How does the crowd sentiment of investors affect international crude oil prices?

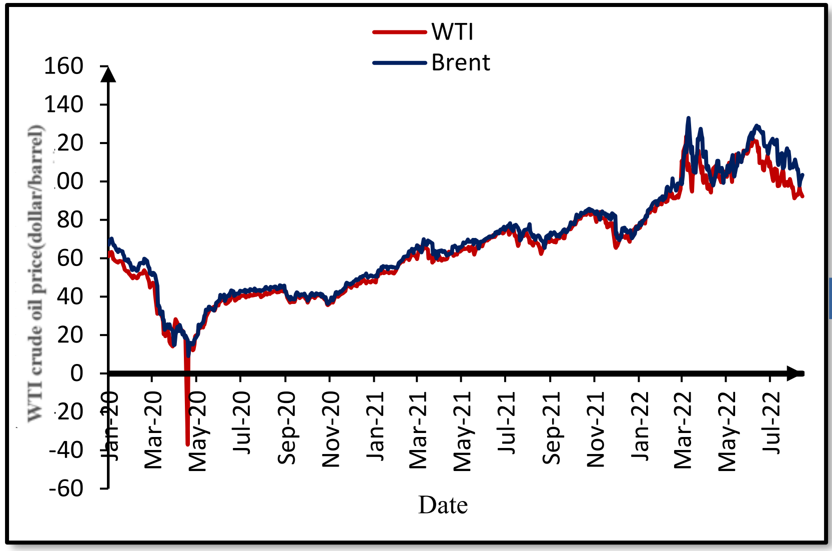
Evidence from COVID-19 pandemic

Lu-Tao ZHAO
Ltzhao@bit.edu.cn

Dec. 14th, 2022

CONTENTS

Background



(Crude oil market)

Influence factors

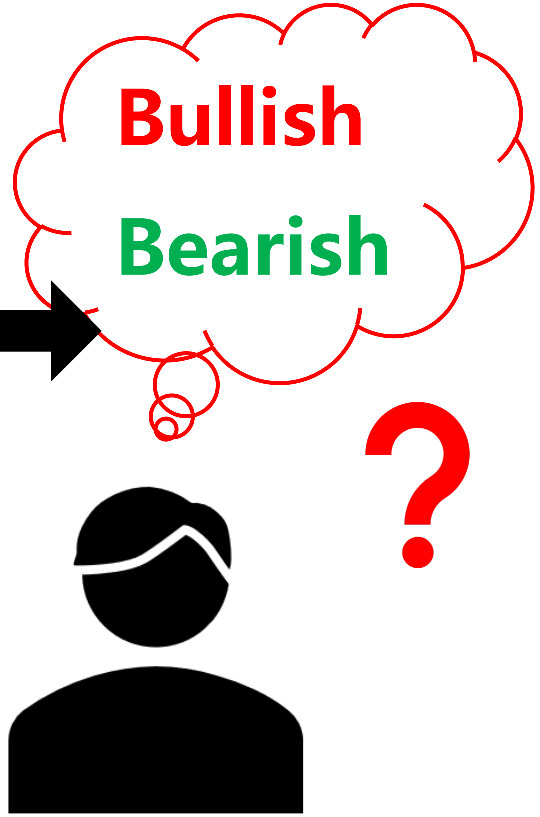
Long-term trends

- Supply
- Demand

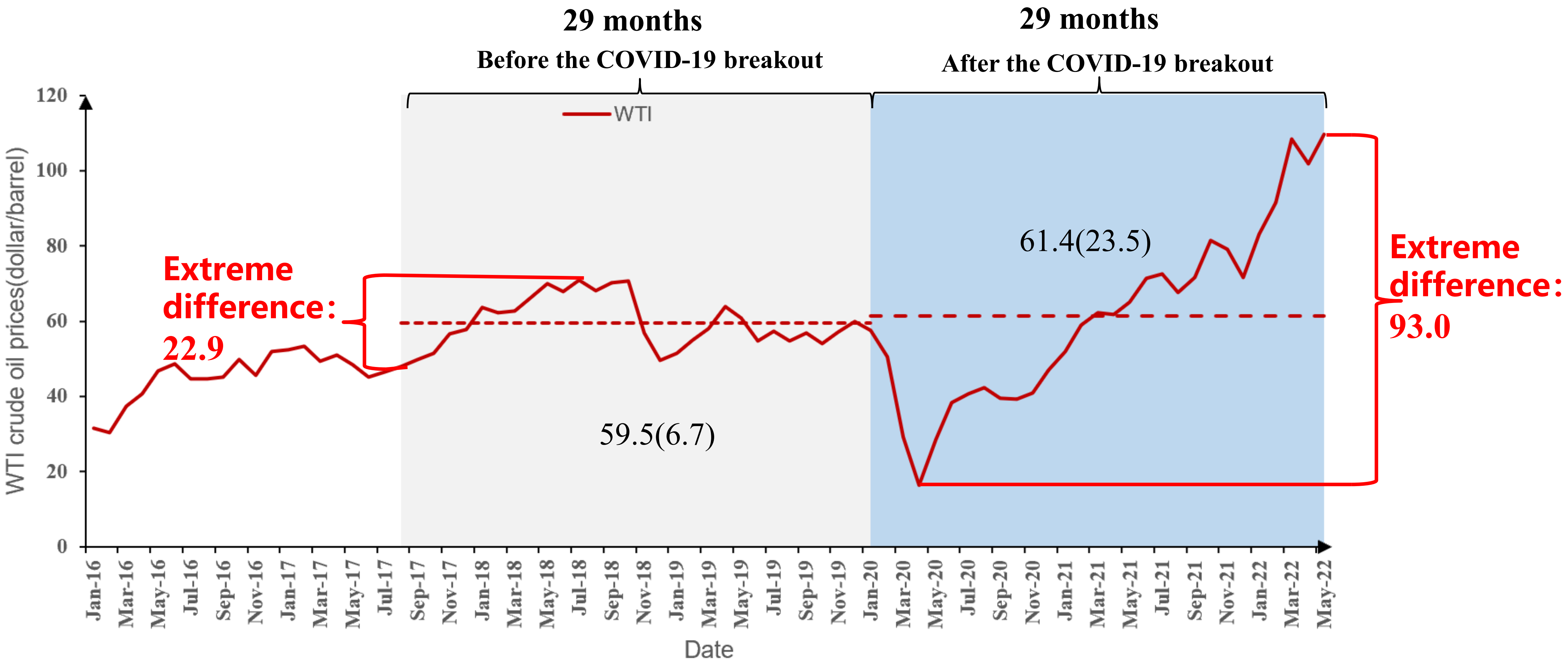
Short-term fluctuations

- USD exchange rate
- Big country games
- Geopolitics
- Emergency
- Market speculation

……



Background



29 months
Before the COVID-19 breakout
29 months
After the COVID-19 breakout
WTI
Extreme difference: 22.9
59.5(6.7)
61.4(23.5)
Extreme difference: 93.0
WTI crude oil prices(dollar/barrel)
0
20
40
60
80
100
120
Jan-16
Mar-16
May-16
Jul-16
Sep-16
Nov-16
Jan-17
Mar-17
May-17
Jul-17
Sep-17
Nov-17
Jan-18
Mar-18
May-18
Jul-18
Sep-18
Nov-18
Jan-19
Mar-19
May-19
Jul-19
Sep-19
Nov-19
Jan-20
Mar-20
May-20
Jul-20
Sep-20
Nov-20
Jan-21
Mar-21
May-21
Jul-21
Sep-21
Nov-21
Jan-22
Mar-22
May-22
Date

Literature Review

CEEP-BIT

★ **Online social networks contain a large number of investment trends and play an increasingly prominent role in disseminating financial market information.**

- **Survey data：** Survey data on bullish and bearish differences of more than 130 stock commentators and consumer confidence in the market, etc. (Gregory and Michael, 2005; Lemmon and Portniaguina, 2006; Zhang et al., 2007).

- **Stock market data：** The number of IPOs, the first-day earnings volume, new account openings, closed-end fund discounts and stock turnover, etc. (Baker and Wurgler, 2006; Ljungqvist et al., 2006; Han and Wu, 2007; Yi et al., 2011).

- **Internet information：** Negative sentiment in media news, extracting investor sentiment from Yahoo stock reviews and Google search index, etc. (Das and Chen, 2007; Tetlock, 2007; Li et al., 2017; Yao et al., 2017; Wang et al., 2018).

Literature Review

CEEP-BIT

★ **The selection of search keywords is subjective and one-sided. Using the textual term mining method to extract search keywords can improve the credibility of sources.**

- **Manual labeling：** Lexical annotation and hierarchical phrase extraction, etc. (Abney, 1991; Lopez et al., 2004; Kazemi et al., 2017).

  × **Consume a lot of manpower, material resources and time, and the economic cost is high.**

- **Statistical method：** Frequency statistics, etc. (El-Kishky, 2014; Parameswaran, 2019).

  × **The quality of the extracted phrases depends on established rules.**

- **Machine learning：** WPDV model, voting algorithm, etc. (Halteren, 2000; Shen and Sarkar, 2005).

  × **Lack of the expansion of text corpora for different domains.**

Literature Review

CEEP-BIT

The sentence-level sentiment analysis is not good at describing the evolution of investor sentiment. The aspect-level sentiment analysis considers the interrelationship between aspect terms and contexts.

Table 1. Summary of aspect-level sentiment analysis methods

Classification		Model	Characteristic	References
Sentiment lexicon		SentiWordNet	Rely on predefined dictionaries, and the accuracy is influenced by the richness of the lexicon.	Deng and Jing, 2019 Federici and Dragoni, 2016 Singh et al., 2013
		Stanford NLP		
		NTUSD		
Supervisory model	statistical method	Daïve Bayes	Weak semantic understanding and poor results in aspect classification.	Alsmadi et al., 2018 Brun et al., 2016 Kumar et al., 2016
		Bayes Networks		
		Decision Trees		
	Deep learning	LSTM、Attention based LSTM	The data sample topics significantly affect the application domain of sentiment classifiers.	Hu et al., 2021 Shuang et al., 2019 Wang et al., 2016
		BPNN,RNN,CNN		

Literature Review

CEEP-BIT

Sentence-level Sentiment Analysis VS Aspect-level Sentiment Analysis

Crude oil prices **rise** on OPEC Plus production **cuts**.

sentimentScore(Sentence)= sentimentScore(**rise**)

+ sentimentScore(**cut**)

= 4.0-4.0 =**0**

Topic=Search keywords="Crude oil"

sentimentScore(Crude oil)=0

Literature Review

CEEP-BIT

Sentence-level Sentiment Analysis VS Aspect-level Sentiment Analysis

Crude oil prices rise on OPEC Plus production cuts.

"aspect1" "opinion1" "aspect2" "opinion2"

sentimentScore(price)= sentimentScore(rise)= 4.0

sentimentScore(production)= sentimentScore(cuts)= -4.0

Literature Review

CEEP-BIT

★ **The significant impact of investor sentiment on international crude oil prices has a rich theoretical basis.**

- **Global explanatory:** The GARCH, SVAR, and risk spillover models are used to verify that investor sentiment can significantly affect the fluctuation of crude oil prices (Baker et al., 2012; Yao et al., 2017; Su et al., 2018; Li et al., 2021).

- **Predictive validity:** Artificial intelligence forecasting models such as SVR, LSTM, and CNN have found that investor sentiment can provide new information to oil price forecasts that is not available in statistics (He and Casey, 2015; Wang et al., 2018; Li et al., 2019; Zárate and Santiago, 2019).

- **Changes in investor sentiment under the extreme events have not been considered.**

Contribution

CEEP-BIT

1

An online text mining method was designed to quantify the aspect-level investor sentiment of international crude oil market.

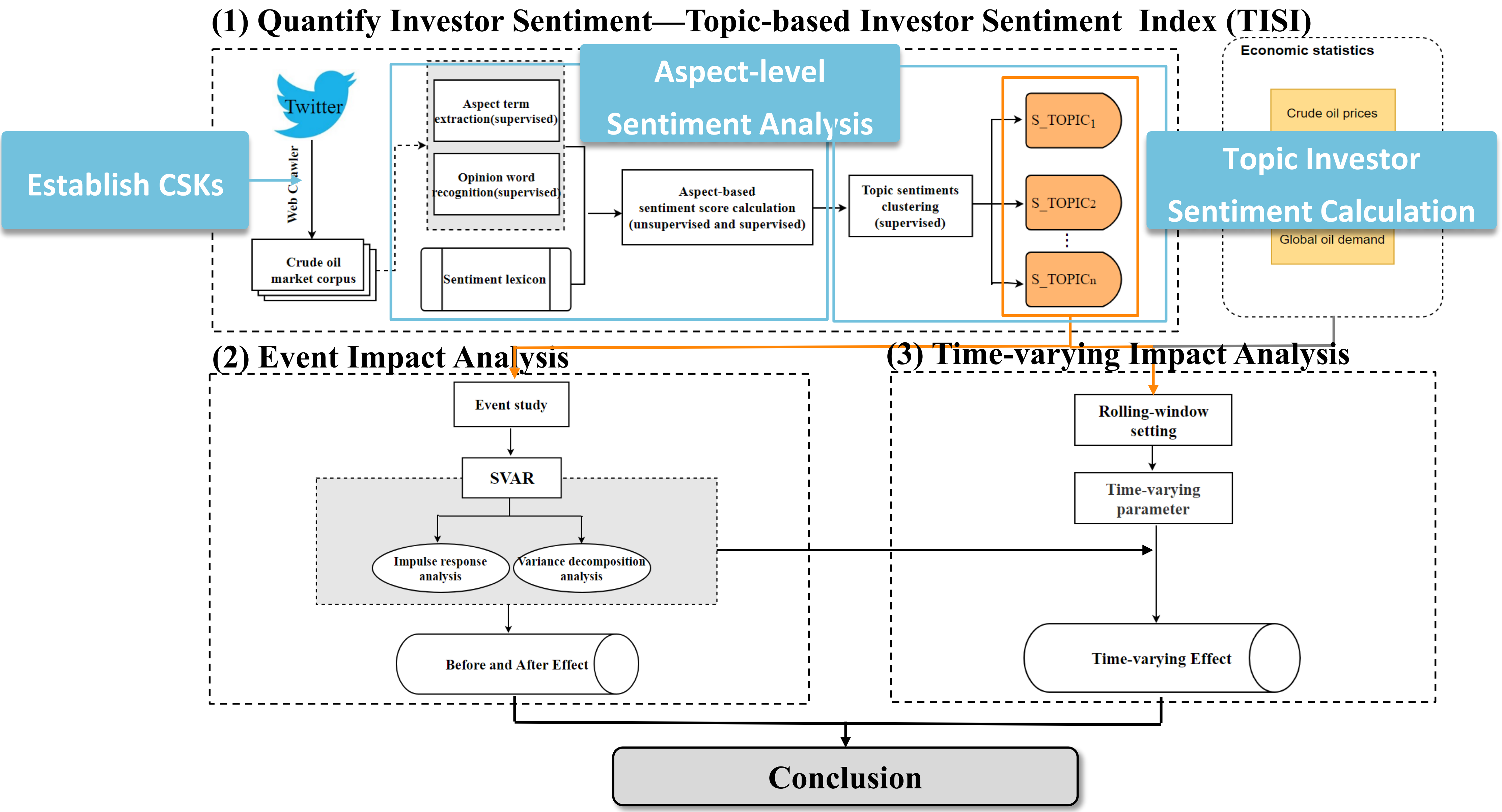
2

Using the event study and rolling window SVAR method, we explored the mechanisms and time-varying effects of investor sentiment before and after the COVID-19 breakout.

CONTENTS

CEEP-BIT

Framework


(1) Quantify Investor Sentiment—Topic-based Investor Sentiment Index (TISI)
Establish CSKs
Twitter
Crude oil market corpus
Aspect-level Sentiment Analysis
Aspect term extraction(supervised)
Opinion word recognition(supervised)
Sentiment lexicon
Aspect-based sentiment score calculation (unsupervised and supervised)
Topic sentiments clustering (supervised)
S_TOPIC1
S_TOPIC2
S_TOPICn
Economic statistics
Crude oil prices
Topic Investor Sentiment Calculation
Global oil demand
(2) Event Impact Analysis
Event study
SVAR
Impulse response analysis
Variance decomposition analysis
Before and After Effect
(3) Time-varying Impact Analysis
Rolling-window setting
Time-varying parameter
Time-varying Effect
Conclusion

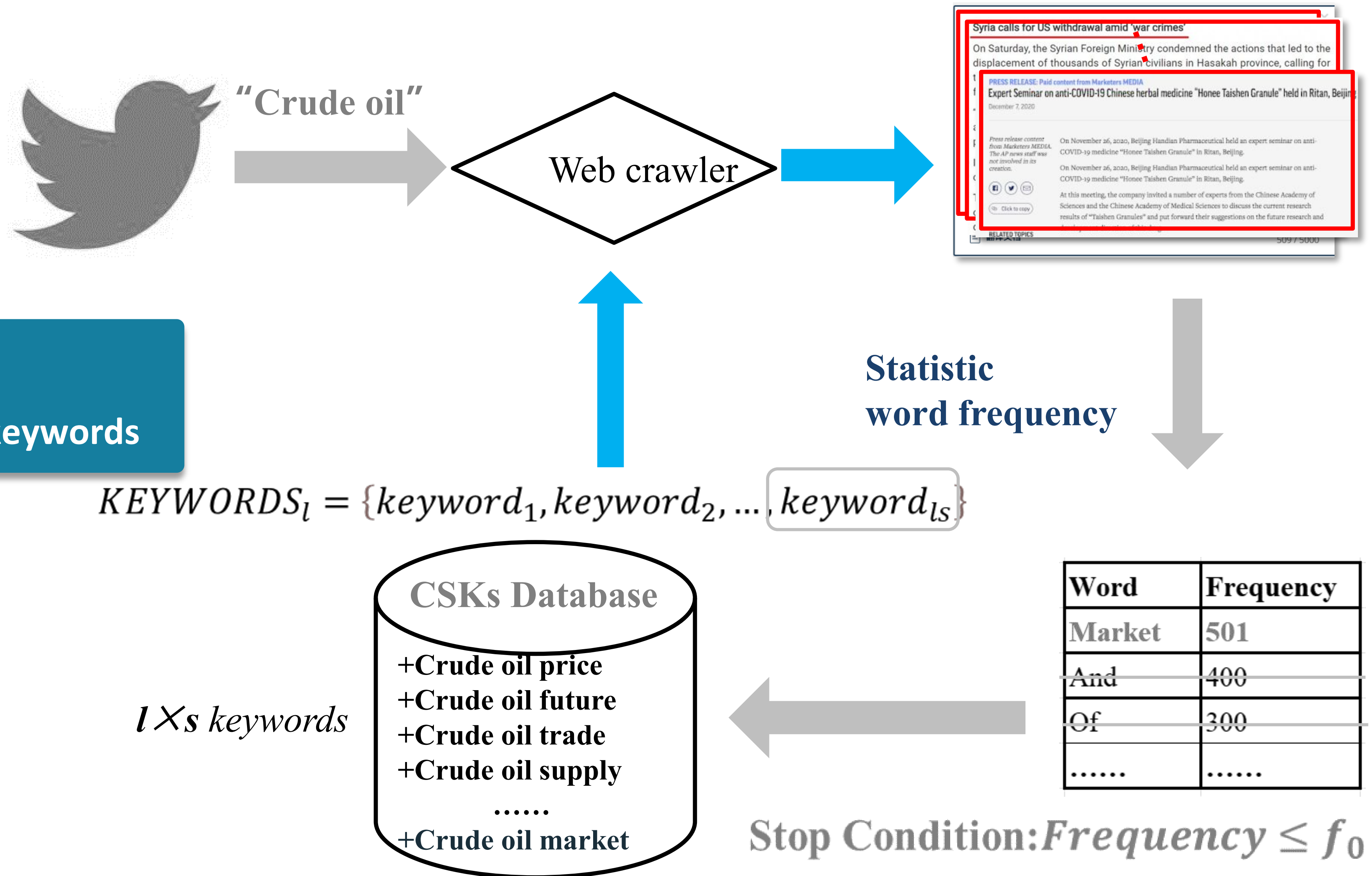
(1) Quantify Investor Sentiment

CEEP-BIT

➢ **Characteristic Search Keywords (CSKs) Establishment**

STEP 1:
Crawl with seed keyword

STEP 2:
Extract features and expand keywords



$$KEYWORDS_l = \{keyword_1, keyword_2, \ldots, \boxed{keyword_{ls}}\}$$

$l \times s$ *keywords*

Word	Frequency
Market	501
~~And~~	~~400~~
~~Of~~	~~300~~
……	……

Stop Condition: $Frequency \le f_0$

(1) Quantify Investor Sentiment

CEEP-BIT

➢ **Aspect-level Sentiment Analysis**

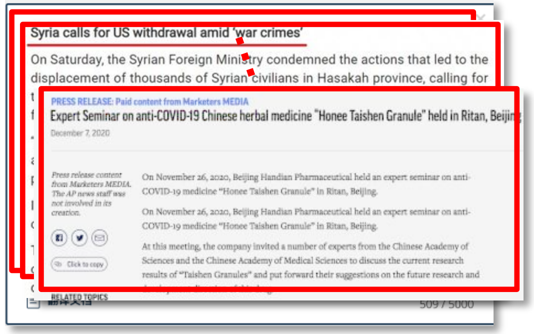
STEP 3:

Label aspect terms and identify sentiment words

STEP 4:

Composite sentiment score calculation

Training set:


Syria calls for US withdrawal amid 'war crimes'
On Saturday, the Syrian Foreign Ministry condemned the actions that led to the displacement of thousands of Syrian civilians in Hasakah province, calling for
PRESS RELEASE: Paid content from Marketers MEDIA
Expert Seminar on anti-COVID-19 Chinese herbal medicine "Honee Taishen Granule" held in Ritan, Beijing
December 7, 2020
On November 26, 2020, Beijing Handian Pharmaceutical held an expert seminar on anti-COVID-19 medicine "Honee Taishen Granule" in Ritan, Beijing.

Preprocessing

$ASPECTs = \{a_1, a_2, \dots, a_i, \dots, a_M\}$

$POSs = \{p_1, p_2, \dots, p_j, \dots, p_N\}$

Brent crude oil price will $\underbrace{\text{rise.}}_{\text{Verb}}$

Crude oil output will $\underbrace{\text{cut.}}_{\text{Verb}}$

……

Test set:

ASPECTs

POSs

Crude oil $\underbrace{\text{price}}_{\text{“aspect”}}$ $\underbrace{\text{recoveries remain capped}}_{\text{“opinion”}}$.

SentiWordNet

pos_score=0.6
neg_score=0.0
opinionScore(recovery)=1.0

pos_score=0.2
neg_score=0.0
opinionScore(remain)=1.0

pos_score=0.2
neg_score=0.6
opinionScore(cap)= -0.5

(1) Quantify Investor Sentiment

CEEP-BIT

- **Aspect-level Sentiment Analysis**

STEP 5:
Calculate sentimentScore of aspect term

sentimentScore(price)=1.0+0.5-0.2=1.3

Crude oil price recoveries remain capped.

"aspect" "opinion"

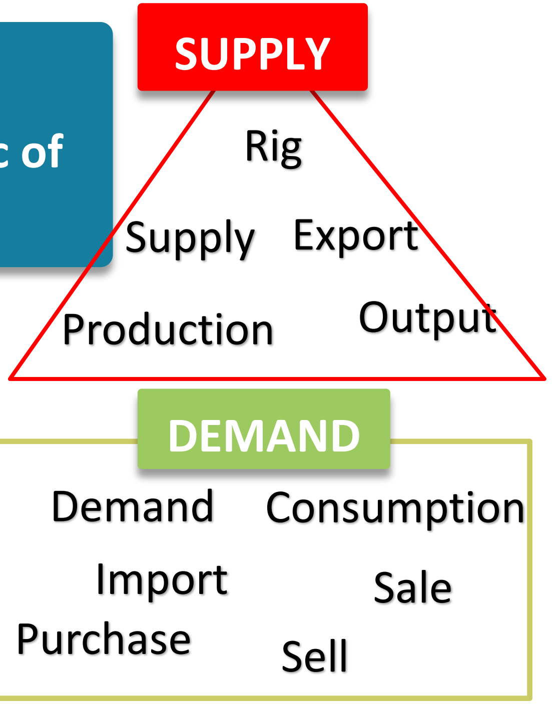
Distance: 1 2 3

$$\text{opinionScore(recovery)}=\frac{1.0}{1}=1.0$$

$$\text{opinionScore(remain)}=\frac{1.0}{2}=0.5$$

$$\text{opinionScore(cape)}=\frac{-0.5}{3}=-0.2$$

STEP 6:
Cluster the topic of aspect terms



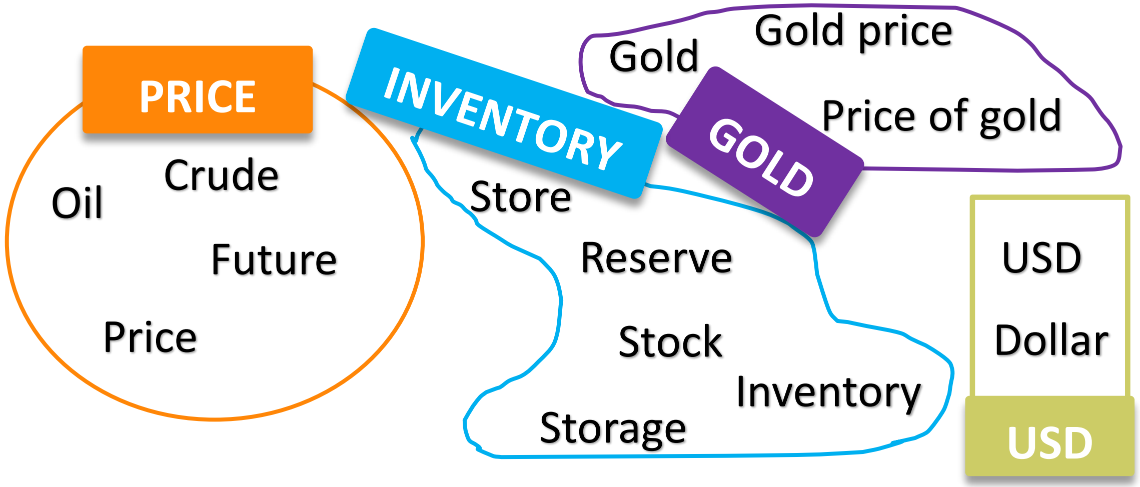
- **Calculate TISI**

STEP 7:
Calculate the Topic-based Investor Sentiment Index (TISI)

The monthly sequence of TISI is calculated as:

$$S\_TOPIC_{i,t}=\frac{\sum_{a_j \in TOPIC_i} sentimentScore(a_j)\ on\ month\ t}{Number\ of\ tweets\ for\ TOPIC_i\ on\ month\ t}$$

where a_j is the aspect term in $TOPIC_i$.



(2)Event Impact Analysis

CEEP-BIT

We define a SVAR model with four variables related to the **crude oil market** as:

$$\varepsilon_t = \begin{bmatrix} \varepsilon_t^{Supply} \\ \varepsilon_t^{Demand} \\ \varepsilon_t^{Senti} \\ \varepsilon_t^{Price} \end{bmatrix} = B_0^{-1} u_t = \begin{bmatrix} b_{11} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} \begin{bmatrix} u_t^{Supply} \\ u_t^{Demand} \\ u_t^{Senti} \\ u_t^{Price} \end{bmatrix}$$

Impulse Response Function(**IRF**):

$$IRF_{ij}^{(q)} = \frac{\partial X_{i,t+q}}{\partial u_{jt}}, q = 0,1,2,\ldots$$

Relative Variance Contribution(**RVC**):

$$RVC_{j\to i}(\infty) = \frac{\sum_{q=0}^{\infty} (b_{ij}^{(q)})^2 \sigma_{jj}}{\mathrm{var}(X_i)} = \frac{\sum_{q=0}^{\infty} (b_{ij}^{(q)})^2 \sigma_{jj}}{\sum_{j=1}^{k} \{\sum_{q=0}^{\infty} (b_{ij}^{(q)})^2 \sigma_{jj}\}}, i, j = 1,2,\ldots,k$$

(3)Time-varying Impact Analysis

CEEP-BIT

Considering the cyclical and regime changes in the crude oil market, the SVAR model is estimated at each fixed window $[k, k+h-1]$ using a rolling window approach:

$$B_0 X_\tau = \beta + \sum_{i=1}^{p} B_i X_{\tau-i} + u_\tau$$

Then, the slide length is set to 1, allowing the parameter to be time-varying.

CONTENTS

CEEP-BIT

(1) Data and Description

CEEP-BIT

- **Twitter sample**

Time horizon:

Total **2,892,589 tweets** from **Jan. 2012 to May. 2022**

Second level high frequency data

Tweet information:

"Tweet id", "Tweet created at", "Tweet full text", "Tweet URL", etc.

Table 2. CSKs for four types.

Type	CSK
Crude oil terms	Petroleum/ Crude oil market/ Crude oil business/ Crude oil energy/ Crude oil trade/ Crude oil trading
Product oil terms	Gas/ Gasoline/ Petrol/ Fuel
Oil aspect terms	Crude oil price/ Crude oil future/ Crude oil demand/ Crude oil supply/ Crude oil production/ Crude oil product/ Crude oil output/ Crude oil inventory/ Crude oil stock/ Crude oil reserve/ Crude oil storage/ Crude oil export/ Crude oil import
Crude oil related terms	WTI/ Brent/ OPEC

(1) Data and Description

➢ Oil market data

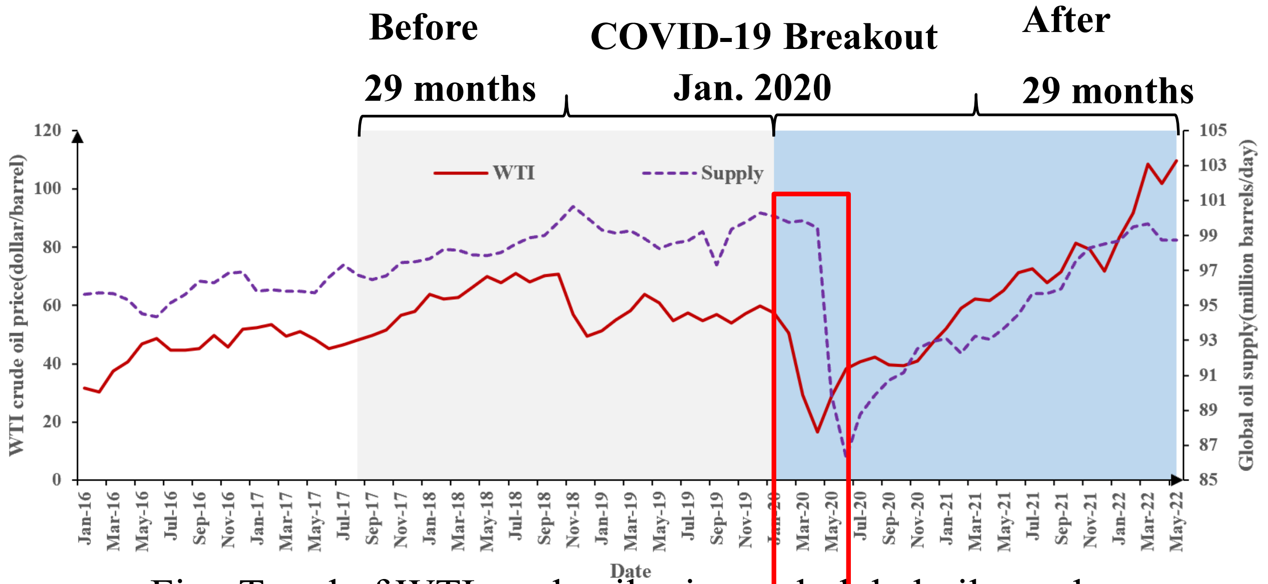


Fig. Trend of WTI crude oil price and global oil supply.

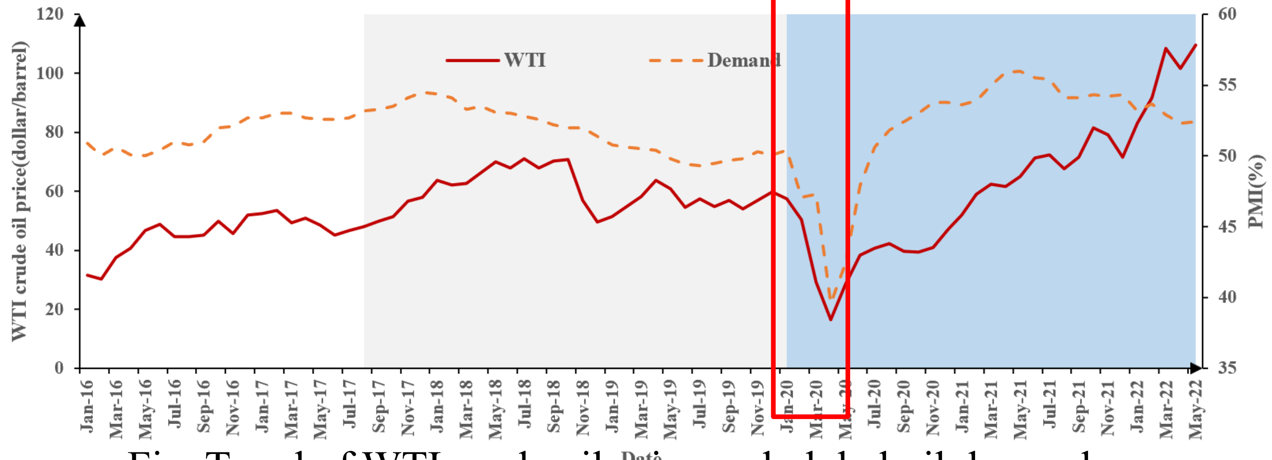


Fig. Trend of WTI crude oil price and global oil demand.

Table 3. Descriptive statistic.

	WTI Spot price	Supply Global oil production	Demand PMI
Mean	67.0417	94.6229	51.6232
Median	59.8800	95.1400	51.8000
Max	109.5500	100.6400	56.0000
Min	16.5500	86.2900	39.6000
Std.	22.9817	3.5970	2.2168
Skewness	0.3122	-0.1107	-1.7790
Kurtosis	1.9169	1.7377	10.6387

Data source: Wind

*Monthly data from Jan. 2012 to May. 2022

(1) Data and Description

CEEP-BIT

➢ **Proxy variables for investor sentiments**

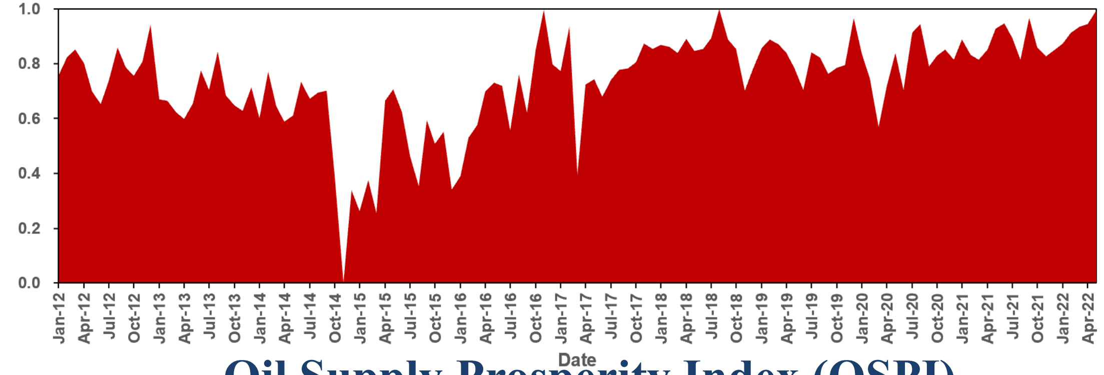
Table 4. Description of investor sentiment proxy variables.

Name	Meaning and Calculation	Data Source
BW	A composite sentiment index is constructed based on six basic sentiment indicators.	http://people.stern.nyu.edu/jwurgler
AAII	The sentiment is calculated as the spread between the percentage of bullish investors and that of bearish investors.	http://www.aaii.com
CCI	Indicators reflecting the development of household consumption and future savings are constructed based on perceptions of the expected financial position of U.S. households, general economic conditions, unemployment, and the ability to save.	https://data.oecd.org/leadind/consumer-confidence-index-cci.htm
DHI	About 10% of daily tweets posted on Twitter are randomly selected for a happiness rating of 1-9.	http://hedonometer.org/index.html
OVX	Presented by the Chicago Board Options Exchange (CBOE), it captures the uncertainty of market participants about the future volatility of oil prices.	http://www.cboe.com
VIX	Proposed by the Chicago Board Options Exchange (CBOE) to measure market expectations of the near-term volatility implied by stock index option prices.	http://www.cboe.com
TEU	A standardized measure of the number of tweets on Twitter containing keywords related to "uncertainty" and "economy".	http://www.policyuncertainty.com/twitter_uncert.html
TMU	A standardized measure of the number of tweets on Twitter that contain keywords related to "uncertainty" and "market".	http://www.policyuncertainty.com/twitter_uncert.html
GS	It is weighted by the Google search volume of 11 related keywords whose correlation coefficient with "crude oil price" is greater than 0.9.	https://trends.google.com/trends

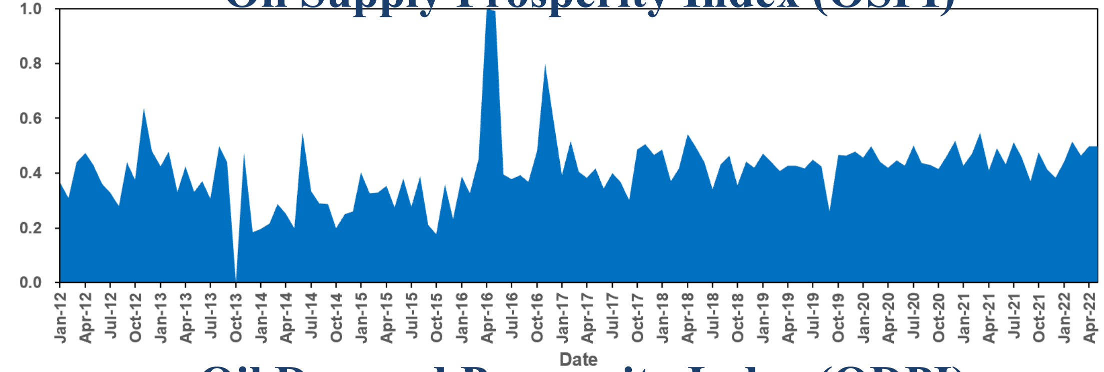
(2) Calculate the TISIs (Topic-based Investor Sentiment Index)

CEEP-BIT

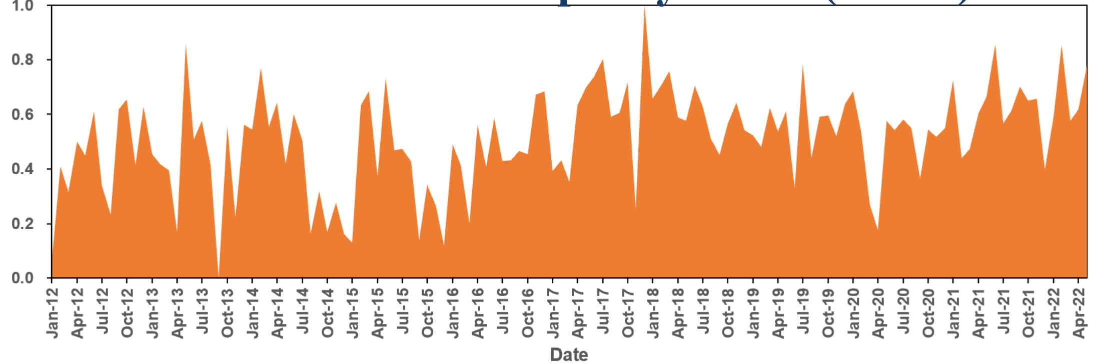
Oil Price Prosperity Index (OPPI)



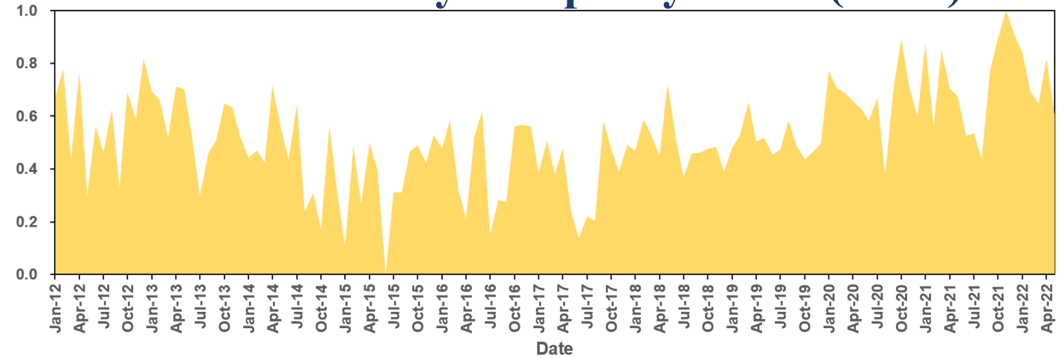
Oil Supply Prosperity Index (OSPI)



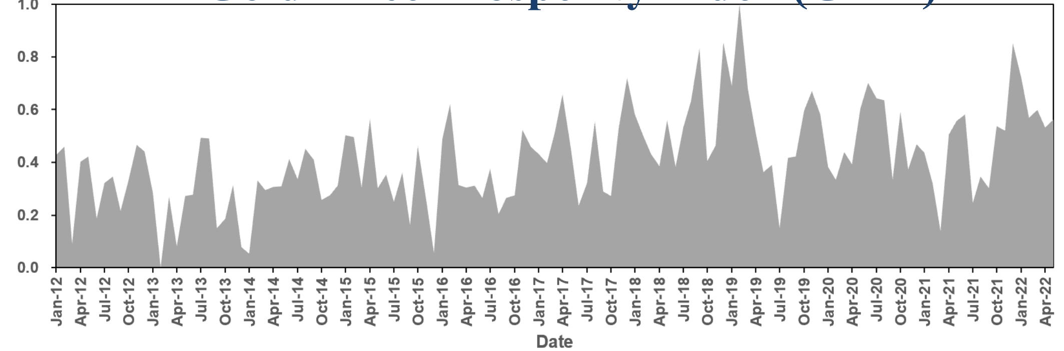
Oil Demand Prosperity Index (ODPI)



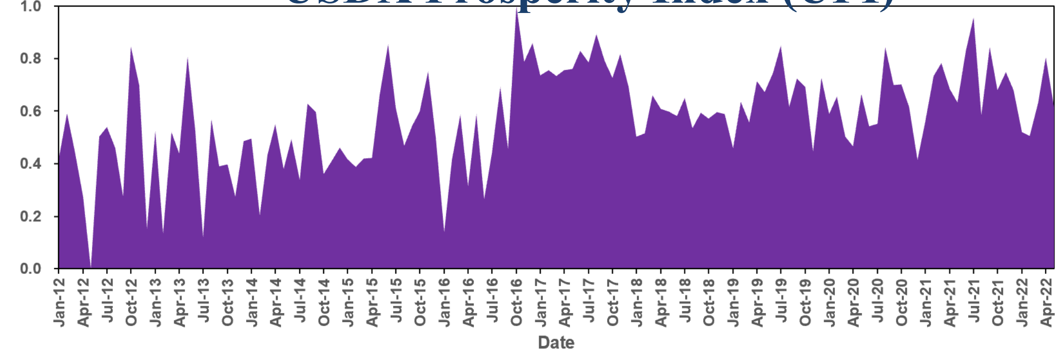
Oil Inventory Prosperity Index (OIPI)



Gold Price Prosperity Index (GPPI)



USDX Prosperity Index (UPI)



(3)How does TISIs affect crude oil prices?

The impulse response results of crude oil prices to one standard deviation of **OPPI, OSPI, ODPI, OIPI, GPPI and UPI shocks** based on a four-variable SVAR model are shown as follows：

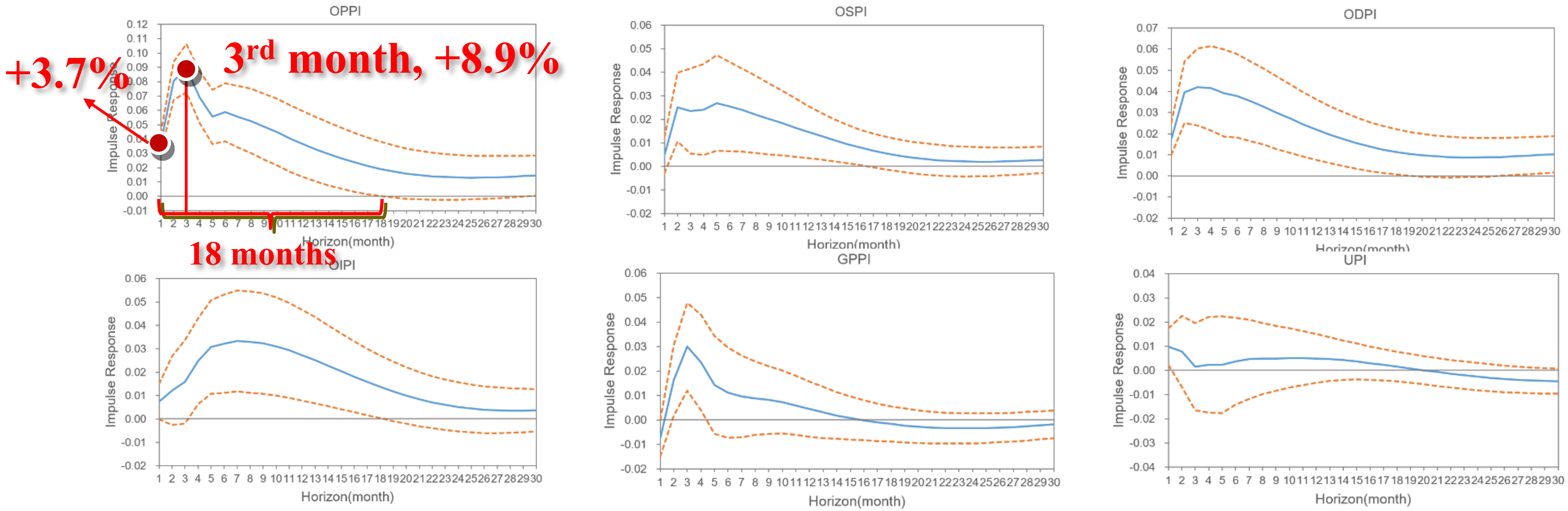


Fig. Responses of WTI crude oil prices to different shocks.

◆**The impact of price-topic investor sentiment index on crude oil prices is the great effective, high timely and long-term correlated.**

(3)How does TISIs affect crude oil prices?

CEEP-BIT

The impulse response results of WTI crude oil prices to one standard deviation of **OPPI, OSPI, ODPI, OIPI, GPPI and UPI shocks** based on a four-variable SVAR model are shown as follows：

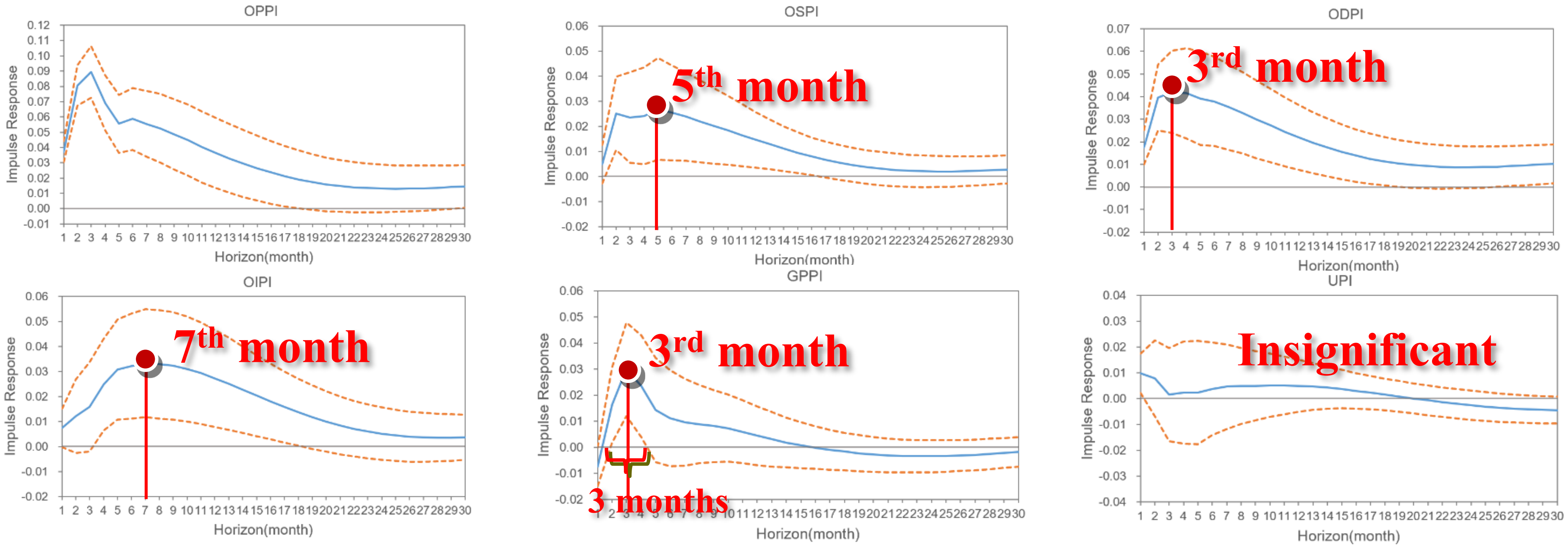


Fig. Responses of WTI crude oil prices to different shocks.

◆**TISIs about fundamental factors(Supply, Demand and inventory) have a stronger impact on oil prices than the other factors(Gold price and USDX).**

(3)How does TISIs affect crude oil prices?

CEEP-BIT

- **Compare with supply and demand factors**

◆ **OPPI can explain the longer term of oil price fluctuations and has a stronger effect than supply and demand factors.**

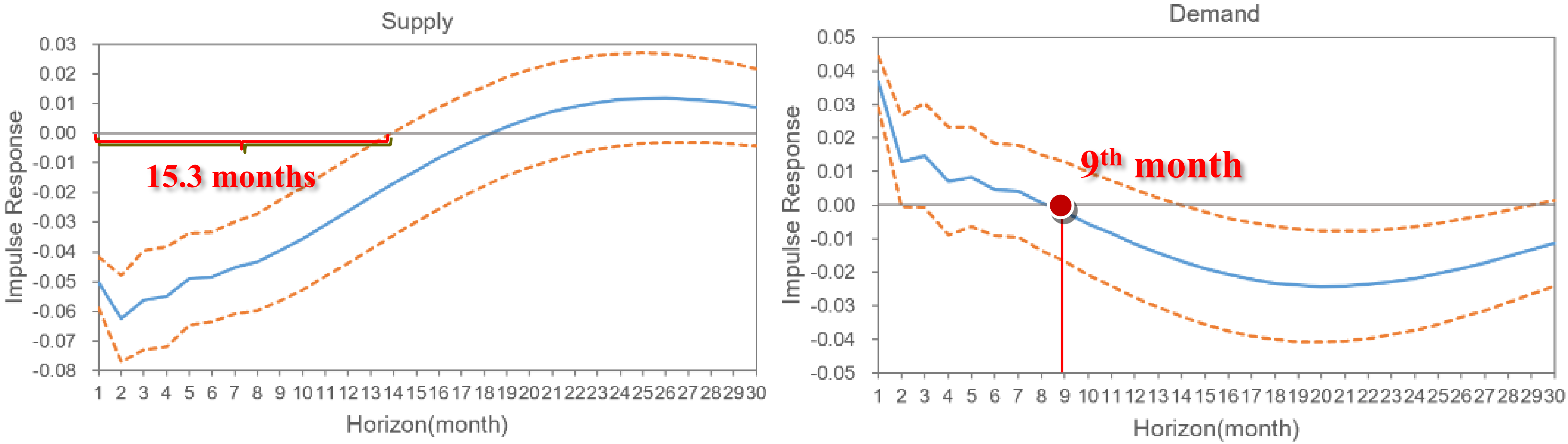


Fig. Responses of WTI crude oil prices to supply and demand shocks.

CEEP-BIT

(3)How does TISIs affect crude oil prices?

➢ **Compare with other proxy variables for investor sentiment**

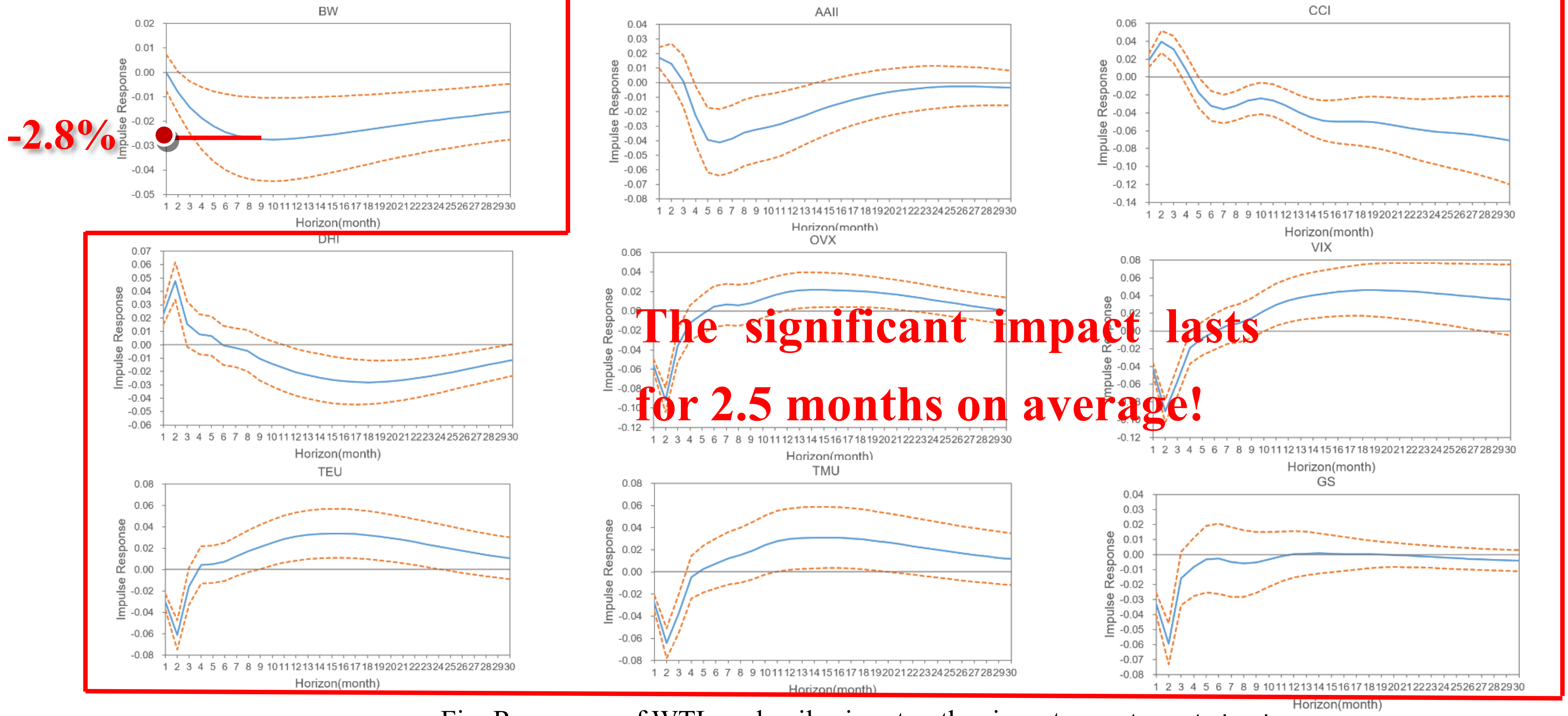


Fig. Responses of WTI crude oil prices to other investor sentiment shocks.

(4) How does the COVID-19 affect the OPPI's performance

CEEP-BIT

➢ **Price Response**

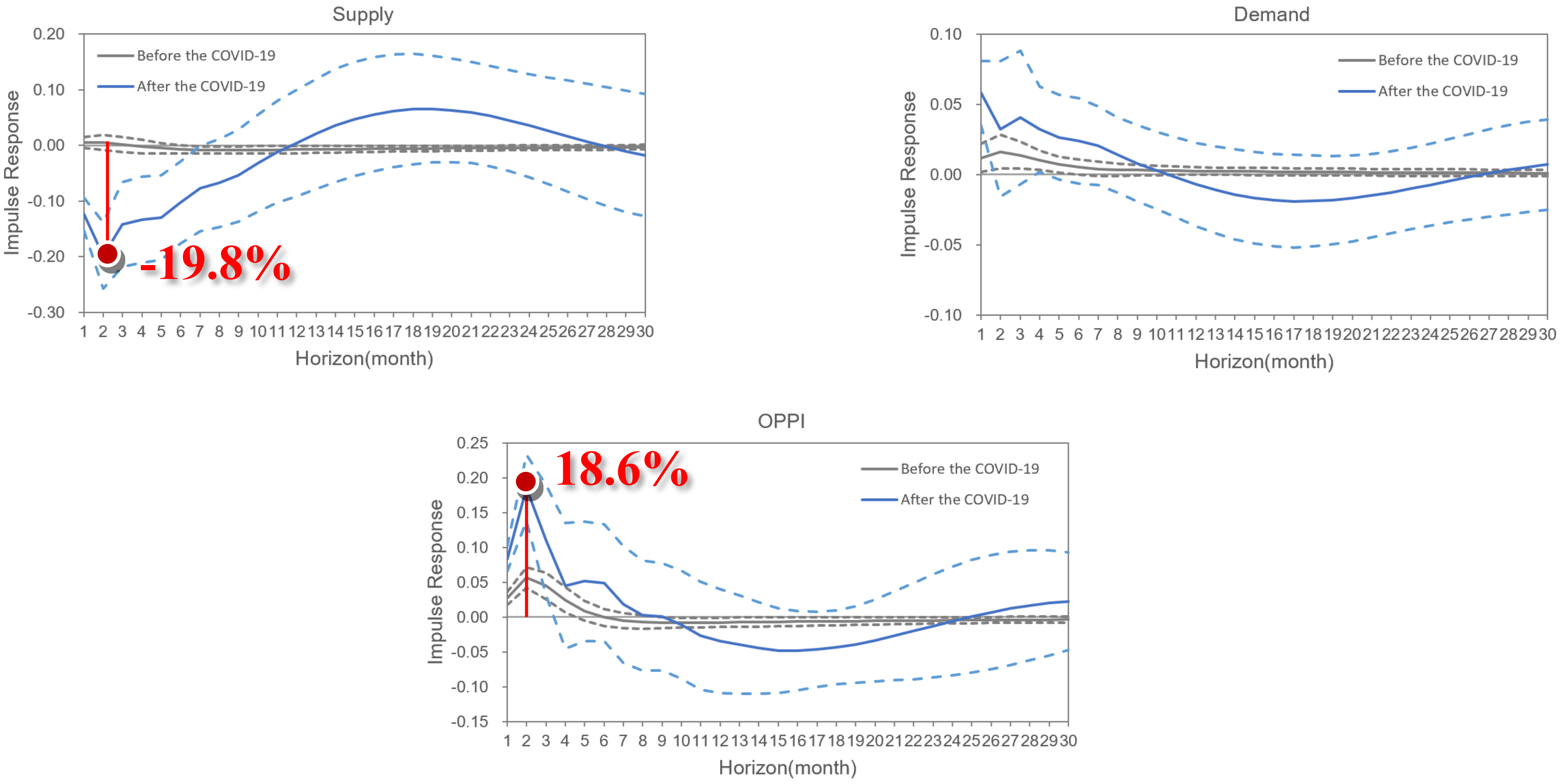
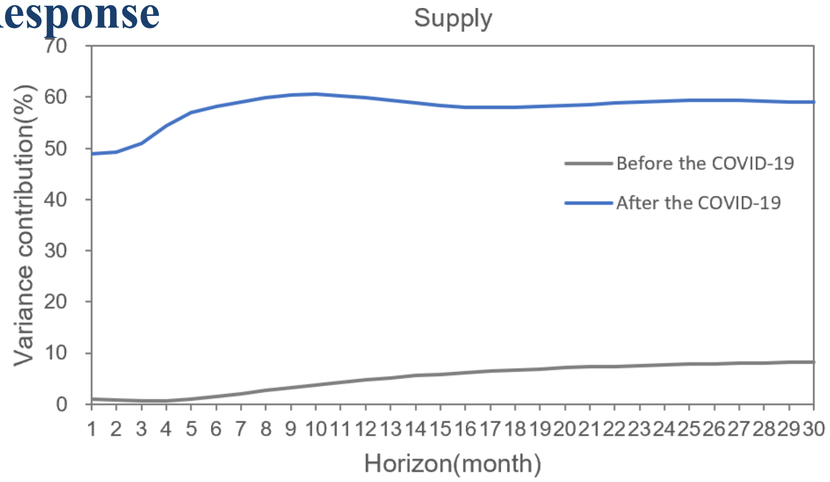


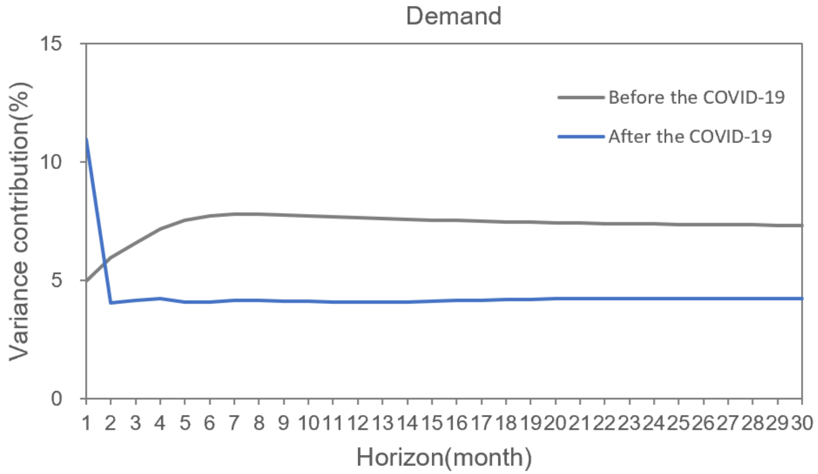
Fig. Responses of WTI crude oil prices to different shocks before and after the COVID-19 breakout.

(4) How does the COVID-19 affect the OPPI's performance

CEEP-BIT

➢ **Price Response**





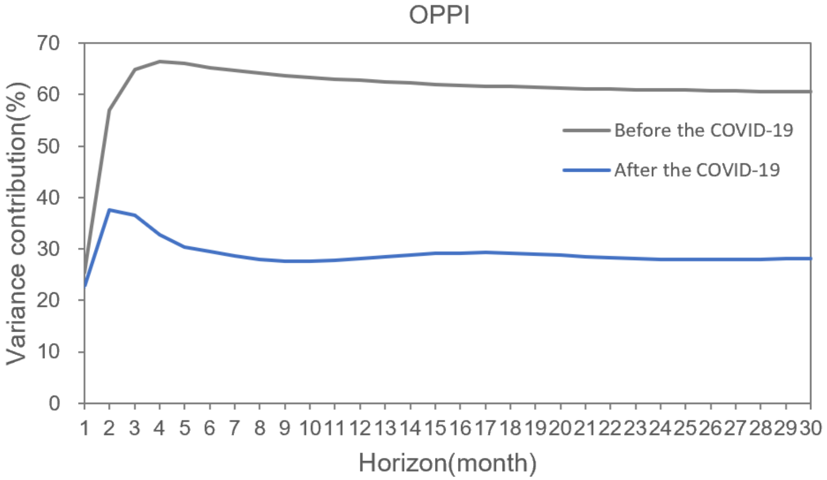


Fig. Contribution of different shocks to WTI crude oil price fluctuations before and after the COVID-19 breakout.

★ WTI crude oil prices were more **closely related to fundamental supply-and-demand factors** during **the COVID-19 pandemic**

CONTENTS

CEEP-BIT

Conclusions

CEEP-BIT

01

Based on the keyword screening rules, we constructed a crude oil market corpus. A fine-grained aspect-level sentiment analysis method was developed to quantify investor sentiment, and the oil price prosperity index (OPPI) was constructed. The empirical results showed that it had a strong, significant and sustained impact on crude oil prices.

02

During the COVID-19 epidemic, the significant positive impact of OPPI on oil prices emerged rapidly with enhanced timeliness and significant short-term effects, but the long-term relationship became weak.

The maximum impulse response of the WTI crude oil price to OPPI shocks increased from 5.7% to 8.3%, and the speed shortened from 3.3 months to 2.0 months. Although OPPI's ability to explain WTI crude oil price fluctuation has weakened, it can still explain 29.4% of oil price fluctuations.

03

Sharp fluctuations of international oil prices triggered market panic, contributed to short-term irrational speculation, and further amplified oil price fluctuations. The changes in investor sentiment would be reasonably guided. The online public opinion on social media platforms would react and be regulated in time to promote the stable and healthy development of the market.

THANKS！

Lu-Tao ZHAO
Ltzhao@bit.edu.cn

Dec. 14th, 2022