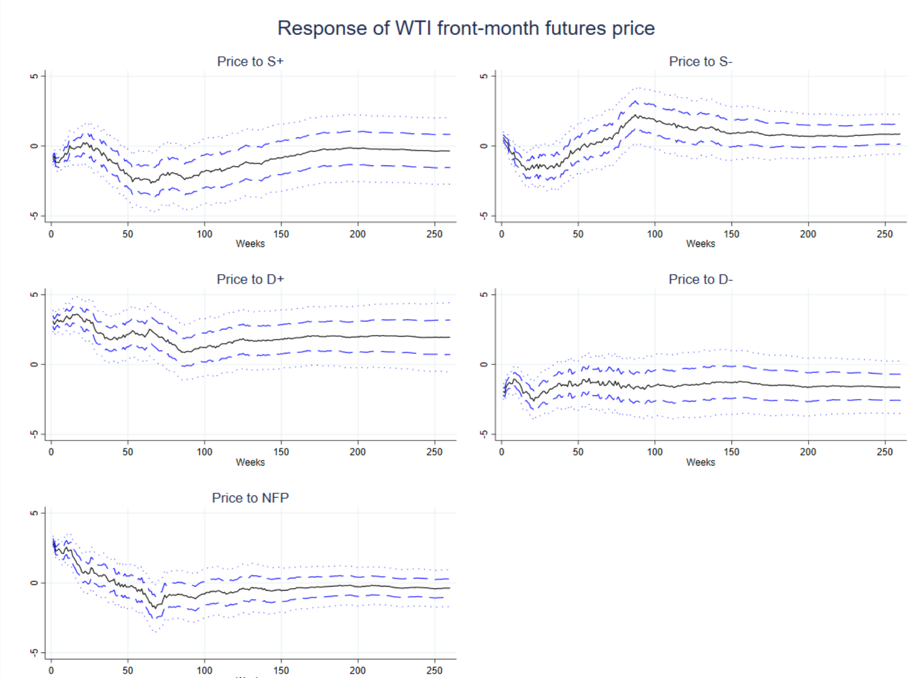
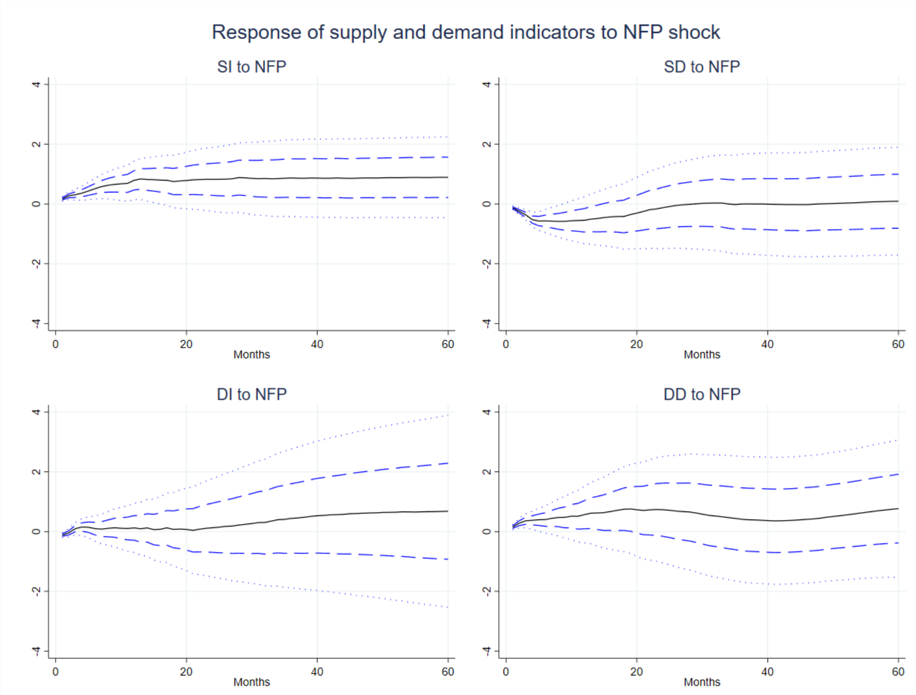


Figure 6: Oil supply and demand indexes responses to a non-fundamentals price shock - monthly and weekly models

(a) Monthly model



(b) Weekly model

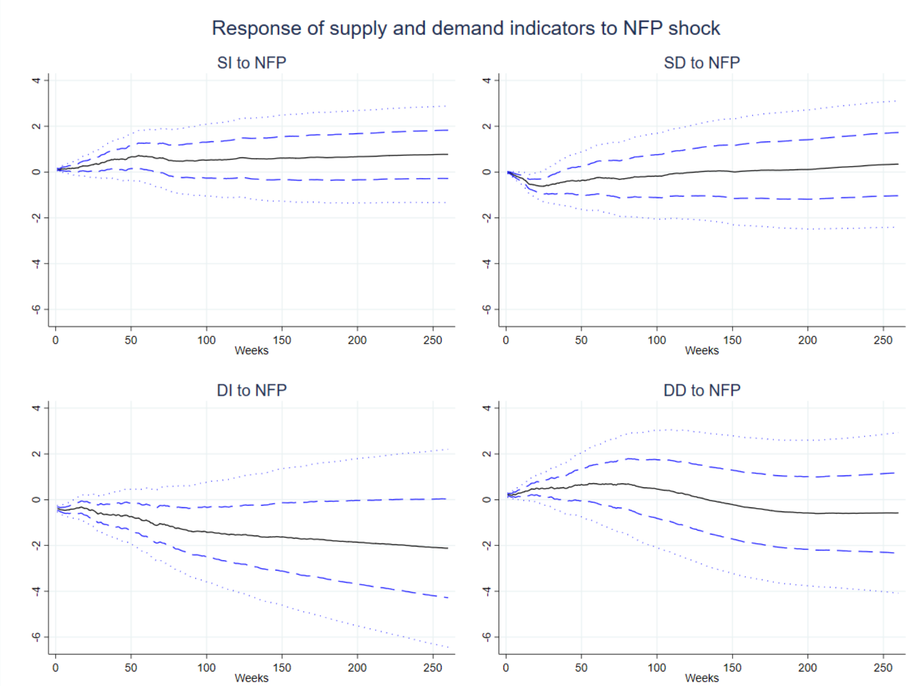


Figure 7: Oil Price Change Decomposition during the 2007-09 Great Recession - June 2008 to December 2008

Note: Each bar corresponds to the sum of all the structural shocks causing oil prices to change between June and December 2008.

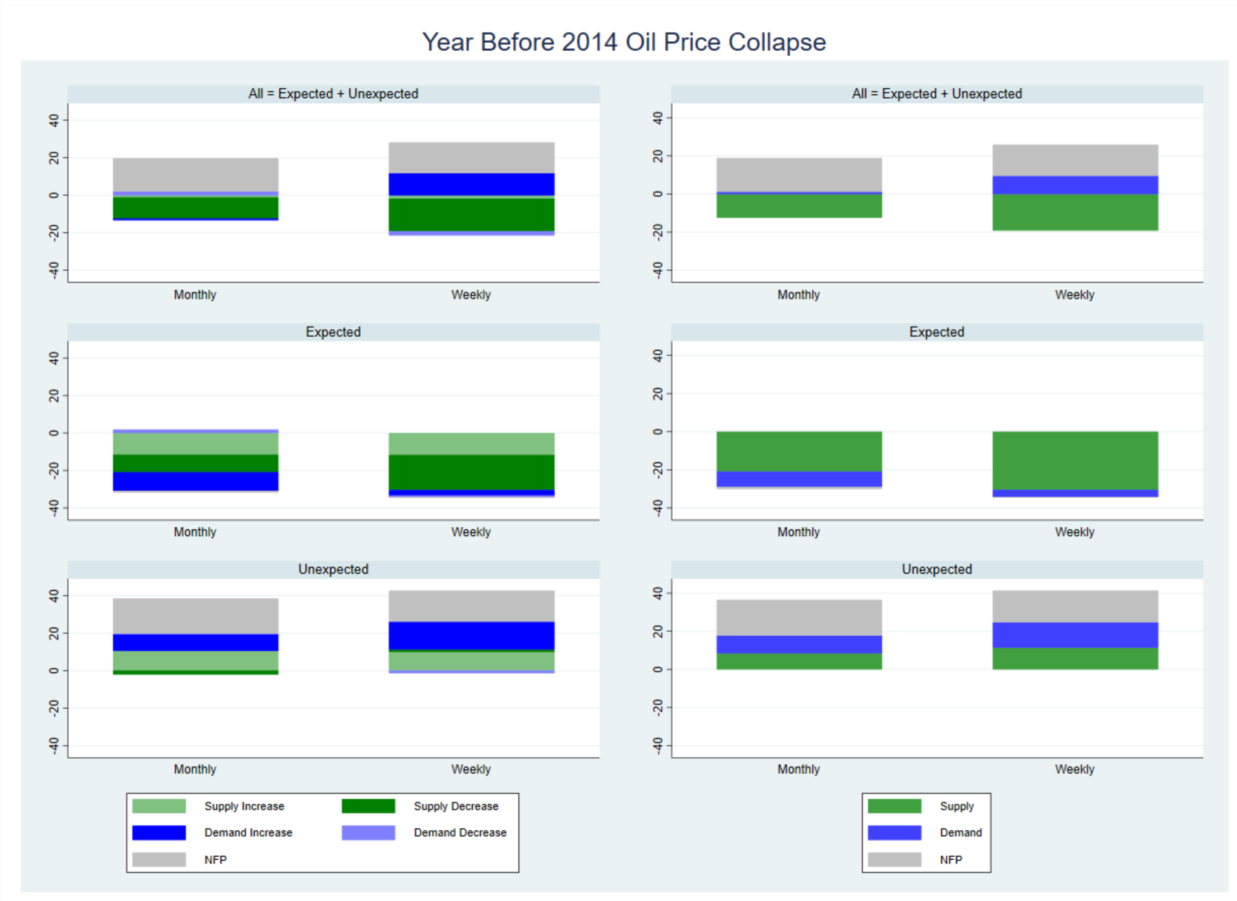
Figure 8: Oil Price Change Decomposition during the 2007-09 Great Recession - November 2007 to June 2008

Note: Each bar corresponds to the sum of all the structural shocks causing oil prices to change between November 2007 and June 2008.

Figure 9: Oil Price Change Decomposition of the 2014 Oil Price Collapse

Note: Each bar corresponds to the sum of all the structural shocks causing oil prices to change between May 2014 and February 2016.

Figure 10: Oil Price Change Decomposition One Year Before the 2014 Oil Price Collapse



Note: Each bar corresponds to the sum of all the structural shocks causing oil prices to change between April 2013 and May 2014.

Table 1: Summary statistics of the "Oil Daily" and "International Oil Daily" publications

	Oil Daily	**International Oil Daily**	**Combined**
Sample start date	Oct. 1992	Apr. 2002	Oct. 1992
Sample end date	Oct. 2018	Oct. 2018	Oct. 2018
Number of issues	6,540	4,189	10,729
Before article filtering			
Number of articles	126,863	90,071	216,934
Articles per year	4,699	5,298	8,035
Articles per issue	19	22	20
Words per article	294	321	305
After article filtering			
Number of articles	86,178	63,163	149,341
Percent of articles retained	67.9	70.1	68.8
Articles per year	3,192	3,715	5,531
Articles per issue	13	15	14
Words per article	345	359	351

Table 2: Top 5 most frequently used supply, demand, increase, and decrease words

Supply			Demand		
word	**freq.**	**share**	**word**	**freq.**	**share**
production	175,007	30.1%	demand	67,612	25.3%
output	61,791	10.6%	refinery	49,602	18.6%
supply	38,394	6.6%	refining	28,649	10.7%
reserves	36,350	6.3%	imports	25,369	9.5%
exports	33,916	5.8%	refineries	21,546	8.1%
Increase			**Decrease**		
word	**freq.**	**share**	**word**	**freq.**	**share**
more	118,322	11.7%	down	54,075	8.7%
up	111,014	10.9%	under	47,717	7.7%
over	87,447	8.6%	lower	34,498	5.5%
high	44,144	4.4%	fell	27,744	4.5%
increase	42,973	4.2%	low	25,964	4.2%

Table 3: Example of an Oil Daily article

"Crude Prices Fall on Opec Doubts, Demand Worries, Strong Dollar"
Oil Daily, 10/22/2008

"Oil prices gave back Monday's gains, falling by more than $3 on Tuesday, as traders expressed doubts that an Opec production cut will be enough to shore up prices. Saudi Arabia, Opec's largest producer, has not publicly commented on the group's upcoming meeting. The kingdom is the only Opec member that can make a substantial cut. While some of the other countries are calling for an output cut, their statements do not carry as much weight as the Saudis. The Opec pre-meeting bounce will not hold, said a broker. It's another case of buy the rumor and sell the fact. Opec Secretary-General Abdullah al-Badri said Tuesday that, There has been no decision to cut production so far. Nevertheless, analysts expect Opec to cut because of the sharp drop in both prices and demand. The real question is how much will the group cut and will its members follow through and implement an agreement. In New York, Nymex light, sweet crude for November, in its last day as prompt month, fell by $3.36 to $70.89 per barrel. In London on ICE Futures, Brent crude lost $2.31 at $69.72/bbl. Besides uncertainty about Opec action, early declines in the US stock market also weighed on prices. Another factor which helped to drive down prices was a rally in the dollar, analysts said. But bearishness regarding demand has overwhelmed everything else lately, even talk of Opec taking as much as 2 million barrels per day off the market. Matt Piotrowski, Washington"

Table 4: Monthly regressions: Oil production, real economic activity, and prices

$$ln(y_t) - ln(y_{t-1})$$

	Oil production		REA		WTI spot		WTI 12-month	
Supply increase	3.14		0.36		-14.45		-8.73	
	(0.83)***		(0.43)		(6.25)**		(5.33)	
Supply decrease	-3.43		-1.38		7.16		4.46	
	(0.72)***		(0.45)***		(5.46)		(3.86)	
Demand increase	-0.52		1.19		24.88		19.95	
	(0.65)		(0.40)***		(7.58)***		(6.27)***	
Demand decrease	-0.18		-1.82		-19.40		-15.30	
	(0.75)		(0.58)***		(6.49)***		(4.71)***	
Net supply		2.82		0.56		-10.66		-6.45
		(0.63)***		(0.37)		(4.95)**		(4.15)
Net demand		-0.62		1.01		22.35		18.24
		(0.66)		(0.43)**		(7.06)***		(5.61)***
R^2	.272	.265	.42	.395	.33	.328	.334	.333
Indexes R^2 share	.161	.138	.209	.14	.094	.085	.165	.152
N	296	296	296	296	296	296	296	296

$$ln(y_{t+12}) - ln(y_t)$$

	Oil production		REA		WTI spot		WTI 12-month	
Supply increase	0.29		-1.31		-17.43		-12.43	
	(0.29)		(0.33)***		(5.36)***		(3.76)***	
Supply decrease	-0.68		0.20		15.09		9.94	
	(0.31)**		(0.38)		(4.13)***		(3.08)***	
Demand increase	0.11		1.04		9.57		8.77	
	(0.24)		(0.32)***		(3.14)***		(2.57)***	
Demand decrease	-0.16		-1.30		-8.68		-6.74	
	(0.27)		(0.61)**		(5.08)*		(3.59)*	
Net supply		0.35		-0.91		-14.75		-10.25
		(0.25)		(0.25)***		(4.07)***		(2.86)***
Net demand		-0.04		0.78		9.68		8.24
		(0.21)		(0.32)**		(2.85)***		(2.23)***
R^2	.333	.297	.38	.327	.367	.366	.269	.268
Indexes R^2 share	.174	.089	.228	.157	.455	.45	.466	.461
N	296	296	296	296	296	296	296	296

Note: *, **, and *** denote statistical significance at 10%, 5%, and 1%. Heteroskedasticity and autocorrelation corrected standard errors in parentheses.

Table 5: Weekly regressions: Spot and futures prices

WTI Spot

	$ln(y_t) - ln(y_{t-1})$		$ln(y_{t+12}) - ln(y_t)$		$ln(y_{t+26}) - ln(y_t)$		$ln(y_{t+52}) - ln(y_t)$	
Supply increase	-14.52		-6.48		-8.29		-11.03	
	(8.92)		(5.43)		(4.52)*		(3.69)***	
Supply decrease	15.65		1.45		5.20		9.80	
	(8.33)*		(4.45)		(3.45)		(2.89)***	
Demand increase	27.60		5.30		4.07		3.45	
	(9.21)***		(3.78)		(3.12)		(2.17)	
Demand decrease	-16.12		-1.41		-1.53		-1.75	
	(9.25)*		(5.96)		(5.00)		(2.86)	
Net supply		-14.25		-4.64		-7.01		-10.23
		(7.18)**		(4.51)		(3.60)*		(2.86)***
Net demand		24.57		3.62		3.08		3.00
		(8.70)***		(4.20)		(3.42)		(1.99)
R^2	.13	.13	.197	.194	.214	.213	.183	.182
Indexes R^2 share	.053	.05	.026	.017	.111	.108	.518	.509
N	1293	1293	1293	1293	1293	1293	1293	1293

WTI 12-Month

	$ln(y_t) - ln(y_{t-1})$		$ln(y_{t+12}) - ln(y_t)$		$ln(y_{t+26}) - ln(y_t)$		$ln(y_{t+52}) - ln(y_t)$	
Supply increase	-2.67		-2.54		-4.39		-6.88	
	(6.47)		(3.16)		(2.81)		(2.39)***	
Supply decrease	6.93		-0.87		1.50		5.47	
	(5.54)		(3.08)		(2.54)		(1.91)***	
Demand increase	20.62		4.14		3.69		3.46	
	(6.90)***		(2.62)		(2.23)*		(1.73)**	
Demand decrease	-14.70		-1.55		-1.66		-1.80	
	(6.68)**		(4.68)		(4.00)		(2.35)	
Net supply		-3.86		-1.36		-3.31		-6.18
		(5.26)		(2.73)		(2.28)		(1.86)***
Net demand		19.56		3.01		2.84		2.93
		(6.52)***		(3.17)		(2.64)		(1.65)*
R^2	.12	.119	.142	.139	.15	.147	.118	.117
Indexes R^2 share	.09	.086	.037	.016	.082	.067	.493	.489
N	1293	1293	1293	1293	1293	1293	1293	1293

Note: *, **, and *** denote statistical significance at 10%, 5%, and 1%. Heteroskedasticity and autocorrelation corrected standard errors in parentheses.

Table 6: Percent deviations from baseline model's residuals standard errors and regression root mean squared error.

	Standard Error		**RMSE**	
RHS variables lags	**Augmented**	**Indexes**	**Augmented**	**Indexes**
L=12	-11.0	-7.8	-1.6	-3.3
L=18	-16.6	-10.6	-0.2	-3.0
L=24	-29.0	-19.1	-4.6	-8.2

Note: Entries in this table are percent deviations from the standard error of the regression residuals and the root mean squared error (RMSE) of each of the models denoted in each column (equations (10) and (11)) relative to those of the baseline model (equation (9)).

Table 7: Structural VAR short-run dynamics parameters estimates.

	SI_t (1)	SD_t (2)	DI_t (3)	DD_t (4)	$\Delta ln(Price_t)$ (5)
Panel A: Monthly model					
SI_t	-	-	-	-	-11.943
					(2.138)***
SD_t	-	-	-	-	11.700
					(2.245)***
DI_t	-	-	-	-	10.599
					(1.589)***
DD_t	-	-	-	-	-10.073
					(1.539)***
$\Delta ln(Price_t)$	0.037	-0.028	-0.026	0.033	-
	(0.008)***	(0.006)***	(0.009)***	(0.01)***	
Sample	1994M04 - 2018M09				
Number observations	294				
Panel B: Weekly model					
SI_t	-	-	-	-	-3.147
					(0.464)***
SD_t	-	-	-	-	2.190
					(0.531)***
DI_t	-	-	-	-	6.979
					(1.011)***
DD_t	-	-	-	-	-5.143
					(0.432)***
$\Delta ln(Price_t)$	0.038	-0.008	-0.134	0.063	-
	(0.007)***	(0.007)	(0.03)***	(0.013)***	
Sample	4/06/1994 - 10/10/2018				
Number observations	1280				

Note: *, **, and *** denote statistical significance at 10%, 5%, and 1%. Analytical standard errors in parentheses.

Table 8: Oil Price Forecast Error Variance Decomposition.

Monthly model							
	S^+	S^-	S	D^+	D^-	D	NFP
$t = 1$	17.3	12.1	29.4	17.5	18.1	35.6	34.9
$t = \infty$	17.2	17.2	34.4	18.2	17.3	35.6	30.0
Weekly model							
	S^+	S^-	S	D^+	D^-	D	NFP
$t = 1$	4.0	1.7	5.7	41.2	15.3	56.5	37.9
$t = \infty$	9.6	9.2	18.7	31.3	17.4	48.6	32.6

Note: Entries in the table are the contribution of each shock, or combination of shocks, to the forecast error variance of oil price in the short run, $t = 1$, and in the long run, $t = \infty$. Columns S and D are the sum of the contribution of shocks S^+ and S^- and D^+ and D^-, respectively.

B Data Cleaning

In this Appendix we explain how we clean the data. In particular, we provide more information on the removal of articles not related to oil market developments, correction of issues related to numbers and punctuation, and adjustments for other idiosyncratic word usage.

First, we seek to exclude articles that are primarily on topics unrelated to the oil industry, including those related to natural gas and gasoline. We aim to exclude articles on natural gas as the oil and natural gas markets are mostly unrelated at this point. We recognize that articles on gasoline can be informative for oil supply and demand. However, while an increase in gasoline supply may be indicative of an increase in oil demand, our text searches would not adjust to this nuance. Rather, our algorithm would count gasoline supply developments as oil supply developments, which would likely introduce noise into the analysis. Consequently, we aim to exclude gasoline-focused articles as well. With this in mind, we remove these articles as follows:

1. For each article, we count the number of times the *oil words* "crude", "oil", and "petroleum" appear in the body of the text.

2. Next, we count the number of times the *gas words* "gasoline", "gas", and "LNG" appear in the body of the text. Notably, the word "gas" captures both articles about natural gas and articles about gasoline.

3. We remove from the corpus the articles in which the number of *gas words* is greater than or equal to twice the number of *oil words*. Notably, this step also eliminates all articles for which the count of *oil words* is 0 because in these cases the number of *gas words* will always be greater than or equal to the number of *oil words*.

We also correct some problems related to numbers and punctuation that stem from the digitization of the article archive. We use the *Sentence Tokenizer* function (part of the *NLTK* suite of libraries for Python) to identify sentences within the article. This is an important step, as our directional word counts rely on word proximity within sentences. Among other things, the tokenizer looks for periods, question marks, and exclamation points to note the ends of sentences. We adjust the punctuation and numerical characters in the text for better application of the sentence tokenizer by taking the following steps:

1. Foreign currency symbols preceding prices often are replaced by a question mark in the raw data. We replace the pattern "?n", where n could be any numerical digit, with the single character "z".

2. The tokenizer often generates mistakes due to improper identification of decimals as sentence-ending periods. To eliminate this problem, we replace all numerical characters with the character "z". We remove decimal numbers from the text by replacing the pattern "z.z" with "zz" each time it appears. This also allows the tokenizer to better recognize, for example, the end of a sentence that concludes with a year or other number (e.g., "...2017."). (We take this opportunity to remind the reader that we never use any numerical information that is included the text; our indexes rely only on textual analysis of words and expressions and not on numerical information.)

3. Frequently, a period at the end of a sentence is (erroneously) not followed by a space. To more clearly delineate sentence endings, we insert a space after all periods.

4. More generally, we find many punctuation marks that are erroneously not followed by a space. In addition to periods, we insert a space following all of the following punctuation characters: commas, double quotes, single quotes/apostrophes, question marks, exclamation points, hyphens, parentheses, brackets, forward and backward slashes, and percent symbol. Since inserting a space after a comma in large numbers would alter our word count, before inserting this spaces, we remove commas from numbers by replacing the pattern "z,z" with "zz" each time it appears.

5. Our vocabulary searches are streamlined by the removal of the possessive forms of words. In particular, we remove the string " 's ", which often appears at the end of singular possessive words and the contraction "it's". This ensures that, e.g. the possessive form of the word "producer's" gets counted along with the word "producer". Similarly, we delete apostrophes at the end of plural possessive words. This ensures that, e.g. the plural possessive word "producers' " gets counted along with the word "producers".

6. We replace the common expression "b\d" with its unabbreviated form, "barrels per day".

Lastly, we also correct some issues related to idiosyncratic word usage. In general, the insertion of the character "z" helps retain the word count in the structure of the original sentence, so that the proximity across words is unchanged after the replacements.

1. The word "shale" is on our supply vocabulary list. However, the common phrase "shale boom" would likely result in supply increase counts which would not be very informative. To reduce the noise arising from this idiosyncratic phrase, we replace the phrase "shale boom" with "shale z". The retention of the word "shale" maintains the supply count while eliminating the likely noisy supply increase count.

2. To reduce false counts of supply, we replace the phrase "shale gas" with "natural gas." This step has no implication for our deletion of gas-related articles, because the deletion is based on the relative count of *oil words* and *gas words*, and this replacement does not replace the word "gas", which is used in the count.

3. To reduce false counts of decrease words, we replace the phrases "tight oil" and "tight crude" with the phrases "shale oil" and "shale crude", respectively. This issue becomes particularly relevant after 2014, when shale oil started its rise into a major source of oil supply.

4. To reduce false counts of supply, we replace the phrase "industrial production" with the string "indprod" and include "indprod" in our demand vocabulary list.

5. To reduce false increase and decrease counts, we replace the phrase "more or less" with the string "z about z".

6. To reduce false increase counts, we replace the phrases "extra heavy" and "extra light" with "z heavy" and "z light".

7. Because the increase word *up* is frequently used within expressions meaning something other than increase, we replace the word "up" with "z" when it appears as one of the phrases listed in Table B1

Table B1: "Up" Phrases Replaced in Text

back(s/ed) up	fill(s/ed) up	look(s/ed) up	size(s/ed) up	tidy up
blown up	follow(s/ed) up	meet(s) up	speak(s) up	tie(s/d) up
break(s) up	gave up	mess(es) up	spoke up	tighten(s/ed) up
broke up	gear(s/ed) up	messed up	stand(s) up	tore up
brush(es/ed) up	give(s) up	met up	start(s/ed) up	true up
buck(s/ed) up	hang(s) up	mix(es) up	stir(s) up	up front
call(s/ed) up	held up	mixed up	stirred up	wake(s) up
came up with	hole(s/d) up	pipe(s) up	stood up	warm(s/ed) up
clean(s/ed) up	hook(s/ed) up	piped up	suck(s/ed) up	wash(es) up
clear(s/ed) up	hung up	put(s) up	sum(s) up	washed up
come(s) up with	leg up	screw(s/ed) up	summed up	whats up
conjure(s/ed) up	light(s/ed) up	set(s) up	tear(s/ed) up	woke up
cover(s/ed) up	line(s) up	shake(s/d) up	threw up	wrap(s) up
draw(s) up	lined up	shook up	throw(s/ed) up	wrapped up
dress(es/ed) up	lighted up	show(s/ed) up	thumbs up	write(s) up
drew up	lit up	shut(s) up	tidied up	wrote up
fed up	lock(s/ed) up	sign(s/ed) up	tidies up	

C Lists of Words and Phrases

In this Appendix we provide the list of regular expressions used to generate the lists of supply, demand, increase, and decrease words and explain how we used these to obtain the list of words we used in the construction of the indexes.

C.1 Supply and Demand Words and Phrases

To generate our list of supply and demand words, we start with the list of regular expressions reported in Table C1. A search using regular expressions will match the many forms of a single root word. That is, searching for the string *suppl* will yield *supply*, *supplied*, *supplies*, etc. The drawback, however, is that our search will also yield words which are are unrelated, like *supplement*. To generate our final vocabulary list, we search the text of the Oil Daily and International Oil Daily articles for all the matches to our list of regular expressions. This procedure is applied to the text that has been cleaned as described in Appendix B. Our list of just 11 supply and 9 demand regular expressions results in about 1300 supply and 500 demand matches, including both unrelated words and many obvious typographical errors.[6] We focus on the most frequent words, including the roughly 100 supply and 50 demand words which appear in the corpus of 150,000 articles at least 50 times. Of these, we include in our final supply vocabulary list– shown in Table C2 –the words which are informative for supply and demand. We exclude both unrelated words (like *right(s)* for supply, which matches the regular expression *rig*, and *consummate* for demand, which matches the regular expression *consum*) and words that human auditing have revealed to be misleading (like *product(s)* for supply and *refiner(s)* and for demand). These criteria yielded 39 supply and 34 demand words.

[6] As noted in Appendix B, we replace the phrase *industrial production* with *indprod*, and include this expression as a demand word.

Some of the above vocabulary words, such as *glut* and *outage* are particularly informative not just for supply, but for our primary interest in the directional concepts of supply increase and supply decrease. These words are included not just in the raw counts for supply words, but also add to the counts for supply increase and supply decrease. Lastly, since directional supply words are particularly informative, we expand our vocabulary list to include the less frequent related words, as long as they appear at least 5 times in the article corpus. In sum, this augments our supply vocabulary list with 10 additional words, to a total of 49 words.

Table C1: Regular expressions

Supply	**Demand**
\b\S*discover\S*\b	\b\S*buy\S*\b
\b\S*export\S*\b	\b\S*consum\S*\b
\b\S*glut\S*\b	\b\S*demand\S*\b
\b\S*outage\S*\b	\b\S*deplet\S*\b
\b\S*output\S*\b	\b\S*import\S*\b
\b\S*produc\S*\b	\b\S*indprod\S*\b
\b\S*reserv\S*\b	\b\S*refin\S*\b
\b\S*rig\S*\b	\b\S*throughput\S*\b
\b\S*shale\S*\b	\b\S*utiliz\S*\b
\b\S*suppl\S*\b	
\b\S*surplus\S*\b	

Table C2: Final vocabulary list

	Supply		
discoveries	overproduce(+)	producers	supplier
discovery	overproduced(+)	produces	suppliers
export	overproducer(+)	producing	supplies
exported	overproducers(+)	production	supply
exporter	overproducing(+)	reserve	supplying
exporters	overproduction(+)	reserves	surplus(+)
exporting	oversupplied(+)	reservoir	surpluses(+)
exports	oversupplies(+)	reservoirs	underproduction(-)
glut(+)	oversupply(+)	rig	undersupplied(-)
gluts(+)	oversupplying(+)	rigs	undersupply(-)
outage(-)	produce	shale	
outages(-)	produced	shales	
output	producer	supplied	
	Demand		
buy	consumes	importation	refinery
buyer	consuming	imported	refinerys
buyers	consumption	importer	refines
buying	demand	importers	refining
buys	deplete	importing	throughput
consume	depleted	imports	throughputs
consumed	depleting	indprod	utilization
consumer	depletion	refine	
consumers	import	refineries	

Note: (+) and (-) denote supply words which add to the counts for supply increase and supply decrease.

C.2 Increase and Decrease Words

The process to generate the list of increase and decrease words used in our analysis is identical to that used to construct the list of supply and demand words. In this case, we search the text using the list of 83 increase and 100 decrease regular expressions shown in Table C3, which results in around 5300 increase and 3600 decrease matches. We again focus on the most frequent words, including the roughly 800 increase and 700 decrease which appear at least 40 times. After excluding those words that were not related to increase and decrease, we ended up with 243 increase and 298 decrease words, listed in Tables C4 and C5.

Table C3: Regular expressions

Increase			
\b\S*above\S*\b	\b\S*cumulat\S*\b	\b\S*more\S*\b	\b\S*rocket\S*\b
\b\S*accelerat\S*\b	\b\S*elevat\S*\b	\b\S*mount\S*\b	\b\S*rose\S*\b
\b\S*accret\S*\b	\b\S*enlarg\S*\b	\b\S*multipl\S*\b	\b\S*shoot\S*\b
\b\S*accru\S*\b	\b\S*escalat\S*\b	\b\S*mushroom\S*\b	\b\S*shot\S*\b
\b\S*accumulat\S*\b	\b\S*expand\S*\b	\b\S*over\S*\b	\b\S*skyrocket\S*\b
\b\S*add\S*\b	\b\S*extra\S*\b	\b\S*perk\S*\b	\b\S*soar\S*\b
\b\S*advanc\S*\b	\b\S*firmed\S*\b	\b\S*pop\S*\b	\b\S*spik\S*\b
\b\S*amass\S*\b	\b\S*gain\S*\b	\b\S*positive\S*\b	\b\S*spring\S*\b
\b\S*amplif\S*\b	\b\S*garner\S*\b	\b\S*propagat\S*\b	\b\S*spurt\S*\b
\b\S*ascend\S*\b	\b\S*grew\S*\b	\b\S*propel\S*\b	\b\S*strengthen\S*\b
\b\S*augment\S*\b	\b\S*grow\S*\b	\b\S*propulsion\S*\b	\b\S*strong\S*\b
\b\S*bloom\S*\b	\b\S*height\S*\b	\b\S*prosper\S*\b	\b\S*surg\S*\b
\b\S*blossom\S*\b	\b\S*high\S*\b	\b\S*push\S*\b	\b\S*surmount\S*\b
\b\S*bolster\S*\b	\b\S*hik\S*\b	\b\S*quicken\S*\b	\b\S*surpass\S*\b
\b\S*boom\S*\b	\b\S*improve\S*\b	\b\S*rais\S*\b	\b\S*swell\S*\b
\b\S*boost\S*\b	\b\S*increas\S*\b	\b\S*rall\S*\b	\b\S*takeoff\S*\b
\b\S*buil\S*\b	\b\S*inflat\S*\b	\b\S*reach\S*\b	\b\S*tookoff\S*\b
\b\S*bullish\S*\b	\b\S*intensif\S*\b	\b\S*recover\S*\b	\b\S*up\S*\b
\b\S*buoyant\S*\b	\b\S*jump\S*\b	\b\S*redoubl\S*\b	\b\S*wax\S*\b
\b\S*burgeon\S*\b	\b\S*leap\S*\b	\b\S*reviv\S*\b	\b\S*zoom\S*\b
\b\S*climb\S*\b	\b\S*lift\S*\b	\b\S*ris\S*\b	
Decrease			
\b\S*abat\S*\b	\b\S*dip\S*\b	\b\S*negative\S*\b	\b\S*slump\S*\b
\b\S*attenuat\S*\b	\b\S*disrupt\S*\b	\b\S*offline\S*\b	\b\S*small\S*\b
\b\S*bearish\S*\b	\b\S*dive\S*\b	\b\S*plummet\S*\b	\b\S*soft\S*\b
\b\S*below\S*\b	\b\S*diving\S*\b	\b\S*plung\S*\b	\b\S*squeeze\S*\b
\b\S*collaps\S*\b	\b\S*dove\S*\b	\b\S*pull\S*\b	\b\S*standstill\S*\b
\b\S*constrict\S*\b	\b\S*down\S*\b	\b\S*quell\S*\b	\b\S*stunt\S*\b
\b\S*contract\S*\b	\b\S*dried\S*\b	\b\S*reced\S*\b	\b\S*subdu\S*\b
\b\S*cool\S*\b	\b\S*droop\S*\b	\b\S*reduc\S*\b	\b\S*subside\S*\b
\b\S*crash\S*\b	\b\S*drop\S*\b	\b\S*restrict\S*\b	\b\S*subsiding\S*\b
\b\S*crimp\S*\b	\b\S*dry\S*\b	\b\S*retreat\S*\b	\b\S*sunk\S*\b
\b\S*crush\S*\b	\b\S*dwindl\S*\b	\b\S*retrench\S*\b	\b\S*suppress\S*\b
\b\S*curtail\S*\b	\b\S*ebb\S*\b	\b\S*sank\S*\b	\b\S*suspend\S*\b
\b\S*cut\S*\b	\b\S*evaporat\S*\b	\b\S*sap\S*\b	\b\S*tank\S*\b
\b\S*damp\S*\b	\b\S*fade\S*\b	\b\S*short\S*\b	\b\S*taper\S*\b
\b\S*decay\S*\b	\b\S*fading\S*\b	\b\S*shrank\S*\b	\b\S*tepid\S*\b
\b\S*declin\S*\b	\b\S*fall\S*\b	\b\S*shrink\S*\b	\b\S*tight\S*\b
\b\S*decreas\S*\b	\b\S*falter\S*\b	\b\S*shrunk\S*\b	\b\S*trickl\S*\b
\b\S*demise\S*\b	\b\S*fell\S*\b	\b\S*shut\S*\b	\b\S*trim\S*\b
\b\S*dent\S*\b	\b\S*flag\S*\b	\b\S*sink\S*\b	\b\S*tumbl\S*\b
\b\S*deplet\S*\b	\b\S*halt\S*\b	\b\S*slack\S*\b	\b\S*under\S*\b
\b\S*depress\S*\b	\b\S*less\S*\b	\b\S*slash\S*\b	\b\S*wane\S*\b
\b\S*destruct\S*\b	\b\S*loosen\S*\b	\b\S*slid\S*\b	\b\S*waning\S*\b
\b\S*deteriorat\S*\b	\b\S*los\S*\b	\b\S*slip\S*\b	\b\S*weak\S*\b
\b\S*diminish\S*\b	\b\S*low\S*\b	\b\S*slow\S*\b	\b\S*wither\S*\b
\b\S*diminution\S*\b	\b\S*maintenance\S*\b	\b\S*sluggish\S*\b	\b\S*worsen\S*\b

Table C4: Increase vocabulary list

above	bolstered	escalating	improve	mounted	raising	rises	strongest
accelerate	bolstering	escalation	improved	mounting	rallied	rising	strongly
accelerated	bolsters	expand	improvement	over	rallies	rocket	surge
accelerates	boom	expanded	improvements	overcame	rally	rocketed	surged
accelerating	boomed	expanding	improves	overcome	rallying	rocketing	surges
acceleration	booming	expands	increase	overcoming	reach	rockets	surging
accrue	booms	extra	increased	overproducing	reached	rose	surpass
accrued	boost	firmed	increases	overrun	reaches	shoot	surpassed
accruing	boosted	flourish	increasing	overruns	reaching	shooting	surpasses
accumulate	booster	gain	inflate	overtake	rebuild	shot	surpassing
accumulated	boosting	gained	inflated	overtaken	rebuilding	skyrocket	swell
accumulating	boosts	gaining	inflating	overtaking	rebuilt	skyrocketed	swelled
accumulation	build	gains	intensified	overtook	recoup	skyrocketing	swelling
accumulations	building	gallup	intensifies	perked	recouped	soar	up
add	builds	grew	intensify	pickup	recouping	soared	upbeat
added	buildup	grow	intensifying	pop	recover	soaring	uplift
adding	built	growing	jump	popped	recovered	soars	upped
addition	bullish	grown	jumped	popping	recovering	spike	upping
additional	bullishness	grows	jumping	positive	recovers	spiked	upsurge
additions	buoyant	growth	jumps	propel	recovery	spikes	upswing
adds	burgeoning	heighten	leap	propelled	regain	spiking	uptick
advance	climb	heightened	leaped	propelling	regained	spring	uptrend
advanced	climbed	heightening	leaping	prosper	regaining	springing	upturn
advances	climbing	high	leapt	prosperity	resurgence	springs	upward
advancing	climbs	higher	lift	push	resurgent	spurt	upwardly
amassed	elevate	highest	lifted	pushed	revival	strengthen	upwards
amassing	elevated	highs	lifting	pushes	revive	strengthened	
amplified	enlarge	hike	liftings	pushing	revived	strengthening	
augment	enlarged	hiked	lifts	raise	reviving	strengthens	
augmented	escalate	hikes	more	raised	rise	strong	
bolster	escalated	hiking	mount	raises	risen	stronger	

Table C5: Decrease vocabulary list

abated	declining	disrupted	fall	negatively	shrinkage	slumped	tightens
abating	decrease	disrupting	fallen	nosedive	shrinking	slumping	tighter
bearish	decreased	disruption	falling	plummet	shrinks	small	tightest
bearishness	decreases	disruptions	falls	plummeted	shrunk	smaller	tightly
below	decreasing	disruptive	falter	plummeting	shut	smallest	tightness
collapse	demise	dive	faltered	plunge	shutdown	soft	trickle
collapsed	dent	dived	faltering	plunged	shutdowns	soften	trickling
collapses	dented	diving	fell	plunging	shuts	softened	trim
collapsing	denting	down	flagged	pullback	shutter	softening	trimmed
contracted	deplete	downgrade	flagging	quell	shuttered	softer	trimming
contraction	depleting	downgraded	halt	recede	shutting	softness	tumble
crash	depletion	downgrades	halted	receded	sink	squeeze	tumbled
crashed	depress	downgrading	halting	receding	sinking	squeezed	tumbling
crashes	depressed	downs	landslide	reduce	sinks	squeezes	under
crashing	depressing	downsize	less	reduced	slack	standstill	underperform
crimp	depression	downsized	lessen	reduces	slackening	stumbled	underperformance
crimped	destruction	downsizing	lessened	reducing	slash	stumbling	underperformed
crimping	destructive	downtrend	lessening	reduction	slashed	stunt	wane
crush	deteriorate	downturn	lesser	reductions	slashing	subdued	waned
crushed	deteriorated	downturns	loosen	restrict	slid	subside	waning
crushing	deteriorates	downward	loosened	restricted	slide	subsided	weak
curtail	deteriorating	downwards	loosening	restricting	slides	sunk	weaken
curtailed	deterioration	drawdown	lose	restrictive	sliding	sunken	weakened
curtailing	diminish	drawdowns	loses	restricts	slip	suppress	weakening
curtailment	diminished	drop	losing	retreat	slippage	suppressed	weakens
curtailments	diminishing	dropped	loss	retreated	slipped	suspend	weaker
cut	dip	dropping	losses	retreating	slipping	suspended	weakest
cutback	dipped	drops	lost	retrenchment	slips	suspending	weakness
cutbacks	dipping	dwindle	low	sank	slow	tanked	weaknesses
cuts	dips	dwindled	lower	shortage	slowdown	tanking	worsen
cutting	disappear	dwindling	lowered	shortages	slowdowns	taper	worsened
dampen	disappearance	ebb	lowering	shorten	slowed	tapered	worsening
dampened	disappeared	ebbed	lowers	shortened	slower	tapering	
dampening	disappearing	evaporate	lowest	shortening	slowing	tepid	
damper	disappoint	evaporated	lows	shortfall	slows	tight	
decline	disappointed	fade	maintenance	shortfalls	sluggish	tighten	
declined	disappointing	faded	meltdown	shrank	sluggishness	tightened	
declines	disrupt	fading	negative	shrink	slump	tightening	

D Additional Results

Figure D1: Structural impulse response functions - monthly SVAR model

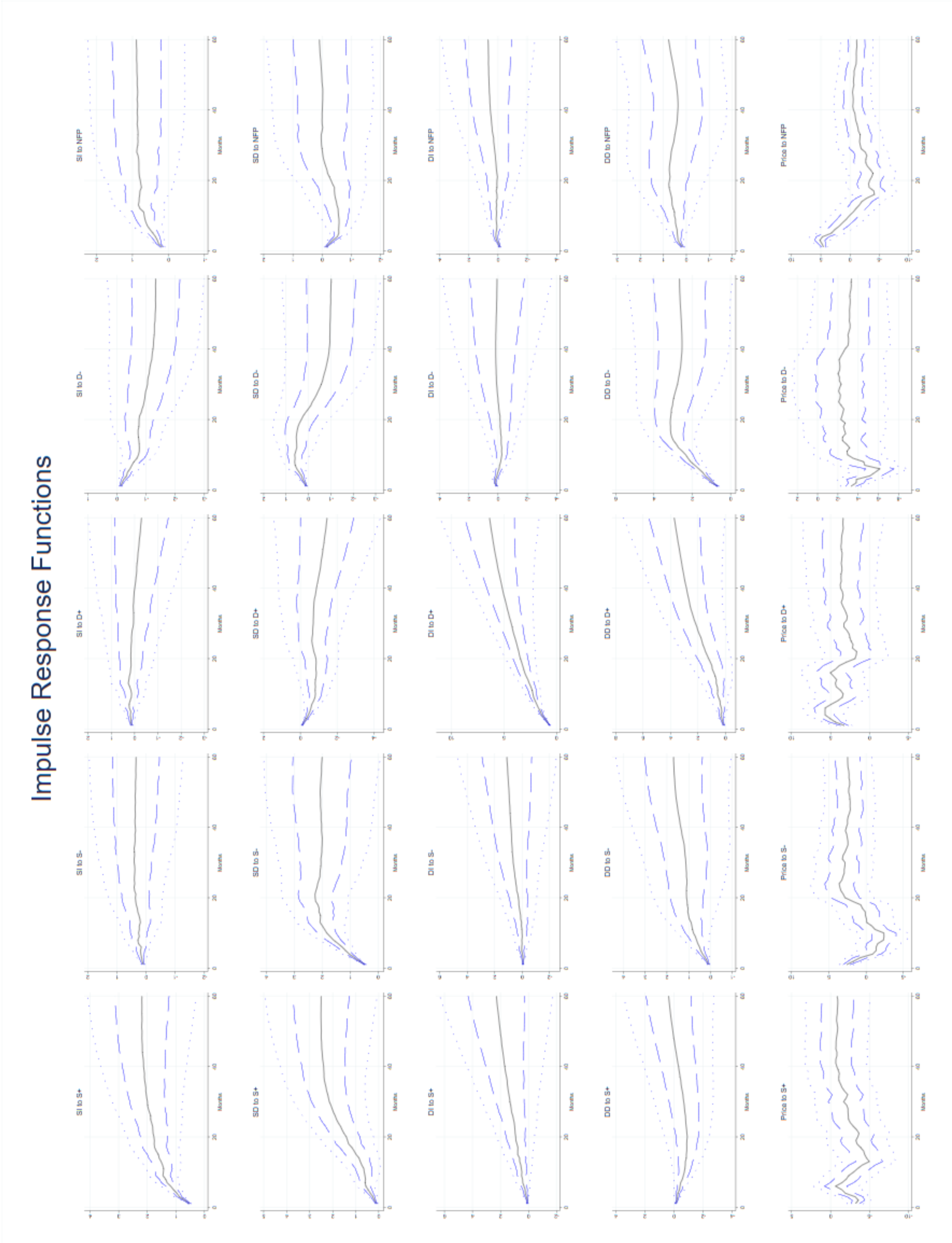


Figure D2: Structural impulse response functions - weekly SVAR model