# Oil Shocks: A Textual Analysis Approach

Deepa D. Datta  Daniel A. Dias  Benjamin K. Johannsen

Federal Reserve Board

August 2022

# Overview

We use textual analysis of a key oil industry news source to measure global supply and demand developments.

1. Develop a systematic and automated process to *read the news*, analogous to the informal approach we employ as Federal Reserve Board oil analysts.
   - Use Energy Intelligence Group's flagship publications "Oil Daily" and "International Oil Daily."
2. Provide a quantitative narrative of market conditions in real-time, as the news develops.
   - Earlier data availability than for oil production or IP.
   - Avoids reliance on alternative markets, e.g. metals, equities.
3. Use the publication archive to relate text-based indexes to supply, demand, and prices.
   - Identify supply- and demand-driven oil price dynamics using a structural VAR model.

# Beyond article counts and sentiment analysis

- Use **phrase counts** to develop a quantitative narrative of the relative importance of supply and demand.
  - Baker, et al. (2016) construct EPU index by counting the number of articles that contain terms related to EPU in 10 international newspapers.
  - MPU, GPR, TPU indexes use similar methods (Husted, et al. (2019); Caldara and Iacoviello (2018); Caldara, et al. (2019)).

- Focus on **direction of supply and demand** to go beyond simple sentiment analysis for oil markets.
  - Loughran, et al. (2019) – short-horizon price movements related to sentiment in DJ Energy Service.
  - Cakir Melek, et al. (2019) – sentiment in Thomson-Reuters oil articles helps forecast oil prices.
  - Brandt and Gao (2019) – sentiment on macroeconomic and geopolitical news in RavenPack affects oil prices.

# Full replicability, Straightforward updating

- Using natural language processing, we extract signals and construct the quantitative narrative from news articles, while **preserving replicability and straightforward updating**.

- Wu and Cavallo (working paper, 2012) similarly combine narrative and quantitative approaches to construct measures of oil price shocks.
  - Narrative approach involves human auditing of Oil Daily, Oil & Gas Journal, and Monthly Energy Chronology.
  - 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.
  - Aggregate select event types to generate exogenous oil shocks series, and show substantial effects.

# "Oil Daily" and "International Oil Daily"

1. Quality: Trade publications from the well-regarded Energy Intelligence Group, whose data is used by OPEC as an official secondary source.
2. Content: One daily price reporting article, additional articles on current and prospective developments.
   - Oil Daily: 26 years of daily articles, from 1992 to the present. 3000 to 5000 articles per year, 25 to 45 articles per day.
   - International Oil Daily: Begins in 2002. 4000 articles per year, 30 to 45 articles per day.
   - We remove non-oil articles, which are those that either omit all oil words or have a greater focus on natural gas or gasoline.
3. Example Headlines:
   - November 29, 2017: "Opec Uncertainty Weighs on Oil Prices", "Aramco, BP Remain Confident About Long-Term Oil Demand"
   - April 8, 2003: "Crude, Products Close Lower Despite Possible Late-April Opec Meeting", "Merger Activity Drives Share Price Gainers for the Week"

# Developing Word Lists

| Supply | | | Demand | | |
|---|---|---|---|---|---|
| **word** | **freq.** | **share** | **word** | **freq.** | **share** |
| production | 173,762 | 30.9% | demand | 65,456 | 22.7% |
| output | 61,571 | 11.0% | refinery | 46,301 | 16.1% |
| supply | 38,113 | 6.8% | refining | 26,828 | 9.3% |
| reserves | 36,166 | 6.4% | imports | 24,848 | 8.6% |
| exports | 33,789 | 6.0% | refineries | 20,144 | 7.0% |

# Developing Word Lists

| Increase | | | Decrease | | |
|---|---|---|---|---|---|
| **word** | **freq.** | **share** | **word** | **freq.** | **share** |
| up | 130,930 | 17.8% | down | 52,563 | 9.6% |
| high | 42,963 | 5.8% | lower | 33,859 | 6.2% |
| increase | 41,705 | 5.7% | fell | 27,029 | 5.0% |
| higher | 39,860 | 5.4% | low | 25,145 | 4.6% |
| growth | 33,803 | 4.6% | cut | 23,872 | 4.4% |

# Index Construction

Directional phrase counts

1. Construct vocabulary lists for supply, demand, increase, and decrease
2. Count the number of times a "supply" word is found in proximity to an "increase" word.
3. Repeat to obtain counts for "supply decrease," "demand increase," and "demand decrease."

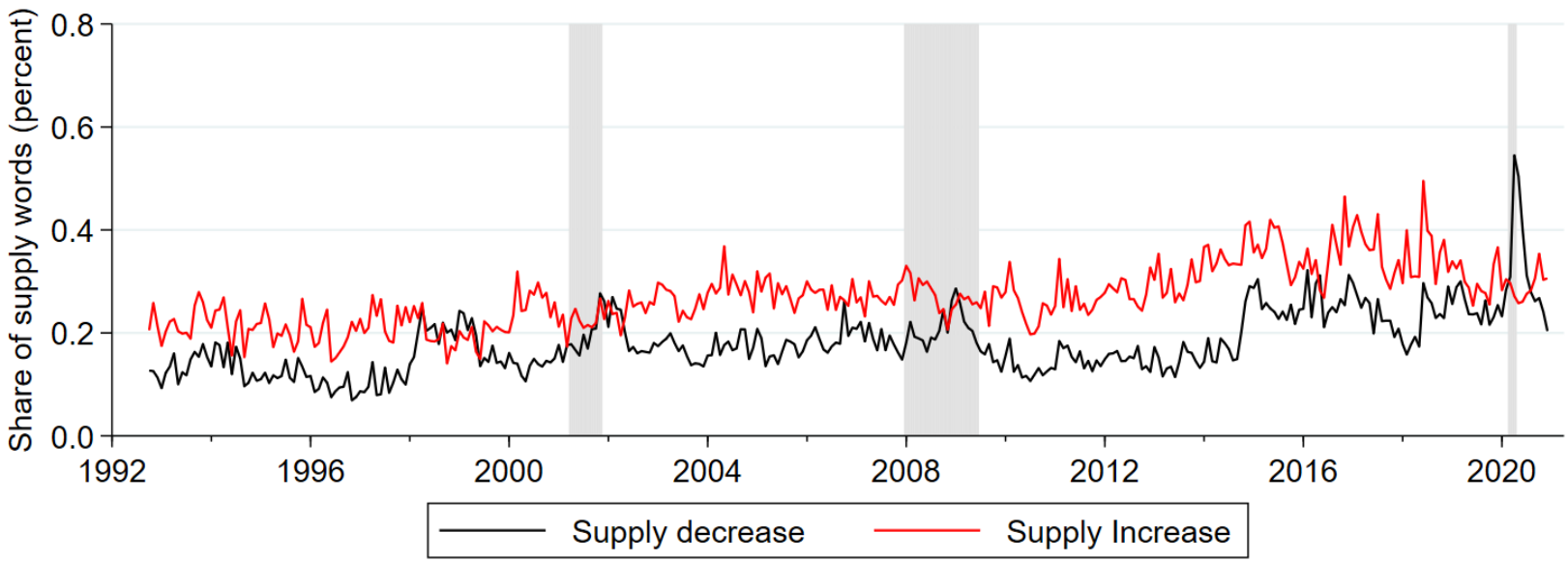
# Index Construction

$$Index_t = \frac{\frac{PhraseCount_t}{WordCount_t} - Mean(\frac{PhraseCount_{1995-2004}}{WordCount_{1995-2004}})}{StDev(\frac{PhraseCount_{1995-2004}}{WordCount_{1995-2004}})}$$

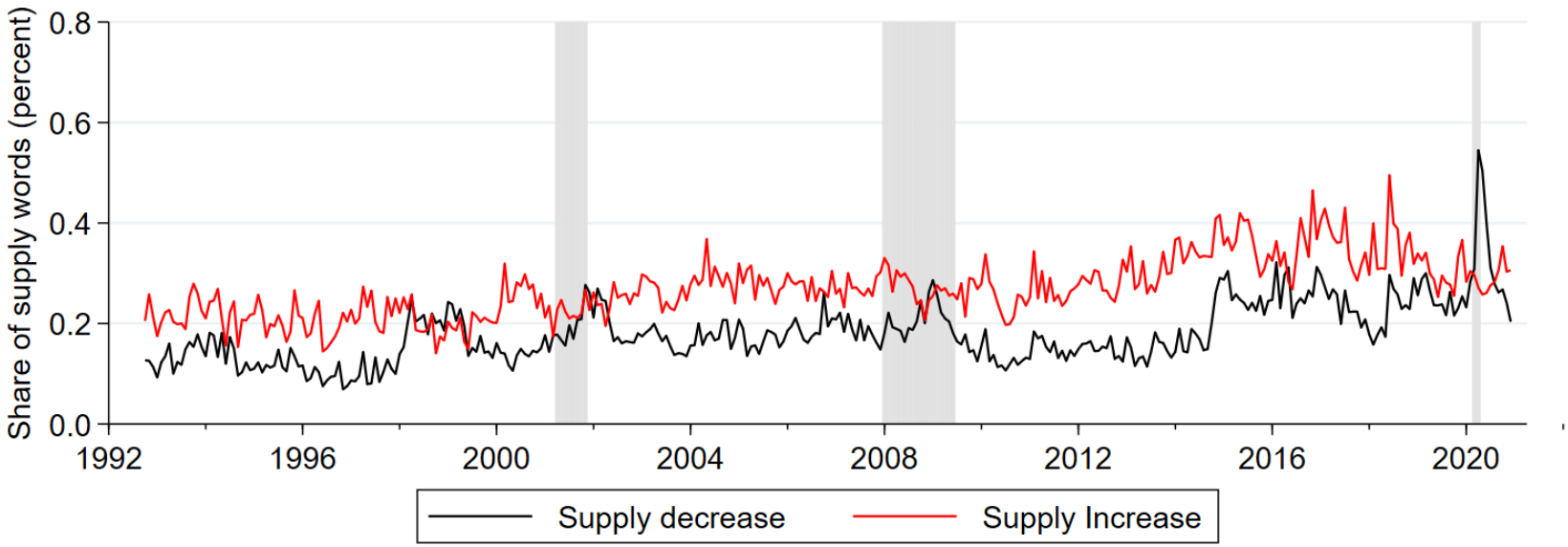
Obtain:

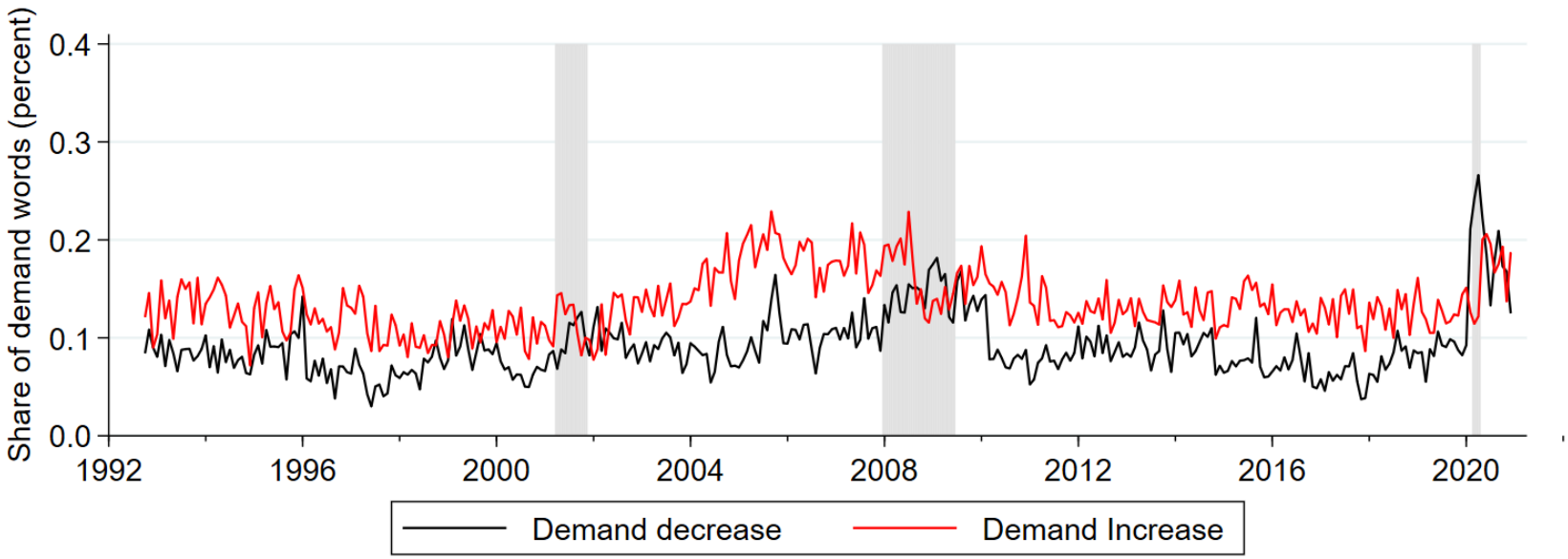
- 4 directional indexes
  1. $SI_t$ - Supply Increase
  2. $SD_t$ - Supply Decrease
  3. $DI_t$ - Demand Increase
  4. $DD_t$ - Demand Decrease
- 2 net indexes
  1. $Net\text{-}S_t$ - Net Supply
  2. $Net\text{-}D_t$ - Net Demand

# Directional Supply - Phrase Count Shares

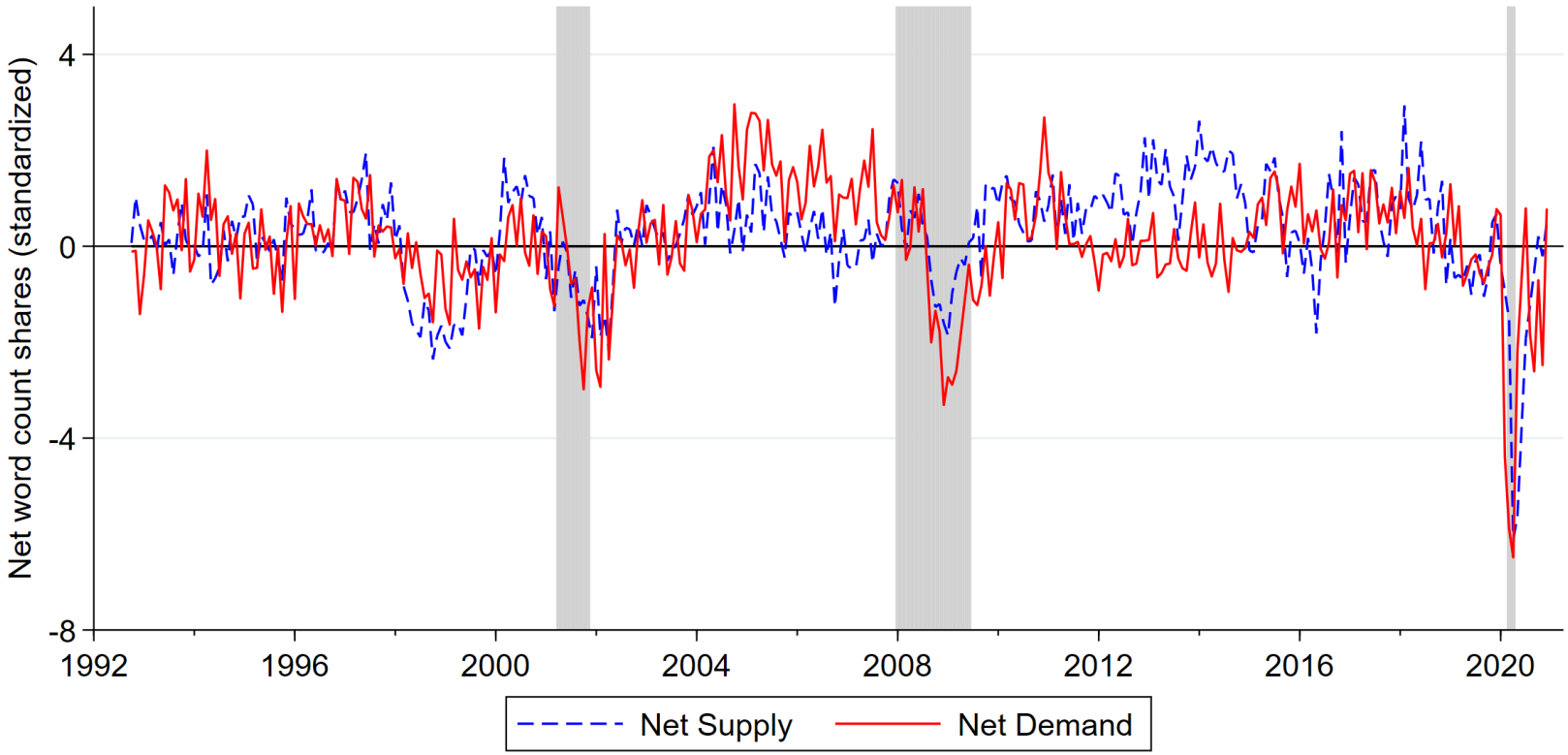


# Directional Demand - Phrase Count Shares



# Net Supply and Net Demand Indexes



# Oil production, real economic activity, and prices

In-sample regression:

$$y_t - y_{t-1} = \beta_0 + \beta_1 SI_t + \beta_2 SD_t + \beta_3 DI_t + \beta_4 DD_t + X_{t-1} + \epsilon_t$$

$$y_t - y_{t-1} = \beta_0 + \beta_1 NetS_t + \beta_2 NetD_t + X_{t-1} + \epsilon_t$$

where $X_{t-1}$ includes lagged values of changes in oil production, real economic activity, and oil prices.

# Oil production, real economic activity, and prices

## Table: Contemporaneous Movements

| | Oil production | | REA | | WTI spot | | WTI 12-month | |
|---|---|---|---|---|---|---|---|---|
| Supply increase | 3.32 | | 0.45 | | -14.75 | | -11.13 | |
| | (0.81)*** | | (0.43) | | (6.27)** | | (5.22)** | |
| Supply decrease | -2.75 | | -1.25 | | 6.58 | | 3.91 | |
| | (0.69)*** | | (0.41)*** | | (4.94) | | (4.08) | |
| Demand increase | -1.19 | | 1.65 | | 30.73 | | 26.37 | |
| | (0.73) | | (0.45)*** | | (7.65)*** | | (6.55)*** | |
| Demand decrease | 0.31 | | -1.38 | | -16.55 | | -14.31 | |
| | (0.59) | | (0.44)*** | | (5.32)*** | | (3.99)*** | |
| Net supply | | 2.50 | | 0.36 | | -10.19 | | -7.51 |
| | | (0.61)*** | | (0.32) | | (4.77)** | | (3.91)* |
| Net demand | | -1.11 | | 0.84 | | 22.55 | | 19.02 |
| | | (0.58)* | | (0.36)** | | (5.85)*** | | (4.85)*** |
| $R^2$ | .272 | .266 | .423 | .391 | .314 | .311 | .312 | .306 |
| Indexes $R^2$ share | .146 | .126 | .245 | .137 | .119 | .109 | .219 | .2 |
| $N$ | 323 | 323 | 323 | 323 | 323 | 323 | 323 | 323 |

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

$$y_t - y_{t-1} = \beta_0 + \beta_1 SI_t + \beta_2 SD_t + \beta_3 DI_t + \beta_4 DD_t + X_{t-1} + \epsilon_t$$

# Out of sample forecasting

Out of sample forecast model:

$$y_{t+h} - y_t = \alpha^h + \beta^h X_t + \epsilon_t^h$$

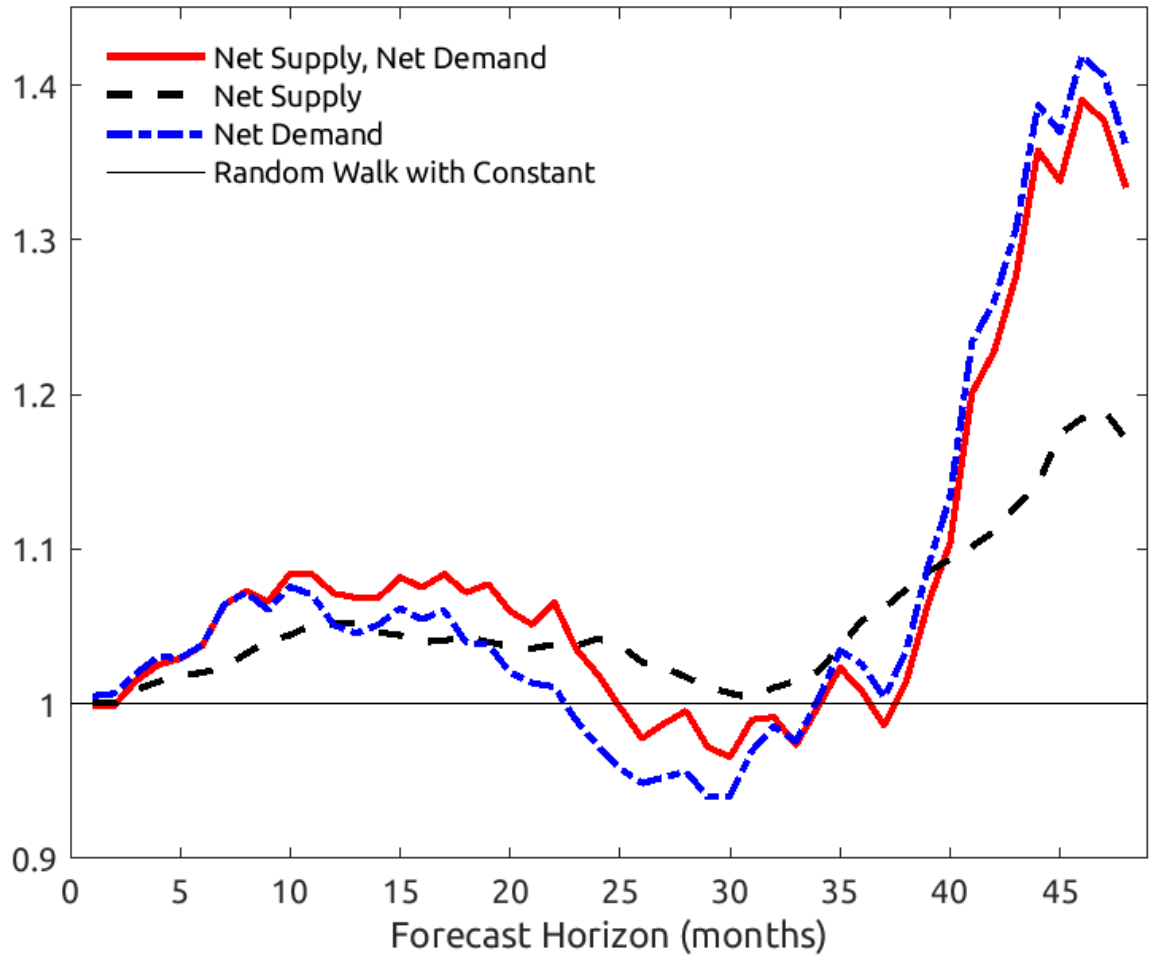
Benchmark random walk with constant:

$$y_{t+h} - y_t = c^h + \varepsilon_t^h$$

Evaluate using RMSFE Ratio:

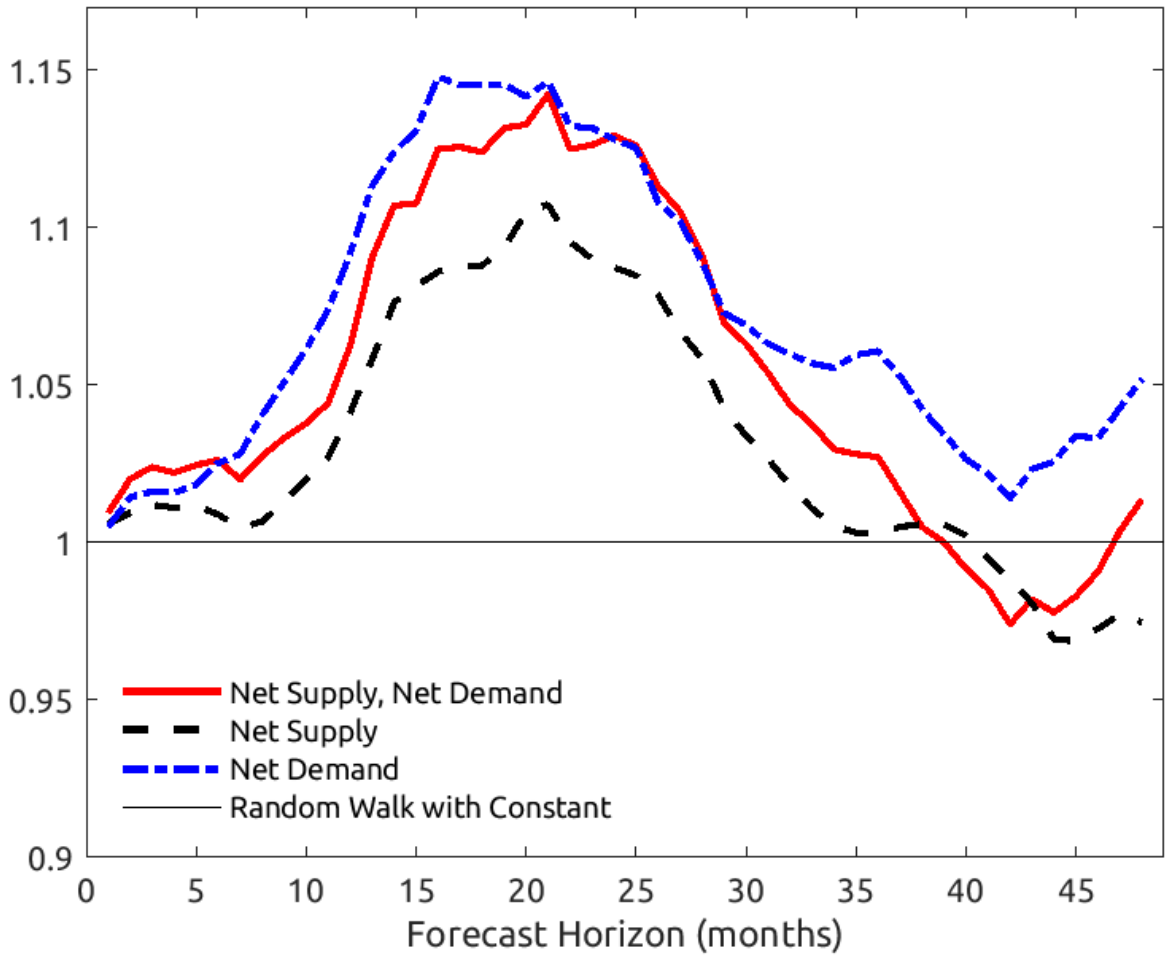
$$RelativeRMSFE^h = \frac{RMSFE^h_{X_t}}{RMSFE^h_{RW}}$$

# Oil production - Relative RMSFE




$$y_{t+h} - y_t = \alpha^h + \beta^h X_t + \epsilon_t^h$$

# Oil prices - Relative RMSFE




$$y_{t+h} - y_t = \alpha^h + \beta^h X_t + \epsilon_t^h$$

# Estimating a structural VAR of the Oil Market

Key advantages:

1. Higher-frequency and more promptly available data
2. Current and prospective information on market conditions

Assume that in the short-run, the supply and demand sides of the oil market only interact with each other via prices.

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

# Weekly SVAR Model

## Table: Short-Run Dynamics

| | $SI_t$ (1) | $SD_t$ (2) | $DI_t$ (3) | $DD_t$ (4) | $\Delta ln(Price_t)$ (5) |
|---|---|---|---|---|---|
| Panel B: Weekly model | | | | | |
| $SI_t$ | - | - | - | - | -7.575 |
| | | | | | (0.864)*** |
| $SD_t$ | - | - | - | - | 6.119 |
| | | | | | (0.608)*** |
| $DI_t$ | - | - | - | - | 4.183 |
| | | | | | (0.523)*** |
| $DD_t$ | - | - | - | - | -2.943 |
| | | | | | (0.438)*** |
| $\Delta ln(Price_t)$ | 0.089 | -0.059 | -0.038 | 0.033 | - |
| | (0.012)*** | (0.008)*** | (0.009)*** | (0.011)*** | |
| Sample | 4/06/1994 - 12/23/2020 | | | | |
| Number observations | 1395 | | | | |

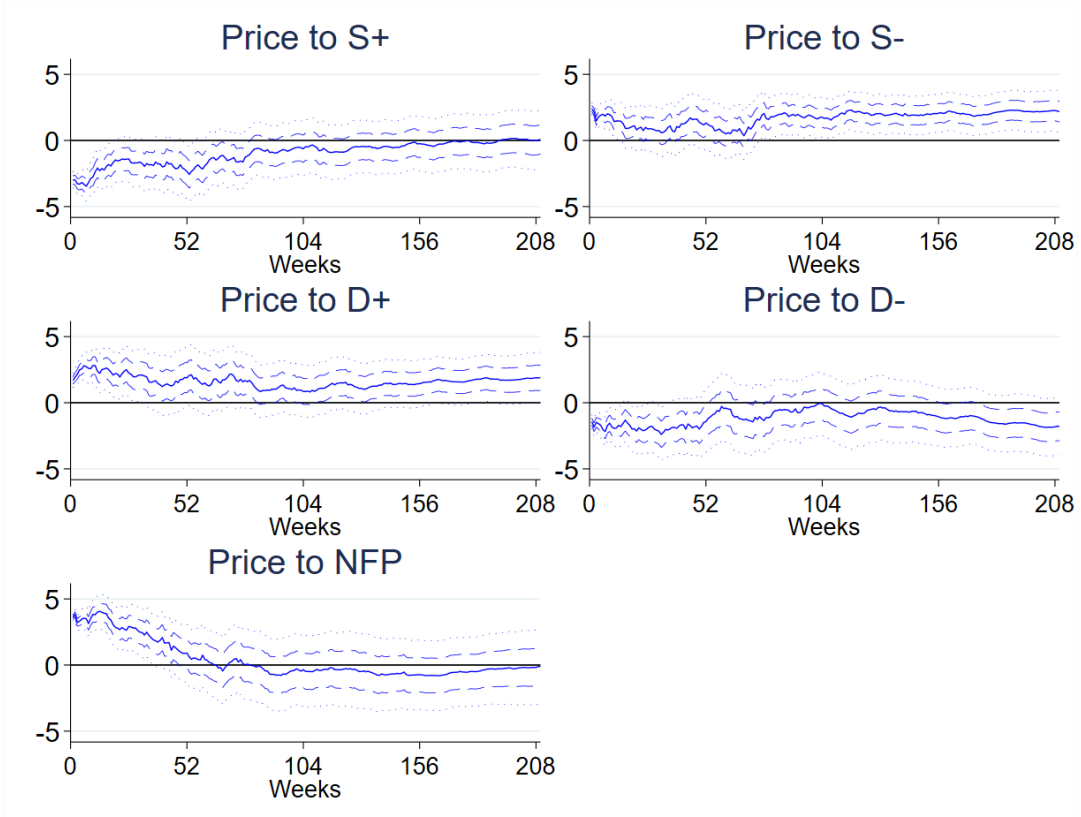
# Weekly SVAR Model

Table: Forecast Error Variance Decomposition

| Weekly model | | | | | | | |
|---|---|---|---|---|---|---|---|
| | $S^+$ | $S^-$ | $S$ | $D^+$ | $D^-$ | $D$ | $NFP$ |
| $t = 1$ | 7.3 | 4.2 | 11.5 | 32.7 | 16.5 | 49.2 | 39.3 |
| $t = \infty$ | 12.9 | 11.0 | 23.8 | 25.9 | 17.8 | 43.6 | 32.6 |

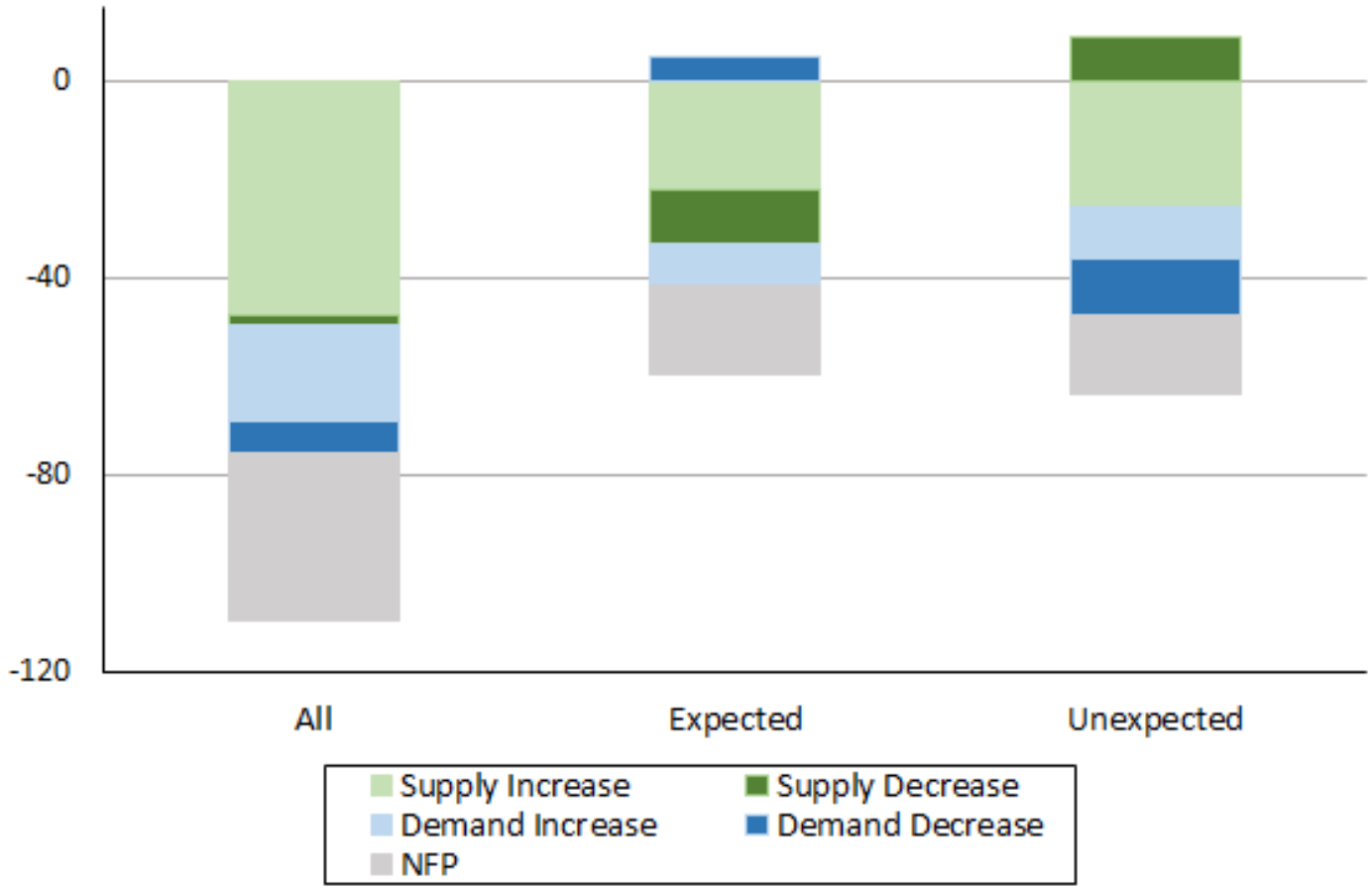
# Weekly SVAR Model - Impulse Responses

Figure: Oil Price Response to Structural Shocks



# The Oil Price Collapse of 2014

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




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

# Conclusions

- Using textual analysis, we construct indexes containing information about supply and demand developments in the oil market.
  - Our indexes correlate well with existing measures of oil supply, demand, and prices.
  - The indexes contain substantial information about current and future oil price movements.

- Used the new indexes to estimate a structural VAR model of the oil market.
  - Results are in line with economic theory and are of plausible magnitudes.
  - Historical decomposition of well-known episodes in the oil market provide further evidence that our indexes contain substantial information about prospective oil price movements.