

Jung, Hail; Kim, Daejin

**Article**

## Predicting crude oil returns and trading position: Evidence from news sentiment

Journal of Derivatives and Quantitative Studies: Seonmul yeon'gu (JDQS)

**Provided in Cooperation with:**

Korea Derivatives Association (KDA)

*Suggested Citation:* Jung, Hail; Kim, Daejin (2025) : Predicting crude oil returns and trading position: Evidence from news sentiment, Journal of Derivatives and Quantitative Studies: Seonmul yeon'gu (JDQS), ISSN 2713-6647, Emerald, Leeds, Vol. 33, Iss. 2, pp. 86-109, <https://doi.org/10.1108/JDQS-12-2024-0050>

This Version is available at:

<https://hdl.handle.net/10419/337291>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# Predicting crude oil returns and trading position: evidence from news sentiment

Hail Jung

*Department of Business Administration,  
Seoul National University of Science and Technology, Seoul, South Korea, and*

Daejin Kim

*Department of FinTech, SKK Business School, Sungkyunkwan University,  
Seoul, South Korea*

## Abstract

We study the effectiveness of textual information in predicting the returns of crude oil futures and understanding the behavior of market participants. Using a machine learning method to extract oil market sentiment from news articles, we find that the computed sentiment is significantly effective in explaining the crude oil futures returns, while existing textual analyses based on pre-defined dictionaries may mislead the contexts in the oil market. Consistent with previous findings that returns help explain the change in traders' positions, the sentiment scores based on the machine learning method are also useful in explaining the behavior of different types of traders. Our empirical findings underscore the fact that accurately identifying textual information can increase the accuracy of oil price predictions and explain divergent behaviors of oil traders.

**Keywords** Crude oil, Prediction, Textual analysis, Machine learning, Trading position

**Paper type** Research paper

## 1. Introduction

Crude oil is an important asset for industrial development and production and plays a pivotal role in all aspects of national planning (Chiroma *et al.*, 2015). In 2015, the market size of crude oil was over US\$ 1.7 trillion, with approximately 4.49 billion barrels of crude oil produced globally. Considering that the annual market size of all major metals and minerals was approximately US\$ 660 million, crude oil constitutes a large proportion of the commodities markets (Desjardins, 2016). Furthermore, crude oil futures contracts are the most actively traded commodity derivatives across derivatives exchanges and the fifth liquid futures contracts, following the S&P 500 E-Mini, 10-Year T-Notes, Nikkei 225 mini, and Euro-bond (Robe and Wallen, 2016).

Predicting crude oil price movements draws ample attention from both governments and industries to facilitate accurate decision-making (Plourde and Watkins, 1998). These demands are increasing further due to changes in oil markets and increased volatility in oil prices caused by the financialization of commodities markets [1]. The involvement of financial institutions along with traditional players in the commodities market is growing rapidly, making it more difficult to predict prices because of the increasing volatility of oil prices. Hence, understanding crude oil movements could be valuable for economic agents to frame plans

## JEL Classification — G11, G12, G17, C53

© Hail Jung and Daejin Kim. Published in *Journal of Derivatives and Quantitative Studies: 선물연구*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Hail Jung thanks to the support of the Research Program funded by the SeoulTech (Seoul National University of Science and Technology).



that account for the hike in crude oil prices and mitigate the associated drawbacks (Xie *et al.*, 2006).

Although previous studies have attempted to predict oil prices, traditional economic variables do not seem to possess robust forecasting abilities. One potential reason is the frequency of the economic variables. The oil market experiences frequent fluctuations compared with other economic variables that are released less frequently. Daily sentiment measures based on news articles can provide immediate responses to timely information on crude oil markets. However, previous studies have typically used pre-defined dictionaries to compute sentiment scores by classifying words as positive, negative, or neutral and averaging them. Using a pre-defined dictionary for classifying words may misrepresent sentence meaning and vary depending on the dictionary chosen. Additionally, the vocabulary used in the oil market differs from that used in other markets and industries.

To address potential limitations of conventional textual analysis, we employ sentiment extraction via screening and topic modeling (SESTM), developed by Ke *et al.* (2019), which leverages a supervised machine learning (ML) technique. The empirical results show that positive scores from the SESTM are positively associated with positive oil returns. However, other sentiment scores demonstrate insignificant explanatory power or wrong signs. We further validate the explanatory powers of SESTM on crude oil returns using various ML algorithms with other macro variables that are commonly used in crude oil prediction literature (Jammazi and Aloui, 2012; Haidar *et al.*, 2008; Guo *et al.*, 2012; Gabralla *et al.*, 2013). We find that various performance metrics improve significantly after including the SESTM-based sentiment score [2]. Thus, these findings imply that the SESTM sentiment score could be an important explanatory variable for forecasting crude oil prices.

Kang *et al.* (2020) find that commercial and non-commercial traders take different positions depending on past returns. Hence, if past returns can explain traders' positions, the sentiment score should also be able to describe traders' positions. Using the position data provided by the Commodity Futures Trading Commission (CFTC), we find that commercial traders tend to reduce their long positions when news sentiment is positive, while non-commercial traders increase their long positions. As commercial traders hedge their business activities, they tend to increase their short positions when the news sentiment is positive or when the return is expected to be positive. By contrast, non-commercial traders, such as financial institutions, have incentives to increase their long positions to increase their profits. These findings align with Kang *et al.* (2020) who find that past crude oil futures returns are positively (negatively) associated with non-commercial traders' (commercial traders') long positions.

Accordingly, this study makes valuable contributions to literature by suggesting that text-based information has high explanatory power in explaining changes in crude oil prices. Moreover, we demonstrate that pre-defined dictionary-based textual analyses may not be effective in oil price prediction because they may fail to capture the correct context of an article. We also find that sentiment score is a significant factor in explaining market participants' positions by confirming the positive (negative) relationship between news sentiment and the positions of commercial (non-commercial) traders. These results imply that the market responds differently to the tone of articles.

The paper is organized as follows: Section 2 summarizes the previous literature. Section 3 describes the data and methodologies. Section 4 presents the main results. Section 5 concludes the paper.

## 2. Literature review

The prediction of crude oil prices has been studied extensively in finance and economics through the development of reliable forecasting models. Predicting energy prices has been a popular area of research for long, and extant studies have used statistical methodologies to identify meaningful factors affecting or forecasting crude oil prices. One popular method is to

use time-series econometrics such as autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH).

Owing to the availability of enhanced hardware performance and abundant granular data, novel ML algorithms have been employed for making predictions. ML models are computational algorithms that identify patterns in data and make predictions without relying on explicit parametric assumptions. While classical parametric econometrics often requires strict theoretical assumptions about functional relationships, ML models can capture complex and nonlinear interactions more flexibly. This adaptability improves predictive accuracy, especially in large and unstructured datasets.

Classic ML models are commonly employed to express complex characteristics like volatility and nonlinearity. Several studies have examined whether ML algorithms can explain price movements in the crude oil market, including neural networks ([Kaboudan, 2003](#); [Moshiri and Foroutan, 2006](#); [Mostafa and El-Masry, 2016](#)), support vector machine ([Xie et al., 2006](#)), and semi-supervised learning ([Shin et al., 2013](#)). Moreover, [Anifowose et al. \(2017\)](#) use ML methods in oil field exploration, and [Fulford et al. \(2016\)](#) employ deep learning models for accurate diagnosis.

While existing ML models for predicting oil prices consider some economic variables, the input variables may lack the necessary frequency to explain daily crude oil prices. This limitation can arise from the infrequent release of macroeconomic data, often on an annual, semi-annual, or monthly basis. Such macroeconomic variables may be less relevant to real-time oil market movements, as the market experiences daily fluctuations driven by the increased participation of financial professionals. Furthermore, [Li et al. \(2019\)](#) argue that relying on macroeconomic statistics may fail to account for qualitative analyses of political, catastrophic, and other emergent events within the scope of quantified data. Moreover, using macroeconomic data for oil price prediction is problematic owing to shallow architectures, which struggle to capture complex patterns and volatile behaviors of oil prices driven by multiple factors ([Bengio, 2009](#); [Zhao et al., 2017](#)). Addressing these concerns, this study employs advanced ML-based text mining techniques to analyze oil market news articles and investigates how oil price predictions can be improved. Improving these predictions requires more frequent data sources, such as daily news articles, to develop more sophisticated forecasting models.

A growing body of literature analyzes textual data to understand the information process in financial markets ([Cowles 3rd, 1933](#); [Tetlock, 2007](#); [Tetlock et al., 2008](#); [Dougal et al., 2012](#); [Boudoukh et al., 2013](#); [Yang, 2023](#)). Most conventional textual analyses rely on selective dictionaries that match a set of words with their sentiment scores [3]. [Tetlock \(2007\)](#) uses the Harvard-IV (HIV) dictionary to measure the sentiment of the Wall Street Journal articles and finds that pessimism in news articles is negatively associated with the Dow Jones Industrial Average (DJIA) on the next day. Instead of using the HIV dictionary, [Loughran and McDonald \(2011\)](#) use a novel sentiment dictionary designed specifically for finance (hereafter referred to as “LM dictionary”) [4]. They measure the sentiment of 10-K filings and show that their dictionary is better at predicting filing returns than the HIV dictionary. However, computing sentiment scores based on pre-defined dictionaries may misinterpret the context of articles, as words classified as positive or negative in advance can have varying meanings depending on the market and asset [5]. Therefore, a potential solution is to classify each word as positive or negative by correlating oil returns with news articles using supervised ML algorithm.

Improvements in technology, especially advanced database software to store large volumes of textual data, enable researchers to incorporate various textual information into prediction models. Newspaper articles are optimal because they are released much more frequently than macroeconomic statistics, and they contain qualitative information, such as the tone and sentiment of texts [6]. However, utilizing newspaper articles is challenging because the content of articles must be converted to quantitative data. Therefore, reliable techniques are required to identify the sentiment of the text. Accordingly, we aim to use specific newspaper

articles and supervised ML to assess the sentiment of articles and predict the corresponding crude oil returns.

We use both textual information and ML model to develop appropriate sentiment scores for the crude oil market. We employ the SESTM developed by [Ke et al. \(2019\)](#). One advantage is that the SESTM algorithm helps understand and learn the sentiment structure of a corpus, allowing researchers to construct a sentiment score model tailored to the context of a particular dataset. Applying the SESTM model to commodity futures markets offer two advantages. First, unlike stock markets, futures markets have fewer restrictions on short positions, enabling a more direct observation of changes in participant positions. Second, extracting firm-level textual information in stock markets is often challenging due to lexical ambiguities [7]. However, using distinct and consistent terminology in oil markets, such as “crude oil,” “West Texas Intermediate,” or “WTI,” enhances the retrieval of comprehensive textual information specific to the crude oil market.

### 3. Data and methodology

#### 3.1 Data

In this study, three main sources are used to retrieve news articles on crude oil: The Wall Street Journal, The Financial Times, and RTTNews. We manually crawl the news articles related to crude oil based on certain keywords, such as “crude oil,” “WTI,” and “drilling.” From the crawled data, duplicated articles reporting the same news are removed, retaining the earliest published version by manually comparing articles from each source.

We obtain the WTI futures prices from Datastream. Daily returns are constructed using nearby futures contracts. These are then matched to new articles published on the corresponding day. The final sample comprises 8,605 articles published from 2014 to 2019. To develop and evaluate the SESTM model, the dataset is divided into two subsets: the first half (2014–2016) is used as a training set, and the second half (2017–2019) serves as a test set.

Following previous energy economics studies, we construct various control variables that may potentially influence crude oil prices ([Jammazi and Aloui, 2012](#); [Haidar et al., 2008](#); [Guo et al., 2012](#); [Gabralla et al., 2013](#)). We use data from the Center for Research in Security Prices (CRSP), International Monetary Fund (IMF), US Energy Information Administration (EIA), Bloomberg, and the Economic Policy Uncertainty (EPU) website. Using these data sources, we calculate returns for the S&P 500 and NASDAQ stock indices, VIX, gold, the US dollar, US 10- and 30-year Treasury bonds, natural gas futures, and the Economic Policy Uncertainty (EPU) index, along with returns and trading volumes for the Invesco Dynamic Energy Exploration and Production (PXE) exchange-traded fund (ETF) [8].

For comparison, we use commonly employed sentiment measures in finance and accounting, based on the HIV and LM dictionaries. We also employ the Valence Aware Dictionary and sEntiment Reasoner (VADER) algorithm, which is a lexicon and rule-based sentiment analysis tool specifically attached to sentiments expressed in texts (social media, blog posts, or articles) [9].

To investigate the relationship between sentiment scores and traders’ positions, we utilize weekly the Commitments of Traders (COT) reports announced by the CFTC, which broadly classifies traders as commercial or non-commercial. Commercial traders hedge business activities through futures markets, while non-commercial traders speculate for profit without direct involvement in the underlying commodity. Following [Kang et al. \(2020\)](#), we measure net trading ( $Q_t$ ) as the weekly change in the net long position, normalized by the open interest at the beginning of the week:

$$Q_t = \frac{Net\ Long_t - Net\ Long_{t-1}}{OI_{t-1}} \quad (1)$$

where  $Net\ Long_t$  refers to net trading, defined as the difference between long and short open interests, and  $OI_{t-1}$  refers to the total open interest in week  $t-1$ .

### 3.2 Methodology

**3.2.1 Sentiment score via textual analysis.** We compute the sentiment scores using the SESTM developed by [Ke et al. \(2019\)](#) because this method could provide a meaningful link between asset returns and textual information. The SESTM is operated in three steps: first, it isolates the most relevant features from a large vocabulary set. The vocabulary is derived from the dictionary of each article as a count vector. The variable selection approach extracts a small number of terms, likely to be informative in explaining crude oil returns. The variable selection through correlation screening is based on ML, which enables fast and simple estimation of the reduced-dimension sentiment list. The underlying mechanism defines whether the term is positively or negatively correlated with returns of the same sign. [Ke et al. \(2019\)](#) argue that this estimation technique is more efficient than other dimension-reduction techniques because it functions well even with high-dimensional data [\[10\]](#).

Second, we assign sentiment scores for each term selected in the first step based on their individual relevance for predictions. Instead of weighting sentiments based on word frequencies, the model uses a supervised learning generative model to address the extreme skewness problem. Moreover, the learning process of SESTM trains the relationship between the list of words and daily returns by assuming an appropriate statistical distribution between the two.

Finally, the sentiment scores of articles are calculated using the consistent likelihood structure of the second step to address potential severe heterogeneity problems in the frequency of words and their sentiment scores. Following [Ke et al. \(2019\)](#), a penalized maximum likelihood estimator with one unknown parameter is estimated for each article. A Bayesian interpretation of the penalization is to impose a Beta-distributed prior on sentiment parameters that are centered at 0.5, ranging from 0 to 1. This Bayesian interpretation of the penalization implies that the estimation begins from the prior when an article has a neutral tone. This procedure allows the beginning of the estimation with the prior that the document is neutral. If the sentiment score after the estimation is above 0.5, we consider the score to indicate a positive tone.

In contrast to conventional sentiment analyses based on pre-defined dictionaries, the SESTM analyzes the context of the article and the corresponding return to predict the return when a new article is used as input. However, note that conventional dictionary-based methods are not ML-based models; instead, they often use a “bag-of-words” including nouns, verbs, or adjectives, which are classified as positive or negative groups or as subjective or objective groups. Consequently, the selection of a dictionary can exert a significant influence on the outcomes of these existing methods. Thus, dictionary-based text analysis should select a dictionary that is suitable for the target domain. However, the SESTM does not depend on the selection of a particular dictionary because it links textual information and asset returns using supervised learning.

[Table 1](#) provides anecdotal evidence to show that the bag-of-words method might misrepresent the news contents. Panel A presents stories on how President Trump threatened Iran with severe outcomes if the latter continued to threaten the US. This news article is related to a sharp increase in crude oil prices on that day, with a daily return of 1.06%. The SESTM score is calculated as 0.9686, which suggests an increase in crude oil prices. However, the HIV dictionary gives a slightly negative score of  $-0.0398$ , and the VADER algorithm gives an extremely negative score of  $-0.9982$ . This demonstrates that if investors trade crude oil futures based on the sentiment score computed from either the HIV or VADER, they might experience losses because the score suggests selling crude oil.

Panel B illustrates another example of how existing sentiment measures can mislead news articles. Panel B presents an article about increasing risks in the Middle East, and on the day of

**Table 1.** Anecdotal example

Panel A. Example 1

... Donald Trump has threatened Iran with severe “consequences” if it makes any more threats towards the US, as administration officials amplify the US president’s toughest rhetoric to date towards the Islamic republic. In a late-night Twitter post on Sunday, Mr. Trump warned President Hassan Rouhani in capital letters: “Never, ever threaten the United States again or you will suffer consequences the likes of which few throughout history have ever suffered before.” He went on: “We are no longer a country that will stand for your demented words of violence and death. Be cautious!” ... (24th of July 2018, Financial Times)

|             |         |
|-------------|---------|
| SESTM score | 0.9686  |
| HIV score   | −0.0398 |
| VADER score | −0.9982 |
| WTI return  | 1.0600% |

Panel B. Example 2

... Officials say intelligence points to Iran as staging ground for strikes, as allies weigh retaliation U.S. intelligence indicates Iran was the staging ground for a debilitating attack on Saudi Arabia’s oil industry, people familiar with the discussions said, as Washington and the kingdom weighed how to respond and oil prices soared. Monday’s assessment, which the U.S. hasn’t shared publicly, came as President Trump said he hoped to avoid a war with Iran and as Saudi Arabia asked United Nations experts to help determine who was responsible for the airstrikes ... (16th of September 2018, The Wall Street Journal)

|             |         |
|-------------|---------|
| SESTM score | 0.8800  |
| HIV score   | −0.9902 |
| VADER score | −0.4300 |
| WTI return  | 1.5100% |

**Source(s):** Authors’ own work

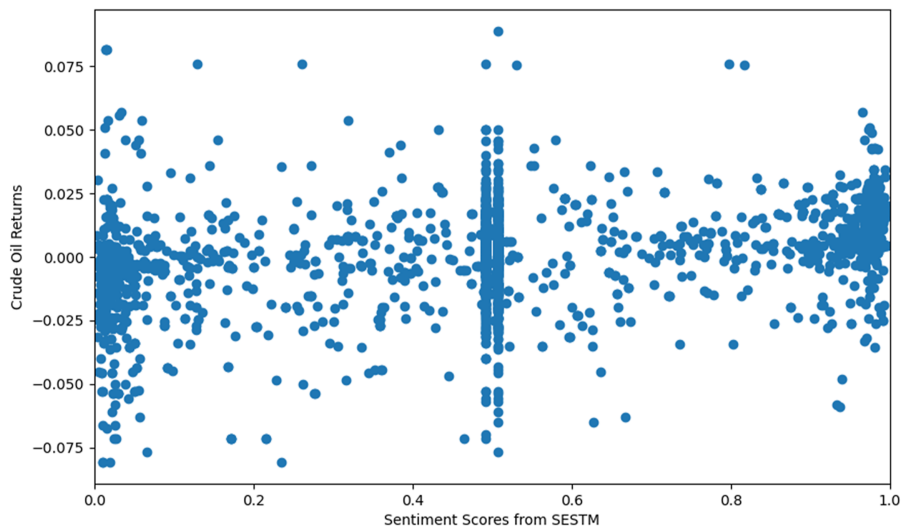
the article, the oil return was 1.51%. This increase in oil returns is consistent with the SESTM value of 0.88, indicating a positive relationship between the SESTM and oil returns. On the other hand, both HIV and VADER scores are −0.9902 and −0.43, respectively, implying that both measures fail to capture the context of the article. Therefore, the examples in [Table 1](#) suggest that sentiment scores calculated by predefined dictionary methodologies may not correctly capture news information in the crude oil market.

[Figure 1](#) illustrates the sentiment-scoring analysis results obtained using the SESTM algorithm. The algorithm analyzes the news article and identifies the relationship between the information extracted from it and daily returns. Note that the sentiment scores are clustered around 0, 0.5, or 1. These results suggest that the SESTM method can effectively classify news articles into negative, neutral, or positive sentiment groups.

[Figure 2](#) displays a list of sentiment-charged words identified using the SESTM. These words are strongly correlated with crude oil price fluctuations and thus surpass the correlation-screening threshold. To illustrate the most impactful sentiment words in our analysis, the word cloud font is drawn in proportion to their average sentiment tone.

[Table 2](#) presents interesting distinctions for extant sentiment dictionaries. For example, compared to the estimated list of the most impactful positive words, words like “political,” “sustained,” “resilient,” “extended,” and “steep” are not visible in either of the LM or HIV dictionaries. Similarly, some impactful negative terms, such as “short” and “excess,” do not appear as negative in those dictionaries. Interestingly, the word “disappointed,” which is selected as a positive word by the SESTM, is classified as a negative term in the LM and HIV dictionaries. Moreover, while the term “excess” is considered negative in the SESTM model, it is neither positive nor negative in the LM dictionary. As this study focuses on the crude oil market, an article that discusses “excess” crude oil stock often leads to a lower return. Thus, the SESTM determines the word as negative.

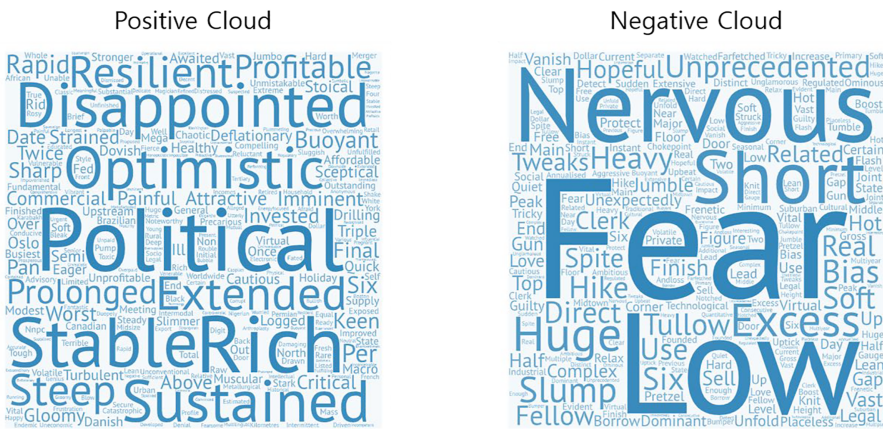
[Figure 3](#) illustrates the time series of the sentiment scores measured using the SESTM algorithm. Considering that sentiment scores reflect the information embedded in news articles, they should be randomly distributed because new information cannot be expected.



**Note(s):** This figure illustrates the relationship between crude oil returns and the sentiment scores computed from the SESTM algorithm. The sentiment scores are derived from parsed news articles, with the algorithm calculating the scores based on words identified through the relationship between crude oil returns and news contents. Note that the sentiment scores are clustered around values of 0, 0.5, or 1

**Source(s):** Authors’ own work

**Figure 1.** Classification results of the SESTM model



**Note(s):** This figure visualizes the sentiment-charged words identified by the SESTM model. These words are strongly correlated with the crude oil returns and thus surpass the correlation screening threshold. To illustrate the most impactful sentiment words in our analysis, the word cloud font is drawn proportional to the average sentiment tone of each word

**Source(s):** Authors’ own work

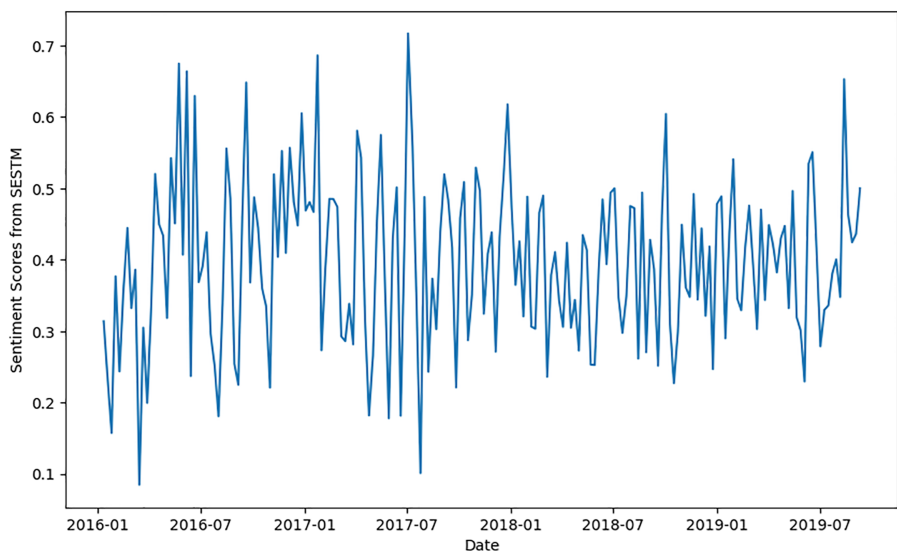
**Figure 2.** Word cloud created by the SESTM model

**Table 2.** Words selected by the SESTM and other sentiment measures

|                                                            | HIV            | LM             |
|------------------------------------------------------------|----------------|----------------|
| Panel A. Top 10 positive words selected by the SESTM model |                |                |
| Disappointed                                               | Negative       | Negative       |
| Political                                                  | Not classified | Not classified |
| Rich                                                       | Positive       | Not classified |
| Stable                                                     | Positive       | Positive       |
| Optimistic                                                 | Positive       | Positive       |
| Sustained                                                  | Not classified | Not classified |
| Resilient                                                  | Not classified | Not classified |
| Extended                                                   | Not classified | Not classified |
| Steep                                                      | Not classified | Not classified |
| Profitable                                                 | Positive       | Positive       |
| Panel B. Top 10 negative words selected by the SESTM model |                |                |
| Fear                                                       | Negative       | Negative       |
| Low                                                        | Negative       | Not classified |
| Nervous                                                    | Negative       | Not classified |
| Short                                                      | Negative       | Not classified |
| Excess                                                     | Negative       | Not classified |
| Huge                                                       | Not classified | Not classified |
| Unprecedented                                              | Not classified | Not classified |
| Slump                                                      | Negative       | Not classified |
| Fellow                                                     | Positive       | Not classified |
| Hopeful                                                    | Positive       | Not classified |

**Note(s):** This table provides key positive and negative words selected by the SESTM, along with their corresponding classifications in the Harvard-IV (HIV) and Loughran and McDonald (LM) dictionaries. Panel A lists the top 10 positive words identified by the SESTM model, while Panel B lists the top 10 negative words

**Source(s):** Authors' own work



**Note(s):** This figure illustrates the time-series plot of the news sentiment measured by the SESTM model

**Source(s):** Authors' own work

**Figure 3.** Time series of news article sentiment score generated by the SESTM model

The time series of sentiments seems to reflect the randomness of the information. In other words, the sentiment variables seem to follow a stationary distribution rather than a certain trend. The randomness shown in [Figure 3](#) reflects these characteristics, and therefore, so does the sentiment score measured by the SESTM.

*3.2.2 Forecasting crude oil returns using ML.* We employ ML models to investigate whether textual information can increase the explanatory power in predicting crude oil returns. This study employs popular methods commonly used in the literature on finance. For example, [Gu et al. \(2020\)](#) use Random Forest (RF), Light Gradient Boosting Machine (LGBM), Categorical Boosting (CB), and linear regression (LR) to explain equity returns. They show that the use of ML models can improve the predictability of equity returns, as compared to traditional regression methods. Among the various techniques, they demonstrate that trees or neural networks perform the best in forecasting equity returns.

The RF is an ensemble learning method for classification and regression that constructs a multitude of decision trees at training time and outputs the class, which is the mode of the classes or average prediction values of individual trees ([Ho, 1995, 1998](#)). The RF corrects the trees' habit of overfitting to the training set. Although the RF generally outperforms decision trees, its accuracy is lower than that of gradient-boosted trees ([Pirayonesi and El-Diraby, 2020](#)). Thus, we employ LGBM and CB because of the comparative advantages of recent gradient-boosted models. A major difference between the RF and gradient-boosted models lies in the construction of trees; the gradient-boosted models do not grow tree-wise but leaf-wise. Moreover, LGBM and CB do not use the sort-based tree algorithm, which searches for the best-split point on sorted feature values ([Mehta et al., 1996](#)). Instead, gradient-boosted models implement a highly optimized histogram-based decision tree-learning algorithm that yields significant advantages in terms of both efficiency and memory consumption.

We use LR to compare the performance of ML forecasting. Due to its extensive use in academic studies to explore relationships among variables, the performance of LR is used as a benchmark to validate whether ML models can improve oil price forecasting.

## 4. Main results

### 4.1 Descriptive statistics

[Table 3](#) reports the summary statistics of the variables used in this study. It reports the weekly averages of all daily measures because the CFTC announces the commitment of traders every week. The estimated SESTM values range from 0 to 1, where the value of 0 (1) indicates extremely negative (positive) tones for the news article. Note that the average sentiment score measured by the SESTM model is 0.502, which is very close to a neutral tone during the sample period. The averages of other methods seem to deviate from neutral tones. That is, other methods based on pre-defined dictionaries seem to interpret news articles in a more positive or negative way. These results suggest that dictionary-based methods might interpret news articles in a specific direction according to the list of words in the dictionary.

In addition to the summary statistics, [Table 4](#) provides a simple correlation matrix of the variables to determine their relationships. First, the SESTM scores are positively and significantly associated with crude oil returns. However, we do not observe any statistical relationship between the sentiment scores calculated by the VADER algorithm and crude oil returns. Moreover, when sentiment scores are computed based on either the HIV or LM dictionaries, the relationships between crude oil returns and the scores are negative. These results suggest that existing methods based on pre-defined dictionaries may fail to find appropriate relationships between textual information and crude oil returns. Finally, the position change of commercial traders (non-commercial traders) is negatively (positively) associated with crude oil returns, supporting the findings of [Kang et al. \(2020\)](#).

[Figure 4](#) provides the scatter plots of various sentiment scores and crude oil returns. The upper-left part of the plot shows that the relationship between the SESTM scores and crude oil returns seems to be well-defined, whereas other methods might not provide strong

**Table 3.** Summary statistics

| Variable                | N   | Mean   | SD    | Min    | Max   |
|-------------------------|-----|--------|-------|--------|-------|
| SESTM                   | 192 | 0.502  | 0.130 | 0.193  | 0.821 |
| VADER                   | 192 | -0.046 | 0.296 | -0.743 | 0.632 |
| HIV                     | 192 | 0.534  | 0.091 | 0.307  | 0.757 |
| LM                      | 192 | 0.398  | 0.100 | 0.162  | 0.685 |
| $Q_t$ (Commercials)     | 192 | 0.000  | 0.014 | -0.034 | 0.043 |
| $Q_t$ (Non-commercials) | 192 | 0.000  | 0.009 | -0.022 | 0.027 |
| WTI return              | 192 | 0.001  | 0.049 | -0.167 | 0.161 |
| US10 return             | 192 | -0.002 | 0.044 | -0.187 | 0.182 |
| US30 return             | 192 | -0.003 | 0.032 | -0.147 | 0.130 |
| VIX return              | 192 | 0.001  | 0.145 | -0.426 | 0.707 |
| Gold return             | 192 | 0.011  | 0.030 | -0.046 | 0.163 |
| S&P500 return           | 192 | 0.002  | 0.016 | -0.080 | 0.043 |
| NASDAQ return           | 192 | 0.003  | 0.022 | -0.089 | 0.057 |
| Dollar return           | 192 | 0.000  | 0.009 | -0.029 | 0.024 |
| Gas return              | 192 | 0.000  | 0.057 | -0.267 | 0.165 |
| EPU return              | 192 | -0.365 | 0.824 | -2.774 | 3.197 |

**Note(s):** This table presents descriptive statistics for various variables. This table include sentiment scores derived from the SESTM, the VADER algorithm, and the HIV and LM dictionaries. We also provide the changes in positions ( $Q_t$ ) of commercial and non-commercial traders. Additionally, we report returns for WTI crude oil, US 10-year and 30-year treasury bonds, the VIX index, gold, the S&P 500 index, the NASDAQ index, the US dollar, natural gas, and the EPU index. Variables are defined in [Appendix](#). For each variable, we provide the number of observations (N), mean (Mean), standard deviation (SD), minimum (Min) and maximum (Max)

**Source(s):** Authors' own work

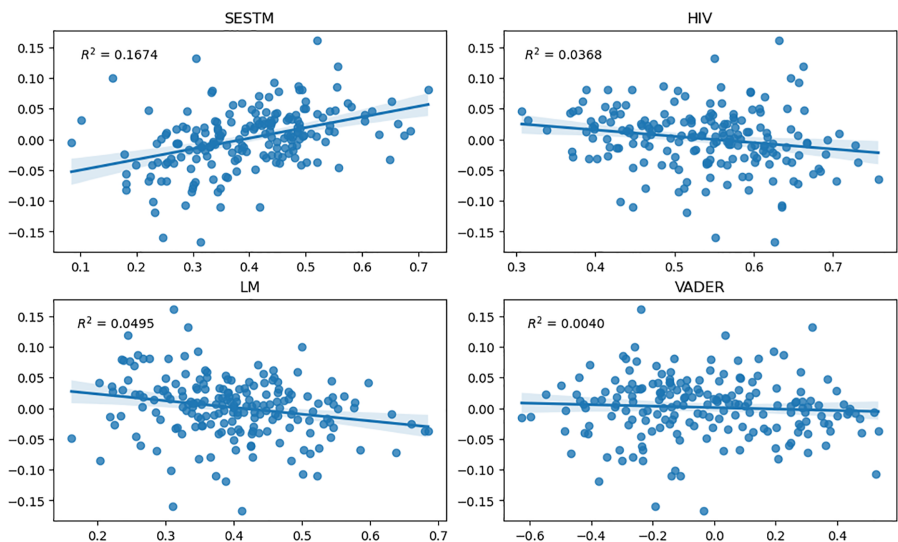
**Table 4.** Correlation matrix

| Variables              | (1) SESTM | (2) VADER | (3) HIV   | (4) LM    |
|------------------------|-----------|-----------|-----------|-----------|
| (1) SESTM              | 1.000     |           |           |           |
| (2) VADER              | -0.128*   | 1.000     |           |           |
| (3) HIV                | -0.189*   | 0.379***  | 1.000     |           |
| (4) LM                 | -0.106    | 0.245***  | 0.355**   | 1.000     |
| (5) Q (Commercial)     | -0.089    | 0.098     | 0.190***  | 0.126     |
| (6) Q (Non-commercial) | 0.143*    | -0.008    | 0.022     | -0.078    |
| (7) WTI return         | 0.440***  | -0.046    | -0.192*** | -0.222*** |
| (8) US10 return        | 0.192*    | 0.023     | -0.058    | -0.125*   |
| (9) US30 return        | 0.169*    | 0.000     | -0.05     | -0.143**  |
| (10) VIX return        | -0.007    | -0.120**  | 0.011     | 0.194***  |
| (11) Gold return       | -0.003    | 0.112     | 0.029     | 0.066     |
| (12) S&P500 return     | 0.200*    | 0.158***  | -0.002    | -0.086    |
| (13) NASDAQ return     | 0.124**   | 0.162*    | 0.026     | -0.089    |
| (14) Dollar return     | -0.033    | -0.081    | -0.032    | 0.007     |
| (15) Gas return        | 0.102     | 0.081     | 0.051     | 0.014     |

**Note(s):** This table provides the correlation matrix of the variables. Variables are defined in [Appendix](#). \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

relationships. On the one hand, the simple linear regression shows a positive relationship between the SESTM sentiment and oil returns, with an R-square of 0.1674. On the other hand, sentiment scores using the LM dictionary and the HIV dictionary show a weak negative relationship ( $R^2 = 0.0495$  and  $0.0368$ , respectively). The sentiment score measured by the VADER algorithm has a low explanatory power to explain oil returns ( $R^2 = 0.004$ ).



**Note(s):** This figure provides scatter plots and the fitted linear regression lines for various sentiment measures and crude oil returns. The upper left plot shows the relationship between the SESTM scores and crude oil returns. The upper right plot shows the results for the sentiment measured by HIV dictionary, the lower left plot shows the results for LM dictionary, and the lower right plot shows the results for VADER algorithm. The R-square for each linear regression model is reported at the top left of each plot

**Source(s):** Authors' own work

**Figure 4.** Various sentiment scores and corresponding crude oil return

4.2 Return predictability

4.2.1 Regression analysis. In this section, we investigate whether sentiment scores are related to crude oil returns using regression analyses. We incorporate various control variables, including returns on the US 10-year and 30-year Treasury bonds, VIX, gold, S&P 500 index, NASDAQ index, US dollar, and Economic Policy Uncertainty (EPU) index, as well as trading volumes and returns of the PXE ETF. Year-month fixed effects are included in all the regression specifications.

Table 5 provides the regression analysis results. Columns from (1) to (4) provide the crude oil return prediction by news sentiment scores calculated by the SESTM, the HIV dictionary, the LM dictionary, and the VADER algorithm, respectively, without control variables. Columns from (5) to (8) provide results after including the control variables. In the following regression analyses, the sample period needs to be limited for comparison purposes. While other dictionary-based sentiment analyses do not necessarily have a training or testing period, the SESTM methods require two separate sample periods. That is, for the implementation of the SESTM, we use half of the dataset (2014–2016) for training and the remaining half (2017–2019) for testing the accuracy. Thus, we use the same test period (2016–2019) to compare SESTM with other methods.

Columns (1) and (5) show that the news sentiment calculated by the SESTM algorithm holds a statistically significant explanatory power. That is, sentiment scores based on SESTM scores are positively associated with crude oil returns, even after controlling for other economic variables. However, although some coefficients are significant, these algorithms are negatively associated with crude oil returns, which show opposite directions. In particular, the

**Table 5.** Return predictability of the SESTM

|                | (1)                 | (2)                   | (3)                   | (4)                | (5)                 | (6)                 | (7)                  | (8)                 |
|----------------|---------------------|-----------------------|-----------------------|--------------------|---------------------|---------------------|----------------------|---------------------|
| SESTM          | 0.034***<br>(6.317) |                       |                       |                    | 0.015***<br>(3.363) |                     |                      |                     |
| HIV4           |                     | −0.129***<br>(−2.624) |                       |                    |                     | −0.048<br>(−1.328)  |                      |                     |
| LM             |                     |                       | −0.103***<br>(−2.649) |                    |                     |                     | −0.061**<br>(−2.111) |                     |
| VADER          |                     |                       |                       | −0.004<br>(−1.483) |                     |                     |                      | −0.002<br>(−0.923)  |
| US10 return    |                     |                       |                       |                    | 0.266<br>(1.194)    | 0.312<br>(1.359)    | 0.342<br>(1.505)     | 0.330<br>(1.431)    |
| US30 return    |                     |                       |                       |                    | −0.440<br>(−1.516)  | −0.487<br>(−1.626)  | −0.533*<br>(−1.796)  | −0.524*<br>(−1.740) |
| VIX return     |                     |                       |                       |                    | 0.042<br>(1.519)    | 0.054*<br>(1.884)   | 0.065**<br>(2.245)   | 0.050*<br>(1.752)   |
| Gold return    |                     |                       |                       |                    | 0.058<br>(0.570)    | 0.042<br>(0.398)    | 0.058<br>(0.557)     | 0.054<br>(0.505)    |
| S&P500 return  |                     |                       |                       |                    | 0.129<br>(0.256)    | 0.231<br>(0.445)    | 0.412<br>(0.793)     | 0.263<br>(0.505)    |
| NASDAQ return  |                     |                       |                       |                    | −0.099<br>(−0.332)  | −0.121<br>(−0.390)  | −0.217<br>(−0.711)   | −0.162<br>(−0.526)  |
| Dollar return  |                     |                       |                       |                    | 0.366<br>(1.098)    | 0.291<br>(0.843)    | 0.347<br>(1.017)     | 0.312<br>(0.902)    |
| Gas return     |                     |                       |                       |                    | 0.036<br>(0.722)    | 0.058<br>(1.129)    | 0.056<br>(1.106)     | 0.054<br>(1.054)    |
| EPU return     |                     |                       |                       |                    | 0.003<br>(0.784)    | 0.004<br>(1.057)    | 0.003<br>(0.950)     | 0.004<br>(1.096)    |
| PXE ETF return |                     |                       |                       |                    | 0.880***<br>(8.769) | 0.950***<br>(9.382) | 0.945***<br>(9.472)  | 0.961***<br>(9.529) |
| PXE ETF volume |                     |                       |                       |                    | −0.001<br>(−0.453)  | −0.001<br>(−0.428)  | −0.001<br>(−0.375)   | −0.001<br>(−0.508)  |

(continued)

**Table 5.** Continued

|                | (1)                   | (2)                 | (3)                 | (4)              | (5)                   | (6)              | (7)                | (8)              |
|----------------|-----------------------|---------------------|---------------------|------------------|-----------------------|------------------|--------------------|------------------|
| Constant       | −0.084***<br>(−6.057) | 0.070***<br>(2.654) | 0.042***<br>(2.674) | 0.000<br>(0.099) | −0.035***<br>(−2.959) | 0.029<br>(1.496) | 0.027**<br>(2.352) | 0.003<br>(0.984) |
| Observations   | 192                   | 192                 | 192                 | 192              | 192                   | 192              | 192                | 192              |
| Adjusted $R^2$ | 0.240                 | 0.076               | 0.076               | 0.046            | 0.559                 | 0.528            | 0.538              | 0.525            |
| Year-month FE  | Yes                   | Yes                 | Yes                 | Yes              | Yes                   | Yes              | Yes                | Yes              |

**Note(s):** This table presents the OLS regression results of crude oil returns on various sentiment measures. Columns from (1) to (4) provide the regression results for the sentiment scores calculated by the SESTM method, HIV dictionary, LM dictionary, and VADER algorithm, respectively, without control variables. Columns from (5) to (8) present the results after including control variables. Variables are defined in [Appendix](#). All regressions include year-month fixed effects. T-statistics are in parentheses. \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

LM dictionary seems to contain information related to crude oil prices. However, the list of words in the predefined dictionaries seems to be interpreted in the opposite direction to the actual movement of crude oil prices [11]. These results suggest that the SESTM can correctly identify the tone of news articles that ultimately affect crude oil prices. Thus, these results suggest that SESTM provides significantly more accurate directions to explain crude oil prices compared to other dictionary-based algorithms.

Table 4 indicates that the SESTM is correlated with other sentiment scores. This finding suggests that we need to show that the SESTM is still positively related to oil returns even when other sentiment scores are combined with the SESTM. Thus, we further investigate the robustness of the SESTM in explaining oil returns by incorporating VADER, LM, and HIV.

Table 6 shows the results of re-estimation by combining other sentiment scores and the SESTM variable. First, the SESTM still shows a statistically significant positive relationship with oil returns. On the other hand, other sentiment scores are statistically insignificant or have

**Table 6.** Return predictability of the SESTM controlling for other sentiment scores

|                | (1)                 | (2)                 | (3)                 | (4)                 | (5)                  | (6)                 | (7)                  |
|----------------|---------------------|---------------------|---------------------|---------------------|----------------------|---------------------|----------------------|
| SESTM          | 0.089***<br>(3.957) | 0.087***<br>(3.802) | 0.083***<br>(3.589) | 0.087***<br>(3.802) | 0.082***<br>(3.663)  | 0.085***<br>(3.681) | 0.082***<br>(3.594)  |
| VADER          | -0.008<br>(-1.415)  |                     |                     | -0.012<br>(-1.362)  | -0.020**<br>(-2.340) |                     | -0.021**<br>(-2.054) |
| HIV            |                     | -0.016<br>(-0.674)  |                     | 0.020<br>(0.572)    |                      | -0.033<br>(-1.080)  | 0.005<br>(0.154)     |
| LM             |                     |                     | 0.003<br>(0.225)    |                     | 0.036*<br>(1.861)    | 0.016<br>(0.874)    | 0.035*<br>(1.763)    |
| US10 return    | -0.076<br>(-0.398)  | -0.069<br>(-0.357)  | -0.044<br>(-0.229)  | -0.061<br>(-0.315)  | -0.078<br>(-0.420)   | -0.078<br>(-0.401)  | -0.074<br>(-0.391)   |
| US30 return    | 0.249<br>(1.150)    | 0.241<br>(1.100)    | 0.223<br>(1.018)    | 0.241<br>(1.105)    | 0.239<br>(1.123)     | 0.239<br>(1.091)    | 0.237<br>(1.105)     |
| VIX return     | -0.010<br>(-0.193)  | -0.011<br>(-0.197)  | -0.021<br>(-0.392)  | -0.017<br>(-0.305)  | -0.028<br>(-0.532)   | -0.015<br>(-0.269)  | -0.030<br>(-0.547)   |
| Gold return    | -0.004*<br>(-1.764) | -0.004<br>(-0.780)  | -0.004<br>(-0.893)  | 0.003<br>(0.645)    | -0.002<br>(-0.463)   | -0.004<br>(-0.593)  | 0.003<br>(0.343)     |
| S&P500 return  | 0.007<br>(0.012)    | -0.128<br>(-0.208)  | -0.154<br>(-0.245)  | 0.069<br>(0.109)    | -0.091<br>(-0.149)   | -0.252<br>(-0.397)  | -0.073<br>(-0.115)   |
| NASDAQ return  | -0.275<br>(-0.582)  | -0.156<br>(-0.333)  | -0.136<br>(-0.286)  | -0.327<br>(-0.677)  | -0.245<br>(-0.525)   | -0.073<br>(-0.153)  | -0.260<br>(-0.542)   |
| Dollar return  | -0.597<br>(-1.537)  | -0.541<br>(-1.385)  | -0.484<br>(-1.227)  | -0.591<br>(-1.515)  | -0.522<br>(-1.357)   | -0.494<br>(-1.253)  | -0.522<br>(-1.348)   |
| Gas return     | -0.010<br>(-0.250)  | -0.013<br>(-0.331)  | -0.016<br>(-0.405)  | -0.010<br>(-0.255)  | -0.013<br>(-0.332)   | -0.015<br>(-0.384)  | -0.013<br>(-0.330)   |
| EPU return     | 0.783<br>(1.212)    | 0.699<br>(1.075)    | 0.616<br>(0.940)    | 0.774<br>(1.193)    | 0.686<br>(1.075)     | 0.636<br>(0.972)    | 0.685<br>(1.068)     |
| PXE return     | -0.177<br>(-0.273)  | -0.079<br>(-0.121)  | 0.008<br>(0.012)    | -0.174<br>(-0.267)  | -0.069<br>(-0.108)   | -0.003<br>(-0.005)  | -0.071<br>(-0.109)   |
| PXE volume     | 0.000<br>(1.286)    | 0.000<br>(1.081)    | 0.000<br>(0.794)    | 0.000<br>(1.153)    | 0.000<br>(1.035)     | 0.000<br>(0.977)    | 0.000<br>(0.992)     |
| Constant       | -0.161*<br>(-1.792) | -0.145<br>(-1.571)  | -0.115<br>(-1.230)  | -0.150<br>(-1.635)  | -0.118<br>(-1.297)   | -0.128<br>(-1.358)  | -0.116<br>(-1.254)   |
| Observations   | 192                 | 192                 | 192                 | 192                 | 192                  | 192                 | 192                  |
| Adjusted $R^2$ | 0.563               | 0.555               | 0.552               | 0.559               | 0.576                | 0.553               | 0.571                |
| Year-month FE  | Yes                 | Yes                 | Yes                 | Yes                 | Yes                  | Yes                 | Yes                  |

**Note(s):** This table presents the OLS regression results of crude oil returns on the SESTM by combining other sentiment scores. Variables are defined in Appendix. All regressions include year-month fixed effects.  $t$ -statistics are in parentheses. \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

low statistical significance. Therefore, we confirm that the SESTM is an essential variable in explaining the variation of oil returns from other sentiment scores.

Table 5 indicates that sentiment scores derived from the SESTM can predict oil returns over at least one week. This result is supported by training the model on data from 2014 to 2016 and testing on data from 2017 to 2019. To provide more robust results, we divide the entire sample randomly by 50% into the training set and use the remaining 50% as the test set to verify whether the sentiment scores calculated by SESTM are still valid.

The random sample results are shown in Table 7. Column (1) shows results without any control variables, and Column (2) shows results with controls. We find a consistent and positive association between alternative SESTM measures and crude oil return, indicating that data sampling is not an issue.

Another interesting question is whether the predictive power of our sentiment score by SESTM on oil returns can persist. One possible answer is that short-term predictive power can be maintained, but predicting over a longer period becomes challenging because market prices should immediately reflect new information and rational traders can exploit such opportunities. Moreover, the predictability in commodity futures markets has recently

Table 7. Return predictability of the SESTEM using randomly trained data

|                | (1)                   | (2)                 |
|----------------|-----------------------|---------------------|
| SESTM          | 0.034***<br>(16.249)  | 0.084***<br>(3.758) |
| US10 return    |                       | −0.047<br>(−0.247)  |
| US30 return    |                       | 0.226<br>(1.042)    |
| VIX return     |                       | −0.019<br>(−0.355)  |
| Gold return    |                       | 1.592*<br>(2.537)   |
| S&P500 return  |                       | −0.130<br>(−0.211)  |
| NASDAQ return  |                       | −0.153<br>(−0.327)  |
| Dollar return  |                       | −0.502<br>(−1.305)  |
| Gas return     |                       | −0.015<br>(−0.387)  |
| EPU return     |                       | 0.641<br>(0.998)    |
| PXE return     |                       | −0.019<br>(−0.030)  |
| PXE volume     |                       | 0.000<br>(0.911)    |
| Constant       | −0.016***<br>(−2.714) | −0.123<br>(−1.430)  |
| Observations   | 192                   | 192                 |
| Adjusted $R^2$ | 0.028                 | 0.558               |
| Year-month FE  | Yes                   | Yes                 |

**Note(s):** This table presents the OLS regression results examining the impact of the SESTM on the crude oil returns using the SESTM scores from randomly selected training data. Column (1) provides the regression result for the sentiment scores calculated by the SESTM method. Column (2) presents the results obtained after including control variables. Variables are defined in Appendix. All regressions include year-month fixed effects.  $t$ -statistics are in parentheses. \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

decreased due to the increased participation of professional financial traders. To test this point, we conduct an additional test by including lagged sentiment scores.

Table 8 shows the regression results by including additional lagged SESTM variables. From Column (1) to (4), we run regressions sequentially by including from one to four lagged SESTM variables, respectively. The results show that lagged variables are not statistically significant, but the SESTM still has the predictive power for oil returns. This result implies that the SESTM has the predictive power for oil returns one week ahead. In other words, news effects generally do not last longer than a week. Thus, this result is consistent with the traditional view of the market efficiency hypothesis.

**4.2.2 ML-based forecasting.** Previously, we provide evidence that sentiment scores based on SESTM are related to crude oil returns and investigate whether text-based information can enhance the predictability of crude oil returns. We employ several ML models to determine whether sentiment scores can enhance forecasting ability. To implement the ML models, 70% of the sample is randomly selected to train the models, and the remaining 30% is used for testing and validation.

The input features include returns from security market indices (S&P 500, NASDAQ, VIX, US dollar, US 10-year and 30-year treasury bonds), natural gas futures, and the EPU index. In addition to these variables, the SESTM and other text-based sentiment measures are used. Crude oil returns are the output variable. We apply ML models both with and without the SESTM and compare the results. The effectiveness of text-based sentiment scores on crude oil returns is evaluated by computing the out-of-sample R-square, root mean square error (RMSE), mean absolute error (MAE), and Quasi-Likelihood (QLIKE). If the SESTM captures the variation in the crude oil return, then the model including the SESTM should exhibit a higher R-square and lower RMSE, MAE, and QLIKE.

Table 9 presents a summary of the performance evaluations of the ML forecasting models. Panel A presents the comparative results for the R-square values. Across various ML models, including the SESTM leads to substantially higher R-square estimates than models that exclude it. Specifically, while R-square values without the SESTM range from 6% to 17%, incorporating the SESTM significantly increases them to between 30 and 37%. Furthermore,

**Table 8.** Forecasting ability of the SESTM

|                         | (1)                   | (2)                   | (3)                   | (4)                   |
|-------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| SESTM                   | 0.028***<br>(5.206)   | 0.028***<br>(5.232)   | 0.027***<br>(5.137)   | 0.028***<br>(5.166)   |
| SESTM lag1              | 0.007<br>(1.235)      | 0.007<br>(1.284)      | 0.007<br>(1.254)      | 0.007<br>(1.306)      |
| SESTM lag2              |                       | 0.007<br>(1.340)      | 0.007<br>(1.292)      | 0.007<br>(1.307)      |
| SESTM lag3              |                       |                       | −0.005<br>(−0.978)    | −0.005<br>(−0.938)    |
| SESTM lag4              |                       |                       |                       | 0.004<br>(0.743)      |
| Constant                | −0.019***<br>(−3.315) | −0.023***<br>(−3.559) | −0.020***<br>(−2.704) | −0.022***<br>(−2.762) |
| Observations            | 192                   | 192                   | 192                   | 192                   |
| Adjusted R <sup>2</sup> | 0.028                 | 0.029                 | 0.029                 | 0.028                 |
| Year-month FE           | Yes                   | Yes                   | Yes                   | Yes                   |

**Note(s):** This table presents the OLS regression results examining the impact of the SESTM on the crude oil returns. Lagged SESTM measures are included to assess the forecasting power of the SESTM measure. SESTM lag $k$  represents  $k$  lagged SESTM variable. All regressions include year-month fixed effects.  $t$ -statistics are in parentheses. \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

**Table 9.** Performance comparison

| Model                        | Without SESTM | With SESTM |
|------------------------------|---------------|------------|
| Panel A. R-square comparison |               |            |
| LGBM                         | 0.060         | 0.300      |
| CB                           | 0.070         | 0.330      |
| RF                           | 0.070         | 0.370      |
| LR                           | 0.170         | 0.310      |
| Panel B. RMSE comparison     |               |            |
| LGBM                         | 0.040         | 0.020      |
| CB                           | 0.040         | 0.010      |
| RF                           | 0.040         | 0.010      |
| LR                           | 0.040         | 0.020      |
| Panel C. MAE comparison      |               |            |
| LGBM                         | 0.030         | 0.030      |
| CB                           | 0.030         | 0.030      |
| RF                           | 0.030         | 0.020      |
| LR                           | 0.020         | 0.020      |
| Panel D. QLIKE comparison    |               |            |
| LGBM                         | 0.009         | 0.007      |
| CB                           | 0.011         | 0.006      |
| RF                           | 0.012         | 0.007      |
| LR                           | 0.008         | 0.005      |

**Note(s):** This table summarizes the performance evaluation of ML forecasting models. Four ML models are used: Light Gradient Boosting Machine (LGBM), Categorical Boosting (CB), Random Forest (RF), and Linear Regression (LR). Panel A represents the R-square results, Panel B displays the Root Mean Squared Error (RMSE), Panel C shows the Mean Absolute Error (MAE), and Panel D reports the Quasi-Likelihood (QLIKE)

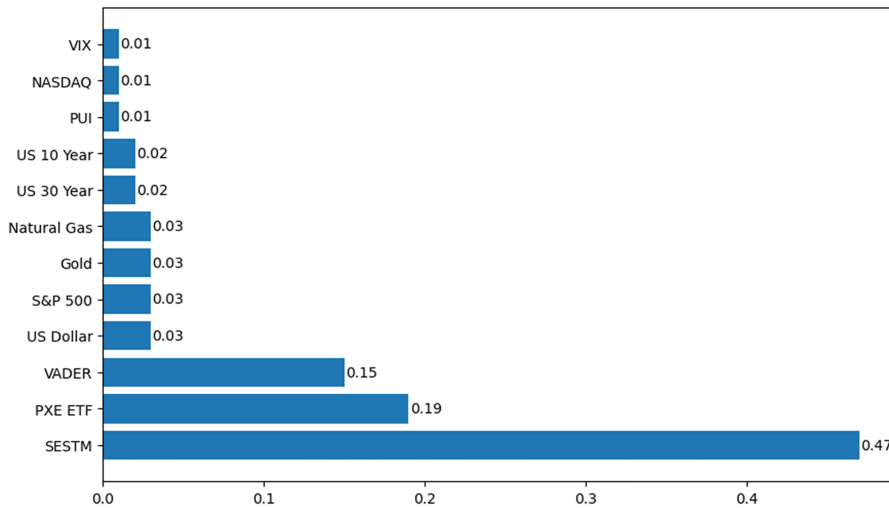
**Source(s):** Authors' own work

Panels B, C, and D examine the changes in forecasting errors when sentiment measures derived from the SESTM are included or excluded. The RMSE, MAE, and QLIKE metrics decrease considerably across all ML models with the inclusion of the SESTM. These findings indicate that the SESTM provides significantly greater explanatory power in capturing crude oil price fluctuations.

Next, we investigate the relative importance of individual variables. For each method, we calculate the reduction in the out-of-sample R-square by setting all values of a given predictor to zero within each training sample and averaging these into a single importance measure for each predictor. The characteristic importance within the model is normalized such that the relative importance of a particular model can be interpreted. Figure 5 shows the overall ranking of the feature importance in the RF model. Consistent with previous findings, the SESTM has the highest importance in the prediction of crude oil returns. Specifically, the characteristic importance of the SESTM model is approximately 0.47, accounting for half of the variables. Other important variables selected by the RF model include PXE returns and the sentiment score by the VADER algorithm. Thus, ML forecasting models suggest that the SESTM could be the most significant explanatory variable for crude oil forecasting. In other words, measuring text information correctly could have significant explanatory power in the prediction of crude oil returns.

4.3 Trading implications

After establishing the significance of the SESTM model in crude oil prediction, we further analyze whether trading strategies based on SESTM sentiment scores are profitable. We design a trading strategy in which we take a long position on WTI index futures if the SESTM sentiment score is above 0.5, and a short position if the sentiment score is below 0.5. We make a trading decision at the moment the article is released to the public, hold the position for one



**Note(s):** This figure presents the ranking of the feature importance as identified by the RF model. For each method, we calculate the reduction in the out-of-sample  $R$ -square by setting all values of a given predictor to zero within each training sample, then averaging these values into a single importance measure for each predictor. The characteristic importance within the model is normalized so that the relative importance of the particular model can be interpreted

**Source(s):** Authors' own work

**Figure 5.** Characteristics importance calculated by the random forest algorithm

day, and clear the position. In other words, we do not make any trading decisions on non-trading days. For articles published on Friday, we clear the position at the open price on the following Monday. After constructing the trading portfolio, we use several benchmark indices to determine whether the trading strategy could have significant alphas. We use the Goldman Sachs Commodity Index (GSCI) to represent the commodities markets. Moreover, we employ two ETFs that follow US crude oil prices: the United States Oil Fund (USO) and the United States 12 Month Oil Fund (USL).

Table 10 presents the summary results of the asset pricing tests for the trading strategies constructed based on various text-based sentiment scores. Panel A shows the results when the GSCI index is a benchmark. Similarly, Panels B and C show the results of using the USO and the USL as benchmarks, respectively. Across all panels, the alphas are positive and statistically significant for the use of the SESTM algorithm. However, negative alphas are observed for other sentiment scores (the HIV and LM dictionaries and the VADER algorithm). While the investment strategies based on the sentiment of information may outperform the benchmark, the strategy based on the SESTM shows a positive alpha, while others show negative alphas. These results imply that the SESTM algorithm can correctly identify the trading direction. That is, text-based methods based on the simple pre-defined dictionary approach may fail to reflect the direction of crude oil returns accurately and, thus, make it infeasible to implement profitable trading strategies.

#### 4.4 Position changes

This subsection further investigates how commercial and non-commercial traders respond to news sentiments. Kang *et al.* (2020) investigate the relationship between past commodities

**Table 10.** Portfolio construction and testing alpha

| Model                          | Alpha                  | Benchmark              | N   | Adjusted R <sup>2</sup> |
|--------------------------------|------------------------|------------------------|-----|-------------------------|
| Panel A. Benchmark: GSCI index |                        |                        |     |                         |
| SESTM                          | 0.008***<br>(12.132)   | −0.237***<br>(−4.192)  | 926 | 0.018                   |
| HIV                            | −0.005***<br>(−8.611)  | 0.820***<br>(−15.486)  | 926 | 0.205                   |
| LM                             | −0.006***<br>(−10.683) | −0.882***<br>(−17.363) | 926 | 0.245                   |
| VADER                          | −0.005***<br>(−7.703)  | −0.228***<br>(−3.864)  | 926 | 0.015                   |
| Panel B. Benchmark: USO ETF    |                        |                        |     |                         |
| SESTM                          | 0.008***<br>(12.146)   | −0.157***<br>(−4.853)  | 926 | 0.024                   |
| HIV                            | −0.005***<br>(−8.513)  | 0.475***<br>(15.563)   | 926 | 0.207                   |
| LM                             | −0.007***<br>(−10.701) | −0.492***<br>(−16.663) | 926 | 0.231                   |
| VADER                          | −0.005***<br>(−7.698)  | −0.133***<br>(−3.926)  | 926 | 0.015                   |
| Panel C. Benchmark: USL ETF    |                        |                        |     |                         |
| SESTM                          | 0.008***<br>(12.211)   | −0.192***<br>(−5.116)  | 926 | 0.026                   |
| HIV                            | −0.005***<br>(−8.494)  | 0.513***<br>(14.245)   | 926 | 0.179                   |
| LM                             | −0.006***<br>(−10.550) | −0.572***<br>(−16.700) | 926 | 0.231                   |
| VADER                          | −0.005***<br>(−7.731)  | −0.163***<br>(−4.158)  | 926 | 0.017                   |

**Note(s):** This table presents the results of portfolio performance tests. Panel A reports the results using the GSCI index as the benchmark return, Panel B presents results using the USO ETF as the benchmark, and Panel C provides results using the USL ETF as the benchmark. The significance of alphas is calculated for various sentiment scores calculated from the SESTM model, HIV dictionary, LM dictionary, and VADER algorithm. *t*-statistics are in parentheses. \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

returns and the traders' positions and find that the speculators' (hedgers') positions are positively (negatively) associated with past commodities returns. They interpret these results as evidence that speculators may employ a momentum strategy, and hedgers could play the role of liquidity providers. Based on their empirical findings, if the SESTM score accurately reflects the information in news articles, similar results should be observed in the relationships between traders' positions and SESTM scores. Specifically, when positive SESTM scores are observed, which is an indicator of an increase in crude oil prices, commercial traders may try to increase their short positions to hedge against price increases, and non-commercial traders are expected to increase their long positions to generate trading profits. Accordingly, we utilize the weekly commitment of traders reported by the CFTC to investigate the relationship between SESTM scores and traders' positions.

Table 11 provides the regression estimation results. The variables of interest include various sentiment scores. Columns (1) and (5) show the results of the SESTM regressions, while the others present those of other text-based sentiment score methods. The empirical results show that when the sentiment of news articles is positive, commercial traders tend to reduce their long positions, while non-commercial traders increase their long positions. As commercial traders include investors who hedge their business activities, they tend to short their positions when crude oil prices are expected to increase, which is reflected by positive

**Table 11.** Position change and the SESTM

|                         | Commercials<br>(1)  | (2)                 | (3)                 | (4)                 | Non-commercials<br>(5) | (6)                | (7)                | (8)                |
|-------------------------|---------------------|---------------------|---------------------|---------------------|------------------------|--------------------|--------------------|--------------------|
| SESTM                   | −0.003*<br>(−1.667) |                     |                     |                     | 0.002*<br>(1.746)      |                    |                    |                    |
| HIV                     |                     | 0.030**<br>(2.399)  |                     |                     |                        | −0.006<br>(−0.750) |                    |                    |
| LM                      |                     |                     | 0.017<br>(−1.632)   |                     |                        |                    | −0.009<br>(−1.272) |                    |
| VADER                   |                     |                     |                     | 0.001<br>(1.522)    |                        |                    |                    | −0.000<br>(−0.927) |
| VIX return              | 0.010<br>(0.921)    | 0.006<br>(0.535)    | 0.004<br>(0.384)    | 0.008<br>(0.717)    | 0.004<br>(0.570)       | 0.007<br>(0.880)   | 0.008<br>(1.037)   | 0.006<br>(0.810)   |
| Gold return             | −0.013<br>(−0.373)  | −0.011<br>(−0.319)  | −0.016<br>(−0.450)  | −0.020<br>(−0.549)  | −0.011<br>(−0.450)     | −0.011<br>(−0.469) | −0.010<br>(−0.400) | −0.009<br>(−0.357) |
| SP&500 return           | 0.256<br>(1.459)    | 0.204<br>(1.176)    | 0.186<br>(1.054)    | 0.212<br>(1.213)    | 0.011<br>(0.105)       | 0.052<br>(0.490)   | 0.067<br>(0.631)   | 0.056<br>(0.524)   |
| NASDAQ return           | −0.119<br>(−1.118)  | −0.119<br>(−1.125)  | −0.095<br>(−0.894)  | −0.111<br>(−1.044)  | 0.034<br>(0.486)       | 0.022<br>(0.319)   | 0.014<br>(0.205)   | 0.021<br>(0.299)   |
| Dollar return           | 0.111<br>(0.935)    | 0.111<br>(0.939)    | 0.095<br>(0.800)    | 0.112<br>(0.939)    | −0.103<br>(−1.292)     | −0.101<br>(−1.258) | −0.095<br>(−1.183) | −0.103<br>(−1.275) |
| Gas return              | 0.017<br>(0.942)    | 0.010<br>(0.558)    | 0.012<br>(0.680)    | 0.012<br>(0.646)    | −0.016<br>(−1.355)     | −0.013<br>(−1.068) | −0.013<br>(−1.075) | −0.013<br>(−1.045) |
| EPU return              | 0.002*<br>(1.960)   | 0.002*<br>(1.789)   | 0.002*<br>(1.882)   | 0.002*<br>(1.815)   | 0.000<br>(0.416)       | 0.001<br>(0.652)   | 0.000<br>(0.570)   | 0.001<br>(0.634)   |
| FSI                     | 0.012***<br>(3.043) | 0.011***<br>(2.694) | 0.012***<br>(2.931) | 0.012***<br>(3.023) | −0.003<br>(−1.271)     | −0.003<br>(−1.064) | −0.003<br>(−1.158) | −0.003<br>(−1.206) |
| OVX return              | 0.044***<br>(3.976) | 0.042***<br>(3.769) | 0.043***<br>(3.845) | 0.046***<br>(4.089) | 0.009<br>(1.210)       | 0.009<br>(1.154)   | 0.009<br>(1.221)   | 0.008<br>(1.050)   |
| Constant                | 0.012***<br>(2.706) | −0.011<br>(−1.553)  | −0.001<br>(−0.256)  | 0.006***<br>(3.238) | −0.005*<br>(−1.802)    | 0.003<br>(0.597)   | 0.003<br>(0.954)   | −0.001<br>(−0.663) |
| Observations            | 192                 | 192                 | 192                 | 192                 | 192                    | 192                | 192                | 192                |
| Adjusted R <sup>2</sup> | 0.083               | 0.098               | 0.083               | 0.081               | 0.017                  | 0.003              | 0.009              | 0.005              |
| Year-month FE           | Yes                 | Yes                 | Yes                 | Yes                 | Yes                    | Yes                | Yes                | Yes                |

**Note(s):** This table presents the OLS regression results examining the impact of various sentiment measures on crude oil returns by different types of investors. Columns from (1) to (4) provide the regression results for the sentiment scores calculated by the SESTM method, HIV dictionary, LM dictionary, and VADER algorithm, respectively, on the position changes of commercial traders. Columns from (5) to (8) present the results for the position changes of non-commercial traders. Variables are defined in [Appendix](#). All regressions include year-month fixed effects. *t*-statistics are in parentheses. \*, \*\*, and \*\*\* denote significance at the 10, 5, and 1% level, respectively

**Source(s):** Authors' own work

SESTM scores. In contrast, non-commercial traders, such as financial institutions, have incentives to increase their long positions to enhance their profits. Thus, our empirical results support the findings of [Kang et al. \(2020\)](#). Moreover, the SESTM algorithm can extract the sentiment score relatively accurately, which reflects the information related to crude oil prices.

## 5. Conclusion

This study uses news articles with information on crude oil to investigate whether sentiment scores can predict crude oil prices and explain the behavior of market participants. To this end, we compute the sentiment scores of the articles to predict crude oil prices by employing a generative, interpretable, and transparent ML model. Specifically, we employ the SESTM algorithm developed by [Ke et al. \(2019\)](#). Moreover, we evaluate whether the SESTM algorithm outperforms the conventional predefined dictionary-based sentiment scores.

The empirical results show that the SESTM algorithm predicts the crude oil returns and position changes of commercial and noncommercial traders more accurately than existing methods. Moreover, the ML forecasting results indicate that incorporating textual information based on the SESTM can dramatically enhance the explanatory power of crude oil returns. This study contributes to the literature by demonstrating that correctly identified news sentiment is a significant factor in explaining market participants' positions. Moreover, this study confirms that the SESTM can explain traders' positions. These findings further support the notion that the SESTM methods can extract accurate information from news articles.

This study contributes to the growing body of textual analysis in empirical finance and economics. Sentiment analysis studies are proliferating in various fields. Conventional textual analyses in finance do not seem to provide strong forecasting powers because existing models compute sentiment scores based on predefined dictionaries that may contain less contents related to finance and accounting ([Loughran and McDonald, 2011](#)). Moreover, [Loughran and McDonald \(2016\)](#) raise concerns about readability measures such as the Fog Index, which is popularly used in finance and accounting. Compared to existing text-based methods, this study provides evidence that supervised learning methods that match texts with asset returns could be more effective in extracting reliable information from news articles. Thus, we conclude that text information can be valuable in explaining asset returns when the proxy for information content is correctly identified.

## Notes

1. The financialization of commodities has significantly increased since 2004 owing to large investment inflows from commodity index traders ([Cheng and Xiong, 2014](#); [Basak and Pavlova, 2016](#)).
2. The results are consistent across different types of ML models. For instance, the explanatory powers of the LGBM, CB, RF, and linear regression without the SESTM score are 0.06, 0.07, 0.07, and 0.17, respectively. Notably, the inclusion of the SESTM score increases the explanatory powers to 0.30, 0.33, 0.37, and 0.31, respectively.
3. The sentiment score is the difference between positive and negative word counts based on a predefined dictionary.
4. [Loughran and McDonald \(2011\)](#) criticize the use of the HIV dictionary, arguing that sentiment analysis in business communications should avoid "classification schemes derived outside the domain of business usage" (p. 62). They suggest instead a list of words designed for business communication by utilizing selective words from 10-K filings.
5. For instance, the word "easy" is classified as a positive word in the LM word list. Depending on the context of the article, however, the word "easy" may imply a negative connotation.
6. Researchers commonly argue that online news articles from reliable sources, such as the Wall Street Journal or Bloomberg, convey better information as compared to general comments on blogs or social media because newspaper articles are more convincing and generally contain less noise.

7. For example, the term “Apple” may refer to either the fruit or the company, creating ambiguity in textual analysis. To address such challenges, previous studies have used ticker symbols to search for firm-level information (Da *et al.*, 2011). However, many news articles do not include ticker symbols, which makes it difficult to utilize the overall textual context.
8. All variables are described in Appendix.
9. The algorithm classifies articles as positive or negative and provides polarity scores to reflect their tone. Using each score, the algorithm produces a compound score. The compound score is a metric that calculates the sum of all lexicon ratings normalized between  $-1$  and  $1$ . Articles are classified as positive if the compound score is  $\geq 0.05$ , neutral if between  $-0.05$  and  $0.05$ , and negative if  $\leq -0.05$ .
10. Ke *et al.* (2019) argue that principal components analysis (PCA) functions poorly in reducing dimensions of text data.
11. The results in Columns (2) and (6) show that coefficients for the HIV are negative. The coefficient in Column (2) is significant but that in Column (6) is insignificant. The results in Columns (3) and (7) indicate that coefficients for the LM are statistically negative and significant. Although significant, the LM is negatively associated with crude oil returns, which show opposite directions. The results for the VADER are insignificant in Columns (4) and (8).

## References

- Anifowose, F.A., Labadin, J. and Abdulraheem, A. (2017), “Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization”, *Journal of Petroleum Science and Engineering*, Vol. 151, pp. 480-487, doi: [10.1016/j.petrol.2017.01.024](https://doi.org/10.1016/j.petrol.2017.01.024).
- Basak, S. and Pavlova, A. (2016), “A model of financialization of commodities”, *Journal of Finance*, Vol. 71 No. 4, pp. 1511-1556, doi: [10.1111/jofi.12408](https://doi.org/10.1111/jofi.12408).
- Bengio, Y. (2009), “Learning deep architectures for AI”, *Foundations and trends in Machine Learning*, Vol. 2, pp. 1-127, doi: [10.1561/2200000006](https://doi.org/10.1561/2200000006).
- Boudoukh, J., Feldman, R., Kogan, S. and Richardson, M. (2013), *Which News Moves Stock Prices? A Textual Analysis*. National Bureau of Economic Research.
- Cheng, I.-H. and Xiong, W. (2014), “Financialization of commodity markets”, *Annual Review of Financial Economics*, Vol. 6 No. 1, pp. 419-441, doi: [10.1146/annurev-financial-110613-034432](https://doi.org/10.1146/annurev-financial-110613-034432).
- Chiroma, H., Abdulkareem, S. and Herawan, T. (2015), “Evolutionary neural network model for West Texas intermediate crude oil price prediction”, *Applied Energy*, Vol. 142, pp. 266-273, doi: [10.1016/j.apenergy.2014.12.045](https://doi.org/10.1016/j.apenergy.2014.12.045).
- Cowles 3rd, A. (1933), “Can stock market forecasters forecast?”, *Econometrica: Journal of the Econometric Society*, Vol. 1 No. 3, pp. 309-324, doi: [10.2307/1907042](https://doi.org/10.2307/1907042).
- Da, Z., Engelberg, J. and Gao, P. (2011), “In search of attention”, *The Journal of Finance*, Vol. 66 No. 5, pp. 1461-1499, doi: [10.1111/j.1540-6261.2011.01679.x](https://doi.org/10.1111/j.1540-6261.2011.01679.x).
- Desjardins, J. (2016), “The oil market is bigger than all metal markets combined”, *Visual Capitalist*.
- Dougal, C., Engelberg, J., Garcia, D. and Parsons, C.A. (2012), “Journalists and the stock market”, *The Review of Financial Studies*, Vol. 25 No. 3, pp. 639-679, doi: [10.1093/rfs/hhr133](https://doi.org/10.1093/rfs/hhr133).
- Fulford, D.S., Bowie, B., Berry, M.E., Bowen, B. and Turk, D.W. (2016), “Machine learning as a reliable technology for evaluating time/rate performance of unconventional wells”, *SPE Economics and Management*, Vol. 8 No. 1, pp. 23-39, doi: [10.2118/174784-pa](https://doi.org/10.2118/174784-pa).
- Gabralla, L.A., Jammazi, R. and Abraham, A. (2013), “Oil price prediction using ensemble machine learning”, *2013 International Conference on Computing, Electrical and Electronic Engineering (ICCEEE)*, IEEE, pp. 674-679.
- Gu, S., Kelly, B. and Xiu, D. (2020), “Empirical asset pricing via machine learning”, *Review of Financial Studies*, Vol. 33 No. 5, pp. 2223-2273, doi: [10.1093/rfs/hhaa009](https://doi.org/10.1093/rfs/hhaa009).
- Guo, X., Li, D. and Zhang, A. (2012), “Improved support vector machine oil price forecast model based on genetic algorithm optimization parameters”, *Aasri Procedia*, Vol. 1, pp. 525-530, doi: [10.1016/j.aasri.2012.06.082](https://doi.org/10.1016/j.aasri.2012.06.082).

- Haidar, I., Kulkarni, S. and Pan, H. (2008), "Forecasting model for crude oil prices based on artificial neural networks", *2008 International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, IEEE, pp. 103-108.
- Ho, T.K. (1995), "Random decision forests", *Proceedings of 3rd International Conference on Document Analysis and Recognition*, IEEE, pp. 278-282.
- Ho, T.K. (1998), "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 No. 8, pp. 832-844, doi: [10.1109/34.709601](https://doi.org/10.1109/34.709601).
- Jammazi, R. and Aloui, C. (2012), "Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modeling", *Energy Economics*, Vol. 34 No. 3, pp. 828-841, doi: [10.1016/j.eneco.2011.07.018](https://doi.org/10.1016/j.eneco.2011.07.018).
- Kaboudan, M.A. (2003), "Short-term compumetric forecast of crude oil prices", *Modeling and Control of Economic Systems 2001*, Elsevier, pp. 365-370.
- Kang, W., Rouwenhorst, K.G. and Tang, K. (2020), "A tale of two premiums: the role of hedgers and speculators in commodity futures markets", *Journal of Finance*, Vol. 75 No. 1, pp. 377-417, doi: [10.1111/jofi.12845](https://doi.org/10.1111/jofi.12845).
- Ke, Z.T., Kelly, B.T. and Xiu, D. (2019), *Predicting Returns With Text Data*. National Bureau of Economic Research.
- Li, X., Shang, W. and Wang, S. (2019), "Text-based crude oil price forecasting: a deep learning approach", *International Journal of Forecasting*, Vol. 35 No. 4, pp. 1548-1560, doi: [10.1016/j.ijforecast.2018.07.006](https://doi.org/10.1016/j.ijforecast.2018.07.006).
- Loughran, T. and McDonald, B. (2011), "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks", *The Journal of Finance*, Vol. 66 No. 1, pp. 35-65, doi: [10.1111/j.1540-6261.2010.01625.x](https://doi.org/10.1111/j.1540-6261.2010.01625.x).
- Loughran, T. and McDonald, B. (2016), "Textual analysis in accounting and finance: a survey", *Journal of Accounting Research*, Vol. 54 No. 4, pp. 1187-1230, doi: [10.1111/1475-679x.12123](https://doi.org/10.1111/1475-679x.12123).
- Mehta, M., Agrawal, R. and Rissanen, J. (1996), "SLIQ: a fast scalable classifier for data mining", *Advances in Database Technology—EDBT'96: 5th International Conference on Extending Database Technology, March 25-29, 1996 Proceedings 5*, Springer, Avignon, France, pp. 18-32.
- Moshiri, S. and Foroutan, F. (2006), "Forecasting nonlinear crude oil futures prices", *Energy Journal*, Vol. 27 No. 4, pp. 81-96, doi: [10.5547/issn0195-6574-ej-vol27-no4-4](https://doi.org/10.5547/issn0195-6574-ej-vol27-no4-4).
- Mostafa, M.M. and El-Masry, A.A. (2016), "Oil price forecasting using gene expression programming and artificial neural networks", *Economic Modelling*, Vol. 54, pp. 40-53, doi: [10.1016/j.econmod.2015.12.014](https://doi.org/10.1016/j.econmod.2015.12.014).
- Pirayonesi, S.M. and El-Diraby, T.E. (2020), "Role of data analytics in infrastructure asset management: overcoming data size and quality problems", *Journal of Transportation Engineering, Part B: Pavements*, Vol. 146 No. 2, 04020022, doi: [10.1061/jpeodx.0000175](https://doi.org/10.1061/jpeodx.0000175).
- Plourde, A. and Watkins, G.C. (1998), "Crude oil prices between 1985 and 1994: how volatile in relation to other commodities?", *Resource and Energy Economics*, Vol. 20 No. 3, pp. 245-262, doi: [10.1016/s0928-7655\(97\)00027-4](https://doi.org/10.1016/s0928-7655(97)00027-4).
- Robe, M.A. and Wallen, J. (2016), "Fundamentals, derivatives market information and oil price volatility", *Journal of Futures Markets*, Vol. 36 No. 4, pp. 317-344, doi: [10.1002/fut.21732](https://doi.org/10.1002/fut.21732).
- Shin, H., Hou, T., Park, K., Park, C.-K. and Choi, S. (2013), "Prediction of movement direction in crude oil prices based on semi-supervised learning", *Decision Support Systems*, Vol. 55 No. 1, pp. 348-358, doi: [10.1016/j.dss.2012.11.009](https://doi.org/10.1016/j.dss.2012.11.009).
- Tetlock, P.C. (2007), "Giving content to investor sentiment: the role of media in the stock market", *The Journal of Finance*, Vol. 62 No. 3, pp. 1139-1168, doi: [10.1111/j.1540-6261.2007.01232.x](https://doi.org/10.1111/j.1540-6261.2007.01232.x).
- Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S. (2008), "More than words: quantifying language to measure firms' fundamentals", *The Journal of Finance*, Vol. 63 No. 3, pp. 1437-1467, doi: [10.1111/j.1540-6261.2008.01362.x](https://doi.org/10.1111/j.1540-6261.2008.01362.x).

- Xie, W., Yu, L., Xu, S. and Wang, S. (2006), "A new method for crude oil price forecasting based on support vector machines", *Computational Science–ICCS 2006: 6th International Conference, Reading, UK, May 28–31, 2006, Proceedings, Part IV* 6, Springer, pp. 444–451.
- Yang, C.-W. (2023), "Investment strategy via analyst report text mining", *Journal of Derivatives and Quantitative Studies*, Vol. 31 No. 2, pp. 98–120, doi: [10.1108/jdqs-09-2022-0022](https://doi.org/10.1108/jdqs-09-2022-0022).
- Zhao, Y., Li, J. and Yu, L. (2017), "A deep learning ensemble approach for crude oil price forecasting", *Energy Economics*, Vol. 66, pp. 9–16, doi: [10.1016/j.eneco.2017.05.023](https://doi.org/10.1016/j.eneco.2017.05.023).

## Appendix

**Table A1.** Variable description

| Variable                            | Description                                                                                                                                                                             |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| SESTM                               | Sentiment score measured by SESTM method (Ke et al., 2019). The sentiment score ranges between 0 to 1 where a higher value implies a positive sentiment                                 |
| HIV                                 | Sentiment score measured by Harvard-IV dictionary. The sentiment score ranges between 0 to 1 where a higher value implies a more positive sentiment                                     |
| LM                                  | Sentiment score measured by Loughran and McDonald dictionary (Loughran and McDonald, 2011). The sentiment score ranges between 0 to 1 where a higher value implies a positive sentiment |
| VADER                               | Sentiment score measured by VADER algorithm. The sentiment score ranges between –1 to 1 where a higher value implies a more positive sentiment                                          |
| US10 return                         | A weekly return of the US 10-year treasury bond                                                                                                                                         |
| US30 return                         | A weekly return of the US 30-year treasury bond                                                                                                                                         |
| VIX return                          | A weekly return of the Chicago Board Options Exchange (CBOE)'s Volatility Index                                                                                                         |
| OVX return                          | A weekly return of the CBOE Crude Oil ETF Volatility Index                                                                                                                              |
| Gold return                         | A weekly return of the gold futures                                                                                                                                                     |
| SP&500 return                       | A weekly return of the S&P 500 index                                                                                                                                                    |
| NASDAQ return                       | A weekly return of the NASDAQ index                                                                                                                                                     |
| Dollar return                       | A weekly return of the US dollar index                                                                                                                                                  |
| Gas return                          | A weekly return of natural gas futures                                                                                                                                                  |
| EPU return                          | A weekly return of the economic policy uncertainty (EPU) change index                                                                                                                   |
| PXE return                          | A weekly return of the Invesco Dynamic Energy Exploration and Production (PXE) ETF                                                                                                      |
| <b>Source(s):</b> Authors' own work |                                                                                                                                                                                         |

## Corresponding author

Daejin Kim can be contacted at: [daejin@skku.edu](mailto:daejin@skku.edu)