Noise or Fundamentals? The Predictive Role of News and Social Media in the Crude Oil Market

Gaoping Ma[a,*], Yahua Xu[a,*], Z. Ivy Zhou[b,*]

[a] China Economics and Management Academy, School of Innovation and Development, Central University of Finance and Economics, China

[b] School of Business, University of Wollongong, Australia

Abstract

This study examines the impact of traditional news and social media platforms on the crude oil market. By using a unique dataset on sentiment scores from both channels, we discover that news media convey fundamental information that leads to a permanent price impact. In contrast, the predictive power of social media exhibits significant reversals yet retains a positive long-term impact, indicating its influence stems from a mix of fundamental information and noise with fundamentals playing a dominant role. We also demonstrate that the information from news and social media does not entirely overlap, with social media offering a unique informational component not present in news content. Furthermore, we show that news and social media can forecast announcements related to changes in oil inventories and reveal that WTI futures market speculators adjust positions based on news media information, but not on social media posts. Our study provides important implications for regulators to enhance their monitoring of misinformation on social media and for market participants to understand the informational elements of news and social media platforms.

Keywords: News sentiment, Twitter sentiment, social media, return predictability, crude oil futures

* Gaoping Ma (gaoping.ma@email.cufe.edu.cn), Yahua Xu (yahua.xu@cufe.edu.cn), Z. Ivy Zhou (izhou@uow.edu.au)

1. Introduction

Investors have long relied on traditional information sources, such as corporate and exchange announcements, press, internet, and analysts' reports , to obtain timely and relevant information. However, the landscape has evolved significantly with the rise of social media platforms over the past decade, in particular Twitter (currently known as X). These platforms have become a prevalent source of information for capital market, offering a real-time dissemination of financial and other information to a broad spectrum of participants. The guidance issued by the US Securities and Exchange Commission (SEC) on April 2, 2013[1], which allows companies to use social media platforms, including Twitter, for broadcasting corporate announcements, has further bolstered Twitter's reputation as a medium for the distribute of financial market information. According to a survey conducted by the Reuters Institute in 2016[2], 51% of respondents use social media to get news each week, with 12% considering it their primary news source. Additionally, Refinitiv and Bloomberg, two leading news platforms for institutional investors, have integrated Twitter content into their systems to provide social media indicators to their users.

Existing literature has examined the impact of news media and social media on financial markets, revealing their role in forecasting stock market returns. These studies not only uncover the predictive power of both traditional news outlets and social media platforms, but also investigate the mechanisms underlying this predictive power. While some studies argue that the sentiment conveyed through the content of news and social media underpins their predictive ability, others attribute it to the fundamental information these channels provide. For instance, Tetlock (2007) find that the

---

[1] Securities and Exchange Commission, Securities and Exchange Act of 1934, Release No. 69279 / April 2, 2013; Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934: Netflix, Inc., and Reed Hastings.,Washington, D.C., April 2, 2013-The Securities and Exchange Commission today issued a report that makes clear that companies can use social media outlets like Facebook and Twitter to announce key information in compliance with Regulation Fair Disclosure (Regulation FD) so long as investors have been alerted about which social media will be used to disseminate such information.

[2] https://reutersinstitute.politics.ox.ac.uk/our-research/digital-news-report-2016

pessimism sentiment obtained from news media content predicts the stock market trend. Similarly, Tetlock, Saar-Tsechansky, and Macskassy (2008) show that firm-specific news pessimism provides novel information about firms' earnings and returns. Conversely, Gu and Kurov (2020) identify the source of Twitter's predictive power as stemming from fundamental information, While Sun, Najand, and Shen (2016), alongside Renault (2017), observe that the intraday predictive ability of news and social media is influenced by sentiment-driven noise trading. Collectively, these studies highlight the impact of both news and social media on stock market returns, demonstrating the various sources through which this influence is exerted.

Aside from the stock market, news and social media also play a crucial role in another important segment of the global markets, the crude oil market. Existing studies, including Li et al. (2021), Jiang et al. (2022), and Abdollahi (2023), have demonstrated the predictive power of qualitative and quantitative insights derived from crude oil news headlines, financial forums, and Twitter feeds on predicting crude oil prices and volatility. However, the debate on whether this predictive power stems from sentiment or fundamental information, a topic previously discussed in the stock market, remains largely unexplored in the crude oil context.

Stock markets and crude oil markets differ significantly in their participant composition and the types of information they require. The stock market is primarily influenced by institutional and retail investors who focus on information related to companies' earnings and earnings growth (Ohlson and Juettner-Nauroth, 2005). On the other hand, the crude oil market involves participants who produce, process, or handle physical commodities and use oil futures to hedge their risks. Aside from information on oil supply and demand, crude oil investors also value external factors such as partisan conflicts, geopolitics and government policy uncertainty (Su et al., 2020; Miao et al., 2017). These differences between the oil and the stock markets indicate a potential variation in the informational role of news and social media across these sectors, highlighting the necessity to investigate and understand the underlying drivers of such predictability for the crude oil market.

With traditional news platforms, most news articles are required to be verified by publishers before they are available to the public. However, the advent of the new social technologies has introduced complexities in information verification, with platforms like Twitter enabling the rapid spread of both accurate and false information. Misinformation can lead to significant market turbulence. For example, a false tweet claiming that Barack Obama was injured in an explosion wiped out $130 billion in US stock value[3]. Similarly in the Bitcoin market, a fake Bitcoin ETF approval tweet caused Bitcoin prices to immediately spike to $47,680 from the $46,800 level, and then fall to $45,400 as the tweets were found to be fake, resulting in $90 million in liquidations[4]. This characteristic of social media, leading to temporary market fluctuations as opposed to the more permanent price movements often associated with traditional news platforms, remains less explored in existing literature. Moreover, traditional news outlets like CNBC and Bloomberg, as well as regulators and exchanges, often disseminate information through their Twitter accounts, raising the question of whether social media provides predictive information beyond those offered by traditional news platforms, a topic yet to be discussed in existing literature. To address these literature gaps, our research focuses on two key research questions: (1) What drives the predictive power of news and social media for crude oil prices - sentiment or fundamentals, and is this predictive ability permanent or temporary? (2) Does social media provide additional information beyond what is already reflected in traditional news, or does it merely duplicate traditional news content?

In this study, we employ a unique dataset, the Thomson Reuters MarketPsych Indices (TRMI), to assess the intraday sentiment scores across news articles and Twitter posts

---

[3] Rapoza, "Can 'fake news' impact the stock market?" Forbes (2017). Available at: https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/#37eb20502fac

[4] Malwa, "Fake Bitcoin ETF Approval Tweet Causes $90M in Liquidations", Coindesk (2024). Available at: https://www.coindesk.com/markets/2024/01/10/fake-bitcoin-etf-approval-tweet-causes-90m-in-liquidations/

respectively. We measure crude oil market returns with WTI futures, yielding significant findings. First, we find that both news and social media sentiments can significantly predict crude oil returns, even after controlling for lagged returns, volatility, day-of-the-week effect, and market conditions. We then investigate whether such predictive ability comes from fundamental information causing a permanent price impact or short-lived noise that leads to a reversal afterwards. Despite social media's predictive power showing significant reversals, its influence remains positive in the long run, suggesting a mix of fundamental information and noise. In contrast, the consistent predictive ability of news media suggests that it primarily conveys fundamental information. Furthermore, we decompose the informational content of social media from that of news media by constructing an orthogonal social component. This unique component is derived from intraday sentiment scores of both platforms, allowing us to isolate the exclusive predictive insights offered by social media. Although the impact of the social orthogonal component is found to be smaller relative to news sentiment, it nonetheless provides significant, additional predictive insights. Collectively, our findings suggest that while social media information may echo fundamentals reflected in news content and contain noise, it also possesses a distinct informational component that contributes to its predictive capacity.

Second, given that crude oil prices are influenced by the dynamics of supply and demand, information on oil inventory changes, which are publicly reported on a weekly basis, represents a crucial information source in this market. We investigate the capacity of news media and social media to forecast information pertaining to these inventory changes. Both news and social media sentiment can forecast changes in American Petroleum Institute (API) and Energy Information Administration (EIA) oil inventories, with social media showing superior predictive ability. This demonstrates that both platforms offer insights on oil inventory changes, reinforcing their value in conveying fundamental information.

Third, we study how speculators, predominantly institutional investors, react to news and social media information in the WTI futures market. Our results reveal a significant disparity in responses to news media versus social media information. Specifically, adjustments in speculators' positions are primarily influenced by news media information, while social media sentiment appears to have minimal impact on their decision-making. This discrepancy further suggests that the information conveyed by news media and social media differs significantly.

Lastly, we assess the financial impacts of news and social media sentiment by constructing a market-timing strategy. This strategy significantly outperforms the benchmark buy-and-hold approach, delivering higher average returns and a superior Sharpe ratio. Thus, leveraging the predictive power of news and social media sentiment in the crude oil market offers investors the opportunity to achieve substantial profits, highlighting its practical importance.

This study makes several contributions to the literature on the informational role of news and social media in financial markets. First, we are the first to investigate the determinants of their predictive power in the crude oil market – a sector distinct from the stock markets, which have been the focus of most previous studies. Although several studies have verified the forecasting power of news and social media in the crude oil market, the specific drivers of this predictive power remain unexplored. Our research reveals that news media sentiment's forecasting power is driven by fundamental information, whereas social media sentiment influences returns through a combination of fundamental information, noise, and a distinct information component not present in news content. Furthermore, our research breaks new ground by investigating how futures market speculators respond to news and social media information, revealing novel evidence that speculators prioritize news media information over social media.

The rest of the paper is structured as follows. Section 2 reviews the related literature. Section 3 introduces the hypotheses of this study. Section 4 describes the data. Section

5 presents the empirical analysis, and Section 6 concludes.

2. Literature review

Our study mainly relates to the role of news and social media in financial markets, and the difference between stock market and crude oil market.

2.1. The role of news media and social media in financial markets

Before the explosion of social media, scholars conducted a detailed analysis of the role of traditional news media in financial markets. For instance, using daily content from a popular Wall Street Journal column, Tetlock (2007) find that high media pessimism predicts downward pressure on market prices followed by a reversion to fundamentals, indicating that news media pessimism is a proxy of investor sentiment in market level. Tetlock, Saar-Tsechansky, and Macskassy (2008) focus on firm-specific news in stock markets. They find that linguistic media content captures otherwise hard-to-quantify aspects of firms' fundamentals and has ability to predict returns. Fang and Peress (2009) study the relationship between mass media coverage and stocks returns. they find that stocks with higher media coverage earn less than stocks without media coverage.

With the rise in popularity of social media, there is a growing body of literature examining the role of social media in financial markets. Chen et al. (2014) investigate the extent to which investor sentiments predict future stock returns and earnings surprises by conducting textual analysis of articles posted on the Seeking Alpha portal and readers' comments in response to these articles. They find that the views expressed in both articles and commentaries predict future long-window stock returns and earnings surprises. Frino, Xu, and Zhou (2022) uncover bidirectional causality between realized volatility and Twitter information, along with causality flowing from implied idiosyncratic volatility to Twitter variables, suggesting that option traders possess an informational advantage over Twitter users. Bartov, Faurel, and Mohanram (2018) and Gu and Kurov (2020) show that twitter sentiments can predict stock returns and provide evidence that twitter content provides new information about analyst recommendations, analyst price targets and quarterly earnings. Sun, Najand, and Shen (2016) extends the predictive ability of news and social media into intraday level. they find that intraday S&P 500 index returns are predictable using lagged half-hour investor sentiment. They also find evidence that the return predictability is most likely driven by the trading activities of noise traders. Similarly, Renault (2017) document the intraday predictive ability of the microblogging platform StockTwits. They also provide direct empirical evidence of sentiment-driven noise trading at the intraday level. Bartov, Faurel, and Mohanram (2023) find that aggregate Twitter sentiment predicts upcoming announcement bond returns and changes in credit default swap spreads, and is associated with future changes in bond yield spreads and credit ratings, thereby providing economically important information to the bond market.

Recently, more and more studies begin to compare the role of news media and social media in financial markets. For example, Jiao, Veiga, and Walther (2020) study the impact of traditional news media coverage and social media coverage on stock volatility and turnover. They find traditional news media coverage negatively predicts subsequent volatility and turnover, while social media coverage positively predicts volatility and turnover. Their findings can be explained by model of “echo chambers”, that is to say, social networks may disseminate repetitive news, but some investors treat repeated signals as new information. Gan et al. (2020) studies the time varying relationship between news media and social media within stock market. Their results suggest that social media is becoming the dominant media source. Dong et al. (2022) analyze nearly a million stock-related news articles from Sina Finance and 12.7 million stock-related social media messages from Weibo. Their findings indicate that social media covers fewer stocks compared to mass media, especially as media volume increases. Additionally, they discovered differing short-term predictive capabilities between the two sources: while mass media outperforms social media in one-day predictions, the opposite holds true for a two-to five-day horizon.

As for crude oil market, it has also seen a surge in related studies. Li et al. (2021) use news related to the oil market to develop a news sentiment index, and show that shifts in news sentiment can predict returns and volatility. Jiang et al. (2022) scrape comments posted on the Eastmoney forum, an online financial forum, and highlight the predictive power of sentiment with regard to China's crude oil prices. Abdollahi (2023) construct a crude oil market Twitter sentiment index, and compare the impact of news media and social media on predicting volatility. In general, these studies all validate the predictive power of the media in the crude oil market. However, although these studies all refer to their quantitative measures of language as a sentiment index, they do not explore whether the predictive power of news media and social media is due to investor sentiment or underlying information about crude oil market. Therefore, whether news media and social media provide fundamental information about crude oil market that has not yet been absorbed into market prices is an open empirical question. In addition, prior related literature in crude oil market does not compare the information role of news media and social media. Whether information in news media and social media overlap is still unexplored.

2.2. The uniqueness of the crude oil market compared to the stock market

The participants in the crude oil market differ from those in the stock market, and their behaviors, as well as their impact on price dynamics, also exhibit distinctions. According to the Commitments of Traders (COT) reports released by the Commodity Futures Trading Commission (CFTC), the categories of traders, such as Producer/Break/Processor/User, Swap Dealers, and Managed Money, vary from those in the stock market. Literature studies the behavior of these participants and their effect on price dynamics. For example, Sanders, Boris, and Manfredo (2004) explore the relationship between trader positions and market prices in various futures markets such as crude oil, unleaded gasoline, heating oil, and natural gas. Their findings indicate a positive (negative) correlation between market returns and noncommercial (commercial) traders, while also revealing that traders' net positions generally do not precede market returns. Similarly, Dedi and Mandilaras (2022) investigate the association between trader positions in the futures market and the one-month futures price of Brent oil. Their research suggests that producers and swap dealers tend to decrease their net positions following a positive price shock, whereas money managers tend to increase theirs. However, there is less conclusive evidence regarding Brent's price response to changes in trader positions. In contract, the behavior of funds in the stock market has predictive implications for stock returns. For example, Wermers, Yao, and Zhao (2012) devise a stock return predictive measure based on an efficient aggregation of portfolio holdings from all actively managed U.S. domestic equity mutual funds, demonstrating the predictive capacity of fund positions on stock returns.

In addition to the types of traders, the crude oil market has many unique features in terms of price determinants. Essentially, crude oil is a commodity, differing from the stock. Factors affecting crude oil prices have been extensively studied in the literature in recent years. In general, the price of crude oil is determined by supply and demand in the long run, and also depends on the impact of inventories and financial markets in the short run (Lang and Auer, 2020). Su et al. (2020) studied the factors driving crude oil prices from the perspective of the United States. They find that partisan conflicts (PCI) and the dollar index (USDX) both had significant impacts on crude oil prices. Miao et al. (2017) highlight the impact of geopolitical factors on crude oil prices. In addition, Pástor and Veronesi (2012) investigate the impact of government policy uncertainty on crude oil prices and find that government uncertainty should lead to a decline in crude oil prices. In recent years, crude oil has begun to be treated as a financial asset, and the financial properties of crude oil have gradually increased. In this context, Juvenal and Petrella (2015) studied the impact of speculation on changes in crude oil prices, and they found that speculation exacerbated the boom and bust in crude oil price.

In recent years, the relationship between crude oil price and stock market has been widely studied. At the beginning, people paid attention to the relationship between crude oil price and aggregate stock returns. In general, most literatures found a negative correlation between the two (e.g., Ferson and Harvey, 1995; Jones and Kaul, 1996; Sadorsky, 1999), while some literatures found no significant relationship between the two (e.g. Huang, Masulis, and Stoll, 1996). The results are so complicated. Subsequent studies begin to explore this issue more deeply from different perspectives, such as whether there is an asymmetric impact of different crude oil shocks (e.g., Park and Ratti, 2008; Chen, 2010; Salisu and Isah, 2017), whether the relationship between crude oil and stocks changes over time (e.g., Miller and Ratti, 2009; Awartani and Maghyereh, 2013; Degiannakis, Filis, and Floros, 2013), and whether the relationship between original stocks is different in crude oil exporting countries and crude oil importing countries (e.g., Apergis and Miller, 2009; Wang, Wu, and Yang, 2013), providing more in-depth insights into the relationship between crude oil prices and stock markets. In general, the relationship between crude oil prices and stock prices is very complex, which indirectly indicates that the determinants of crude oil prices and stock prices are not the same.

Taken together, these studies point to major dissimilarities between the equity and crude oil markets, and highlight considerable differences in the information required by investors in the two markets. Given these differences, it is unclear what information news and social media sentiment provides in the crude oil market, although previous findings have shown that media contains underlying information that affects stocks.

3. Hypothesis

Our initial idea was to test the power of news media and social media sentiments to predict oil market returns at a daily frequency. In other words, we test the following hypothesis:

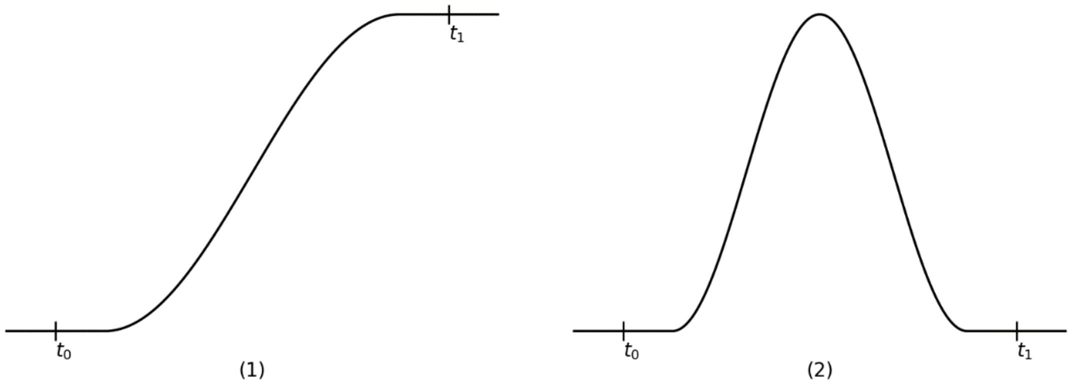
Hypothesis 1: News media sentiments and social media sentiments can predict returns of crude oil market.

If Hypothesis 1 holds, we also wonder the driving force of the predictive ability of news media and social media sentiments. Referring to previous studies (Tetlock, 2007; Gu and Kurov, 2020), we propose two hypotheses. One hypothesis is that news media and social media sentiments contain useful fundamental information that has yet to be incorporated into prices. In this scenario, the impact of news and social media sentiment should be permanent (as shown in Figure 1 A). Since the forecast is driven by genuine changes in market fundamentals, there is no reason for prices to reverse back to their previous levels. Another hypothesis is that news and social media sentiment simply represents the mood of uninformed noise traders. If this hypothesis is confirmed, the effect of news and social media sentiment should be reversed over the following periods, since arbitrageurs will step in to correct the anomaly (as shown in Figure 1 B). In general, the two hypotheses are as follows:

Hypothesis 2a: The predictive ability of news media sentiments and social media sentiments is driven by fundamental information contained in these media.

Hypothesis 2b: The predictive ability of news media sentiments and social media sentiments is driven by sentiment of noise traders conveyed by these media.

Figure 1.



If hypothesis 2a is true, there are several interesting questions. First, whether news media and social media contain the same predictive information? On the one hand, articles in traditional news media are usually completed by professional authors and need to be reviewed before they can be published, while in social media, everyone can express opinions based on their own judgment, which may lead to different information contained in news media and social media. On the other hand, most traditional news outlets have their own social media accounts. These news outlets disseminate news through both news media and social media. Sentiments in social media may just be a discussion of these news. These may lead to not much difference between predictive information in news media and social media. This question has not been explored in the oil market. But in the stock market, Gu and Kurov (2020) find that Twitter and the news media convey different predictions. Therefore, referring to Gu and Kurov (2020), we propose the following hypothesis:

Hypothesis 3: Predictive information in social media is incremental to that contained in traditional news media.

Second, what fundamentals are reflected in news and social media sentiments? According to the EIA, crude oil prices are primarily determined by supply, demand, inventories, and financial markets. In this article, we will only focus on the factors of the crude oil market itself, namely supply, demand and inventories. In these three factors, the impact of supply and demand is long-term, while the impact of inventory is short-term. Given that the data we used is at a daily frequency, the change in inventories may be more reflected in news media views and social media views. Therefore, we propose the fourth hypothesis:

Hypothesis 4: News media sentiment and social media sentiment can reflect changes in crude oil inventories.

Third, do speculators react differently to information in news media and social media? If Hypothesis 3 holds true, it implies a disparity in the information presented by social media and news media. Additionally, it's crucial to acknowledge the distinct characteristics of information found in these platforms. For instance, news media typically undergoes moderation, whereas social media content often reflects personal opinions and tends to be more fragmented. Consequently, we hypothesize that traders, particularly speculators, may exhibit varied reactions to information sourced from news media versus social media:

Hypothesis 5: Speculators react differently to information from the news media and social media.

At last, we also want to know the economic significance of return predictability. We propose the following hypothesis:

Hypothesis 6: A market-timing trading strategy based on news and social media sentiments can earn significant positive profits in crude oil market.

4. Data

4.1. Data source

We employ Thomson Reuters MarketPsych Indices (TRMI) to measure news and social media sentiment in the crude oil market. The TRMI offers a real-time assessment of news and social media sentiment index and updates at minutely, hourly, and daily intervals. According to TRMI, this index is defined as a measure of investor Sentiment. Investor sentiment is often viewed as the level of noise traders' beliefs relative to Bayesian beliefs (Tetlock, 2007). But in fact, the index measures the attitude of investors towards the crude oil market. This attitude may not be an irrational emotion, but a rational expectation based on underlying information.

The RMA assesses sentiment across three distinct content categories: news, social media, and the combined content. Notably, the news dataset of TRMI encompasses a wide range of influential sources, including Reuters news and a host of mainstream news sources collected by MarketPsych Data. Furthermore, the inclusion of Internet news content from LexisNexis starting from 2005 further enriches the news dataset. The collection of social media data is more diverse, starting from 1998 with some small Internet forums and expanding significantly in recent years due to the development of various social media platforms. LexisNexis social media content and Twitter were added in the dataset in 2008 and 2009, respectively. Using popularity ranks based on incoming links, the social media collection includes the top 20% of blogs, microblogs, and other social media content. These data allow us to compare differences between social media and news media.

For each asset tracked, the TRMI offers not only sentiment scores, but also media activity measure: *Buzz*. The option score is calculated as the difference between the proportion of positive references and the proportion of negative references concerning the underlying asset within a given time period. The option score, ranging from -1 to 1, reflects the level of market optimism or pessimism towards the underlying asset. A positive score indicates bullish expectations and investor optimism, while a negative score suggests bearish sentiments and pessimistic outlooks. A neutral sentiment is represented by a score of zero. As for media activity measure, *Buzz*, it counts the total number of words and phrases referring to the underlying asset in the collected news and social media sources. Therefore, higher Buzz implies more media coverage, more intense discussion and higher investor attention. Our data sample ranges from January 4, 2016 to May 31, 2022.[5]

We calculate daily returns based on WTI futures contract, sourced from LSEG Tick History. By taking the most actively traded future contract, the liquidity fluctuations caused by expiry cycles, are minimized.

[5] The relatively short sample period is attributed to the limited availability of RMA sentiment data. For comparison, Broadstock and Zhang (2019) utilized Twitter data to construct social media sentiment scores for the month of August 2018. Bollen, Mao, and Zeng (2011) analyzed data spanning a ten-month period in 2008.

Besides news media and social media data and WTI prices data, we also utilize traders position data, from Commitment of Traders (COT) reports released by the Commodity Futures Trading Commission (CFTC)[6]. Specifically, the COT report, issued weekly by the CFTC every Friday at 3:30 p.m. Eastern Time, offers a comprehensive overview of the collective positions held by various participants in the U.S. futures market. This report provides a snapshot of the commitments made by different classified trading groups as of the preceding Tuesday within the same week. In the COT report, reporting traders are divided into two main categories: commercials and noncommercials. Commercials typically represent entities engaged in cash-related businesses and are commonly referred to as hedgers. On the other hand, noncommercials, who do not have direct involvement in cash-related businesses, are often labeled as speculators. In this paper, we mainly focus on the net positions changes of speculators.

4.2. Data aggregation process

Daily update frequency refers to TRMI windows that end at 3:30 p.m. Eastern time. However, WTI futures trade Sunday to Friday from 6 p.m. to 5 p.m. Eastern time. To align with return data, we utilize minute-level news and social media data, and aggregate them to daily frequency. Specifically, following the guidelines provided in the User Guide of Refinitiv MarketPsych Analytics, we calculate sentiment scores over a daily window by taking Buzz-weighted averages of minutely sentiment data:

$$Opinion_t = \frac{\sum_{i=1}^{1440} Buzz_i * Opinion_{i,t}}{\sum_{i=1}^{1440} Buzz_i} \quad (1)$$

where $Buzz_i$ represents Buzz, a sum of entity-specific words and phrases, at minute i within the trading day t and $Opinion_{i,t}$ denotes the sentiment score at minute i within the same trading day t.

[6] https://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm

In TRMI, two asset codes are designated for the crude oil market: CRU (crude oil) and USCRU (US crude oil). As outlined in the User Guide of Refinitiv MarketPsych Analytics, these asset codes align with the topic codes found in news articles. Consequently, news articles and social media tweets associated with CRU and USCRU may both contain predictive insights for the crude oil market. As we mentioned earlier, TRMI provides a media activity measure, Buzz. Previous relevant literature uses this index to measure the intensity of discussion and media coverage of a particular topic (Jiao, Veiga, and Walther, 2020; Gan et al., 2020). A higher Buzz value indicates greater media coverage, discourse and discussion, implying the presence of more information. Therefore, in order to capture media information more comprehensively, we combine the sentiment scores of CRU and USCRU by comparing their Buzz values:

$$
\begin{aligned}
Opinion_{News,t} &= \begin{cases} Opinion_{News,t}^{CRU}, if \;\; Buzz_{News,t}^{CRU} \geq Buzz_{News,t}^{USCRU} \\ Opinion_{News,t}^{USCRU}, if \;\; Buzz_{News,t}^{CRU} < Buzz_{News,t}^{USCRU} \end{cases} \\
Opinion_{Social,t} &= \begin{cases} Opinion_{Social,t}^{CRU}, if \;\; Buzz_{Social,t}^{CRU} \geq Buzz_{Social,t}^{USCRU} \\ Opinion_{Social,t}^{USCRU}, if \;\; Buzz_{Social,t}^{CRU} < Buzz_{Social,t}^{USCRU} \end{cases}
\end{aligned}
\tag{2}
$$

4.3. Summary statistics

Table 1 reports summary statistics for our data set. It reveals that the average sentiment score for news media is -0.04, which is slightly negative. As a comparison, the average sentiment score for social media is 0.07, indicating that on average the content of social media tweets concerning the crude oil market is slightly positive. The difference in the mean values of sentiment scores between news media and social media implies that the two may contain different information. The mean daily WTI return is about 7 basis points, consistent with the general upward trend in the crude oil market during our sample period. The extreme values of WTI returns, significantly deviating from the sample mean, were observed in March and April 2020, indicating the heightened volatility experienced in the crude oil market during that period.

Table 1.

This table reports summary statistics for variables in our analysis. The *Sentiment* scores are constrained within the range of [-1, 1]. Within this scale, negative and positive values signify optimistic and pessimistic sentiment, respectively, while a score of zero indicates a neutral stance. The *Sentiment* index is acquired from TRMI. It aggregates the sentiment scores of CRU and USCRU according to Equation (2). *Returns* are the daily returns of WTI futures.

	Mean	Std	Max	Min	P25	Median	P75
News media sentiment	-0.04	0.08	0.26	-0.32	-0.10	-0.04	0.02
Social media sentiment	0.07	0.05	0.24	-0.23	0.05	0.07	0.10
WTI returns (%)	0.07	3.26	34.53	-48.08	-1.10	0.17	1.35

5. Empirical results

5.1. The forecasting power of news and social media sentiment

Following García (2013) and Obaid and Pukthuanthong (2022) who investigate the news sentiment in stock market, we conduct the regression below, with Newey and West (1987) adjusted t-statistics:

$$R_t = a + bOpinion_{t-1} + cR_{t-1} + dVolatility_{t-1} + eWeekday_t + fRecession_t + \varepsilon_t \qquad (3)$$

Where R_t denotes log returns of WTI futures on day *t*. $Opinion_{t-1}$ is the sentiment score of news media or social media on trading day *t*-1. $Volatility_{t-1}$ represents the volatility calculated from 5-minute returns on day *t*-1. In addition, we add the lag of WTI futures returns as control variable, since return autocorrelation in conjunction with contemporaneous correlation of returns and media sentiment can generate spurious evidence of lead-lag relation. $Weekday_t$ is a series of dummy variables for day-of-the-week for potential weekday anomalies. $Recession_t$ is a dummy variable for whether trading day t belongs to a recession. During our sample period, the only recession period was from February to April of 2020[7], when the crude oil market was experiencing severe turbulence. We include this dummy variable in Equation (1) to ensure the results

[7] https://www.nber.org/research/business-cycle-dating

are not driven by this special period.

Panel A of Table 2 presents the results of the regression analysis. Notably, both news media sentiment and social media sentiment display positive and statistically significant coefficients, as evident in Column 1 and Column 2, respectively. On average, a one standard deviation increase in news media (social media) sentiment score is followed by 25 (35) basic points increase in the crude oil market return. In addition, the coefficients of dummy variable for day-of-the-week are mostly significantly positive, suggesting abnormal negative returns on Monday, consistent with finding of Li et al. (2022). The significantly negative coefficient of dummy variable for recession is consistent with the large decline in crude oil prices during that period.

In sum, the analysis demonstrates that both news media sentiment and social media sentiment possess predictive capability for subsequent returns on a daily frequency. Therefore, Hypothesis 1 is supported by our results. These results are consistent with previous studies of Tetlock (2007) and Gu and Kurov (2020) that provide evidence that news sentiment and social media sentiment can predict daily returns in stock market.

Table 2.

This table reports results of the following regression:

$$R_t = a + bOpinion_{t-1} + cR_{t-1} + dVolatility_{t-1} + eWeekday_t + fRecession_t + \varepsilon_t$$

Where R_t is the returns of WTI futures on day t, and $Opinion_{t-1}$ is either news media sentiment or social media sentiment on day t-1. Control variables include the lag of returns and volatility, and dummy variables for day-of-the-week and recession period. The data sample spans from January 4, 2016 to May 31, 2022. Newey and West (1987) adjusted t-statistics are shown in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% level, respectively.

	News media	Social media
$Intercept$	-0.011***	-0.018***
	(-4.08)	(-4.17)
$Opinion_{t-1}$	**0.031*****	**0.070*****
	(3.05)	(3.04)
R_{t-1}	0.024	0.033
	(0.64)	(0.89)
$Volatility_{t-1}$	7.520***	7.813***
	(5.73)	(5.74)
$Tuesday_t$	0.002	0.004*
	(0.99)	(1.66)
$Wednesday_t$	0.006**	0.007**
	(2.11)	(2.31)
$Thursday_t$	0.004	0.005*
	(1.58)	(1.79)
$Friday_t$	0.006**	0.006**
	(2.36)	(2.57)
$Recession_t$	-0.040**	-0.042**
	(-2.08)	(-2.15)
R^2(%)	11.60	12.04

5.2. Do sentiments in news media and social media have a permanent effect on crude oil market?

In the previous section, we have provided evidence of the predictive ability of news and social media sentiment in crude oil market. However, whether the sentiment of the news media and social media within crude oil market represent the irrational investors sentiment or the fundamental information not absorbed by the price still needs further investigation.

Prior investigations into the origins of media's predictive ability, conducted across various media platforms, have yielded divergent conclusions (e.g., Tetlock, 2007; Bartov, Faurel, and Mohanram, 2018; Gu and Kurov, 2020). However, previous studies primarily concentrate on the stock market, leaving the crude oil market unexplored. Compared with these studies, our media sentiment indicators are constructed from a wide range of information sources, which makes our analysis more general and comprehensive. Therefore, in this section, we conduct a test to examine whether the positive effect of news and social media sentiment is permanent or temporary within the crude oil market. Before delving into these tests, it is important to acknowledge that the two hypotheses (Hypothesis 2a and Hypothesis 2b) frame media sentiment as either pure noise mood or pure fundamental information, whereas, in reality, media sentiment often comprises a blend of noise mood and fundamental information. Our forthcoming tests aim to determine which of these factors serves as the principal driving force behind the predictive ability of media sentiment.

To test the Hypothesis 2a and 2b, we include more lags of media sentiment into the predive regression, referring to García (2013) and Obaid and Pukthuanthong (2022):

$$R_t = a + \sum_{k=1}^{5} b_k Opinion_{t-k} + \sum_{k=1}^{5} c_k R_{t-k} + \sum_{k=1}^{5} d_k Volatility_{t-k} + eWeekday_t + fRecession_t + \varepsilon_t \tag{4}$$

We conduct this analysis based on the following considerations. If the predictive power of news and social media sentiment stems from the fundamental information within them, then these sentiments should have the capacity to forecast returns over longer periods. In other word, fundamental information within news and social media sentiment can lead to lasting price changes, as illustrated in Figure 1 A. In contrast, if the predictive power of news and social media sentiment is driven solely by the mood of noise traders, any temporary price changes will likely be reversed, as depicted in Figure 1 B.

The first two columns of Table 3 present the results of regression analysis (4), highlighting two key findings. Firstly, the predictive power of news media sentiment mainly comes from fundamental information. As shown in the first column of Table 3, the coefficients on t-1 and t-5 news media sentiment are significantly positive. Although the coefficients on t-2, t-3, and t-4 news media sentiment are negative, none of these coefficients are significant. In addition, their magnitudes are much smaller than the coefficients on t-1 and t-5 news media sentiment. More importantly, the chi-square test shows that the sum of the news sentiment coefficients between t-1 and t-5 is significantly distinguishable from zero, suggesting that the positive predictive power of t-1 news media sentiment has not disappeared over the remainder of the trading week. Second, the predictive power of social media may come from both investor sentiment and fundamental information. Similar to news media sentiment, the coefficient on $t-1$ social media sentiment is also significantly positive. The difference is that the coefficient on t-4 social media sentiment is significantly negative and its magnitude is just slightly smaller than the coefficient on t-1 social media sentiment. This means that the positive forecasting power of social media is reversed on day t+4, supporting the investor sentiment hypothesis. On the other hand, the chi-square test shows that the sum of the social media sentiment coefficients between t-1 and t-5 is significantly larger than zero, indicating that the reversal effect only weakens the positive predictive power of social media sentiment, but does not make it disappear. This supports the fundamental information hypothesis. This finding is different from that in stock market (e.g., Tetlock, 2007; Bartov, Faurel, and Mohanram, 2018; Gu and Kurov, 2020; Dong and Gil-Bazo, 2020), highlighting the distinctive characteristics of the crude oil market.

In summary, our results for news media support Hypothesis 2a, while the results for social media support both Hypothesis 2a and Hypothesis 2b. In other words, the predictive power of news media is mainly driven by fundamental information, while the predictive power of social media comes from both fundamental information and investor sentiment.

Table 3.

This table shows the result of the following regression:

$$R_t = a + \sum_{k=1}^{5} b_k Opinion_{t-k} + \sum_{k=1}^{5} c_k R_{t-k} + \sum_{k=1}^{5} d_k Volatility_{t-k} + eWeekday_t + fRecession_t + \varepsilon_t$$

Where R_t is the returns of WTI futures on day t, and $Opinion_{t-k}$ is either news media sentiment or social media sentiment on day t-k. Control variables include the lags of returns and volatility, and dummy variables for day-of-the-week and recession period. The data sample spans from January 4, 2016 to May 31, 2022. Newey and West (1987) adjusted t-statistics are shown in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% level, respectively.

	News media	Social media	Social media orthogonal
Intercept	-0.009***	-0.014***	-0.010
	(-3.03)	(-2.99)	(-3.00)
$Opinion_{t-1}$	**0.023****	**0.074*****	**0.064****
	(2.08)	(2.63)	(2.24)
$Opinion_{t-2}$	-0.007	-0.014	-0.007
	(-0.45)	(-0.51)	(-0.26)
$Opinion_{t-3}$	-0.013	0.014	0.027
	(-0.84)	(0.44)	(0.90)
$Opinion_{t-4}$	-0.010	**-0.062****	**-0.058****
	(-0.75)	(-2.21)	(-2.22)
$Opinion_{t-5}$	**0.031*****	0.038	0.013
	(2.73)	(1.47)	(0.55)
Controls	Yes	Yes	Yes
R^2(%)	14.55	15.03	14.85
Sum $Opinion_{t-1}$ to $Opinion_{t-5}$	0.025	0.051	0.040
χ^2	**3.34***	**5.93****	**3.28***

5.3. Does social media provide information incremental to news media?

In the previous section, we find that news media and social media are able to predict WTI futures returns for different reasons. This seems to imply that social media and news media provide different predictive information. In this section, we explore this issue a little further.

There are many articles giving different views on this issue. For example, Jiao, Veiga, and Walther (2020) study the impact of news media and social media on stock market volatility and trading volume, and find that the results were consistent with the "echo chambers" theory, that is, social networks repeat news, but some investors interpret repeated signals as truly new information. However, Sprenger et al. (2014) believe that in the news media only a limited number of authors can publish articles and they only report on major events. In social media, everyone can create and disseminate information at any time. Therefore, social media contains not only major news stories, but also minor news items. Gu and Kurov (2020) compare the predictive power of Twitter sentiment and news sentiment for stock returns and finds that Twitter provides information that is not contained in the news media. These studies all focus on the stock market, paying little attention to the crude oil market. As recounted in Section 2.2, the information that affects the return in the crude oil market is different from that in the stock market, which may lead to different conclusions obtained in the crude oil market than in the stock market. Another reason that motivates us to care about this issue is that quite a few news outlets have opened accounts on social media, which may cause TRMI to capture the same information in news media and social media. Therefore, it is necessary to investigate whether the information of news media and social media are exactly the same in the crude oil market.

To eliminate the influence of news media in social media sentiment, we firstly regress social media sentiment on news media sentiment. Then, we take the residuals of this regression as a proxy of the unique information of social media sentiment, denoted as social media orthogonal sentiment. Next, we evaluate the predictive ability of social media orthogonal sentiment through Equation (4). If the predictive ability of news media sentiment is subsumed by news media, we would expect that the residuals of social media sentiment, i.e., social media orthogonal sentiment, to have no predictive power for the price of WTI futures. Otherwise, we would expect it to remain a significant predictor of returns in the crude oil market.

The regression results are reported in the first two columns of Table 3, revealing several noteworthy insights. Firstly, the social media orthogonal sentiment exhibits performance similar to social media sentiment in predicting returns of the crude oil market. Specifically, the coefficient of t-1 (t-4) social media orthogonal sentiment is significantly positive (negative), similar to the results in the second column of Table 3. In addition, the chi-square test also shows that the sum of the social media orthogonal sentiment coefficients between t-1 and t-5 is significantly larger than zero. Notably, when comparing the results in the last two columns of Table 3, we observe that the size of coefficients and R^2 in the third column is just slightly smaller than that in the second column, underscoring that the predictive information in social media exhibits a low degree of overlap with that in traditional news media. In other words, the predictive abilities of social media and news media stem from different sources.

Overall, we provide evidence that fundamental information in social media is incremental to that contained in traditional news media, supporting Hypothesis 3.

5.4. Inventory announcements reflected in news media and social media

In section 5.2, we test Hypothesis 2a and Hypothesis 2b, and find that news media and social media contain some fundamental information that leads news media sentiment and social media sentiment to predict returns. In this section, we will explore what information related to crude oil may be contained in the news media and social media.

The price of crude oil is influenced by many factors. The EIA provides a very simple model that summarizes the main factors driving changes in crude oil prices. According to the EIA, crude oil prices are determined by demand and supply in the long run and by inventories and financial markets in the short run. Given that the frequency of the data we use is daily, news sentiment and social media sentiment may be more sensitive to short-term influencing factors. In this section, focusing on the crude oil market itself, we only consider whether information about inventory changes is included in the news media and social media. We will study the relationship between media and speculations in the next section.

There are two significant weekly reports in the United States that offer insights into the inventory levels of crude oil. One report, titled the "Weekly Statistical Bulletin," is released by the industry-supported API on Tuesdays at 4:30 pm EST. The other, known as the "Weekly Petroleum Status Report," is issued by the US Department of Energy's EIA agency on Wednesdays at 10:30 am EST. Both reports detail the inventory levels of crude oil as of the previous Friday. While the methodologies for these reports are similar, participation in the EIA survey is mandated by the government, whereas participation in the API survey is voluntary. Consequently, the EIA report is typically considered the primary market influencer.

To examine whether news media sentiment and social media sentiment contain information about inventories changes, we perform the following regressions:

$$ReportChange_t = a + bOpinion_{t-1} + \varepsilon_t, \tag{5}$$

Where $ReportChange_t$ is the changes of crude oil inventories (million barrels) reported by API or EIA on Tuesday and Wednesday, respectively. $Opinion_{t-1}$ refers to social media sentiment or news media sentiment on the day before reports are released.

Table 4 reports the results. In Panel A, when forecasting changes of crude oil inventories reported by API, the coefficient of news media sentiment and social media sentiment are significantly negative at the 10% and 1% level, respectively, indicating that both news media sentiment and social media have some predictive power for API reports. In Panel B, the results are similar to that in Panel A. Not only news media sentiment, but social media sentiment shows significant forecasting power on the EIA reports. The negative coefficient of news and social media sentiment is reasonable because the direction of change in oil stocks is usually opposite to the direction of change in oil prices, and media sentiment can positively predict oil price changes. In addition, by comparing the coefficients of news media sentiment and social media sentiment and the R-squares in Panel A and Panel B, we find that social media sentiment seems have a stronger predictive power for changes of crude oil inventories than news media sentiment.

In sum, our results support the Hypothesis 3. We provide evidence that news media sentiment and social media sentiment can reflect changes in crude oil inventories.

Table 4

This table reports results for the following regression:

$$ReportChange_t = a + bOpinion_{t-1} + \varepsilon_t,$$

where $ReportChange_t$ is the changes of crude oil inventories (million barrels) reported by API or EIA on Tuesday and Wednesday, respectively. $Opinion_{t-1}$ refers to social media sentiment or news media sentiment on the day before reports are released. Panel A reports results for API and Panel B reports results for EIA. The data sample spans from January 4, 2016 to May 31, 2022. Newey and West (1987) adjusted t-statistics are shown in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% level, respectively.

Panel A: API		
	News media	Social media
$Intercept$	-0.475	1.003*
	(-1.14)	(1.66)
$Opinion_t$	**-9.640***	**-17.834*****
	(-1.82)	**(-2.78)**
R^2	2.54	3.38
Panel B: EIA		
	News media	Social media
$Intercept$	-0.462	1.817**
	(-1.06)	(2.26)
$Opinion_t$	**-11.237****	**-25.080*****
	(-2.13)	**(-3.14)**
R^2	3.26	5.37

5.5. How do speculators respond to information in the news media and social media?

In the preceding section, we confirmed that both news media and social media contain information that can forecast inventory reports. Here, we delve into the connection between media information and another short-term factor influencing crude oil prices: financial markets.

Commodity markets have undergone notable structural changes over time. Prior to the early 2000s, commodity markets operated with some degree of segregation from financial markets (Bessembinder, 1992). However, post-2002, financial institutions began regarding commodities as a fresh asset class, strategically integrating them into their portfolios for diversification purposes (Li, 2018). These structural shifts have sparked interest in examining hedge fund activity, often termed speculative activity, and its impact on commodity markets (Kolodziej and Kaufmann, 2013; Awan, 2019; Dedi and Mandilaras, 2022). Nevertheless, there remains a lack of exploration into the specific information on which speculators rely to guide their actions.

Considering that we verified in section 5.3 that the relevant information of the crude oil market contained in social media is different from that in the news media, we wonder whether speculators react to the information in social media and the news media in the same way. To explore this issue, we first obtain weekly position changes from the CFTC for speculators. Then, we take the weekly average of social media sentiment and news media sentiment. Next, we study the relationship between speculators' position changes and news media and social media through the following regression:

$$PositionChange_{t,weekly} = a + bOpinion_{t,weekly} + cR_{t,weekly} + dVolatility_{t,weekly} + eRecession_{t,weekly} + \varepsilon_{t,weekly} \tag{6}$$

Where $PositionChange_{t,weekly}$ represents the change in the net long position of speculators (non-commercial traders) at week t compared to week t-1.

Table 5 reports the regression results. Comparing the first and second columns of Table 5, we find that speculators have completely different reactions to information in the news media and social media. Specifically, the coefficient of news media sentiment is 109,906.8 with a t-statistics of 3.99, which is significantly greater than 0 at the 1% level. This means that when the news media sentiment is positive, speculators will increase the net long position. In other words, speculators defer the future direction of crude oil prices movement, referring to the information published by the news media and speculate on the future direction of crude oil prices based on it. As a comparison, the coefficient of social media sentiment is less than 0 and not significant, which means that speculators rarely refer to information on social media when making trades.

This finding not only verifies Hypothesis 5, but also indirectly shows that news media and social media contain different information related to the crude oil market, supporting Hypothesis 3. In addition, this finding also has good practical value. Because we have verified the predictive power of social media, it may be beneficial for institutional investors to properly refer to the information in social media.

Table 5.

This table reports results of the following regression:

$$PositionChange_{t,weekly} = a + bOpinion_{t,weekly} + cR_{t,weekly} + dVolatility_{t,weekly} + eRecession_{t,weekly} + \varepsilon_{t,weekly}$$

Where $PositionChange_{t,weekly}$ represents the percent change of the net long position of speculators (non-commercial traders) at week t compared to week t-1. Control variables include the weekly returns and volatility, and dummy variable for recession period. The data sample spans from January 4, 2016 to May 31, 2022. Newey and West (1987) adjusted t-statistics are shown in parentheses. *, ** and *** denote significance at 10%, 5%, and 1% level, respectively.

	News media	Social media
$Intercept$	0.008	0.011
	(1.36)	(0.86)
$Opinion_{t,weekly}$	**0.242*****	-0.125
	(3.50)	(-0.88)
$R_{t,weekly}$	0.200	0.247
	(1.20)	(1.24)
$Volatility_{t,weekly}$	0.382***	0.338***
	(3.07)	(2.57)
$Recession_{t,weekly}$	0.041***	0.007
	(4.69)	(0.69)
R^2(%)	12.68	8.47

5.6. Economic value

The value of a predictor can be assessed by examining its performance in market timing. Referring to Obaid and Pukthuanthong (2022), we construct a market timing strategy based on news media sentiment and social media sentiment. To ensure that the returns on our trading strategy are not driven by bid-ask bounce or day-of-the-week effects, we first get residuals from recursively regressing news media sentiment or social media sentiment on lagged returns and volatility (five lags) and day-of-the-week dummies available at time t-1, denoted by $News^{\perp}$ and $Social^{\perp}$, respectively. Then we use these residuals as trading signals. Specifically, we initiate a long or short position in the market at the start of trading day t, based on the positivity or negativity of the $News^{\perp}$ and $Social^{\perp}$ obtained on trading day t-1. Subsequently, we close the position when the trading day t concludes. As a result, the payoff of the market timing strategy on the nth half-hour interval can be represented as:

$$y_t = \begin{cases} R_t, \ if \ News_{t-1}^{\perp} (or \ Social_{t-1}^{\perp}) \geq 0 \\ -R_t, \ if \ News_{t-1}^{\perp} (or \ Social_{t-1}^{\perp}) < 0 \end{cases}$$

To provide a point of comparison, we also consider a benchmark trading strategy known as the "buy-and-hold" strategy. This approach means simply taking a long position in the market at the beginning of the sample period and maintaining it until the end of the entire sample duration.

Table 6 presents the outcomes of the market-timing strategy, which relies on news and social media sentiment, alongside the results of the hold-on strategy for reference. The strategy based on news (social) media generates an annual return of 51.35% (31.16%), significantly surpassing the benchmark's 13.86%. The Sharpe ratio for the market timing strategy, based on news (social) media sentiment, stands at 0.06 (0.04), while the benchmark lags behind at 0.02. These results support Hypothesis 6, suggesting that the predictive ability of news media and social media can generate substantial economic value from the market timing perspective.

Table 6.

This table presents the performance of the market timing strategy. In the market-timing strategy, we initiate long or short positions at the start of each trading day based on the $News^{\perp}$ and $Social^{\perp}$ obtained last day. A positive $News^{\perp}$ or $Social^{\perp}$ triggers a long position, while a negative one prompts a short position, and we liquidate these positions by the end of each trading day. As a benchmark, we compare this approach to a buy-and-hold strategy, which establishes a position at the beginning of the sample period and retains it until the end. For both strategies, we provide four performance metrics, including the average return (Avg ret), standard deviation (Std dev), Sharpe ratio (SRatio), skewness (Skew) and Kurtosis (Kurt). The returns are annualized and expressed as percentages. Standard deviations are computed from daily returns generated by each strategy, also annualized and presented as percentages.

	Avg ret (%)	Std dev (%)	Sratio	Skew	Kurt
market timing based on	51.35	841.86	0.06	4.43	54.23
market timing based on	31.16	842.85	0.04	2.92	54.62
buy and hold	13.86	843.31	0.02	-1.59	55.04

6. Conclusion

As important channels for the dissemination of information, the impact of news media and social media on financial markets has attracted the attention of researchers. However, most of the related researches focus on the stock market. Related studies in crude oil market only investigate the predictive ability of news and social media. The informational role of news and social media in crude oil market is still not well understood.

By using intraday news and social media sentiment scores provided by TRMI, we find that both news media and social media sentiments can significantly and positively predict returns in the crude oil market. And this predictive ability has a different driving force in social media and news media. Specifically, the positive predictive ability of news media shows considerable persistence without reversals, which proves that its predictive ability comes from fundamental value-related information, rather than the subjective beliefs from noise investors. As for social media, its influence on returns originates from a mix of noise and fundamental information, leading to significant return reversals, yet it still sustains a positive impact over the long run. Furthermore, we discover that the fundamental information contained in media and social media does not overlap, suggesting that social media provides a distinct informational component that is not covered by news content. In addition, we examine the predictive ability of news and social media sentiment on oil inventory changes reported by API and EIA. We show that news and social media sentiments can predict inventory changes, reinforcing our statement that the predictive power of news and social media is driven by fundamental information. Moreover, we observe that speculators in the WTI futures market respond differently to news v.s. social media information. They only adjust their positions based on the information reflected in the news media. Notably, we develop a market timing trading strategy based on the sentiment scores extracted from news and social media. This strategy outperforms the buy-and-hold benchmark in terms of both average returns and Sharpe ratio.

Our research provides participants in the crude oil market enhanced understanding into the effectiveness of social media platforms as channels for information disclosures. More importantly, it highlights the necessity for regulators to continuously monitor social media platforms, given their varied informational components, to curb the spread of misinformation that could lead to unnecessary liquidations or impair investors' capacity to access capital efficiently.

References

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

Apergis, Nicholas, and Stephen M. Miller, 2009, Do structural oil-market shocks affect stock prices?, *Energy Economics* 31, 569–575.

Awan, Obaid A., 2019, Price discovery or noise: The role of arbitrage and speculation in explaining crude oil price behaviour, *Journal of Commodity Markets* 16, 100086.

Awartani, Basel, and Aktham Issa Maghyereh, 2013, Dynamic spillovers between oil and stock markets in the Gulf Cooperation Council Countries, *Energy Economics* 36, 28–42.

Bartov, Eli, Lucile Faurel, and Partha Mohanram, 2023, The Role of Social Media in the Corporate Bond Market: Evidence from Twitter, *Management Science* 69, 5638–5667.

Bartov, Eli, Lucile Faurel, and Partha S. Mohanram, 2018, Can Twitter Help Predict Firm-Level Earnings and Stock Returns?, *The Accounting Review* 93, 25–57.

Bessembinder, Hendrik, 1992, Systematic Risk, Hedging Pressure, and Risk Premiums in Futures Markets, *The Review of Financial Studies* 5, 637–667.

Bollen, Johan, Huina Mao, and Xiaojun Zeng, 2011, Twitter mood predicts the stock market, *Journal of Computational Science* 2, 1–8.

Broadstock, David C., and Dayong Zhang, 2019, Social-media and intraday stock returns: The pricing power of sentiment, *Finance Research Letters* 30, 116–123.

Chen, Hailiang, Prabuddha De, Yu (Jeffrey) Hu, and Byoung-Hyoun Hwang, 2014, Wisdom of Crowds: The Value of Stock Opinions Transmitted Through Social Media, *The Review of Financial Studies* 27, 1367–1403.

Chen, Shiu-Sheng, 2010, Do higher oil prices push the stock market into bear territory?, *Energy Economics* 32, 490–495.

Dedi, Valentina, and Alex Mandilaras, 2022, Trader positions and the price of oil in the futures market, *International Review of Economics & Finance* 82, 448–460.

Degiannakis, Stavros, George Filis, and Christos Floros, 2013, Oil and stock returns: Evidence from European industrial sector indices in a time-varying environment, *Journal of International Financial Markets, Institutions and Money* 26, 175–191.

Dong, Hang, and Javier Gil-Bazo, 2020, Sentiment stocks, *International Review of Financial Analysis* 72, 101573.

Dong, Hang, Jie Ren, Balaji Padmanabhan, and Jeffrey V. Nickerson, 2022, How are social and mass media different in relation to the stock market? A study on topic coverage and predictive value, *Information & Management* 59, 103588.

Fang, Lily, and Joel Peress, 2009, Media Coverage and the Cross-section of Stock Returns, *The Journal of Finance* 64, 2023–2052.

Ferson, WE, and CR Harvey, 1995, Predictability and Time-Varying Risk in World Equity Markets, Chen,-Andrew-H., ed. Research in finance, *Research in Finance* 13, 25–88.

Frino, A., Xu, C. and Zhou, Z.I., 2022. Are option traders more informed than Twitter users? A PVAR analysis. *Journal of Futures Markets*, 42(9), pp.1755-1771.

Gan, Baoqing, Vitali Alexeev, Ron Bird, and Danny Yeung, 2020, Sensitivity to sentiment: News vs social media, *International Review of Financial Analysis* 67, 101390.

García, Diego, 2013, Sentiment during Recessions, *The Journal of Finance* 68, 1267–1300.

Gu, Chen, and Alexander Kurov, 2020, Informational role of social media: Evidence from Twitter sentiment, *Journal of Banking & Finance* 121, 105969.

Huang, Roger D., Ronald W. Masulis, and Hans R. Stoll, 1996, Energy shocks and financial markets, *Journal of Futures Markets* 16, 1–27.

Jiang, Zhe, Lin Zhang, Lingling Zhang, and Bo Wen, 2022, Investor sentiment and machine learning: Predicting the price of China's crude oil futures market, *Energy* 247, 123471.

Jiao, Peiran, André Veiga, and Ansgar Walther, 2020, Social media, news media and the stock market, *Journal of Economic Behavior & Organization* 176, 63–90.

Jones, Charles M., and Gautam Kaul, 1996, Oil and the Stock Markets, *The Journal of Finance* 51, 463–491.

Kolodziej, Marek, and Robert K. Kaufmann, 2013, The role of trader positions in spot and futures prices for WTI, *Energy Economics* 40, 176–182.

Lang, Korbinian, and Benjamin R. Auer, 2020, The economic and financial properties of crude oil: A review, *The North American Journal of Economics and Finance* 52, 100914.

Li, Bingxin, 2018, Speculation, risk aversion, and risk premiums in the crude oil market, *Journal of Banking & Finance* 95. Commodity and Energy Markets, 64–81.

Li, Wenhui, Qi Zhu, Fenghua Wen, and Normaziah Mohd Nor, 2022, The evolution of day-of-the-week and the implications in crude oil market, *Energy Economics* 106, 105817.

Li, Yuze, Shangrong Jiang, Xuerong Li, and Shouyang Wang, 2021, The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach, *Energy Economics* 95, 105140.

Miao, Hong, Sanjay Ramchander, Tianyang Wang, and Dongxiao Yang, 2017, Influential factors in crude oil price forecasting, *Energy Economics* 68, 77–88.

Miller, J. Isaac, and Ronald A. Ratti, 2009, Crude oil and stock markets: Stability, instability, and bubbles, *Energy Economics* 31, 559–568.

Newey, Whitney K., and Kenneth D. West, 1987, A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

Obaid, Khaled, and Kuntara Pukthuanthong, 2022, A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news, *Journal of Financial Economics* 144, 273–297.

Ohlson, James A., and Beate E. Juettner-Nauroth, 2005, Expected EPS and EPS Growth as Determinantsof Value, *Review of Accounting Studies* 10, 349–365.

Park, Jungwook, and Ronald A. Ratti, 2008, Oil price shocks and stock markets in the U.S. and 13 European countries, *Energy Economics* 30, 2587–2608.

Pástor, Ľuboš, and Pietro Veronesi, 2012, Uncertainty about Government Policy and Stock Prices, *The Journal of Finance* 67, 1219–1264.

Renault, Thomas, 2017, Intraday online investor sentiment and return patterns in the U.S. stock market, *Journal of Banking & Finance* 84, 25–40.

Sadorsky, Perry, 1999, Oil price shocks and stock market activity, *Energy Economics* 21, 449–469.

Salisu, Afees A., and Kazeem O. Isah, 2017, Revisiting the oil price and stock market nexus: A nonlinear Panel ARDL approach, *Economic Modelling* 66, 258–271.

Sanders, Dwight R., Keith Boris, and Mark Manfredo, 2004, Hedgers, funds, and small speculators in the energy futures markets: an analysis of the CFTC's Commitments of Traders reports, *Energy Economics* 26, 425–445.

Sprenger, Timm O., Philipp G. Sandner, Andranik Tumasjan, and Isabell M. Welpe, 2014, News or Noise? Using Twitter to Identify and Understand Company-specific News Flow, *Journal of Business Finance & Accounting* 41, 791–830.

Su, Chi-Wei, Meng Qin, Ran Tao, Nicoleta-Claudia Moldovan, and Oana-Ramona

Lobonţ, 2020, Factors driving oil price —— from the perspective of United States, *Energy* 197, 117219.

Sun, Licheng, Mohammad Najand, and Jiancheng Shen, 2016, Stock return predictability and investor sentiment: A high-frequency perspective, *Journal of Banking & Finance* 73, 147–164.

Tetlock, Paul C., 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance* 62, 1139–1168.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More Than Words: Quantifying Language to Measure Firms' Fundamentals, *The Journal of Finance* 63, 1437–1467.

Wang, Yudong, Chongfeng Wu, and Li Yang, 2013, Oil price shocks and stock market activities: Evidence from oil-importing and oil-exporting countries, *Journal of Comparative Economics* 41, 1220–1239.

Wermers, Russ, Tong Yao, and Jane Zhao, 2012, Forecasting Stock Returns Through an Efficient Aggregation of Mutual Fund Holdings, *The Review of Financial Studies* 25, 3490–3529.