Comprehending the Influence of Oil Shock News[1]

Huiming (Minna) Zhang[2]

March 10, 2024

Abstract

This study investigates the impact of financial news tones, particularly regarding "oil shocks", on market responses using the advanced machine learning model FinGPT. It reveals the difference in the impacts of sentiment on oil returns and stock returns between the GPT model and the conventional dictionary-based language model (Harvard-Lasswell general dictionary and McDonald and Loughran financial dictionary) mainly exists on the "positive" side of the news. The difference exacerbates when text readability declines, contains more numerical words, and features firmer tones. We analyze "oil shock" news spanning from 1986 to 2019 sourced from financial newspapers and wire feeds, examining impacts of sentiments on oil and firm stock returns and showing the impact on oil returns and individual firms' abnormal returns. The study finds that readability of news and sureness in tones amplifies the impact of sentiments on returns and that the significant impacts prevail only during periods of high investor attention. FinGPT's sentiment is stronger in predicting oil returns when news has worse readability and more uncertain tones. The study finds that LLM may not always produce a higher predicting power in linking news tones to market impacts when gauging information from human natural language news but contingent on the comprehensibility of the text.

Keywords: Textual analysis, Oil price, Large language model, Sentiment

JEL classification: G14, G40, G41.

---

[1] I appreciate the comments from OU BrownBag seminar participants.

[2] Huiming (Minna) Zhang (minna.zhang@okstate.edu, 405-885-8548) is at Oklahoma State University, Spears School of Business, Stillwater, OK 74078.

1. Introduction

The use of machine learning techniques has become a prominent topic in financial technology, particularly with the rapid development of large language models (LLMs). Many scholars recognize the potential of these tools for processing and interpreting natural human language from extensive sources, including Garcia et al. (2023), who find machine learning superior to traditional dictionary methods by showing their machine learning technique extracts dictionaries better than the bag-of-words approach in predicting stock prices. Financial news, in particular, represents a valuable resource for investors seeking information about companies and financial markets. The question of how to comprehend the news and link it to the price impact is an ongoing debate. The conventional dictionary-based language model relies on using the percentage of words from the category of "negativity" or "positivity" to gauge the tones of the news to comprehend whether the news brings a positive or negative impact including Loughran and McDonald (2011), Tetlock (2007), and Feldman et al. (2010). This method is created by extracting the keywords (lexicon-based) from a large database of human language. Vaswani et al. (2017) in "Attention is All You Need" demonstrate that the effectiveness of deep learning models improves with the size of the model and show that the model surpasses all previous techniques in performing translation tasks. Hence, *Scalability matters.* The transformer model in natural language processing, as shown by Brown et al. (2020), indicates that the power of the model increases dramatically with an increase in parameters. The power of large language models has been demonstrated in showing superior ability in comprehending human natural language by AI such as ChatGPT and Gemini, which enables good performance in maintaining dialogues in human natural languages. The existing literature indicates that the tone or sentiment expressed in news articles can influence stock returns (e.g., Tetlock, 2007; Garcia, 2013; Loughran and McDonald, 2020). Some findings suggest that news coverage affects investors' awareness of information and can alter how they react to that information (Tetlock, 2011).

Most financial studies traditionally employ methods that rely on counting the appearance of negative words in text to measure the "negativity" in the news tone. Common dictionaries such as the Harvard-Lasswell and Loughran and McDonald dictionaries help identify these negative words. However, it is important to note that human natural language is context-sensitive and can convey different tones depending on the context. For example, consider the word "strike," which can have a negative connotation in a labor context but a positive one in a sports context. This highlights the need for a more nuanced approach to measuring tone in financial news.

FinGPT (Yang et al., 2023), an advanced machine learning model, offers a solution. Trained on an extensive corpus of text data, including Dow Jones Institutional News and The Wall Street Journal articles, it applies a technique that utilizes feedback from trainers to enhance the model's capabilities. Yang et al. (2023) have shown that FinGPT outperforms BloombergGPT and other models, including FinBERT, in generating more accurate representations. By doing so, FinGPT comprehends human natural language content, enabling it to effectively tokenize and model words within a text. Unlike dictionary-based methods, FinGPT distinguishes itself by considering the relevance of the text surrounding a word, allowing for a more efficient and precise understanding of human natural language.[3] This is particularly beneficial for analyzing sentiment, distinguishing between negative and positive tones in the news.

This study focuses on news coverage related to "oil shocks" collected from various widely followed sources, including the most circulated newspapers for financial news: The Wall Street Journal (including online and printed versions, U.S. Eastern issue, European issue, and Asian issue), The Financial Times, The New York Times, Dow Jones Institutional News serving institutional investors, and Energy Weekly News, an energy-news focused magazine. Our dataset, obtained from ProQuest for Dow Jones Institutional News and Factiva for the others, includes detailed information on each piece of news such as title, full text, date, timestamp, location, companies, NAICS codes, authors, category, topics, etc. Unlike news that covers individual firms, which is more sparse and concentrated around several high-attention periods, the coverage of oil news is a high-attention sector with a rich content feed. For instance, The Wall Street Journal has a section on “Energy” that covers news on world energy markets, and The New York Times has a section under Business for “Energy and Environment” (previously “Energy”). The oil market is a high-attention, highly covered, and rapidly moving market that receives a large amount of news coverage every day, unlike the stock market. Additionally, we also use a sample of individual firms’ stock returns to test the impact of oil shock news on the related firms’ market responses by extracting information on companies mentioned in each piece of news related to their abnormal returns on the coverage days. However, the news will be a smaller sample covering the “oil shocks” news on related firms (about 2,000 articles) We use the daily returns on West Texas Intermediate (WTI) crude oil to measure investor market responses. WTI is a critical commodity and financial market benchmark closely tied to economic fundamentals. The advantage of using oil prices, in addition, is that in the U.S., it serves as a benchmark to measure national crude oil prices, with

[3] Figure 1 is a picture showing the comparison among different machine learning techniques.

WTI prices traded on the Chicago Mercantile Exchange (CME). In total, our collection of “oil shock” news covers the years 1986–2019 and encompasses 24,000 articles and 2,000 articles related to individual firms.

Our findings reveal that sentiment (the negative tone minus positive tone) has a significant impact on oil returns. A more negative sentiment (tone) (increase by 100% probability) from “oil shock” news is associated with a -0.248 % change in the WTI oil return on a daily basis on average. Notably, we find that WSJ Online significantly influences changes in oil returns, whereas the traditional printed edition of the WSJ did not. The Dow Jones Institutional News, primarily targeting institutional investors, impacts oil returns, while other sources such as the Financial Times and New York Times do not show such effects. Furthermore, some studies suggest that financial literacy (e.g., Van Rooij et al., 2011 and Guiso and Viviano, 2014) may influence how investors respond to similar information. However, financial news readability has received little attention; such literacy includes Loughran and McDonald (2014). The ease with which investors comprehend news, as reflected in the complexity of grammar, text structure, and word choices, may have different impacts on their decision-making. We integrated the sentiments extracted from FinGPT to be interacted with the readability of the news and find that the readability of news on oil shocks moderates the impact of tone. We observe that, when controlling for sentiment and readability, the interaction of these two factors remains highly significant in affecting oil returns. Specifically, when we have more negative news tone and better readability, as measured by Fog Index, the return-reducing effect is stronger.

We further explore the impact of "oil shock" news by focusing on the individual firms' price responses to the sentiment of the news, concentrating on the firms that are mentioned in the news articles. We focus on a subsample of the Dow Jones Institutional News collected from ProQuest that contains detailed information on the names of the companies appearing in each piece of news, showing that having oil shock news coverage on the companies will lead to no significant changes in the abnormal returns of the individual firms; consistent results are shown when focusing on the three-day and five-day cumulative abnormal returns of the individual firms. But we demonstrate that the sentiment of the news change the impacts of the news coverage on abnormal returns of the individual firms. Those covered in oil shock news and are related to the shocks in oil markets also experience significant changes in their firm values, and the worse the tone of the news coverage, the greater the stickiness on the firm values.

We then compare FinGPT to dictionary-based models, analyzing the sentiment generated by FinGPT against that generated by traditional dictionary-based models. Each sentiment is computed as the "negative" tone of the news subtracted by the "positive" tone. When comparing the FinGPT sentiment with the dictionary-based sentiment, we find a difference exists on the "positive" side, while the "negative" side is very much aligned. Furthermore, we find that when news has lower readability (difficult to be quickly comprehended), FinGPT generates lower prediction errors for oil returns. Additionally, when news has a more uncertain tone, FinGPT generates better results in predicting oil returns. However, when numbers are higher, FinGPT's performance in predicting oil returns worsens. This suggests that FinGPT is particularly good at conducting sentiment analysis for human language text that is more difficult to comprehend, contains more non-numerical/non-quantitative content, and presents more ambiguous content. It shows superior ability in correctly gauging the information from "oil shock" news for complicated tasks, though its predicting power is generally less than that of the dictionary-based model of the Harvard-Lasswell dictionary. When examining the relationship between readability, numbers, and certainty in tones and the sentiments, we find that sentiment extracted from FinGPT correlates with these factors, but the Harvard-Lasswell dictionary is more stable across different sets of news. Interestingly, we find that the Harvard-Lasswell dictionary sentiment is negatively correlated with the other two sentiments, and this difference in sentiment mainly comes from the positive side of the sentiment. i.e., the negative component in the sentiment is pretty much aligned among the three sentiments, but the positive component differs.

The development of large language models in human natural languages has significantly advanced how we comprehend news information and interact with AI techniques. There is an ongoing debate about how these tools will change the way we interpret financial information and how quickly we can absorb this information. However, the utilization of these tools highly depends on how we apply them in specific cases. Using the "oil shock" news collected from financial newspapers and newsfeeds, we find that FinGPT is better at generating sentiment analysis for more complicated tasks in human natural languages. However, on average, sentiment analysis is not a complicated task, and the dictionary-based textual analysis technique may outperform the LLM in this task. But we also find that high investor attention in the oil market reinforces the impact of the GPT-generated sentiment from news on oil returns, suggesting that the results may be driven by the limited attention of investors to the complicated text of the news on "oil shock". To predict the market response to news, on average, a simple model may perform better in reflecting investors' limits on comprehending news information.

2. Literature review

2.1. Impact of Media Coverage Tone on Security Prices

Studies on commodity markets reveal that movements in oil prices are related to macroeconomic surprises (Rigobon and Sack, 2008), storage surprises (Halova, Kurov, and Kucher, 2014), storage forecasts (Ederington, Lin, Linn, and Yang, 2019), and weather forecasts (Fink and Fink, 2014, and Fink, Fink, and Russell, 2010). In this study, we analyze the impact of news releases including detailed narrations and analyses of surprises on oil prices.

Empirical studies on the impact of news tone suggest that the sentiment content of news coverage, including the tone and emotional language used in news reports, predicts price, volume, and volatility in the short run. Such studies include Sprenger et al. (2014) uses Tweets to analyze the sentiments in investors in microblogging forums to be related to trading volume and stock returns. Brown and Cliff (2004) use many investor sentiment measures to find they are not directly related to short-term stock market returns. Loughran and McDonald (2020) provide a comprehensive analysis of the literature on textual analysis in finance. Tetlock (2007) presents evidence that pessimistic Wall Street Journal articles lead to adverse price reactions among firms covered by the media. His findings contradict information theory as he observes price reversals in the short run following media coverage. In a subsequent study (Tetlock, 2011), he offers further evidence of price reversals for "stale news." Empirical studies of market benchmarks and news coverage suggest that sentiment in capital market coverage predicts changes in market indices (Garcia, 2013). Klibanoff, Lamont, and Wizman (1998) demonstrate the impact of news coverage on country fund returns, whereas Burt (2018) establishes that content similarity within information networks forecasts the covariances of turnovers and returns. Analyses based on television news coverage (Busse and Green, 2002; Meschke, 2003; Fehle, Tsyplakov, and Zdorovtsov, 2005) and studies based on online trading chat-room messages (Wysocki, 1998; Antweiler and Frank, 2005; Das and Chen, 2007) suggest that information content beyond authoritative announcements on fundamentals influences investor sentiment and conveys messages about potential investors' tendencies. Coval and Shumway (2001) demonstrate that noise levels in a trading pit forecast the volume, depth, and volatility of bonds.

Several studies concentrate on the impact of commodity coverage tone on security prices. For instance, Smales (2014) demonstrates that gold market news sentiment indices can forecast gold prices, Narayan and Narayan (2017) examine the impact of oil-related news tone on stock returns, and Loughran, McDonald, and Pragidis (2019) find that oil-specific words can forecast oil returns. However, no prior study examines the relationship between tone, investors’ attention, and commodity prices.

2.2. Machine learning and Security Prices

The application of AI and *machine learning* in finance is not a new topic. For instance, Varetto (1998) used an artificial intelligence technique, the Genetic Algorithm, one of the neural networks, to classify bankrupt firms and make predictions. Though the study found that the technique is not superior to the traditional linear discriminant analysis. Another such example is Altman et al. (1994), who used neural networks to diagnose financial distress in firms. Altman (1968) used multiple discriminant analysis to predict a firm's financial distress, compared with traditional ratio analysis used for group classification to extract common characteristics. The application of machine learning models more recently includes Sigrist and Hirnschall (2019), who used the grabit model, which can solve for nonlinear, discontinued, and complex interactions for the default prediction. Cheng and Cirillo (2018) used a reinforced urn process model to detect the recovery rates and times for counterparty.

Heaton et al. (2017) provide a detailed descriptions of the application of *deep learning* in finance, and Ozbayoglu et al. (2020) conduct a literature review on financial applications of deep learning. More recent studies include Sirignano (2019), who used a spatial neural network to analyze the limit order book distribution of bid and ask prices on U.S. stocks, whose results outperform logistic regression. In the specific area of predicting oil prices, He et al. (2018) used a Quantile Regression Neural Network to predict oil price fluctuations using fundamental and transient risk factors, showing this approach is particularly effective in predicting downside risks in oil prices.

Ghoddusi et al. (2019) provided a detailed literature review on the application of machine learning in *energy* economics and finance. The applications in the field of oil markets particularly focused on neutral network, one of the deep learning techniques, including Zhao et al. (2017), which used a deep learning ensemble approach to predict oil prices, including deep neural network bootstrap aggregation, both showing superior forecasting powers. Yu et al. (2008) used empirical mode decomposition-based neural network ensemble to forecast WTI oil prices. Similarly, Moshiri et al. (2006) applied neural network to forecast nonlinear oil prices. On the other hand, Dogah and Premaratne (2018) used Random Forest technique in the traditional VAR model to examine the impact of oil risk factors on sectoral stocks. However, most existing literature focuses on using fundamentals in the oil market to make forecasts, and none of these studies focus on using the sentiment extracted from "oil shock" news to predict oil prices.

Natural Language Processing's role is to analyze the words or word combinations in text. As Aziz et al. (2021) describes, AI is a large area, whereas Natural Language Processing is an area that could be intertwined with areas of machine learning. The Loughran-McDonald dictionary contains the positive and negative words (364 positive and 2329 negative words) extracted from 10-K. Studies that used the dictionary include Feldman et al. (2010), who analyzed the management discussion and analysis section in

10-Q and 10-K to analyze the sentiments in the text, finding that short window market reactions around SEC filings, and Garcia (2013). Loughran and McDonald (2016) and Kearney and Liu (2014) provided detailed literature reviews on textual analysis. Such studies include Renault (2017), who used field-specific lexicons to analyze messages on social media for sentiment analysis, correlating with intraday market returns of the S&P 500 index. For literature review on the application of AI in the subarea of natural language processing and in particular textual analysis, please see Mishev et al. (2020). Such studies applying the deep learning techniques in natural language processing include Peng and Jiang (2015) used a deep neural network (DNN) to analyze financial news to predict stock price changes, and Zhuge et al. (2017) used long short-term memory neural network to predict market returns. Huang et al. (2017) used a topic modeling approach to extract information from analyst reports, different from the conference calls, using Latent Dirichlet Allocation. A similar study is Dyer et al. (2017), who used the Latent Dirichlet Allocation to analyze 10-K. In the subarea of using techniques to extract information from oil news, Abdollahi (2023) uses BERT to extract information from news and Twitter sentiments in predicting oil volatility. However, none of the literature uses a GPT model to extract information on oil shock news and make predictions for oil and stock returns, and explore the difference between the dictionary-based model and the large language model.

Hypotheses development

Hypothesis 1.a: Media coverage effectively communicates messages regarding shifts in commodity fundamentals.

In the sentiment analysis of news reporting and its impact on stock returns, the existing literature shows a range of findings. Some studies arrived at positive conclusions, whereas others did not. For example, Baumeister, Bratslavsky, and Finkenauer (2001) discover that bad news tends to exert a more pronounced and quicker influence compared to good news. Baker and Wurgler (2006) and with Baker, Wurgler, and Yuan (2012) establish a connection between sentiment indices and subsequent returns.

Hypothesis 1.b: LLMs can extract sentiment information from news reports

Prior studies use LLMs to assess natural human languages in various domains, such as healthcare (Cascella et al., 2023). Llama, in particular, stands out as one of the most advanced LLMs for evaluating human natural language, as indicated by the "L" in LLM, signifying a large, fine-tuned model with approximately billions of parameters dedicated to understanding natural language information.

FinGPT is a specialized model fine-tuned using financial news sources including Yahoo Finance, Bloomberg, and The Wall Street Journal, along with financial announcements and social media feeds. Compared with BloombergGPT, FinGPT is further enhanced with Reinforcement Learning from Human Feedback (RLHF), which incorporates human risk aversion and preferences into predicting the tone of human natural language text. It provides better supervision than other LLMs such as OpenAI's ChatGPT, FinBERT, or BloombergGPT. While LLMs have applications in various domains, it remains uncertain whether FinGPT can effectively extract information from "oil shock" news and accurately assess the tone of such news.

Hypothesis 2.a: The readability of news coverage influences the effectiveness of its conveyed sentiment.

Li's (2006) study explored the impact of "uncertainty" words in 10-K reports, revealing their predictive power for stock returns. Subsequent research indicates that an individual's financial literacy influences their response to similar financial news reports. For example, Van Rooij et al. (2011) discover that financial literacy affects participation in financial markets, whereas Guiso and Viviano (2015) demonstrate its effect on financial market performance. Despite these insights, it remains unclear how the readability of financial texts influences the effectiveness of sentiment. This study addresses whether "readability," as measured by the Gunning Fog Index, affects the impact of sentiment in financial news.

Hypothesis 2.b Visibility moderates the impact of news tone on commodity prices, either amplifying or mitigating these effects.

Nofsinger (2001) finds that the "visibility" of stocks significantly influences the trading behavior of individual investors, yet this factor does not exert a comparable influence on institutional investors. Expanding on these insights, Baber and Odean (2008) reveal that individual investors tend to exhibit a marked preference for trading stocks that capture their attention more effectively than others. Fang and Peress (2009) contend that enhanced news coverage amplifies investor awareness of covered stocks, which, in turn, diminishes information asymmetry and subsequently leads to lower returns. Additionally, Peress and Schmidt (2019) document a noticeable reduction in noise traders' activities during "distraction days, " when non-economic events dominate the news. These studies collectively suggest that the "visibility" or "attention" among investors may be a crucial determinant in assessing whether sentiment in news impacts oil prices.

3. Empirical Method

3.1. Oil returns and stock abnormal returns

The oil returns were computed from West Texas Intermediate (WTI) nearest-term futures prices, and these futures prices were collected from the Chicago Mercantile Exchange (CME). WTI is the most traded crude oil commodity futures contract in the world. It is also the benchmark for oil prices in the U.S., known for its significant trading volume. We obtain the daily settlement prices of WTI futures traded on the CME and track the nearest-term futures prices as a proxy for oil prices. The daily oil return is calculated as

$$Oil\ Return_t = \frac{P_t}{P_{t-1}} - 1$$

where P_t represents the price on trading day *t*.

We collected the abnormal returns (alphas) from the WRDS Beta Suite using the Fama-French four-factor model and utilized days as the window to estimate the alphas. We then matched the abnormal returns with the ticker and company names, and used the company names to be matched with the news articles mentioning the names of the companies and the sentiments linked to the news article dataset. Thus, we can link the individual firms' abnormal returns to the sentiment of the news covering "oil shock" mentioning the companies in the news for that day. We calculate the cumulative abnormal returns as follows:

$$CAR_{i,t} = \sum_{t=1}^{T} AR_{i,t}$$

Where $CAR_{i,t}$ is the cumulative abnormal return for firm *i* on day *t* and $AR_{i,t}$ is the abnormal return for firm *i* on day *t*. We calculate the three-day and five-day abnormal returns for each firm by summing the daily returns, multiplied by 100, over the respective period.

3.2. News

The sample includes 24,000 news articles containing the terms "oil" and "shock" gathered from the Wall Street Journal, The Financial Times, and The New York Times, a specialized magazine, Energy Weekly News, and an institutional newswire, Dow Jones Institutional News. Our data were sourced from

Factiva and ProQuest and cover the period from July 8, 1996, to November 6, 2019 by manually downloading the full text of the news. [4]

Figure 1 shows the statistics of the articles in our sample. With 68.51% from Dow Jones Institutional and 9.53% from the Wall Street Journal Eastern edition (printed version), 3.87% from Wall Street Journal Online. We have a collection of news from the Financial Times amounting to 1,328 pieces of news and from the New York Times, 843 pieces of news. The magazine Energy Weekly News has contributed 1,305 news articles. The figure illustrates the number of articles from each of the sources over the years 1986 to 2019. We have 92 pieces of news on "oil shock" in 1986, and the volume of news gradually increased up to 2008, when Dow Jones Institutional News joined our sample, and the total number of news in 2008 amounted to 2,678. In 2019, we have slightly more news, with a total of about 1,500.

(Insert Figure 1 here)

3.3. Sentiment

The sentiments "Negative," "Positive," and "Neutral" are extracted from news texts using FinGPT, a machine learning LLM specifically fine-tuned with diverse financial news, announcements, and social media feeds. "Negativity" is the differential between the "Negative" and "Positive" scores in news articles, analogous to the "Disagreement" metric utilized in prior research.[6]

(Insert Figure 2 here)

FinGPT is one of the most advanced Large Language Models (LLMs) in the field, with a parameter size of approximately 2GB. (Additional details about the technology in FinGPT would follow please see appendix A). See the Figure 2 for a comparison between FinGPT and other LLMs, as well as sentiment analysis tools. We input the text of news articles, and the model subsequently generates estimates for three sentiment measures: "Negative," "Positive," and "Neutral." The "Negative" output provides a numerical estimation of the likelihood that a news text is classified as "Negative," while the "Positive" output similarly

---

[4] The Dow Jones Institutional News service, which we use here, is recognized as one of Dow Jones' premier information services, catering specifically to institutional professionals. In the case of The Wall Street Journal, we meticulously selected various editions to ensure a comprehensive coverage. This included the New York-based WSJ Eastern Edition (since 1986), the WSJ Online Edition (since 2000), the Brussels-based WSJ Europe Edition (since 1983), and the Hong Kong-based WSJ Asia Edition (since 1976), ensuring a diverse, global perspective on the financial impact of oil shocks.

[6] For details on the training of FinGPT, please see Appendix A.

assesses the likelihood of a news report being "Positive." The "Neutral" output denotes a neutral tone in the news text. We standardized these numerical outputs to ensure that their sum was equal to one.

For example, Figure 3 illustrates a news article from Dow Jones Institutional News. The probability of being classified as positive news is 0.3560806, the probability of being classified as negative news is 0.1721893, and the probability of being classified as neutral news is 0.4717301.

(Insert Figure 3 here)

We calculate *Sentiment_GPT* as the negative probability minus the positive probability. For the dictionary-based sentiments, we calculate the sentiments as the difference between the negative percentage of words minus the positive percentage of words. Then, we aggregate the news-level sentiment to daily sentiment by summing up the individual news' sentiments. We scored the days with no news reports on "oil shock" as having a "zero" sentiment score.

$Sentiment_t = \sum_{n=1}^{N} Sentiment_{n,t}$, where n = 1, 2, … N for the n-th pieces of news on day t.

In addition, media coverage may offer valuable insights by presenting business journalists' analyses, recommendations, and detailed documentation of market events that are not evident in quantitative data. Supporting the information hypothesis, which posits that news coverage enhances information processing in capital markets rather than merely influencing investor sentiment, we analyze the use of numerical figures in articles (numbers). This approach, in line with the findings of Bai, Dong, and Hu (2019) tests how numerical data, such as tables, improve the readability of financial documents. In the analysis of text complexity, we used Gunning Fog Index, which calculates as $0.4 \times \left[\left(\frac{words}{sentences}\right) + 100 \times \left(\frac{complex\ words}{words}\right)\right]$[7]. Each of these scores was normalized based on the word count of the text. Additionally, to assess the certainty of the text's tones, the Sureness Dictionary (Harvard-Lasswell) was utilized.

3.4. Comprehensive moderators

We use three sentiment indices to gauge investor sentiment: the News-Implied Volatility Index (NVIX), as proposed by Manela and Moreira (2017); the Financial and Economic Attitudes Revealed by Search (FEARS) index developed by Da, Engelberg, and Gao (2015); and the oil-specific sentiment index

[7] Details on the computation of the index please see https://en.wikipedia.org/wiki/Gunning_fog_index.

(Oil Specific) introduced by Qadan and Nama (2018).[8] NVIX, a measure of market uncertainty derived from WSJ coverage, has been demonstrated by Manela and Moreira (2017) to be correlated with shifts in market returns and volatility and to be predictive of economic downturns.[9] The FEARS employs Google Search Trends for sentimental term queries to forecast returns and volatility[10]. The Oil-Specific Index, which includes search trends for terms like "oil price," "crude oil," and "price of oil," provides insights into investor sentiment in the oil market.

3.5. Other variables

We collected the daily trading volume for the WTI daily trade in units from the Chicago Mercantile Exchange (CME). We obtained the option prices of WTI from CME and zero rates from WRDS OptionMetrics. Individual firms' Fama-French four-factor model alphas were collected from WRDS Beta Suite.

We use the following empirical model to examine the relationship between the oil returns and negativity in "oil shock" news.

$$R_t = \beta_0 + \beta_1 Sentiment_t + \Delta_t + \varepsilon_t \tag{1}$$

And we use a more roubst model by including the lags of returns and negativity:

$$R_t = \beta_0 + \beta_1 Sentiment_t + \Delta_t + \sum_{i=1}^{4} \theta_i R_{t-i} + \sum_{i=1}^{3} \gamma_i Sentiment_{t-i} + \varepsilon_t \tag{2}$$

Sentiment is calculated as the sum of the *Sentiment* scores in single news reports on that day. Δ_t are weekday dummies (Monday is the baseline category); weekday dummies control for the effects of U.S. Energy Information Administration (EIA) announcements on crude oil inventories on Wednesdays at 10:30 a.m. Linn, Ederington, Fernando, and Guernsey (2017) find that EIA releases important fundamental

---

[8] Data source: https://www.sciencedirect.com/science/article/pii/S0140988317303766

[9] Data source: Asaf Manela's personal website at: https://asafmanela.github.io/data/

[10] Data source: Zhi Da's personal website at: https://www3.nd.edu/~zda/

changes that lead to significant subsequent oil price movements. In addition, we account for the lags of *Sentiment* up to four days (five weekdays per week for news reports) and three lags of oil returns to address the serial correlation effect in both *Sentiment* and oil returns.

4. Empirical Results

4.1. Summaries of sentiments

In the Table 1, we summarize the statistics of the Sentiment_GPT, along with the correlation matrix of the sentiment across different newspapers (newsfeeds). In Panel A, we summarize the three components of the sentiments extracted from FinGPT. "Positive" is the probability of classifying the news as "positive" news by FinGPT, "negative" refers to the probability of classifying news as negative, and "neutral" denotes neutral news. Later, we summarize the average of the Sentiment_GPT across different news sources and find that most sources give a positive sentiment for the "oil shock" news. In Panel B, the negative and positive components in the Sentiment_GPT are positively and highly correlated, as is the neutral component.

(Insert Table 1 here)

(Insert Figure 4 here)

The Figure 4 displays the FinGPT sentiment spanning from 1986 to 2019. To better observe the trends of change in sentiment, we compute the 91-day (three-month) moving average (MA91) of the daily sentiment. The moving averages clearly indicate that Sentiment_LMP (MA91) and Sentiment_GPT (MA91) more closely track each other and show a reverse trend with Sentiment_HL (MA91). However, when we graph the moving average of the negativities of the three sentiments (Negativity_HL (MA91), Negativity_LMP (MA91), and Negativity_GPT (MA91)) in another graph, the three lines for negativities actually closely relate and show comovement. This indicates that the negative tones are measured quite persistently across the three methods of language models. The histogram on the right side (figure) shows a similar distribution of the negative tones across the three sentiments. When we depict the positivity in the three sentiments (the percentage of positive words in an article for the dictionary-based sentiments and the probability of being classified as positive news by GPT), the distribution of positivity obviously differs a lot across the three sentiments.

(Insert Figure 5 here)

In Figure 5, we show the WTI crude oil prices and the oil returns from 1986 to 2019 on a daily basis. We observe some left-skewness in oil returns around 1991, less volatility around the 1990s, and strong volatility in 2008. The oil price has an increasing trend since 1999, but there is a sharp decline around 2008 and another sharp drop around 2015. These periods clearly show some shocks in the market. We see corresponding sharp changes in the sentiments in Figure 4, mirroring the correlation between the sentiments and the oil price changes.

(Insert Figure 6 here)

We report the preliminary result in Table A. Coverage = 1 for any news report on that day; we find that *Coverage* does not have a significant impact on oil returns. When we control for the weekday of reports but not lags in sentiment and oil returns, we obtain similar results, indicating that *Sentiment_GPT* does not affect oil returns. However, when we account for the lags, *Sentiment_GPT* has a significant impact by reducing oil prices' returns on that day. We segmented our news sample into various publications, including news collected from the WSJ (Eastern version), WSJ Online, Financial Times, New York Times, Dow Jones Institutional News, and Energy Weekly News. We conducted regression analysis of oil returns on each publication's news tone in relation to "oil shocks," where we calculate the daily tone as the sum of sentiment in single news articles for that publication. We find that WSJ Online has a significant impact on oil returns, reducing daily oil returns by 0.52% for every unit increase in negativity. Additionally, we observe that the Dow Jones Institutional News has a significant impact at the 10% level.

4.2. The impact of sentiments on oil returns

We further explore whether text readability affects the likelihood of news tone regarding oil shocks influencing daily oil returns. Previous studies indicate that readability impacts individuals' comprehension and processing of information (Rennekamp, 2012; Song and Schwarz, 2008). If investors are able to digest news content more swiftly and effectively gather crucial information, their responses to news content might be more pronounced. Hence, we examine the influence of readability measures on oil returns. The findings suggest that the interactions between readability measures and negativity are highly significant, underscoring the impact of either sentiment or readability. Moreover, we have introduced two new variables: *Sureness* (percentages of definite words in the Harvard-Lasswell dictionaries, such as "always" and "certain") and *Number* (percentages of numerical figures, e.g., 0-9). Our results indicate that both *Sureness* and *Number* reinforce the impact of *Sentiment_GPT*.

(Insert Table 2 here)

Investors' Attention

Furthermore, we assess whether investors' awareness affects their responses to news coverage containing relevant information. Merton's (1987) visibility theory suggests that awareness moderates the impact of information on responses. To measure the level of awareness or recognition of investors in the financial or oil markets, we employ three sentiment indices. First, we control for the awareness measures and the interactions of awareness with *Sentiment_GPT* in oil shock news. None of the variables (interactions) are statistically significant, except for *Oil Specific* (which shows a -0.016% reduction in oil returns when *Oil Specific* increases by one unit).

We further divide our sample into two subperiods—high awareness periods and low awareness periods. We find that when Oil Specific is above its time series median, *Sentiment_GPT* has a significant impact on oil returns, reducing daily returns by approximately and 0.15%. We do not find such an effect when *FEARS35* is lower or higher than its average over time.

Oil Specific is a measure specifically related to the attention to the oil market's responses from investors, and *NVIX* is another news-implied index but measures volatility in the financial market from textual analysis. Meanwhile, *FEARS35* is an index measuring financial and economic search trends. The results suggest that investors' attention to the oil market (and the uncertainty in news reports) causes this impact on the news' sentiments on the responses in the oil prices. However, during periods when investors' attention is low (on the oil market or financial market volatility), the news sentiment does not show an impact on oil returns. These findings suggest that the visibility of the specific market is very important in determining whether the sentiment of oil coverage could be associated with changes in the security prices.

(Insert Table 3 here)

Region

Our collection of Wall Street Journal news comprises several different issues: we have the online edition that was available after year 2009, and the traditional printed edition, which includes three different versions corresponding to three geographical locations: U.S. Eastern, Europe, and Asia. The latter three issues of the Wall Street Journal were available after 1986, 1991, and 2005 respectively, in our sample. First, we divide our news sourced from the traditional WSJ (in contrast to the WSJ online edition) into three categories: WSJ Asia, WSJ Europe, and WSJ U.S. Eastern. Notably, the WSJ releases different versions of its subscription services in various regions. We were interested in whether coverage from different regions has varying impacts on oil returns, considering that WTI oil is a globally traded commodity. Note that, the traditional WSJ version does not have a significant impact on oil returns as shown in previous results. Similarly, when we split it into different locations, we find that none of the printed versions, though in different locations, show significant impact on the WTI oil returns. This is consistent with our previous finding that only the WSJ online version seems to be related to the concurrent oil returns of the news day.

(Insert Table 4 here)

Liquidity

We collect the total daily trading volumes of WTI crude oil nearest-term futures from the CME End-of-Day Dataset. We calculate the option-implied risk-neutral volatility on oil returns over 30 days, following the methods outlined in Bakshi, Kapadia, and Madan (2003). We chose 30 days for option-implied volatility since this is the most liquid term of option contract, and it looks forward into the volatility in the oil returns in the future 30 days. We also obtain the option and futures prices for oil from the CME. To compute the option-implied moments, we gather data on risk-free rates from OptionMetrics. We estimate the daily illiquidity as

$$Illiq_t = \frac{|R_t|}{Price_t \, Trading\ Volume_t},$$

which is similar to Amihud's (2002) stock liquidity measure. We winsorize the volume, liquidity measures, and implied volatility at 5% to drop the outliers in the sample. We find no significant impact of oil shock sentiment on liquidity measures. The daily sentiment extracted from FinGPT does not seem to be associated with the daily trading volume of WTI crude oil, Amihud's liquidity measure, or the 30-day implied volatility of WTI prices. The sentiment in the news on "oil shock" does not lead to changes in the liquidity in the U.S. oil market.

Qualitive Analysis

We further categorize the news sample into three categories: good, bad, and no news. As previously discussed, *the tone of the news may not fully align with the implications of the information contained within the news itself.* The sentiment may be more related to the writing style of the author of the news or how the event was covered. Additionally, the sentiment in the news may not indicate the implications for oil price changes but to the consumers. Therefore, we qualitatively classify the news into three categories, based on whether the news contains generally more positive or more negative tones. Good news is defined as 1 for negativity < 0, bad news as 1 for negativity > 0, and no news when negativity is missing for the day. The dummies for tone do not significantly impact daily oil returns. This indicates that the quantitative measure for the sentiment of the news is important to gauge the impact of the news report on the oil market price responses. We cannot simply rely on the general categorization of good or bad from the news for oil return prediction; instead, we must comprehend the detailed information contained in the "tones" of the news coverage. Furthermore, FinGPT can extract these "tones" from the news.

5. Impacts of Sentiments on Stock Returns

To explore the impact of "oil shock" news and the sentiments, we further collected stock prices of companies mentioned in the "oil shock" news. The Dow Jones Institutional News, sourced from ProQuest, comes with the names of companies mentioned in the news. We have a number of pieces of news on "oil shock" from DJI to extract information about the companies mentioned. Figure 3 shows how we extracted the company names from each piece of news. Note that in the row "*Company / Organization: Name:*", it reports the names of the companies mentioned in the news. In this figure, it shows that Costco Wholesale Corp and Walmart Inc were mentioned with NAICS 452910 and 452112, 452910, 454111. This is particularly related to the sentence in the report: "*However, in the U.S., big-box discount retailers and warehouse clubs including Wal-Mart, Target, Costco and BJ's Wholesale could benefit as consumers economize...*". Since the database does not provide the company's ticker/permno, we had to manually clean the company names and then merge them with our sentiment dataset. We identified about 2,000 news pieces with company names with stock abnormal returns. We collected the Fama-French four-factor abnormal returns (alphas) for these companies identified from the DJI sample from WRDS Beta Suite. The companies are merged between the CRSP stock prices and DJI news sentiments by company names.

We denote the days with news that mentions the firms' names on DJI as "news days" and the days for the firms that have no news mentions as "no news days." The sample of abnormal returns contains only the firms that have been mentioned in the DJI sample. We then calculate the cumulative abnormal returns over three days (CAR3) and over five days (CAR5), along with the alphas (AR) across different relative days (Day=0 for news day). Table 5 shows the average AR from Day -2 to Day +2 and the CAR3 and CAR5 on news days and no news days for those firms. It suggests that AR is slightly higher on news days than on no news days, mirroring the fact that the average *Sentiment_GPT* in the subsample is slightly lower than 0 (-0.117), suggesting that on average, news coverage on the firms is generally positive. We find that on average, the abnormal returns within two days prior to the news day (AR(-2) and AR(-1)) are slightly lower, and the news days have the lowest average (AR (news days) = 0.038). The two days after the news, average abnormal returns are slightly higher; both AR(+1) and AR(+2) are higher than the AR prior to and on the news days. The average three-day and five-day cumulative abnormal returns for news days (CAR3 and CAR5 for news days) are slightly higher than the cumulative abnormal returns for no news days, suggesting a positive price impact on the firms mentioned in the news coverage on "oil shock."

(Insert Table 5 here)

When we perform a Wald test on the null hypothesis that the abnormal return on news days (AR (news days)) is equal to the abnormal return on no news days (AR (no news days)), we cannot reject the null hypothesis given a p-value of 0.4353. We perform a similar Wald test on the null hypothesis that the cumulative abnormal returns are equal between news days and no news days (CAR3 and CAR5), but we draw a similar conclusion that the news days' cumulative abnormal returns are not significantly different from the no news days' returns. This finding suggests that news coverage on "oil shock" does not lead to a significant change in the related individual firms' abnormal returns, though there is an insignificant slight increase in returns.

Next, we include the sentiments and graph the AR (CARs) on news days based on sentiments. In Figure 7, we find that prior to the news days, the average abnormal returns of the firms in the subsample are already higher than the average abnormal returns on no news days. It shows that prior to the release of the news, there is a higher price return than average, which declines a bit on the news days before increasing after news days. On the contrary, the days around the no news days have no significant changes in abnormal returns, and the lines are very flat as depicted in the figure. This suggests that there is some slight change in the firm's value around the news days, but these changes are never statistically significant. In Panel C of Table 5, We run regressions of AR (CARs) on the sentiment with a simplified model and day of the week in the figure. It suggests that in none of the regressions is *Sentiment_GPT* a significant cause of price changes in the stock returns (in the regression for CAR5, the p-value for Sentiment_GPT is slightly significant at 0.098), and the dummy that equals one for news days is never statistically significant. This suggests that the news coverage for "oil shock" on the individual firms associated hardly leads to changes in firm values.

(Insert Figure 7 here)

6. Comparison Between FinGPT and Dictionaries

6.1. Statistical summaries and correlation matrix of sentiments

In Table 6, we summarize the statistics for the three sentiments that were extracted using the tools of the Harvard-Lasswell dictionary, the McDonald and Loughran financial dictionary, and FinGPT, respectively. We labeled the sentiment generated from the Harvard-Lasswell dictionary as *Sentiment_HL*, the sentiment generated from the McDonald and Loughran financial dictionary as *Sentiment_LMP*, and the sentiment created by FinGPT as *Sentiment_GPT*. Panel A's correlation matrix shows that *Sentiment_HL* is negatively correlated with *Sentiment_LMP* and *Sentiment_GPT,* which was a surprising finding for us, indicating that the Harvard-Lasswell dictionary, which extracts from general human language text, is negatively tracking the other sentiments. This suggests that using different dictionaries may yield different results for the impact of news tones on oil prices. *Sentiment_LMP* is positively correlated with *Sentiment_GPT*, suggesting that the financial dictionary created by McDonald and Loughran can track the tones of financial news regarding "oil shock" just as well as the FinGPT model. In Panel A, we summarized the statistics of the three sentiments, showing that on average, *Sentiment_HL* returns with a positive tone (a negative number indicates a positive tone), while the other two sentiments, on average, show negative tones (positive values). This mirrors the negative correlation between *Sentiment_HL* and the others. Interestingly, we also find that *Sentiment_HL* is far more right-skewed, and *Sentiment_LMP* and *Sentiment_GPT* are more left-skewed, revealing that tones of "oil shock" news have more extreme bad news than good news. We have 3,328 observations for daily sentiments spanning from 1986 to 2019 (missing values for days with no "oil shock" news). Panel B delves into the "negative" side of the sentiments – extracting only the series of the "negativity" in the tones of the sentiments, which is defined by the negative words for the dictionary-based sentiments and by the probability of "negative" for the GPT sentiment. All three sentiments are now positively correlated; *Sentiment_HL* is positively related to both *Sentiment_LMP* and *Sentiment_GPT*, contrary to the correlation for the sentiments shown in Panel A. Panel B shows that the three negativity series are similar in their mean, standard deviation, min, max, and skewness. Panels C give similar results for the "positivity" side of the sentiments. Later, we do a robust summary of the statistics for the three sentiments with the sentiments for only the news of "oil shock" from Dow Jones Institutional News (DJI in Panel D) and from the Wall Street Journal Online version (WSJ Online in Panel E). The sentiments for the DJI news show similar statistics as with the results in Panels A-C, suggesting that *Sentiment_HL* is negatively correlated to the other two sentiments and tends to be more positive than the latter. However, Panel E suggests that the WSJ Online news sentiments are showing *Sentiment_LMP* negatively related to the other two measures, but we still find that *Sentiment_HL* is more positive than the other two sentiments.

(Insert Table 6 here)

We then run the main regressions similar to those in the Table A, assessing the impact of daily oil return on sentiment and weekday dummies, controlling for the day of the week. In a more complete model, we add lags of the sentiment for up to four-order lags, and lags of the oil returns for three lags. We report the results in Table 7. The robust model shows consistent results that sentiment is significant. However, the model in Panel B for *Sentiment_LMP* shows no impact from the sentiment of oil shock news. In Panel B, the sentiment derived from the Harvard-Lasswell dictionary (negativity minus positivity, each defined by the percentage of tone words) significantly affects oil returns, showing a significant impact of 0.021 and a p-value smaller than 0.01. The robust model presents consistent results that sentiment is significant, with an impact of -0.025. However, the LMP model in Panel C indicates that the sentiment from the oil shock news does not impact oil returns.

(Insert Table 7 here)

In exploring the relationship between sentiments and readability, numerical content, and sureness in the tones of the news in Figure 8, we find varying relationships. *Readability* is measured by the Fog Index, indicating that higher Fog Index values, indicating more complex texts, do not have a consistent effect on sentiment scores across the different models. *Sentiment_GPT* tends to give a positive sentiment when text is more difficult to understand, contains more numbers, or has a firmer tone. We compare the distribution of sentiments for cases when variables are above or below their medians, finding different distributions for *Sentiment_LMP* when these variables are above their medians, but similar patterns for all sentiments when they are below medians.

(Insert Figure 8 here)

6.2. Sentiments and readability, numerical, and sureness

The negative relationship between *Sentiment_HL* and *Sentiment_LMP* and *Sentiment_GPT* remains unresolved. While the negativity of the three sentiments closely comoves, the positivity (percentage of positive tones or the probability of being classified as positive news by GPT) shows opposite trends. To further investigate this, we examine the relationship between the readability of the news, the percentage of numerical words in the news articles, and the sureness in the tones of the news. These variables are defined as follows: readability is measured by the Fog Index, with a higher index indicating more complex text; sureness is quantified by the percentage of words indicating firmness or certainty in tone, aggregated to a daily level; and numerical content is calculated as the percentage of numerical characters in the text, also aggregated daily.

We then run the main regressions, similar to those outlined in the Table 2, analyzing the impact of daily oil return on sentiment, alongside weekday dummies to control for the day of the week in Panel A of Table 7. In a more comprehensive model, we incorporate lags of the sentiment up to four-order lags, and lags of the oil returns for three lags. Our analysis in Table 7 reveals that *Sentiment_LMP* does not correlate clearly with "Numerical," "Readability," and "Sureness," showing no evident trend along these variables. We observe no significant interaction effects between dictionary-based sentiments and these text characteristics.

We further decompose the news sources into Dow Jones Institutional News (DJI) and Wall Street Journal Online (WSJ Online). In Figure 9, daily sentiments, their 91-day moving averages, and sentiment distributions are analyzed. WSJ Online articles are sparse before 2010, limiting our ability to graph their 91-day moving average for earlier years. For DJI, sentiments follow the general trend of *Sentiment_HL* moving inversely to *Sentiment_LMP* and *Sentiment_GPT*, with histograms reinforcing the distinction between *Sentiment_HL* and the other two sentiments. This pattern is less pronounced in WSJ Online, suggesting a closer relationship among sentiments for this source.

(Insert Figure 9 here)

6.3. Predictions from sentiments

We use the three models to predict the oil returns and report these results in a table. First, we split our sample into two halves: the first half of the sample is used to fit the model with the specification in the equation (2). The sentiment in each regression is LMP, HL, or GPT. Then we use the second half of the sample to make predictions and summarize the mean prediction error (residuals). We find that GPT produces similar average prediction errors with LMP, mirroring the finding that GPT tracks closely with LMP but is less related to HL. HL produces the lowest prediction error.

(Insert Table 8 here)

Then in the column "Readability," we split our sample further into four halves. In the first half of the sample, we further split the observations into when readability (as measured by the Fog index) is greater than the median and lower than its median. We use the subsample with greater than median readability to fit the model. Then we use the second half's observations with greater than median readability to make predictions. We find that in this subsample with greater readability than the sample median, GPT produces the best prediction for oil returns as it has the lowest prediction error on average (-1.246). In the column of "Sureness," we split the sample similarly but based on the median of sureness (the percentage of words measuring sureness in tones) and find that GPT still procures the lowest errors (4.983). But when numerical content is greater, we find that LMP procures the best prediction results. In the last column of "Oil Sentiment," we split the sample based on the measure for the investor attention in the oil market. As in previous results, we find that only when investors' attention in the oil market is greater does the sentiment have a significant impact on oil returns. We find that in this subsample with greater than median Oil_Sentiment, GPT procures much lower average prediction errors (0.005). We conclude that GPT's power in generating greater prediction for oil returns on measuring the tones in the "oil shock" news depends on the specific content within the news; when news is more difficult to comprehend and when investors' attention is better, it procures better results.

7. Robustness

7.1. Simple average of daily sentiments

We employ robust regression for the main model by computing the sentiments alternatively as the simple average of the article sentiments. It is calculated as follows:

$$Sentiment_t = \frac{1}{N}\sum_{n=1}^{N} Sentiment_{n,t}, \text{ where n = 1, 2, … N for the n-th pieces of news on day t.}$$

We graph the sentiments (mean) in Figure A and report the regression results in Table D. We find that the *Sentiment_GPT* is not significant for the oil returns, and *Sentiment_HL* is significant for oil returns; *Sentiment_LMP* is not significantly related to oil returns either. The main conclusions are similar to these.

8. Conclusions

This paper utilizes FinGPT to assess sentiment in oil-related news articles spanning from the 1980s to 2019. Sentiment refers to the contrast between negativity and positivity in tone. The sentiment analysis conducted through NLP techniques reveals a significant impact of news sentiment on oil returns, leading to a reduction in daily oil price returns. Furthermore, when we segment our news dataset by their publication sources, we observe significant effects on oil returns. Specifically, WSJ Online and Dow Jones Institutional News both contribute to the decrease in daily returns.

By examining the relationship between oil returns and comprehension measures, such as readability, numerical content, and tone certainty, we find intriguing findings. These comprehension measures act as moderating factors, influencing the impact of sentiments. When news articles are easier to comprehend, negative tones tend to exert a more pronounced effect. On the other hand, our analysis reveals no significant impacts on oil returns from WSJ regional editions (Asia, Europe, Eastern), and sentiment does not seem to affect oil volatility, volume, or illiquidity. Moreover, when FinGPT is employed to estimate news sentiment, investor awareness measures demonstrate significant effects.

In summary, this study demonstrates the substantial influence of sentiment in oil news articles on daily oil price returns, particularly when considering the source of the news. It also highlights the moderating role of interpretability measures, with easier-to-read news intensifying the impact of negative sentiments.

References

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2018). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*, 64(6), 2833–2855.

Abdollahi, H. (2023). Oil price volatility and new evidence from news and Twitter. *Energy Economics*, 122, 106711.

Adams, Z., & Glück, T. (2015). Financialization in commodity markets: A passing trend or the new normal?. *Journal of Banking & Finance*, 60, 93-111.

Agmon, T., & Lessard, D. R. (1977). Investor recognition of corporate international diversification. *The Journal of Finance*, 32(4), 1049-1055.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*, 18(3), 505-529.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), 31-56.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59(3), 1259-1294.

Asquith, P., Pathak, P. A., & Ritter, J. R. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics*, 78(2), 243-276.

Aziz, S., Dowling, M., Hammami, H., & Piepenbrink, A. (2022). Machine learning in finance: A topic modeling approach. *European Financial Management*, 28(3), 744-770.

Baber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional in. *Review of Financial Studies*, 21(2), 785-818.

Bai, X., Dong, Y., & Hu, N. (2019). Financial report readability and stock return synchronicity. *Applied Economics*, 51(4), 346-363.

Bailey, W., Chung, Y. P., & Kang, J. K. (1999). Foreign ownership restrictions and equity price premiums: what drives the demand for cross-border investments?. *Journal of Financial and Quantitative Analysis*, 489-511.

Baker, H. K., Nofsinger, J. R., & Weaver, D. G. (2002). International cross-listing and visibility. *Journal of Financial and Quantitative Analysis*, 37(3), 495-521.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645-1680.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129-152.

Baker, M., Wurgler, J., & Yuan, Y. (2012). Global, local, and contagious investor sentiment. *Journal of Financial Economics*, 104(2), 272-287.

Bakshi, G., Kapadia, N., & Madan, D. (2003). Stock return characteristics, skew laws, and the differential pricing of individual equity options. *The Review of Financial Studies*, 16(1), 101-143.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307-343.

Basu, P., & Gavin, W. T. (2010). What explains the growth in commodity derivatives*?. Federal Bank of St. Louis Review*., 93(1), 37-48.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323-370.

Bekaert, G., Harvey, C. R., & Lumsdaine, R. L. (2002). Dating the integration of world equity markets. *Journal of Financial Economics*, 65(2), 203-247.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1), 1-27.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.

Burt, A. (2018). Information Networks and Market Segmentation. Available at *SSRN* 3229324.

Cascella, M., Montomoli, J., Bellini, V., & Bignami, E. (2023). Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *Journal of Medical Systems*, 47(1), 33.

Chang, B. Y., Christoffersen, P., Jacobs, K., & Vainberg, G. (2012). Option-implied measures of equity risk. *Review of Finance*, 16(2), 385-428.

Cheng, D., & Cirillo, P. (2018). A reinforced urn process modeling of recovery rates and recovery times. *Journal of Banking & Finance*, 96, 1-17.

Coval, J. D., & Shumway, T. (2001). Is sound just noise?. *The Journal of Finance*, 56(5), 1887-1910.

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all FEARS investor sentiment and asset prices. *The Review of Financial Studies*, 28(1), 1-32.

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375-1388.

Dogah, K. E., & Premaratne, G. (2018). Sectoral exposure of financial markets to oil risk factors in BRICS countries. *Energy Economics*, 76, 228-256.

Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, 64(2-3), 221-245.

Ederington, L. H., Lin, F., Linn, S. C., & Yang, L. Z. (2019). EIA storage announcements, analyst storage forecasts, and energy prices. *The Energy Journal*, 40(5).

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, 66(1), 67-97.

Fang, L., & Peress, J. (2009). Media coverage and the cross‐section of stock returns. *The Journal of Finance*, 64(5), 2023-2052.

Fehle, F., Tsyplakov, S., & Zdorovtsov, V. (2005). Can companies influence investor behaviour through advertising? Super bowl commercials and stock returns. *European Financial Management*, 11(5), 625-647.

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2010). Management’s tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, 15, 915-953.

Ferreira, M. A., & Matos, P. (2008). The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics*, 88(3), 499-533.

Fink, J. D., Fink, K. E., & Russell, A. (2010). When and how do tropical storms affect markets? The case of refined petroleum. *Energy Economics*, 32(6), 1283-1290.

Fink, J., & Fink, K. (2014). Do seasonal tropical storm forecasts affect crack spread prices?. *Journal of Futures Markets*, 34(5), 420-433.

Foerster, S. R., & Karolyi, G. A. (1999). The effects of market segmentation and investor recognition on asset prices: Evidence from foreign stocks listing in the United States. *The Journal of Finance*, 54(3), 981-1013.

Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2006). Institutional investors and stock market volatility. *The Quarterly Journal of Economics*, 121(2), 461-504.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3), 1267-1300.

Garcia, D., Hu, X., & Rohrer, M. (2023). The colour of finance words. *Journal of Financial Economics*, 147(3), 525-549.

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review. *Energy Economics*, 81, 709-727.

Guiso, L., & Viviano, E. (2015). How much can financial literacy help?. *Review of Finance*, 19(4), 1347-1382.

Halova, M. W., Kurov, A., & Kucher, O. (2014). Noisy inventory announcements and energy prices. *Journal of Futures Markets*, 34(10), 911-933.

He, K., Tso, G. K., Zou, Y., & Liu, J. (2018). Crude oil risk forecasting: New evidence from multiscale analysis approach. *Energy Economics*, 76, 574-583.

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry*, 33(1), 3-12.

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171-185.

Klibanoff, P., Lamont, O., & Wizman, T. A. (1998). Investor reaction to salient news in closed‐end country funds. *The Journal of Finance*, 53(2), 673-699.

Linn, S. C., Ederington, L. H., Fernando, C. S., & Guernsey, S. B. (2017). (Information, Price Levels and Price Volatility of Petroleum Product Futures Price). Retrieved September 19, 2017.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, 69(4), 1643-1671.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.

Loughran, T., & McDonald, B. (2020). Measuring Firm Complexity. Available at *SSRN* 3645372.

Loughran, T., McDonald, B., & Pragidis, I. (2019). Assimilation of oil news into prices. *International Review of Financial Analysis*, 63, 105-118.

Manela, A., & Moreira, A. (2017). News implied volatility and disaster concerns. *Journal of Financial Economics*, 123(1), 137-162.

Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information.

Meschke, F. (2004). CEO interviews on CNBC. Available at *SSRN* 302602.

Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: from lexicons to transformers. *IEEE* access, 8, 131662-131682.

Moshiri, S., & Foroutan, F. (2006). Forecasting nonlinear crude oil futures prices. *The Energy Journal*, 27(4), 81-96.

Narayan, S., & Narayan, P. K. (2017). Are oil price news headlines statistically and economically significant for investors?. *Journal of Behavioral Finance*, 18(3), 258-270.

Nofsinger, J. R. (2001). The impact of public information on investors. *Journal of Banking & Finance*, 25(7), 1339-1366.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, 93, 106384.

Peng, Y., & Jiang, H. (2015). Leverage financial news to predict stock price movements using word embeddings and deep neural networks. arXiv preprint arXiv:1506.07220.

Peress, J., & Schmidt, D. (2019). Glued to the TV Distracted Noise Traders and Stock Market Liquidity. *The Journal of Finance*.

Qadan, M., & Nama, H. (2018). Investor sentiment and the price of oil. *Energy Economics*, 69, 42-58.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance*, 84, 25-40.

Rennekamp, K. (2012). Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research*, 50(5), 1319-1354.

Rigobon, R., & Sack, B. (2008). Noisy macroeconomic announcements, monetary policy, and asset prices. *In Asset prices and monetary policy* (pp. 335-370). University of Chicago Press.

Ruan, X., & Zhang, J. E. (2018). Risk-neutral moments in the crude oil market. *Energy Economics*, 72, 583-600.

Sias, R. W., & Starks, L. T. (1997). Return autocorrelation and institutional investors. *Journal of Financial Economics*, 46(1), 103-131.

Sigrist, F., & Hirnschall, C. (2019). Grabit: Gradient tree-boosted Tobit models for default prediction. *Journal of Banking & Finance*, 102, 177-192.

Sirignano, J. A. (2019). Deep learning for limit order books. *Quantitative Finance*, 19(4), 549-570.

Smales, L. A. (2014). News sentiment in the gold futures market. *Journal of Banking & Finance*, 49, 275-286.

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, 19(10), 986-988.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926-957.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.

Tetlock, P. C. (2011). All the news that's fit to reprint: Do investors react to stale information?. *The Review of Financial Studies*, 24(5), 1481-1512.

Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 101(2), 449-472.

Varetto, F. (1998). Genetic algorithms applications in the analysis of insolvency risk. *Journal of Banking & Finance*, 22(10-11), 1421-1439.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wysocki, P. D. (1998). Cheap talk on the web: The determinants of postings on stock message boards. *University of Michigan Business School Working Paper*, (98025).

Yang, H., Liu, X. Y., & Wang, C. D. (2023). FinGPT: Open-Source Financial Large Language Models. arXiv preprint arXiv:2306.06031.

Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30(5), 2623-2635.

Zhao, Y., Li, J., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, 66, 9-16.

Zhuge, Q., Xu, L., & Zhang, G. (2017). LSTM Neural Network with Emotional Analysis for prediction of stock price. *Engineering Letters*, 25(2).

TABLE 1 Summary Statistics

Table 1 summarizes the average of daily sentiment, text readability, and liquidity of oil shock news by news sources in each panel. The negativity in tone (Sentiment_GPT) is measured as the difference between negative and positive probabilities, as determined by FinGPT. Readability is assessed using the Fog index score; a lower index indicates higher text complexity. The numerical figure (Number) represents the percentage of numbers in the text. Tone sureness (Sureness) is measured as the percentage of words indicating certainty according to the Harvard-Lasswell dictionaries, such as "absolute" or "always". Each variable is calculated for every oil shock news article, then aggregated for each oil futures trading day; subsequently, the average is taken. The news from each source is segmented to summarize their means and correlation coefficients. Oil volatility refers to the WTI option implied volatility of the future price, calculated using the method by Bakshi, Madan, and Kapadia (2003). Amihud's liquidity measure is defined as $Illiq_t = \frac{|Return_t|}{Price_t \, Trading \, Volume_t}$. Trading volume is measured in daily units.

Panel A: Summaries

Variable: *Sentiment_GPT*	Obs	Mean	Median	Min	Max	Std Dev.
Positive	4,217	2.132	1.112	0.000	24.987	2.735
Negative	4,217	1.444	0.676	0.000	22.027	2.065
Neutral	4,217	1.768	0.997	0.000	18.987	2.069
			Sentiment_GPT			
All	4,217	-0.688	-0.378	-10.960	1.976	0.903
WSJ Online	767	-0.109	-0.082	-1.744	0.923	0.260
WSJ Europe	1,110	-0.134	-0.111	-1.222	0.567	0.198
WSJ Eastern	1,652	-0.168	-0.139	-1.436	0.560	0.182
WSJ Asia	456	-0.110	-0.100	-1.133	0.537	0.196
New York Times	437	-0.042	-0.026	-0.713	0.616	0.143
Financial Times	648	-0.091	-0.075	-1.543	1.258	0.275
Energy Weekly News	8	-0.444	-0.354	-1.062	0.000	0.373
Dow Jones Institutional News	2,553	-0.885	-0.648	-10.438	1.976	0.950
			Readability Measures			
Fog index	5,982	62.031	31.895	3.647	1550.901	82.194
Number	5,982	0.559	0.070	0.000	18.039	1.300
Sureness	5,982	0.059	0.028	0.000	1.627	0.087
			Liquidity Measures			
Oil Return (×100)	7,830	0.190	0.529	-400.478	164.097	24.542
Oil Volatility	5,401	0.316	0.281	-1.963	4.290	0.316
Amihud's Illiquidity	7,634	70.392	28.039	0.000	3278.076	142.555
Trading Volume	7,644	178,004.20	70,212	0.000	1,404,916	410,184.90

Panel B: Correlation in *Sentiment_GPT*

	Positive	Negative
Negative	0.9675	
Neutral	0.9176	0.8860

	Dow Jones Institutional News	Financial Times	New York Times	WSJ Asia	WSJ Eastern	WSJ Europe
Financial Times	-0.7592					
New York Times	-0.4522	-0.1468				
WSJ Asia	0.1038	-0.6447	0.806			
WSJ Eastern	0.5751	-0.7081	0.0764	0.6157		
WSJ Europe	-0.2614	-0.3317	0.9737	0.8645	0.1377	
WSJ Online	-0.0876	-0.3640	0.2830	0.4743	0.5665	0.2357

TABLE 2 The Impacts of Sentiment and Comprehensiveness on Oil Returns

Table 2 shows the impact of news tones, proxies for interpretability, and interaction of tones and interpretability on daily oil returns. Fog Index measures *Readability*; The Harvard-Lasswell dictionaries measure *Sureness* that describes sureness in tones; *Number* is the percentage of numerical figures in news articles. Oil returns are multiplied by 1000. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

dependent variable: Oil Return	Readability				Sureness				Number			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Negativity	-1.776	0.673	0.008	***	-1.665	0.061	0.006	***	-0.102	0.057	0.075	*
Moderator	-0.043	0.010	0.000	***	-4.307	9.785	0.004	***	-0.263	0.060	0.000	***
Interaction	-0.006	0.003	0.030	**	-0.631	0.220	0.004	***	-0.575	0.177	0.001	***
Tuesday	-0.066	0.912	0.942		-0.008	0.091	0.928		0.009	0.091	0.918	
Wednesday	0.684	0.939	0.466		0.068	0.094	0.468		0.082	0.094	0.384	
Thursday	1.249	0.943	0.185		0.122	0.094	0.196		0.139	0.094	0.140	
Friday	1.379	0.899	0.125		0.109	0.089	0.220		0.118	0.089	0.188	
Lags of Sentiment												
L1.Negativity	0.263	0.510	0.606		0.024	0.051	0.632		0.037	0.051	0.471	
L2.Negativity	-0.181	0.498	0.716		-0.017	0.050	0.733		-0.006	0.050	0.911	
L3.Negativity	-0.416	0.503	0.408		-0.043	0.050	0.396		-0.038	0.050	0.446	
L4.Negativity	0.467	0.489	0.339		0.043	0.049	0.377		0.053	0.049	0.272	
Lags of Oil Return (divided by 100)												
L1.Oil Return	-2.526	2.011	0.209		-2.594	2.005	0.196		-2.485	2.014	0.217	
L2.Oil Return	-3.907	1.781	0.028	**	-3.972	1.780	0.026	**	-3.921	1.783	0.028	**
L3.Oil Return	-2.114	2.139	0.323		-2.122	2.137	0.321		-2.070	2.136	0.332	
Constant	0.325	0.724	0.654		0.033	0.072	0.647		-0.008	0.072	0.911	
obs	7,826				7,826				7,826			
R-squared	0.0086				0.0095				0.0090			

TABLE 3 Interactions of News Sentiments and Investor Attention

Table 3 panel A shows the impact of sentiment of oil shocks on daily WTI crude oil futures price returns, interacting with proxies for investor sentiments. The dependent variable is the day-to-day percent changes in WTI crude oil nearest term futures prices. The main variable of interest, *Sentiment* is the disagreement (*Negative* minus *Positive*) in tones of news coverages. FinGPT measure *Negative* and *Positive*. *NVIX* (News-Implied Volatility Index), *FEARS*, and *Oil-Sentiment* are investor sentiment proxies. *NVIX* is derived from news coverage in The Wall Street Journal (Manela and Moreira, 2017). *FEARS35* represents an index that measures the search volume related to investor concerns (Da et al., 2015), while *Oil Sentiment* tracks the search trends pertaining to the oil markets (Qadan and Nama, 2018). In panel B we split the sample into the periods when oil-specific is below or above its median and reports the results of the examination of returns on sentiment. Oil returns are multiplied by 1000. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

Panel A:

dependent variable: Oil Return	*NVIX*				*FEARS35*			*Oil Sentiment*			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value	
Sentiment_GPT	-0.336	0.314	0.283		-0.076	0.078	0.329	-0.025	0.089	0.779	
Attention	-0.010	0.008	0.167		-0.230	0.266	0.386	-0.016	0.006	0.009	***
Sentiment_GPT×Attention	0.008	0.011	0.460		0.382	0.319	0.231	-0.004	0.004	0.349	
					Week Day						
Tuesday	0.013	0.100	0.900		-0.073	0.192	0.704	0.045	0.117	0.700	
Wednesday	0.074	0.102	0.468		0.113	0.205	0.580	0.112	0.125	0.368	
Thursday	0.167	0.104	0.107		0.153	0.200	0.446	0.159	0.121	0.188	
Friday	0.113	0.097	0.246		0.099	0.187	0.599	0.148	0.117	0.206	
					Lags of Sentiment						
L1.Negativity	0.056	0.060	0.349		0.042	0.078	0.588	0.049	0.051	0.334	
L2.Negativity	-0.023	0.058	0.696		-0.032	0.075	0.669	0.019	0.050	0.708	
L3.Negativity	0.011	0.059	0.850		0.015	0.077	0.843	-0.033	0.051	0.511	
L4.Negativity	0.051	0.056	0.364		0.087	0.071	0.223	0.057	0.049	0.247	
					Lags of Oil Return (divided by 10)						
L1.Oil Return	-2.106	2.154	0.328		-4.881	3.540	0.168	-7.049	2.481	0.005	***
L2.Oil Return	-4.513	1.920	0.019	**	-2.750	3.731	0.461	1.492	2.569	0.562	
L3.Oil Return	-1.672	2.328	0.473		4.830	3.572	0.176	0.798	2.429	0.743	
Constant	0.201	0.180	0.266		0.011	0.150	0.941	0.210	0.127	0.099	*
obs	6,884				1,871			3,981			
R-squared	0.52%				1.44%			1.25%			

Panel B:

dependent variable: Oil Return	Oil Sentiment>median				Oil Sentiment<median			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Negativity	-0.149	0.076	0.050	*	-0.060	0.051	0.233	
		Week Day						
Tuesday	-0.008	0.210	0.968		0.141	0.117	0.229	
Wednesday	0.199	0.224	0.376		0.079	0.124	0.524	
Thursday	0.301	0.218	0.167		0.065	0.120	0.587	
Friday	0.224	0.211	0.290		0.079	0.113	0.487	
		Lags of Sentiment						
L1.Negativity	0.070	0.077	0.364		0.052	0.050	0.294	
L2.Negativity	0.038	0.073	0.604		-0.014	0.055	0.793	
L3.Negativity	-0.021	0.078	0.785		-0.020	0.053	0.717	
L4.Negativity	0.072	0.075	0.337		0.048	0.050	0.342	
		Lags of Oil Return (divided by 10)						
L1.Oil Return	-7.146	3.317	0.031	**	-5.580	2.512	0.026	**
L2.Oil Return	-1.767	3.440	0.608		-0.281	2.525	0.911	
L3.Oil Return	2.568	3.252	0.430		-1.827	2.580	0.479	
Constant	-0.225	0.169	0.183		0.034	0.092	0.708	
obs	1,927				1,968			
R-squared	1.11%				0.57%			

TABLE 4 Impact of News Sentiment on Illiquidity

Table 4 presents the effects of oil shock news tones on the daily growth rates of WTI crude oil futures trading volume, focusing on The Wall Street Journal's traditional printed edition, which includes three different versions corresponding to three geographical locations: U.S. Eastern, Europe, and Asia. The table also covers Amihud's illiquidity measure and BKM's 30-day implied volatility. The primary variable of interest, *Sentiment_GPT*, quantifies the difference in tones of news coverage (Negative minus Positive probabilities). Growth rates for volume, illiquidity, and implied volatility are winsorized at the 5th percentile to mitigate the effect of extreme values. News is categorized as good when negativity is less than 0, as bad when negativity is greater than 0, and as no news when negativity data is missing for the day. Oil returns are multiplied by 1000. *, **, and *** indicate significance at 10%, 5%, and 1%.

Panel A		WSJ Asia			WSJ Europe			ESJ US Eastern	
	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value
Negativity	0.545	0.477	0.253	0.47	0.34	0.167	-0.423	0.267	0.113
Week Day	Yes			Yes			Yes		
Lags of Sentiment	Yes			Yes			Yes		
Lags of Return	Yes			Yes			Yes		
obs	7,826			7,826			7,826		
R-squared	0.31%			0.40%			0.34%		

Panel B		Volume			Illiquidity			Volatility	
	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value
Negativity	0.008	0.006	0.204	-0.086	0.08	0.280	0.001	0.003	0.746
Week Day	Yes			Yes			Yes		
Lags of Sentiment	Yes			Yes			Yes		
Lags of Return	Yes			Yes			Yes		
obs	7,611			7,254			5,394		
R-squared	8.10%			7.31%			18.12%		

Panel C		Good News			Bad News			No News	
	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value
Negativity	0.059	0.064	0.359	-0.22	0.141	0.120	-0.011	0.066	0.873
Week Day	Yes			Yes			Yes		
Lags of Sentiment	Yes			Yes			Yes		
Lags of Return	Yes			Yes			Yes		
obs	7,826			7,826			7,826		
R-squared	0.43%			0.36%			0.39%		

TABLE 5 The Impacts of News and Sentiments on Stock Abnormal Returns

Table 5, Panel A, provides a summary of the statistics for abnormal returns and cumulative abnormal returns over three days and five days around the news days, along with *Sentiment_GPT*. We define news days as those on which "oil shock" news is reported, mentioning the names of firms. For these analyses, we match firm names to both abnormal returns and sentiments. We also report the abnormal returns for relevant event days, defining Day 0 as the news day. In Panel B, we examine the equality of (cumulative) abnormal returns between no news days and news days, reporting the p-values for the Wald tests. Panel C presents results from regressions of the (cumulative) abnormal returns on *Sentiment_GPT*, a dummy variable for news days, and controls for days of the week. We adjust the reported abnormal returns (AR) and cumulative abnormal returns (CAR), multiply each by 1000. **, and *** indicate significance at 10%, 5%, and 1%.

Panel A: Statistics	Obs	Mean	Standard Deviation	Min	Max
AR (no news days)	783,079	0.026	1.249	-11.811	15.946
AR (news days)	7,296	0.038	1.378	-5.515	6.944
CAR3 (no news days)	782,509	0.078	3.736	-33.846	47.687
CAR3 (news days)	7,296	0.118	4.120	-18.066	20.271
CAR5 (no news days)	781,939	0.129	6.214	-56.120	78.398
CAR5 (news days)	7,296	0.202	6.861	-31.158	33.740
AR(-2)	4,002	0.050	1.351	-6.290	6.055
AR(-1)	5,459	0.044	1.357	-6.400	6.520
AR(+1)	6,189	0.054	1.408	-6.150	6.808
AR(+2)	4,767	0.060	1.369	-4.218	6.934
Sentiment_GPT	7,296	-0.117	0.132	-0.913	0.625

Panel B: Wald Test (p-values)					
Test if equal AR	0.4353	Test if equal CAR3	0.3610	Test if equal CAR5	0.3149

Panel C: Regressions	AR				CAR3				CAR5			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Sentiment_GPT	0.169	0.114	0.137		0.531	0.334	0.118		0.937	0.566	0.098	*
News Day	0.031	0.022	0.165		0.102	0.067	0.128		0.182	0.111	0.102	
Day of Week		Yes				Yes				Yes		
Constant	0.026	0.003	0.000	***	0.078	0.010	0.000	***	0.130	0.016	0.000	***
Obs		790,375				789,805				789,235		
R-squared		0.0000				0.0000				0.0000		

TABLE 6 Summary Statistics: Comparison Between GPT and Dictionary Models

In Table 6, we report the statistical summaries of sentiments estimated using FinGPT, the Harvard-Lasswell dictionary, and the Loughran and McDonald financial dictionary, denoted as *Sentiment_GPT*, *Sentiment_HL*, and *Sentiment_LMP*, respectively. Panel B is dedicated to analyzing the negative aspect of sentiments. For *Sentiment_GPT*, this is the probability of classifying news as negative. For dictionary-based sentiments (*Sentiment_HL* and *Sentiment_LMP*), this is the percentage of negative words and aggregating them to a daily level. Panel C shifts the focus to the positive aspects of oil shock news. In Panels D and E, the analysis is exclusively on the news from *Dow Jones Institutional News* and *The Wall Street Journal Online*, respectively.

Panel A: Sentiments

	Sentiment_HL	Sentiment_LMP	Sentiment_GPT
Sentiment_HL	1		
Sentiment_LMP	-0.556	1	
Sentiment_GPT	-0.334	0.444	1

	Mean	Standard Deviation	Min	Max	Skewness	Obs
Sentiment_HL	-0.340	0.566	-2.516	7.813	3.167	3,328
Sentiment_LMP	0.333	0.291	-8.094	1.366	-5.500	3,328
Sentiment_GPT	0.455	0.661	-6.466	2.357	-2.511	3,328

Panel B: Negative Tones

	Sentiment_HL	Sentiment_LMP	Sentiment_GPT
Sentiment_HL	1		
Sentiment_LMP	0.573	1	
Sentiment_GPT	0.803	0.433	1

	Mean	Standard Deviation	Min	Max	Skewness	Obs
Sentiment_HL	-0.550	-0.550	-1.072	6.575	2.955	3,328
Sentiment_LMP	-0.387	-0.387	-0.859	7.325	2.897	3,328
Sentiment_GPT	-0.545	-0.545	-1.007	4.928	2.930	3,328

Panel C: Positive Tones

	Sentiment_HL	Sentiment_LMP	Sentiment_GPT
Sentiment_HL	1		
Sentiment_LMP	0.760	1	
Sentiment_GPT	0.796	0.626	1

	Mean	Standard Deviation	Min	Max	Skewness	Obs
Sentiment_HL	-0.576	-0.576	-1.194	5.484	2.479	3,328
Sentiment_LMP	-0.393	-0.393	-0.796	8.064	5.041	3,328
Sentiment_GPT	-0.570	-0.570	-1.106	4.188	2.649	3,328

Panel D: Mean Sentiments (DJI)

	Sentiment_HL	Sentiment_LMP	Sentiment_GPT
Sentiment_HL	1		
Sentiment_LMP	-0.547	1	
Sentiment_GPT	-0.346	0.434	1

	Mean	Standard Deviation	Min	Max	Skewness	Obs
Sentiment_HL	-0.374	0.569	-1.924	7.844	3.444	2,445
Sentiment_LMP	0.338	0.555	-7.329	1.375	-5.249	2,445
Sentiment_GPT	0.443	0.670	-6.214	2.458	-2.397	2,445

Panel E: Mean Sentiments (WSJ Online)

	Sentiment_HL	Sentiment_LMP	Sentiment_GPT
Sentiment_HL	1		
Sentiment_LMP	-0.345	1	
Sentiment_GPT	0.123	-0.011	1

	Mean	Standard Deviation	Min	Max	Skewness	Obs
Sentiment_HL	-0.080	0.913	-3.124	4.664	0.600	693
Sentiment_LMP	0.124	0.856	-6.603	3.316	-2.523	693
Sentiment_GPT	0.051	0.929	-5.594	3.634	-0.657	693

TABLE 7 The Impacts of Sentiments: Comparison Between GPT and Dictionary Models

Table 7 presents the regression results of sentiments, the comprehension moderators (readability, sureness, and numerical content), and their interactions on oil returns. *Readability* and *Sureness* are measured using the Harvard-Lasswell dictionary. The percentage of numerical text in the news, labeled as *Number*, is quantified as the proportion of numerical content. Each of these factors is aggregated to a daily level by summing up. Controls for the day of the week and lags of sentiment and oil returns are included, consistent with the main results in Table 2. In Panel B, we conduct similar regressions as detailed in Appendix A to examine the impact of sentiments on oil return. Oil returns are multiplied by 1000. **, and *** indicate significance at 10%, 5%, and 1%.

Panel A: With Interactions
dependent variable: Oil Return

	Readability								
		Sentiment_HL				*Sentiment_LMP*			
	Coef.	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Sentiment		-24.510	6.518	0.000	***	-39.286	36.602	0.283	
Moderator		-0.007	0.006	0.230		-0.027	0.009	0.003	***
Interaction		0.006	0.020	0.746		-0.006	0.111	0.959	
Week Day			Yes				Yes		
Lags of Sentiment			Yes				Yes		
Lags of Returns			Yes				Yes		
Constant		0.086	0.718	0.904		-0.078	0.728	0.914	
obs		7,826				7,826			
R-squared		0.0100				0.0055			

	Sureness								Number							
	Sentiment_HL				*Sentiment_LMP*				*Sentiment_HL*				*Sentiment_LMP*			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Sentiment	-23.209	6.090	0.000	***	-31.863	34.484	0.356		-19.089	7.228	0.008	***	-8.255	39.102	0.833	
Moderator	-7.181	6.041	0.235		-27.780	9.090	0.002	***	0.004	0.043	0.993		-1.275	0.553	0.021	**
Interaction	3.139	13.692	0.819		-41.086	86.045	0.633		-1.306	0.995	0.189		-2.634	5.865	0.653	
Week Day		Yes				Yes				Yes				Yes		
Lags of Sentiment		Yes				Yes				Yes				Yes		
Lags of Oil Return		Yes				Yes				Yes				Yes		
Constant	0.073	0.072	0.919		-0.055	0.073	0.940		-0.130	0.705	0.853		-0.341	0.713	0.632	
obs	7,826				7,826				7,826				7,826			
R-squared	0.0100				0.0060				0.0102				0.0051			

Panel B: Preliminary Results
dependent variable: Oil Return

	Sentiment_HL								*Sentiemtn_LMP*							
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Negativity	-0.021	0.006	0.000	***	-0.025	0.006	0.000	***	0.014	0.024	0.557		0.011	0.033	0.734	
Week Day																
Tuesday	0.000	0.001	0.644		0.000	0.001	0.892		0.000	0.001	0.650		0.000	0.001	0.732	
Wednesday	0.001	0.001	0.251		0.001	0.001	0.432		0.001	0.001	0.245		0.001	0.001	0.290	
Thursday	0.002	0.001	0.083	*	0.001	0.001	0.115		0.002	0.001	0.078	*	0.002	0.001	0.101	
Friday	0.001	0.001	0.208		0.001	0.001	0.382		0.001	0.001	0.097	*	0.001	0.001	0.132	
Lags of Sentiment																
L1.Negativity					0.009	0.005	0.074	*					0.014	0.023	0.534	
L2.Negativity					-0.004	0.006	0.496						-0.004	0.022	0.857	
L3.Negativity					-0.005	0.005	0.273						0.011	0.020	0.591	
L4.Negativity					0.011	0.005	0.033	**					-0.015	0.018	0.394	
Lags of Oil Return																
L1.Oil Return					-0.023	0.020	0.257						-0.023	0.020	0.251	
L2.Oil Return					-0.041	0.018	0.021	**					-0.038	0.018	0.034	**
L3.Oil Return					-0.023	0.021	0.273						-0.019	0.021	0.365	
Constant	0.000	0.001	0.644		0.000	0.001	0.799		-0.001	0.001	0.349		-0.001	0.001	0.466	
obs	7,830				7,826				7,830				7,826			
R-squared	0.51%				0.99%				0.08%				0.32%			

TABLE 8 Predicting Oil Returns

Table 8 presents the average prediction errors for sentiment predictions based on the model specified in Equation (2). Our sample is divided into two main segments: the first half is used to fit the model, and the second half is for making predictions, as shown in the "full sample" column. Additionally, the sample is split into four parts for a more detailed analysis in the remaining four columns, based on whether the value of the moderator is above or below its median in the first half of the data. The predictions in the second half are then made depending on whether the moderator's value is above or below the median. The sentiments are estimated using three different models: LMP (Loughran and McDonald dictionary), HL (Harvard-

Median prediction errors	Full sample	*Readability* > median	*Sureness*< median	*Number*> median	*Oil_Sentiment*> median
Observations	3,915	2,638	1,296	2,643	2,022
LMP	1.244	-6.229	6.075	-0.887	0.177
HL	0.854	-2.195	5.970	1.956	-0.221
GPT	1.212	-1.246	4.983	1.504	0.005

FIGURE 1 Number of News and Sources Over Years

Figure 1 shows the number of news articles from each of the specified news sources—The Wall Street Journal (WSJ), Financial Times, The New York Times, Dow Jones Institutional News, and Energy Weekly News—plotted over a timeline by year.

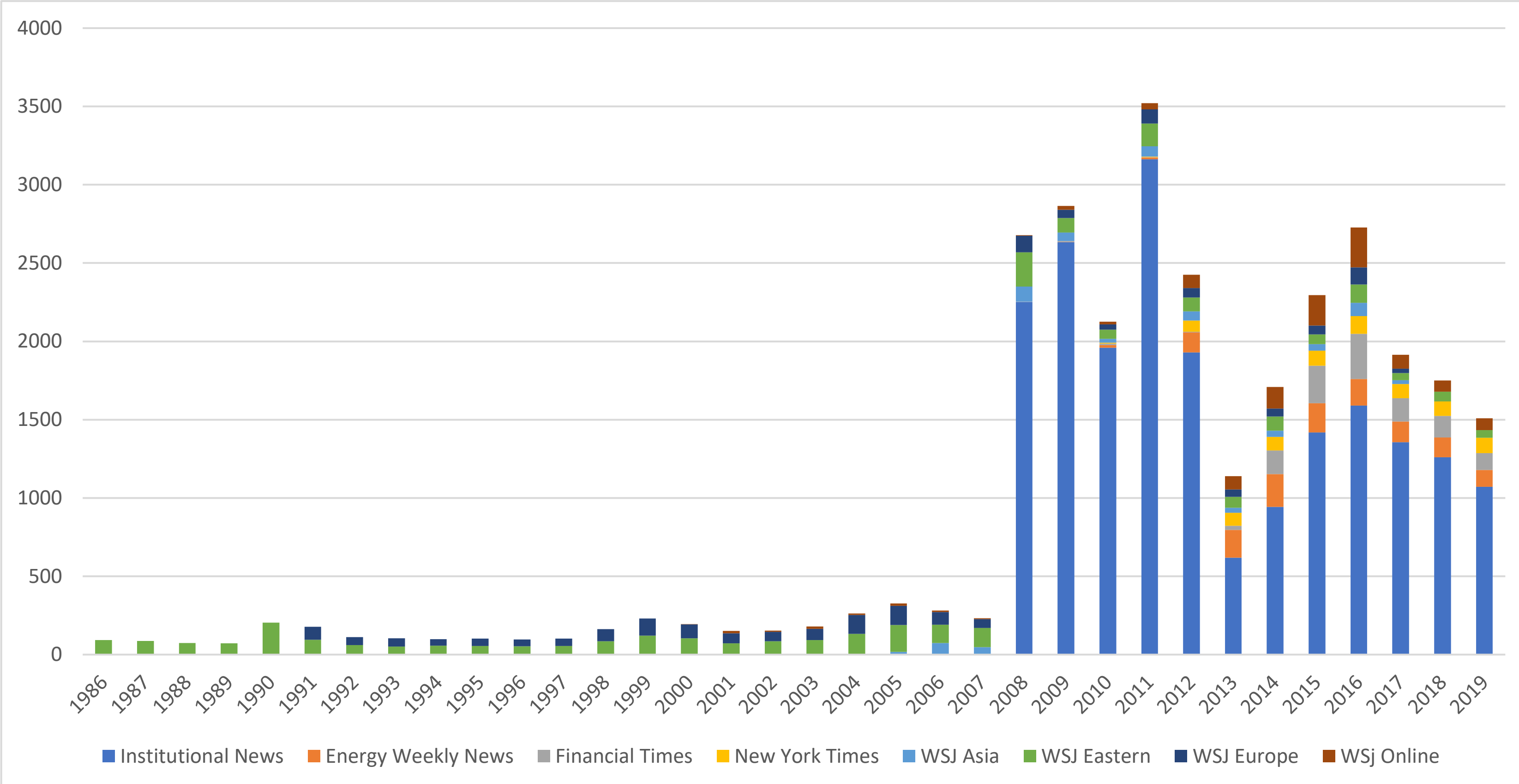


FIGURE 2 Natural Language Processing and Machine Learning

Figure 2 illustrates the hierarchical relationship and intersections among key concepts in artificial intelligence, specifically focusing on machine learning, deep learning, natural language processing (NLP), and large language models (LLMs). NLP incorporates methods ranging from lexicon-based models, which rely on predefined word lists (like the Harvard-Lasswell dictionary for sentiment analysis), to advanced deep learning models like BERT and FinGPT.

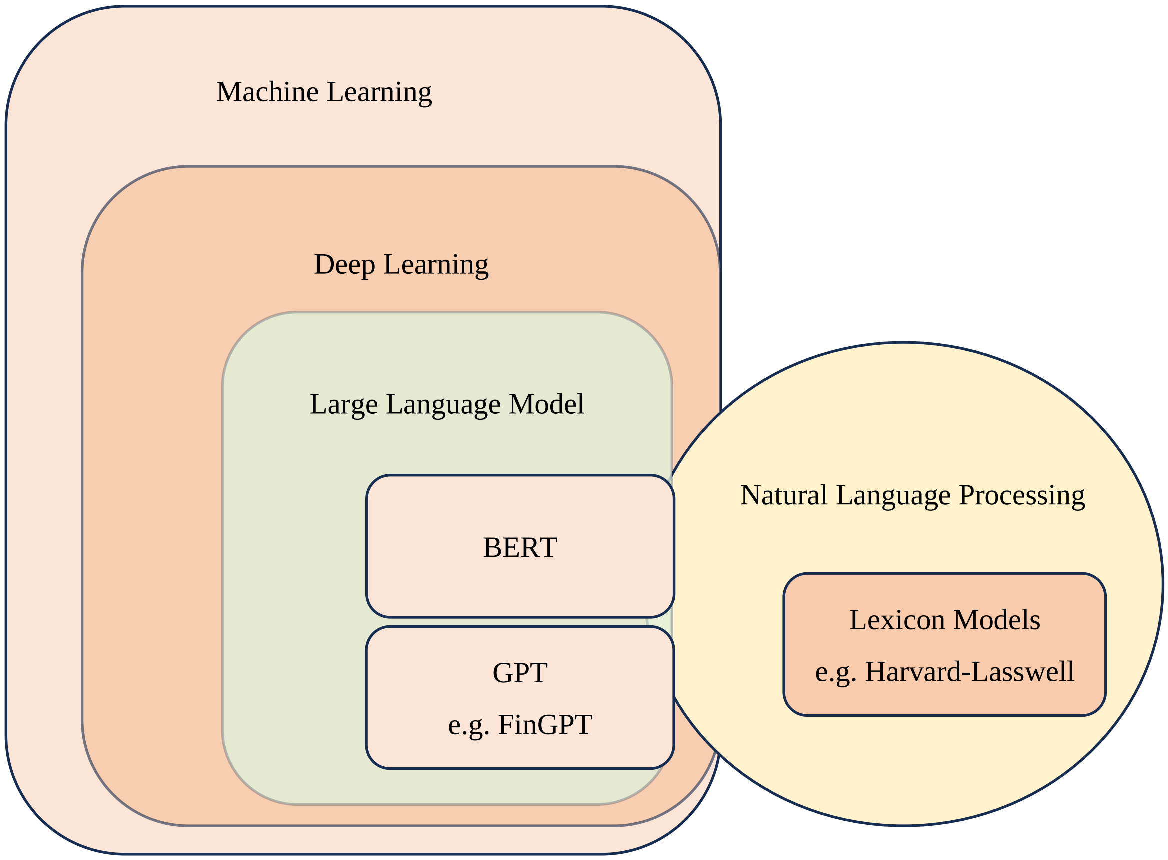


FIGURE 3 Example of A Sample News

Figure 3 is a sample news article collected from ProQuest.

Higher Oil Prices Due To Iran Sanctions To Hurt Businesses -Moody's

Publication info: Dow Jones Institutional News ; New York [New York]22 May 2012.

https://login.ezproxy.lib.ou.edu/login?url=https://search.proquest.com/docview/2125615667?accountid=12964

Abstract: None available.

Links:

http://libraries.ou.edu/eresources/resolver.aspx?rft.genre=article&rft.atitle=Higher+Oil+Prices+Due+To+Iran+Sanctions+To+Hurt+Businesses+-Moody%27s&rft.au=&rft.volume=&rft.issue=&rft.spage=&rft.date=2012&rft.btitle=&rft.jtitle=Dow+Jones+Institutional+News&rft.issn=&rft.isbn=&sid=ProQ%3Aabidateline_

http://libraries.ou.edu/eresources/resolver.aspx?rft.genre=article&rft.atitle=Higher+Oil+Prices+Due+To+Iran+Sanctions+To+Hurt+Businesses+-Moody%27s&rft.au=&rft.volume=&rft.issue=&rft.spage=&rft.date=2012&rft.btitle=&rft.jtitle=Dow+Jones+Institutional+News&rft.issn=&rft.isbn=&sid=ProQ%3Aabidateline_

Full text: SINGAPORE (Dow Jones)--Higher oil prices due to sanctions on Iranian oil exports will benefit international upstream companies but will be detrimental for the airlines sector, downstream oil businesses, and Europe's automobile and retail sectors, Moody's Investors Service said Tuesday.

These industries will be hardest-hit in the event of an *oil shock* when oil prices rise to $150.00 a barrel for several months, which could also derail the global economic recovery, Moody's said.

High oil prices will push up fuel costs for airlines, hurting passenger demand as air travel becomes unaffordable, hit discretionary spending in sectors like retail and restaurants and push up raw material costs for manufacturing companies.

While airlines and carmakers on both sides of the Atlantic will be affected, European companies face a greater risk, it said.

End-users for Iranian crude exports are mostly in Europe and Asia and will bear the brunt of a supply disruption.

"European airlines face more exposure to any political instability in the Middle East than their U.S. or Asian counterparts," Moody's said.

"However, in the U.S., big-box discount retailers and warehouse clubs including Wal-Mart, Target, Costco and BJ's Wholesale could benefit as consumers economize," it added.

"At this stage, it is unclear to what degree Iran's top customers will reduce their imports in order to avoid the sanctions," Moody's said.

-By Eric Yep, ; +65 6415 4063; eric.yep@dowjones.com

(END)

May 22, 2012 03:23 ET (07:23 GMT)

Subject: Airlines; Sanctions; Price increases

Location: Iran Middle East United States--US Asia Europe

Company / organization: Name: Moodys Investors Service Inc; NAICS: 522110, 523930, 561450; Name: Costco Wholesale Corp; NAICS: 452910; Name: Walmart Inc; NAICS: 452112, 452910, 454111

Title: Higher Oil Prices Due To Iran Sanctions To Hurt Businesses -Moody's
Publication title: Dow Jones Institutional News; New York
Publication year: 2012
Publication date: May 22, 2012
Publisher: Dow Jones & Company Inc
Place of publication: New York
Country of publication: United States, New York
Publication subject: Business And Economics
Source type: Wire Feeds
Language of publication: English
Document type: News
ProQuest document ID: 2125615667
Document URL:
https://login.ezproxy.lib.ou.edu/login?url=https://search.proquest.com/docview/2125615667?accountid=12964

Last updated: 2018-10-27
Database: Business Premium Collection

FIGURE 4 Oil Shock News Sentiments

Figure 4 displays the pooled daily tones of sentiment from FinGPT (*Sentiment_GPT*), sentiment analyzed using the Harvard-Lasswell dictionary (*Sentiment_HL*), and sentiment derived from the Loughran and McDonald dictionary (*Sentiment_LMP*) for "oil shock" news in The Wall Street Journal, Financial Times, The New York Times, Dow Jones Institutional News, and Energy Weekly News from 1986 to 2019. The sentiments measure the disagreement (Negativity minus Positivity) in tones using FinGPT, and the percentage of negative words minus the percentage of positive words for *Sentiment_HL* and *Sentiment_LMP* using the Harvard-Lasswell dictionary and the Loughran and McDonald dictionary, respectively. The news sentiments are aggregated daily by summing up. To better illustrate the sentiments, "zero" values are not assigned to days with no news. All sentiments are standardized at the news level to have a zero mean and one standard deviation, then aggregated daily. In the graphs, navy represents *Sentiment_HL*, dark green represents *Sentiment_LMP*, and light blue represents *Sentiment_GPT*.

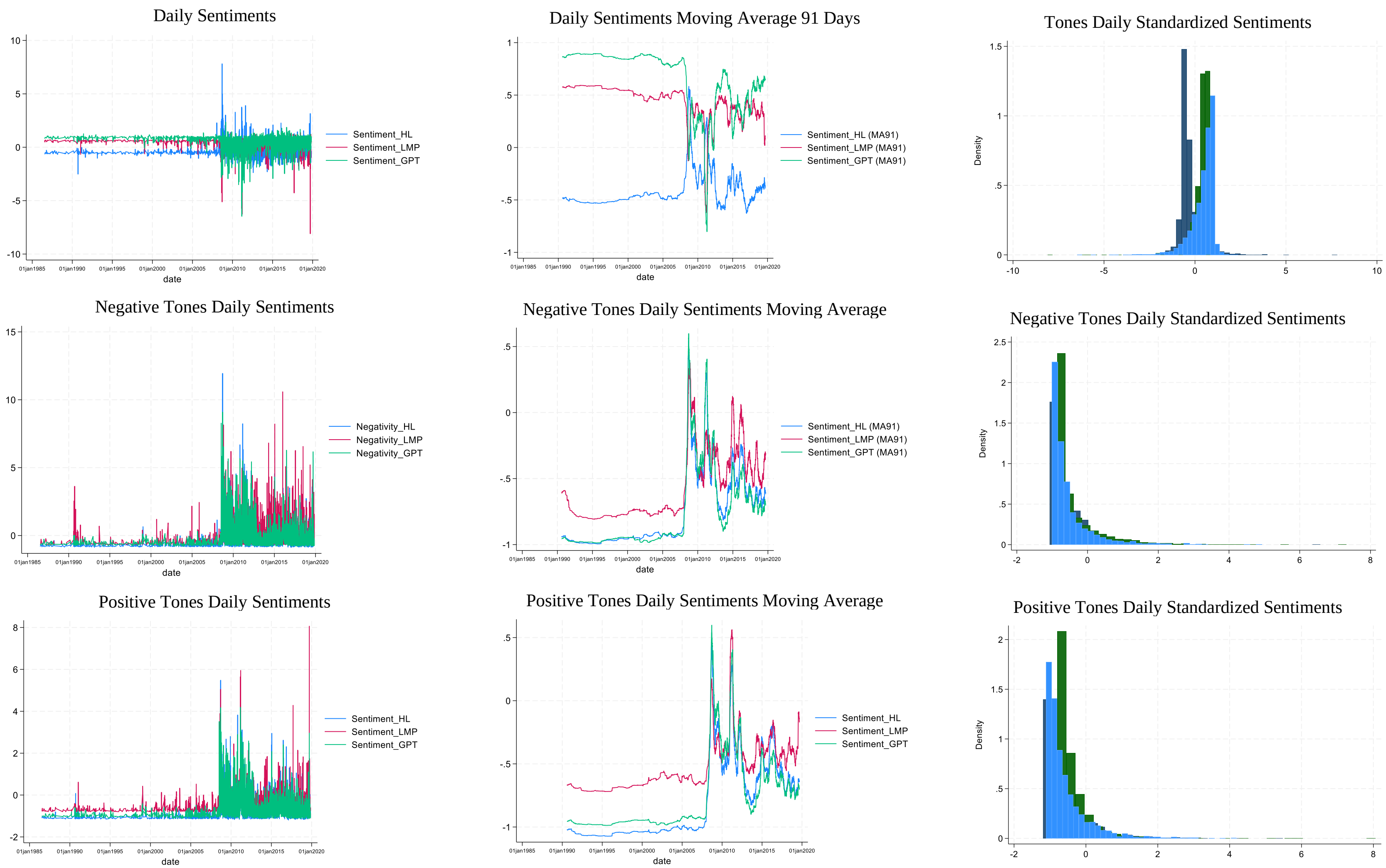


Figure 5 Oil Price and Returns

Figure 5 displays the trends in oil prices and oil returns, with the oil prices depicted by a blue line and the oil returns by red lines. The oil return is calculated as the percentage change in oil prices on a daily basis. The data for oil prices is collected from the Chicago Mercantile Exchange (CME) settlement prices of West Texas Intermediate (WTI) for the 1st front-month futures contract.

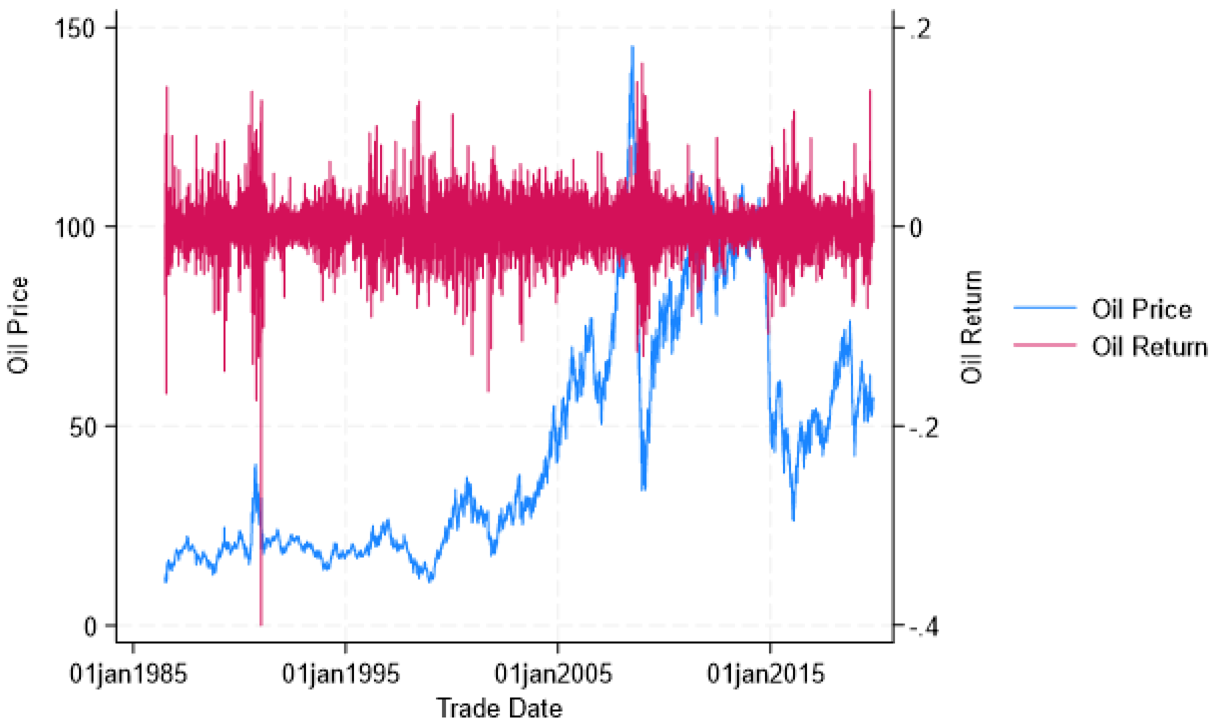


FIGURE 6 Sentiments by Source, Geographic Location, Media Attention, and Liquidity of Oil

Figure 6.1 visualizes the pooled tones of oil shock news in the Financial Times, New York Times, Dow Jones Institutional News, and Energy Weekly News from 2008 to 2019. The primary variable, Negativity, represents the difference between negative and positive tones in news coverage, using FinGPT to measure both negative and positive probabilities. Figure 6.2 focuses on the pooled tones of oil shock news in WSJ Asia, WSJ Eastern, WSJ Europe, and WSJ Online from 1986 to 2019. Figure 6.3 presents trends in sentiment measures including NVIX, FEARS, and Oil-Sentiment from 1986 to 2019. NVIX, or News-Implied Volatility Index, measures market sentiments using the tones of WSJ front-page titles, as in Manela and Moreira (2017). FEARS, indicating Financial and Economic Attitudes Revealed by Search, through Google search trends for terms associated with fear, according to Da et al. (2015). Oil-Sentiment is an index that tracks search trends for three oil-related terms (“oil price,” “price of oil,” and “crude oil”), as developed by Qadan and Nama (2018). Figure 6.4 displays the daily WTI crude oil futures returns, daily trading volume, and BKM’s 30-day option-implied volatility from 1986 to 2020.

Figure 6.1

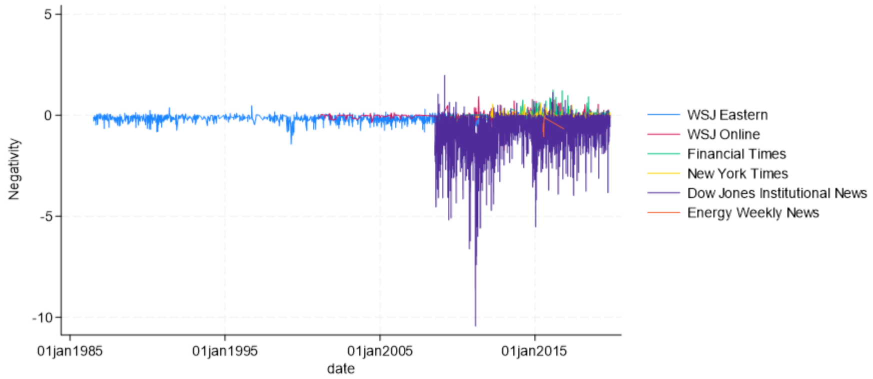


Figure 6.2

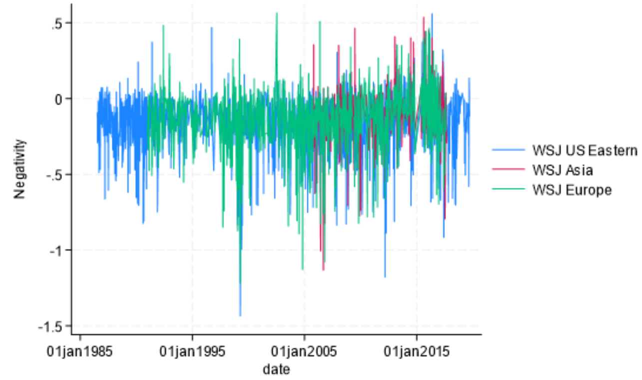


Figure 6.3

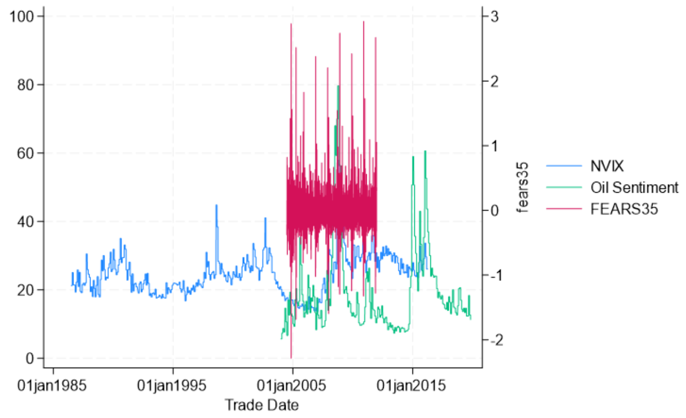


Figure 6.4

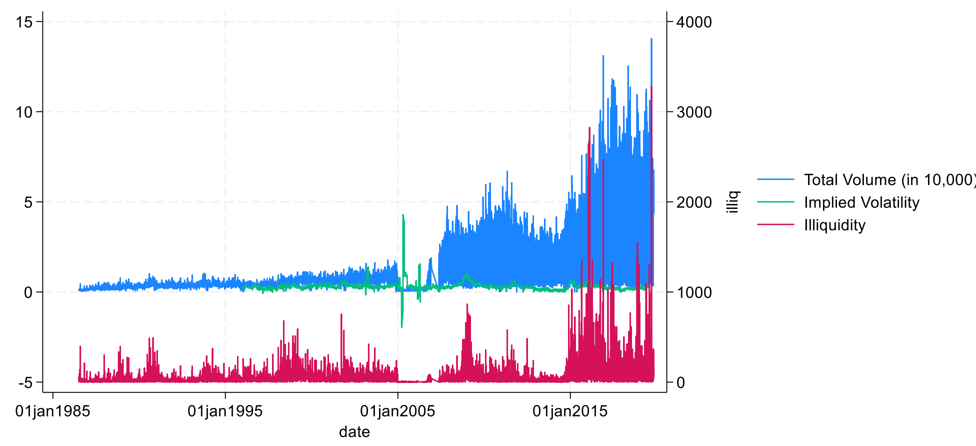
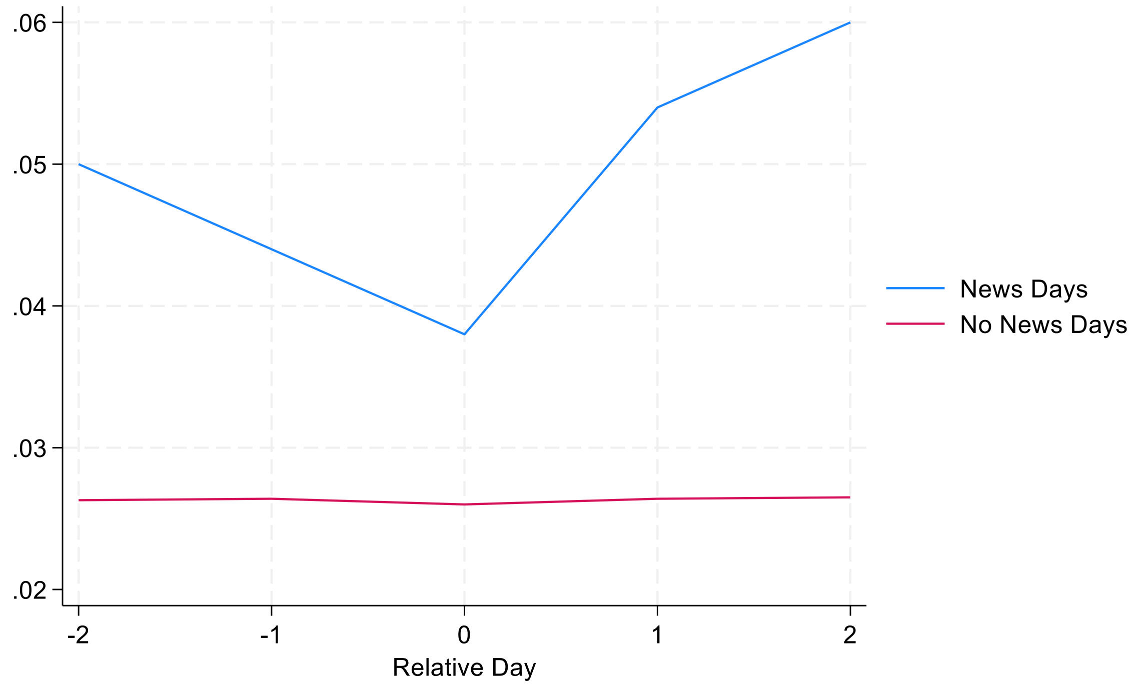


FIGURE 7 Impacts of Reported Oil Shock News on Stock Returns

We depict the average of the abnormal returns on the graph for the days from -2 to +2, where day 0 is identified as the report day of oil shock news mentioning the names of the firms, illustrated with blue lines. These abnormal returns are calculated as Fama-French four-factor abnormal returns, multiplied by 100 for clarity, and gathered from the WRDS beta suite. The red lines display the average of the abnormal returns for days with no news mentioning the names of the firms.



Relative Day	News Days	No News Days
-2	0.0500	0.0263
-1	0.0440	0.0264
0	0.0380	0.0260
+1	0.0540	0.0264
+2	0.0600	0.0265

FIGURE 8 Sentiments and Comprehension Moderators: Comparison Between FinGPT and Dictionaries

Figure 8 illustrates the relationship between sentiments (estimated using GPT, HL for the Harvard-Lasswell dictionary, and LMP for the Loughran and McDonald dictionary) and three comprehension moderators (numerical text, readability, and sureness). *Numerical text* is the percentage of numbers in the news; *Readability* is measured by the Fog index; *Sureness* is the percentage of words included in the Sureness dictionary from Harvard-Lasswell. The figure depicts histograms of the distributions of sentiments when splitting the moderators by their medians. In the graphs, navy color represents *Sentiment_HL*, dark green indicates *Sentiment_LMP*, and light blue symbolizes *Sentiment_GPT*.

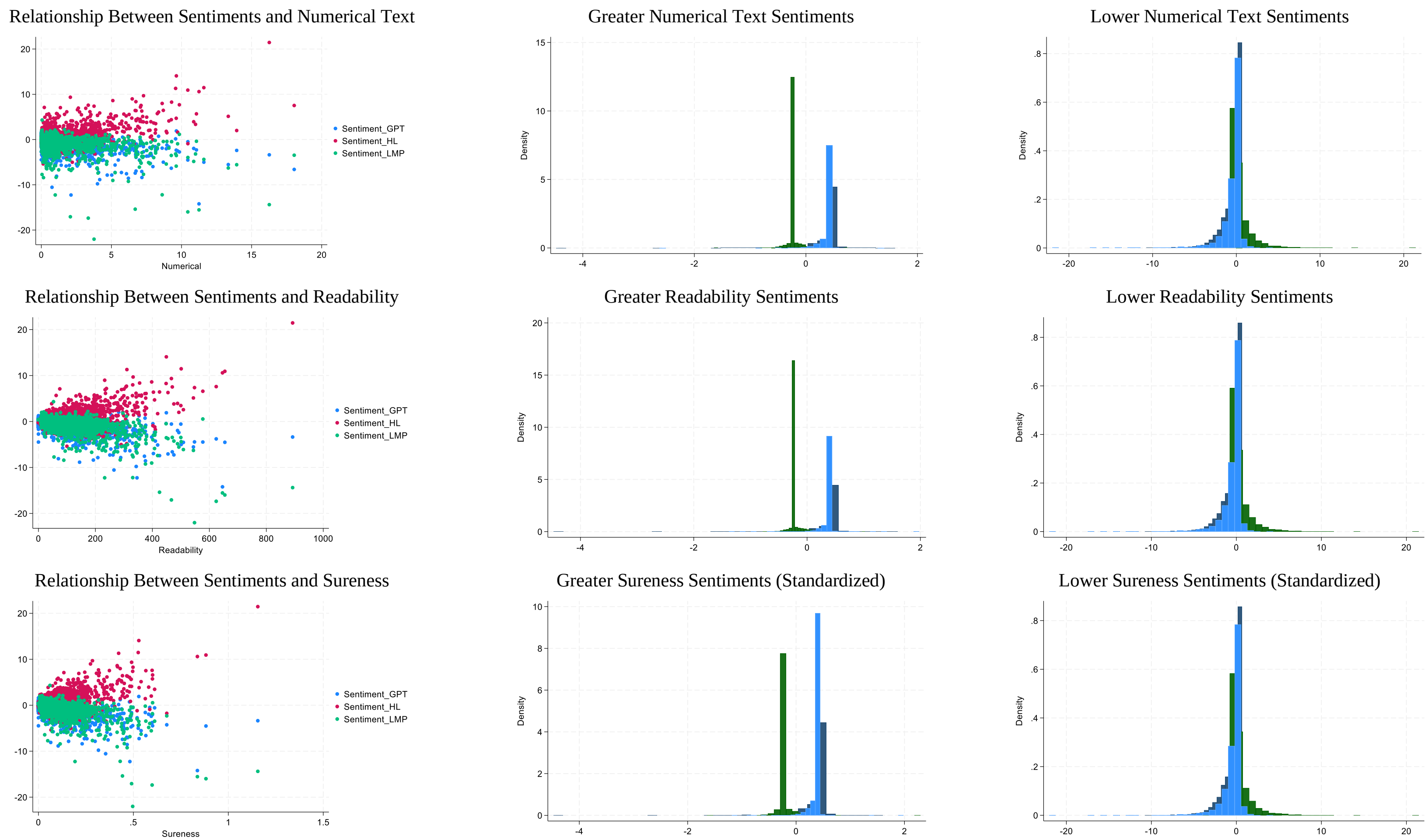
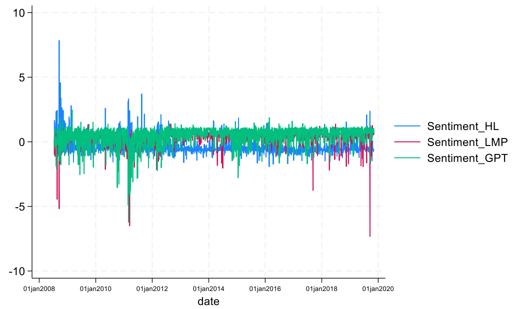


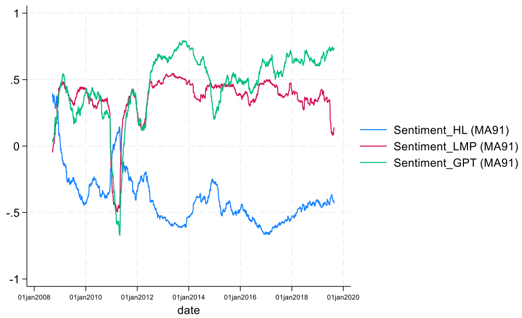
FIGURE 9 Sentiments from Dow Jones Institutional News and Wall Street Journal Online

Figure 9 presents the sentiments analyzed by GPT, Harvard-Lasswell, and Loughran and McDonald specifically for news exclusively from Dow Jones Institutional News (*DJI*) and The Wall Street Journal Online (*WSJ Online*). Additionally, the figure includes the 91-day moving average of the sentiments to show trends over time. The histograms use navy color to represent *Sentiment_HL*, dark green for *Sentiment_LMP*, and light blue for *Sentiment_GPT*.

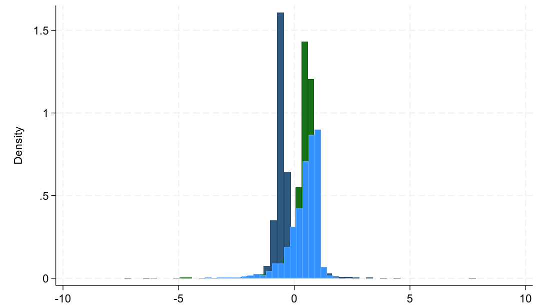
Daily Sentiments DJI Only



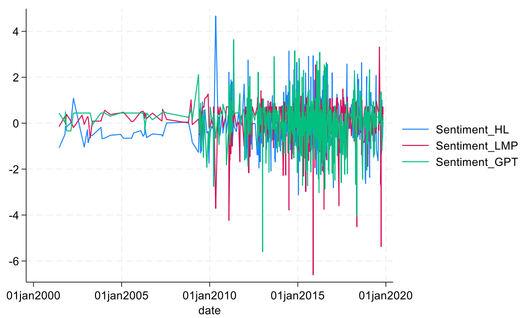
Daily Sentiments DJI Moving Average



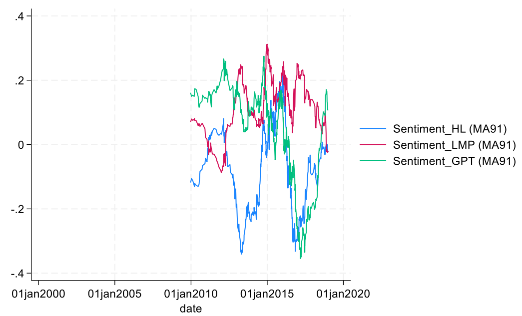
Histogram Daily Sentiments DJI Only



Daily Sentiments WSJ Online



Daily Sentiments WSJ Online Moving Average



Histogram Daily Sentiments WSJ Online Only

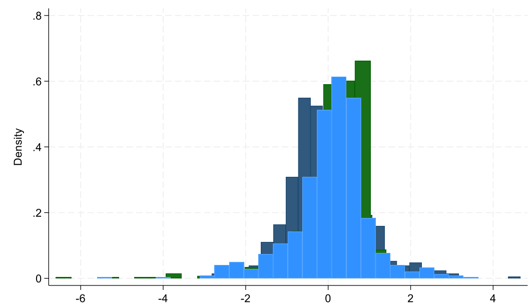


TABLE A Impact of Media Coverage and Tones on Daily Oil Returns

Table A shows the impact of coverage of oil shocks from WSJ, Financial Times, New York Times, Dow Jones Institutional News, and Energy Weekly News on daily WTI crude oil futures price returns. The dependent variable is the day-to-day decimal changes in WTI crude oil nearest term futures prices. The main variable of interest, *Sentiment* is the disagreement (*Negative* minus *Positive*) in tones of news coverages. FinGPT measure *Negative* and *Positive*. Tuesday, Wednesday, Thursday, and Friday, as well as coverage, are dummy variables. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

dependent variable: Oil Return	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value
Coverage	-0.031	0.039	0.417						
Sentiment				-0.032	0.039	0.417	-0.094	0.050	0.062 **
Week Day									
Tuesday	0.036	0.091	0.694	0.036	0.091	0.694	0.013	0.092	0.891
Wednesday	0.104	0.094	0.267	0.104	0.094	0.267	0.084	0.094	0.371
Thursday	0.162	0.094	0.085 *	0.162	0.094	0.085 *	0.143	0.094	0.130
Friday	0.144	0.088	0.106	0.144	0.089	0.106	0.114	0.089	0.202
Lags of Sentiment									
L1.Negativity							0.065	0.051	0.200
L2.Negativity							0.007	0.049	0.894
L3.Negativity							-0.012	0.051	0.810
L4.Negativity							0.078	0.048	0.106
Lags of Oil Return									
L1.Oil Return							-2.277	2.012	0.258
L2.Oil Return							-3.801	1.785	0.033 **
L3.Oil Return							-1.932	2.147	0.368
Constant	-0.083	0.071	0.242	-0.083	0.071	0.242	-0.033	0.072	0.650
obs	7,830			7,830			7,826		
R-squared	0.09%			0.07%			0.39%		

TABLE B Impact of News Tones on Daily Oil Returns, by News Source

Table B shows the impact of coverage of oil shocks from WSJ, WSJ Online, Financial Times, New York Times, Dow Jones Institutional News, and Energy Weekly News on daily WTI crude oil futures price returns. The dependent variable is the day-to-day percent changes in WTI crude oil nearest term futures prices. The main variable of interest, *Sentiment* is the disagreement (*Negative* minus *Positive*) in tones of news coverages. FinGPT measure *Negative* and *Positive*. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

dependent variable: Oil Return	WSJ Eastern				WSJ Online				Financial Times			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Sentiment	-0.423	0.267	0.113		-0.519	0.304	0.087	*	-9.90×$^{-8}$	0.343	1.000	
Week Day												
Tuesday	0.013	0.093	0.885		0.029	0.092	0.747		0.031	0.091	0.739	
Wednesday	0.091	0.095	0.342		0.998	0.095	0.291		0.099	0.094	0.294	
Thursday	0.132	0.095	0.166		0.151	0.094	0.108		0.152	0.094	0.107	
Friday	0.117	0.089	0.191		0.132	0.089	0.137		0.134	0.089	0.132	
Lags of Sentiment												
L1.Negativity	-0.095	0.284	0.737		-0.012	0.291	0.967		0.530	0.326	0.104	
L2.Negativity	-0.259	0.269	0.337		-0.028	0.287	0.922		0.195	0.422	0.644	
L3.Negativity	-0.019	0.291	0.948		-0.541	0.292	0.064		0.471	0.324	0.145	
L4.Negativity	0.087	0.303	0.774		-0.019	0.292	0.949		0.184	0.362	0.611	
Lags of Oil Return												
L1.Oil Return	-2.338	2.010	0.245		-2.317	2.009	0.249		-2.340	2.016	0.245	
L2.Oil Return	-3.824	1.785	0.032	**	-3.768	1.789	0.035	**	-3.844	1.786	0.031	**
L3.Oil Return	-1.200	2.148	0.352		-1.938	2.149	0.367		-1.949	2.146	0.364	
Constant	-0.074	0.072	0.304		-0.073	0.071	0.301		-0.051	0.070	0.467	
obs	7,826				7,826				7,826			
R-squared	0.34%				0.36%				0.36%			

Table B Impact of News Tones on Daily Oil Returns, by News Source

(Continued)

dependent variable: Oil Return	New York Times				Dow Jones Institutional News				Energy Weekly News			
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value	
Sentiment	0.404	0.891	0.650		-0.100	0.058	0.083	*	0.535	0.684	0.434	
Week Day												
Tuesday	0.031	0.091	0.735		0.026	0.092	0.774		0.025	0.092	0.786	
Wednesday	0.100	0.094	0.291		0.090	0.094	0.340		0.097	0.095	0.307	
Thursday	0.152	0.094	0.106		0.151	0.094	0.109		0.152	0.094	0.108	
Friday	0.136	0.089	0.127		0.125	0.089	0.160		0.132	0.089	0.139	
Lags of Sentiment												
L1.Negativity	-0.535	0.724	0.459		0.067	0.056	0.228		1.806	0.705	0.010	**
L2.Negativity	0.773	0.760	0.309		0.009	0.056	0.876		-1.406	0.414	0.001	***
L3.Negativity	-0.192	0.661	0.771		-0.026	0.057	0.647		1.525	1.554	0.326	
L4.Negativity	-0.576	0.674	0.393		0.096	0.056	0.084		1.456	0.860	0.090	*
Lags of Oil Return												
L1.Oil Return	-2.281	2.007	0.256		-2.276	2.011	0.258		-2.290	2.010	0.254	
L2.Oil Return	-3.790	1.789	0.034	**	-3.780	1.784	0.033	**	-3.810	1.789	0.033	**
L3.Oil Return	-1.926	2.146	0.369		-1.909	2.146	0.374		-1.966	2.147	0.360	
Constant	-0.063	0.070	0.372		-0.044	0.071	0.540		-0.058	0.070	0.411	
obs	7,826				7,826				7,826			
R-squared	0.32%				0.39%				0.34%			

TABLE C Interactions of News Sentiments and Investor Attentions

Table C shows the impact of coverage of oil shocks on daily WTI crude oil futures price returns, interacting with proxies for investor sentiments. Samples are divided into higher-than-median and lower-than-median sentiment subsamples. The dependent variable is the day-to-day percent changes in WTI crude oil nearest term futures prices. The main variable of interest, *Sentiment* is the disagreement (*Negative* minus *Positive*) in tones of news coverages. FinGPT measure *Negative* and *Positive*. *NVIX* (News-Implied Volatility Index) and *FEARS* are investor sentiment proxies. *, **, and *** indicate significance at 10%, 5%, and 1%, respectively.

dependent variable: Oil Return	NVIX>median				NVIX<median		
	Coef.	Robust S.E.	p-value		Coef.	Robust S.E.	p-value
Sentiment	-0.107	0.064	0.091	*	0.107	0.159	0.501
			Week Day				
Tuesday	0.079	0.157	0.615		-0.024	0.128	0.850
Wednesday	0.096	0.165	0.559		0.091	0.124	0.462
Thursday	0.265	0.172	0.125		0.106	0.120	0.378
Friday	0.206	0.158	0.193		0.051	0.116	0.664
			Lags of Sentiment				
L1.Negativity	0.052	0.065	0.422		0.155	0.144	0.282
L2.Negativity	-0.006	0.063	0.923		-0.128	0.150	0.391
L3.Negativity	0.030	0.065	0.648		-0.113	0.150	0.450
L4.Negativity	0.060	0.061	0.330		0.063	0.147	0.668
			Lags of Oil Return				
L1.Oil Return	-1.649	3.203	0.607		-2.560	2.087	0.220
L2.Oil Return	-6.161	2.759	0.026	**	-1.210	2.035	0.552
L3.Oil Return	-1.884	3.523	0.593		-0.631	1.944	0.745
Constant	-0.141	0.129	0.274		0.019	0.095	0.837
obs	3,375				3,509		
R-squared	0.68%				0.21%		

Table C Interactions of News Sentiments and Investor Attentions

(Continued)

dependent variable: Oil Return	FEARS>median			FEARS<median			
	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value	
Sentiment	-0.030	0.114	0.795	-0.109	0.111	0.327	
Week Day							
Tuesday	-0.128	0.283	0.651	0.033	0.265	0.900	
Wednesday	0.205	0.290	0.479	0.129	0.285	0.650	
Thursday	0.017	0.286	0.951	0.364	0.281	0.196	
Friday	0.160	0.262	0.542	0.144	0.272	0.596	
Lags of Sentiment							
L1.Negativity	0.019	0.113	0.867	0.016	0.119	0.893	
L2.Negativity	-0.076	0.112	0.498	0.003	0.106	0.975	
L3.Negativity	0.172	0.138	0.213	-0.120	0.087	0.169	
L4.Negativity	0.080	0.097	0.408	0.110	0.108	0.309	
Lags of Oil Return							
L1.Oil Return	-0.870	5.111	0.865	-8.897	4.972	0.074	*
L2.Oil Return	0.626	4.724	0.895	-5.959	5.632	0.290	
L3.Oil Return	7.023	4.349	0.107	3.336	5.723	0.560	
Constant	-0.078	0.223	0.728	0.001	0.199	0.995	
obs	935			935			
R-squared	1.21%			1.95%			

Table D The Impacts of Simple Average Sentiments on Oil Returns

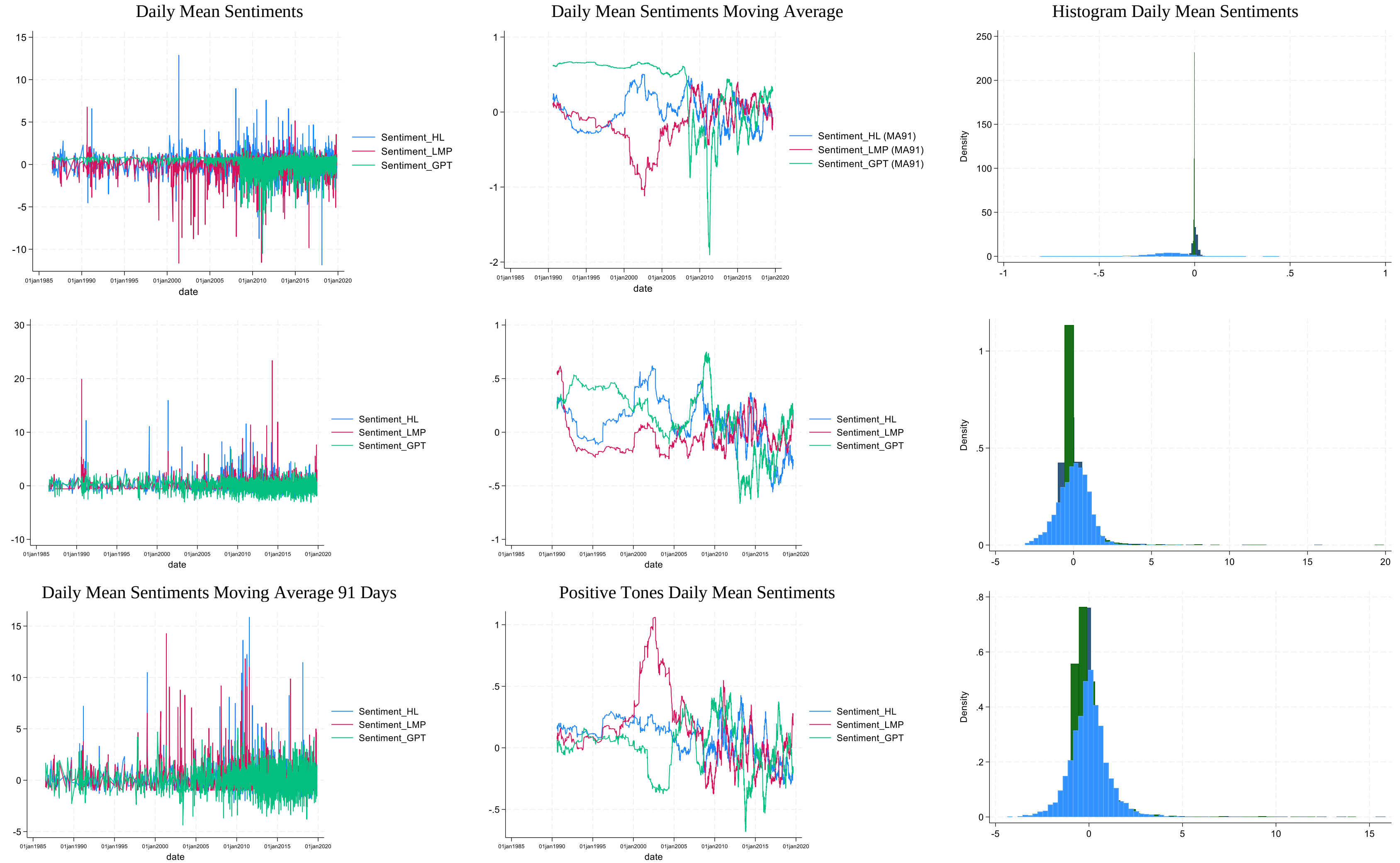
Table D details the impact of the simple average of news-level sentiments on oil returns, offering a distinct approach from the main model, which aggregates news-level sentiments by summing them up. This method involves taking a simple average of news-level sentiments for daily aggregation. The significance levels are marked with asterisks: *, **, and *** signify statistical significance at the 10%, 5%, and 1% levels, respectively.

Panel A: FinGPT						
dependent variable: Oil Return						
	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value
Sentiment	0.002	0.003	0.480	-0.001	0.003	0.815
Week Day						
Tuesday	0.000	0.001	0.635	0.000	0.001	0.780
Wednesday	0.001	0.001	0.234	0.001	0.001	0.312
Thursday	0.002	0.001	0.071 *	0.002	0.001	0.107
Friday	0.002	0.001	0.086 *	0.001	0.001	0.154
Lags of Sentiment						
L1.Negativity				0.002	0.003	0.470
L2.Negativity				0.001	0.003	0.793
L3.Negativity				0.001	0.003	0.738
L4.Negativity				0.004	0.003	0.149
Lags of Oil Return						
L1.Oil Return				-0.023	0.020	0.244
L2.Oil Return				-0.038	0.018	0.032 **
L3.Oil Return				-0.020	0.021	0.361
Constant	-0.001	0.001	0.352	0.000	0.001	0.852
obs	7,830			7,826		
R-squared	0.01%			0.35%		

Panel B: HL							Panel C: LMP					
	dependent variable: Oil Return											
	Coef.	Robust S.E.	p-value	Coef.	Robust S.E.	p-value						
Sentiment	-0.076	0.027	0.004 ***	-0.087	0.038	0.022 **	-0.057	0.138	0.680	-0.351	0.235	0.136
Week Day												
Tuesday	0.000	0.001	0.797	0.001	0.002	0.745	0.000	0.001	0.807	0.001	0.002	0.663
Wednesday	0.001	0.002	0.697	0.001	0.002	0.558	0.001	0.002	0.708	0.001	0.002	0.530
Thursday	0.001	0.002	0.610	0.001	0.002	0.500	0.001	0.002	0.588	0.001	0.002	0.519
Friday	0.001	0.001	0.688	0.001	0.002	0.697	0.001	0.001	0.527	0.001	0.002	0.450
Lags of Sentiment												
L1.Negativity				0.038	0.038	0.318				0.075	0.201	0.709
L2.Negativity				-0.013	0.048	0.783				0.006	0.247	0.981
L3.Negativity				-0.001	0.039	0.990				-0.116	0.200	0.561
L4.Negativity				0.053	0.040	0.184				0.137	0.192	0.477
Lags of Oil Return												
L1.Oil Return				-0.066	0.035	0.062 *				-0.063	0.035	0.069 *
L2.Oil Return				-0.017	0.037	0.639				-0.016	0.037	0.661
L3.Oil Return				0.030	0.034	0.377				0.032	0.033	0.336
Constant	-0.001	0.001	0.330	-0.001	0.001	0.408	-0.002	0.001	0.185	-0.002	0.002	0.248
obs	7,830			7,826			7,830			7,826		
R-squared	0.24%			0.57%			2.45%			0.35%		

Figure A Simple Average of News-Level Sentiments

Figure A displays the simple average of news-level sentiments (GPT, Harvard-Lasswell, and Loughran and McDonald dictionaries) aggregated to daily levels. Unlike summing up news-level sentiments in the main model, this approach takes a simple average of the news-level sentiments for daily aggregation. The figure also features the 91-day moving average of the sentiments to illustrate trends over time. In the histograms, navy color represents *Sentiment_HL*, dark green indicates *Sentiment_LMP*, and light blue symbolizes *Sentiment_GPT*.



Appendix A: Introduction to FinGPT

FinGPT, developed by Yang et al. (2023), builds on the foundational GPT models, featured for their interactivity and based on transformer technology. As a generative model, FinGPT extracts information from vast data feeds and create new content from it. Instances of this technology include GPT-3 and GPT-4, and ChatGPT. While many large language models (LLMs) like BERT have garnered attention, GPT models have proven well adept at understanding and extracting information from natural language.

A distinctive feature of GPT models is their ability to be fine-tuned, meaning additional, domain-specific information can be integrated into the framework to enhance their capability for understanding specialized tasks. For FinGPT, this entails refining the model with a substantial financial news data.

Special Features of FinGPT:

Reinforcement Learning from Stock Prices: FinGPT uses stock price responses as a feedback mechanism to augment its comprehension of financial information.

Lora (Low-Rank Adaptation Tensors): Utilizing stock price responses, FinGPT optimizes the use of pre-trained data through low-rank adaptation tensors, enabling a reduction in parameters.

The development process of FinGPT involves several steps:

Base Model Preparation: Starting with a pre-trained GPT model capable of understanding and responding in human natural language.

Fine-Tuning: The GPT model is then specialized with financial news data from sources such as Reuters, CNBC, Yahoo Finance, social media platforms (Twitter, Facebook, Weibo, Reddit), SEC filings, and Google trends.

Application: In this study, FinGPT is used to analyze oil shock news from DJI, The Wall Street Journal, Financial Times, The New York Times, and Energy Weekly News. It evaluates news content, assigning scores based on the likelihood of being classified as *positive*, *negative*, or *neutral*.

FinGPT is made available as an open-source tool under the MIT license, allowing for widespread use and adaptation.