Research in International Business and Finance 64 (2023) 101881

Contents lists available at ScienceDirect

Research in International Business and Finance

journal homepage: www.elsevier.com/locate/ribaf

Full length article

Machine learning sentiment analysis, COVID-19 news and stock market reactions☆

Michele Costola[a,*], Oliver Hinz[b], Michael Nofer[b], Loriana Pelizzon[c,a]

[a] *Ca' Foscari University of Venice, Italy*
[b] *Goethe University Frankfurt, Germany*
[c] *Leibniz Institute for Financial Research SAFE, Frankfurt, Germany*

A R T I C L E   I N F O

JEL classification:
G10
G14
G15

Keywords:
COVID-19 news
Sentiment analysis
Stock markets

A B S T R A C T

The recent COVID-19 pandemic represents an unprecedented worldwide event to study the influence of related news on the financial markets, especially during the early stage of the pandemic when information on the new threat came rapidly and was complex for investors to process. In this paper, we investigate whether the flow of news on COVID-19 had an impact on forming market expectations. We analyze 203,886 online articles dealing with COVID-19 and published on three news platforms (MarketWatch.com, NYTimes.com, and Reuters.com) in the period from January to June 2020. Using machine learning techniques, we extract the news sentiment through a financial market-adapted BERT model that enables recognizing the context of each word in a given item. Our results show that there is a statistically significant and positive relationship between sentiment scores and S&P 500 market. Furthermore, we provide evidence that sentiment components and news categories on NYTimes.com were differently related to market returns.

1. Introduction

News can have an impact on the stock market either because it provides financial information related to asset values or simply because it is linked to investor psychology, as described by theoretical models of noise and liquidity traders (DeLong et al., 1987). The release of public news reduces information asymmetry, and the new flow of information is quickly absorbed in market prices (i.e., Marty et al., 2020).

Several studies provided empirical evidence of the relationship between market reactions and different types of news such as macroeconomic, environmental, corporate governance, and earning news (see, among others, Hamilton, 1995; Zhang et al., 2013; Huang, 2018; Caporin et al., 2019; Carlini et al., 2020). The ability to investigate the impact of news on stock prices has recently increased thanks to the use of natural language processing (NLP) in finance and economics (Xing et al., 2018; Wan et al., 2021). In times of market distress, the reaction of stock prices to news is even more pronounced, and the recent COVID-19 pandemic that started in 2020 is among the first instances of global, unforeseen financial turmoil triggered by a non-financial event that could be studied with NLP and sentiment analysis. Some financial economists have defined COVID-19 as an "exogenous shock" or even a

☆ The authors thank two anonymous Reviewers and the Guest Editor, Sabri Boubaker, for their insightful comments and suggestions. This research has received funding from the Volkswagen Foundation, Germany as part of the initiative "Artificial Intelligence and its Impact on Tomorrow's World". Loriana Pelizzon gratefully acknowledges research and financial support from the Leibniz Institute for Financial Research SAFE, Frankfurt. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: michele.costola@unive.it (M. Costola), hinz@wiwi.uni-frankfurt.de (O. Hinz), michael@nofer.net (M. Nofer), pelizzon@safe.uni-frankfurt.de (L. Pelizzon).

https://doi.org/10.1016/j.ribaf.2023.101881
Received 30 August 2021; Received in revised form 23 December 2022; Accepted 6 January 2023
Available online 16 January 2023

"black swan"; i.e., a rare event that has substantial influences on stock markets and could not reasonably have been predicted (e.g., Ahmad et al., 2021; Yarovaya et al., 2021).

An increasing amount of research has investigated the impact of COVID-19 on the stock market. For instance, Baker et al. (2020) show that no previous pandemics, like the Spanish flu and SARS, affected the financial markets as profoundly as COVID-19. The effect of the pandemic has been also documented in the Chinese stock market (Liu et al., 2022). Assessing public perception as proxied by internet users, Costola et al. (2020) and Smales (2021) provide evidence of the relationship between Google search volumes on the new coronavirus and financial markets. Gormsen and Koijen (2020) show that equity dividend futures represent a forward-looking measure to extract investor expectations about growth during the COVID-19 pandemic. Rebucci et al. (2022) analyze the market reaction to monetary policy interventions due to the COVID-19 pandemic crisis and show the effectiveness of quantitative easing. While the relationship between public measures of COVID-19 attention and financial markets has been demonstrated in previous research, evidence for the influence of news sources on the broader market is rather scarce.

A notable exception is represented by Akhtaruzzaman et al. (2022) and Huynh et al. (2021). Akhtaruzzaman et al. (2022) study the relationship between the COVID-19 media coverage index (MCI) and the European Market Union ESG leader index. The authors show that news play an important role in transmitting financial contagion during the pandemic.[1] Using principal component analysis, Huynh et al. (2021) propose the "feverish" sentiment built on six behavioral indicators such as media coverage, fake news, panic, sentiment, media hype, and infodemics retrieved by the RavenPack database. Findings show that the proposed sentiment is a meaningful predictor of stock volatility and return in the largest economies during the outbreak of COVID-19. Table 1 summarizes the main findings of the aforementioned empirical studies related to the COVID-19 pandemic and the financial markets.

In general, the prevailing view in the literature is that online content diffuses insights and that information extracted from financial news, online stock message boards, or social networks can predict stock market developments (e.g., Mitchell and Mulherin, 1994; Antweiler and Frank, 2004; Casarin and Squazzoni, 2013). For instance, Tetlock (2007) observes an influence of negative media sentiment as measured by *Wall Street Journal* articles on returns of the Dow Jones Industrial Average. Das and Chen (2007) study stock message boards and find a relationship between online sentiment about technology companies and their stocks. Xu et al. (2022) propose a news-based manager sentiment that is built on the tone of managers' news reports and find strong predicting ability in returns both in- and out-of-sample. Koch et al. (2022) study the impact of news sentiment in more than 34,000 news articles concerning Brexit and show the existence of limited spillover from news sentiment to equity markets. Nofer and Hinz (2015) show that weighted social mood levels on Twitter also have predictive value to stock market developments.

However, the recent pandemic represents an extremely rare event and thus provides previously unknown information to investors and the public. In the case of COVID-19, it is unclear whether asset prices are affected by online specialized financial and business news and how market participants consider these types of news sentiment. In this paper, we focus on COVID-19 related news in financial and business online media and study the influence of the correspondent sentiment on stock market developments. We consider the outbreak of the pandemic during the period from January until June 2020 and analyze 203,886 COVID-19-related articles published on three media platforms: MarketWatch.com, Reuters.com, and NYTimes.com. Given that all three news sources are based in the United States, we investigate whether we observe any statistically significant reaction in S&P 500 index returns, realized volatility, and changes in trading volumes after news on COVID-19 was published on the platforms.

We analyze three components of news sentiment: (i) the sentiment score, which reflects the positive and negative news sentiment on a particular day; (ii) the variance of this measure; and (iii) the number of COVID-19-related news items on that day. We thus focus on the early stages of the pandemic when COVID-19-related information emerged rapidly and was difficult to interpret for market participants. Although the pandemic is now more than two years old and various medical (e.g., vaccinations) and non-medical (e.g., lockdowns and travel limitations) interventions have been implemented by governments around the world, analyzing the long-term impact of COVID-19 on stock markets is out of this paper's scope. Instead, we aim to investigate whether sentiment analysis by machine learning models can help to predict the financial market reaction in case of a rare event that emerges suddenly, such as the contagious virus that causes the COVID-19 disease. For the sentiment analysis, we apply a financial market-adapted BERT model, that was recently developed by Google (Devlin et al., 2019).

Unlike previous models, BERT is able to recognize the context of each word due to its bidirectional architecture. However, whether BERT models can be applied to reveal COVID-19-related sentiment and thus predict broader stock market developments remains unclear. To the best of our knowledge, our paper is the first study that focuses on the outbreak of the pandemic applying a financial domain adapted BERT model for analyzing sentiment collected from major online news sources. Other non-BERT machine learning approaches have been applied in economics, energy economics, finance and time-series forecasting (see for a review Athey and Imbens, 2019; Ghoddusi et al., 2019; Masini et al., 2021; Ahmed et al., 2022). For instance, De Spiegeleer et al. (2018) apply Gaussian process regression in quant finance problems such as curve fitting, derivative pricing and hedging. Bianchi et al. (2021) show strong predictability of bonds using extreme trees and neural networks. We extend these previous approaches in the literature by applying the BERT model for stock market sentiment analysis.

Our study makes a threefold contribution to the literature. First, we show that there is a statistically significant and positive relationship between sentiment scores and market returns. This indicates that an increase (decrease) in the sentiment score implies a rise in positive (negative) news and corresponds to positive (negative) market returns. We also find that the variance of the sentiment and the volume of the news sources for Reuters and MarketWatch are negatively associated with market returns, indicating that

[1] Furthermore, the presence of financial contagion can affect stock market reactions (see for instance Akhtaruzzaman et al., 2021).

Table 1
A chronological list of the literature on COVID-19 and financial markets.

Authors	Study	Main data source	Empirical strategy	Main finding
Baker et al. (2020)	The Unprecedented Stock Market Reaction to COVID-19	Newspapers	Text-based methods developed in Baker et al. (2018)	No previous pandemic affected the financial markets as COVID-19
Chiah and Zhong (2020)	Trading from home: The impact of COVID-19 on trading volume around the world	Trading volume of international stock markets	Turnover measures for trading activities	Intensified share trading activities during the COVID-19 pandemic
Costola et al. (2020)	Google search volumes and the financial markets during the COVID-19 outbreak	Google trends data on COVID-19	Econometric approach	GT-COVID-19 data are related to the financial markets returns
Gormsen and Koijen (2020)	Coronavirus: Impact on Stock Prices and Growth Expectations	Data from stock and dividend futures markets	Econometric approach	Dividend futures are forward-looking measures for the economy
Haroon and Rizvi (2020)	COVID-19: Media coverage and financial markets behavior—A sectoral inquiry	Panic , global, sentiment and media coverage indexes from Ravenpack	Univariate GARCH type models	COVID-19 affected the media sentiment shock that caused higher volatility in the financial market
Salisu and Vo (2020)	Predicting stock returns in the presence of COVID-19 pandemic: The role of health news	Google trends data on COVID-19 and data on COVID-19 disease	Econometric approach	Health news combined with financial improve stock return forecast
Ahmad et al. (2021)	Black swan events and COVID-19 outbreak: Sector level evidence from the US, UK, and European stock markets	Data on COVID-19 disease and data on stock prices	Structural break models and factor-augmented event study methodology	The outbreak of COVID-19 as a black swan event in March 2020
Akhtaruzzaman et al. (2021)	Financial contagion during COVID-19 crisis	Data on COVID-19 disease and data on stock prices	Multivariate GARCH models and dynamic connectedness	Financial firms are more transmitters of contagion than nonfinancial ones
Biktimirov et al. (2021)	Sentiment and hype of business media topics and stock market returns during the COVID-19 pandemic	Wall Street Journal articles of the printed edition in 2020	Text-based methods developed in Hu and Liu (2004)	The breadth and intensity of coverage are significantly related to the S&P 500 returns
Huynh et al. (2021)	Feverish sentiment and global equity markets during the COVID-19 pandemic	RavenPack database	Feverish sentiment using PCA, dynamic connectedness	Feverish sentiment positively (negatively) predicts the stock volatility (return) during the COVID-19 pandemic
Le et al. (2021)	Different firm responses to the COVID-19 pandemic shocks: machine-learning evidence on the Vietnamese labor market	Survey on 16,000 firms belonging to 82 industries in Ho Chi Minh City	ML approaches such Logistic regression, decision tree, random forest and boost classifiers	COVID-19 affected the labor market heterogeneously across industries and areas
Rebucci et al. (2022)	An Event Study of COVID-19 Central Bank Quantitative Easing in Advanced and Emerging Economies	30 QE announcements made by 21 central banks	Event study	QE remains an effective measures on bond yields
Smales (2021)	Investor attention and global market returns during the COVID-19 crisis	Google trends data on COVID-19	Econometric approach	GT-COVID-19 data are associated with stock markets across G7 and G20 countries
Yarovaya et al. (2021)	The effects of a “black swan” event (COVID-19) on herding behavior in cryptocurrency markets	Hourly cryptocurrency data	Cross-sectional absolute deviation with an econometric approach	COVID-19 did not amplify herding in cryptocurrency markets
Akhtaruzzaman et al. (2022)	COVID–19 media coverage and ESG leader indices	Ravenpack Media Coverage index and ESG leader indices	Dynamic connectedness	MCI transmits financial contagion during the COVID–19 especially on March and April 2020
Dey et al. (2022)	Impacts of COVID-19 local spread and Google search trend on the US stock market	Google trends data on COVID-19 and data on COVID-19 disease	Econometric and random forest approaches	COVID-19 cases and deaths and GT-COVID-19 impacted abnormal stock prices in the period January–May 2020

Table 2
Summary statistics on articles collected from MarketWatch.com, NYTimes.com, and Reuters.com.

Source	Overall	COVID-19 articles	Average words	Max words	Method
MarketWatch.com	65,336	588	706	3857	Crawling
NYTimes.com	43,172	1710	381	5859	API
Reuters.com	95,378	4801	461	4607	Crawling

Note: Source refers to the corresponding media source, Overall indicates the total number of collected articles, COVID-19 articles indicates coronavirus-related topics, and Average words and Max words the average and maximum number of words per article, respectively. Method refers to the method employed to obtain the articles in a given source.

an increase in the uncertainty of the sentiment and an increase in the arrival of news have adverse impacts on the stock market. NYTimes.com is the only news source that provides a statistically significant result with realized volatility.

Our findings show that high sentiment is associated with low volatility in the market. The NYTimes sentiment volume is also statistically significant and negatively associated with market volatility. Because the NYTimes is a generalist publication, this link might simply reflect a different timing represented by COVID-19. In fact, at the start of the pandemic, the market responded promptly and negatively due to a shift in investor expectations, while the number of news items on the issue inevitably rose as time passed.

Regarding the changes in the S&P 500 trading volume, we find that sentiments and control variables have almost null explanatory power. Also, in this case, NYTimes represents the only news source that provides a statistically significant result that is negatively associated with positive variation of the volumes. This might indicate that an increase of positive COVID-19 news could have mitigated the number of trades with respect to the previous trading day. Second, we further analyze the specific nature of the NYTimes news. By disentangling the positive, negative, and neutral sentiment components, we show that a reduction in bad news eventually leads to a statistically significant impact on financial returns, but that is not true for an increase in good news.

Finally, we focus on the type of news released by the NYTimes and show that the business news category represents the main sentiment driver in explaining stock market returns. This result confirms that the information flow that contributes to forming market expectations depends on business news, even during an unprecedented event like the outbreak of COVID-19. The other news category that provides a statistically significant outcome is science. Given that this category is important in science dissemination, it is expected that articles responding to the COVID-19 pandemic may have influenced public opinion on the disease. The results hold after controlling for a set of control variables such as the volatility index (VIX), which is a forward-looking estimate of future stock market volatility, the OFR Financial Stress Index as a measure of the stress in the global market, the growth rate for the COVID-19 cases which measures the pandemic trend, and worldwide Google web searches for the coronavirus topic as a proxy for public attention.

The paper is organized as follows. Section 2 describes COVID-19 news sentiment construction through the data collection and chosen methodology. We also present previous research on sentiment analysis, especially with regard to recently developed machine learning models such as BERT. Section 3 investigates the relationship between the extracted COVID-19 news sentiment and the financial markets. We study both the aggregate view and a fine-grained level that considers positive and negative sentiment values and the different news categories. Finally, Section 4 concludes the paper with a discussion and suggestion for future research projects.

2. COVID-19 news sentiment construction

In this section, we present the data collection procedure, related sources and the implemented methodology to extract COVID-19 news sentiment indicators.

2.1. Data collection

Overall, we collected 203,886 online articles that were published on the three media platforms between 23 January 2020 and 22 June 2020. Table 2 presents summary statistics of the collected articles. Reuters.com and NYtimes.com are the websites of the respective international news companies owned by Thomson Reuters and New York Times Company. The topics covered include business, politics, financial markets, science and health. In addition, we collected data from MarketWatch.com, which focuses solely on financial news and stock market data. On average, the MarketWatch articles contain the most words (706) and the lowest maximum word count in a single article (3857).

The data collection process consisted of three steps. First, we gathered the URLs of online articles through an application programming interface (API) for NYTimes and web crawling for Reuters and MarketWatch. The web crawlers were developed using a link extractor written in Python Scrapy. The main goal of web scraping is to extract structured data from unstructured web pages. Scrapy contains the Spider class, which can be used to define how to crawl and parse pages to extract items from a particular site (e.g., by specifying the links).

In addition, the Item class supports the creation of a container to collect the scraped data. First, we stored the metadata such as headline, author, published date, and URL in a database. Through the NYTimes API, we were also able to retrieve the category of each news item (e.g., business, health). We then searched for COVID-19 URLs by focusing on the relevant keywords, like "COVID" and "Corona". In the last step, we collected all text elements (p-tags) from the remaining URLs; that is, date, title, author, and text. Fig. 1 depicts the weekly number of articles collected from NYTimes, Reuters, and MarketWatch during the first six months of the pandemic.

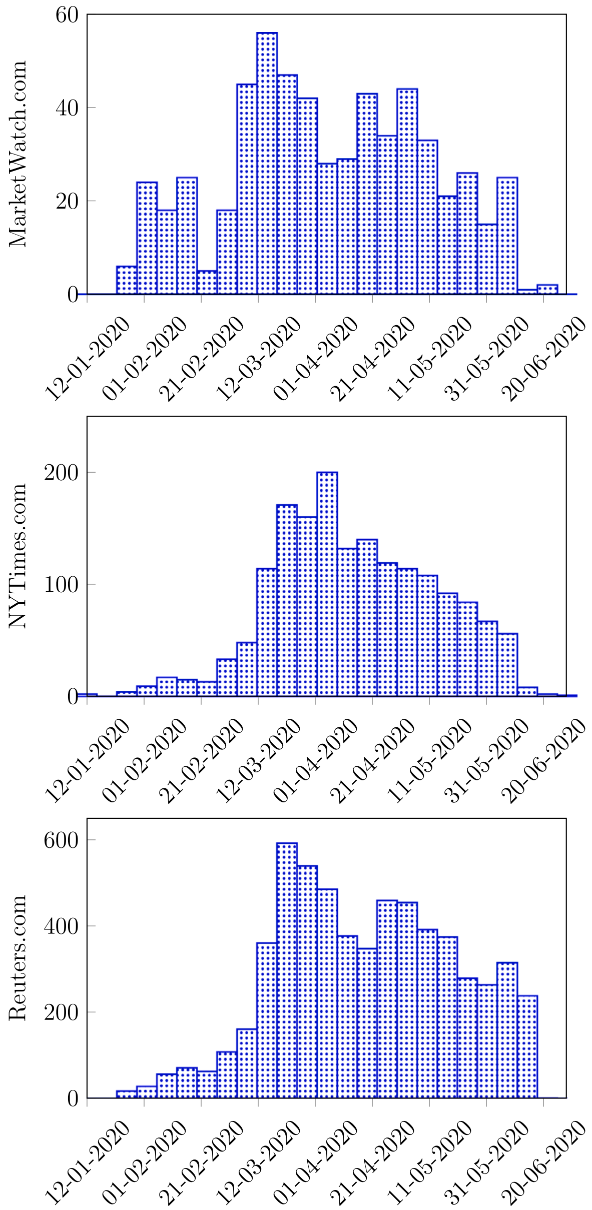

Fig. 1. Weekly number of articles collected from the three news sources over time: MarketWatch.com, NYTimes.com, and Reuters.com.

2.2. News sentiment analysis

Sentiment analysis starts with preprocessing; we describe each step displayed in Fig. 2. After the data collection, we tokenized all articles using the Natural Language Toolkit (NLTK), a leading platform for building Python programs to work with human language data. In general, tokenization means dividing a large quantity of text into smaller parts called tokens. First, the text is split into single sentences.

Machine learning models need numerical data to be trained and make predictions. Afterwards, each sentence was further split into words known to the model. The beginning and end of each sentence are marked with the [CLS] (''Classification'') and [SEP] (''Separation'') tokens after which the tokens are converted into IDs that are readable for the BERT model.

BERT is an open-source model that was pre-trained with millions of words from the entire Wikipedia corpus, employing a bidirectional transformer encoder. A transformer is a self-attention mechanism capable of understanding the context of sentences and making sense of the relationships between words. Attention mechanisms evolved from simple encoder–decoder frameworks such as Long Short-Term Models (LSTM) that create an output word from the input without taking the context of other close words into consideration. Thus, these pre-existing approaches were not able to focus on the most relevant word (e.g., Chen et al., 2015).

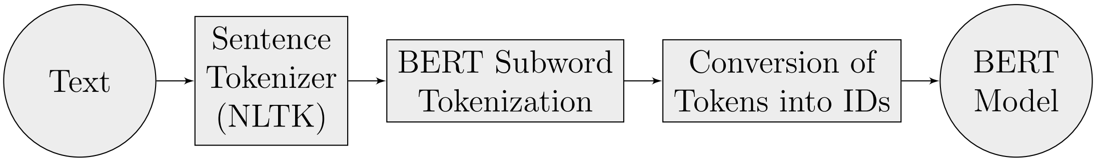

Fig. 2. The Preprocessing steps using the Natural Language Toolkit (NLTK).
Notes: Text refers to the raw text, which consists of all articles collected from NYTimes, Reuters, and MarketWatch; Sentence Tokenizer (NLTK) indicates the splitting of the text into single sentences using the NLTK; BERT Subword Tokenization refers to the splitting of sentences into words known to the model (or subwords and characters if the entire word is unknown); Conversion of Tokens into IDs indicates the creation of readable IDs; BERT Model is the final step, where BERT is able to train the model based on the preprocessing input.

They rather used two neural networks, the first one for encoding the source language and the second one for decoding the source sequence. These models have difficulties translating the entire sequence, which especially becomes apparent with long sentences.

In contrast, attention mechanisms repetitively consider small sub-parts of the sentence and assign weights and attention to the words. In this way, they focus on the context of surrounding words and thus better align the output to the input. The novelty of transformer models is self-attention, which means that the encoder does not only recognize the context of the directly surrounded words but also of all the other words in the sentence (e.g., Vaswani et al., 2017). In contrast to directional models, which only read the text step by step from one end to the other, the BERT transformer reads the entire input from both sides; it is thus bidirectional using the preceding and succeeding words for the prediction purposes.

During training, the model performs two tasks. First, BERT randomly masks a fraction of the words present and predicts the words that have been masked out. Furthermore, the model aims to predict whether the second sentence follows the first one based on a pair of sentences as input. In half of all cases, the second sentence is positioned directly after the first one. For the other half, random sentences of the corpus are used (Devlin et al., 2019). We can thus formalize the characteristics of the BERT model as follows:

$$\max_{\theta} \log p_{\theta}(\overline{\mathbf{x}} \mid \hat{\mathbf{x}}) \approx \sum_{t=1}^{T} m_t \log p_{\theta}\left(x_t \mid \hat{\mathbf{x}}\right) = \sum_{t=1}^{T} m_t \log \frac{\exp\left(H_{\theta}(\hat{\mathbf{x}})_t^{\top} e\left(x_t\right)\right)}{\sum_{x'} \exp\left(H_{\theta}(\hat{\mathbf{x}})_t^{\top} e\left(x'\right)\right)} \tag{1}$$

where x is the text sequence; $\overline{\mathrm{x}}$ are the masked tokens constructed from $\hat{\mathrm{x}}$, a corrupted version which results from a fraction of tokens in the text sequence x. $H_{\theta}(\mathbf{x}) = \left[H_{\theta}(\mathbf{x})_1, H_{\theta}(\mathbf{x})_2, \dots, H_{\theta}(\mathbf{x})_T\right]$ are hidden vectors mapped from x by the Transformer H_{θ}.

Thanks to this approach, BERT has outperformed state-of-the art NLP systems in various tasks. Two of these systems are highlighted by Devlin et al. (2019) due to their relatively similar approaches. GPT (Generative Pre-Trained Transformer) is a model created by OpenAI (Radford et al., 2018). Before the release of GPT, the most sophisticated NLP models typically relied on supervised learning, in which algorithms are trained with labeled data. Complex and large amounts of data are difficult to handle with this method since tagging the correct output can become expensive and time-consuming. GPT follows a semi-supervised approach that starts with an unsupervised pre-training stage using unlabeled text and is followed by supervised fine-tuning to learn a specific task. The transformer architecture enables the model to be highly context-aware. Therefore, GPT has been found to be superior to previous NLP benchmarks (Radford et al., 2018).

In addition to GPT, Devlin et al. (2019) also compare BERT's performance with ELMo ("Embeddings from Language Model"), a word embedding method proposed by the Allen Institute for Artificial Intelligence (Peters et al., 2018). ELMo is a deep contextualized word representation model capable of recognizing the dynamic nature of words. However, GPT and ELMo use unidirectional language models for pre-training; that is, the algorithm goes from left to right in a sentence or vice versa. By contrast, BERT is bidirectional and jointly considers the context on both sides of the words. Devlin et al. (2019) show that BERT outperforms GPT and ELMo in various tasks. For instance, BERT achieves an accuracy of 94.9% in the completion of the Stanford Sentiment Treebank (SST-2), a General Language Understanding Evaluation Benchmark, surpassing GPT (91.3%) and ELMo (90.4%) by several percentage points. Devlin et al. (2019) also investigate the F1 score, which combines the precision and recall of a model. The authors report BERT F1 scores of 72.1 for Quora Question Pairs and 89.3 for the Microsoft Research Paraphrase Corpus, outperforming both GPT and ELMo. Moreover, BERT achieves higher Spearman correlations than the other systems for the semantic textual similarity benchmark task.

The major reason for this outperformance is that BERT applies a bidirectional architecture while the other models are unidirectional, meaning that every token can only be directed to the context of the previous token. BERT is able to incorporate both sides, which opens up the possibility of recognizing more contexts. Various domain-adapted versions of BERT have been developed in recent years, and the finance area is of particular interest to our study. These FinBERT models have been applied to different financial tasks, including financial sentiment analysis (e.g., Araci, 2019; Liu et al., 2021), the prediction of stock returns (e.g., Sinha et al., 2022) and Forex price movements (Xing et al., 2020).

Previous research found superior accuracy and F1 Scores (Liu et al., 2021) of models adapted to financial tasks than in the original BERT model, indicating that continuous pre-training with financial texts is beneficial for sentiment analysis. While domain adaptation tends to increase accuracy, one major drawback of fine-tuning existing models are fluctuations that result from smaller datasets (e.g., Peng et al., 2021). Furthermore, in some domains such as biomedicine, it has been shown that pre-training from scratch might be even superior than continuously pre-training a general BERT model. Since previous research in the financial domain

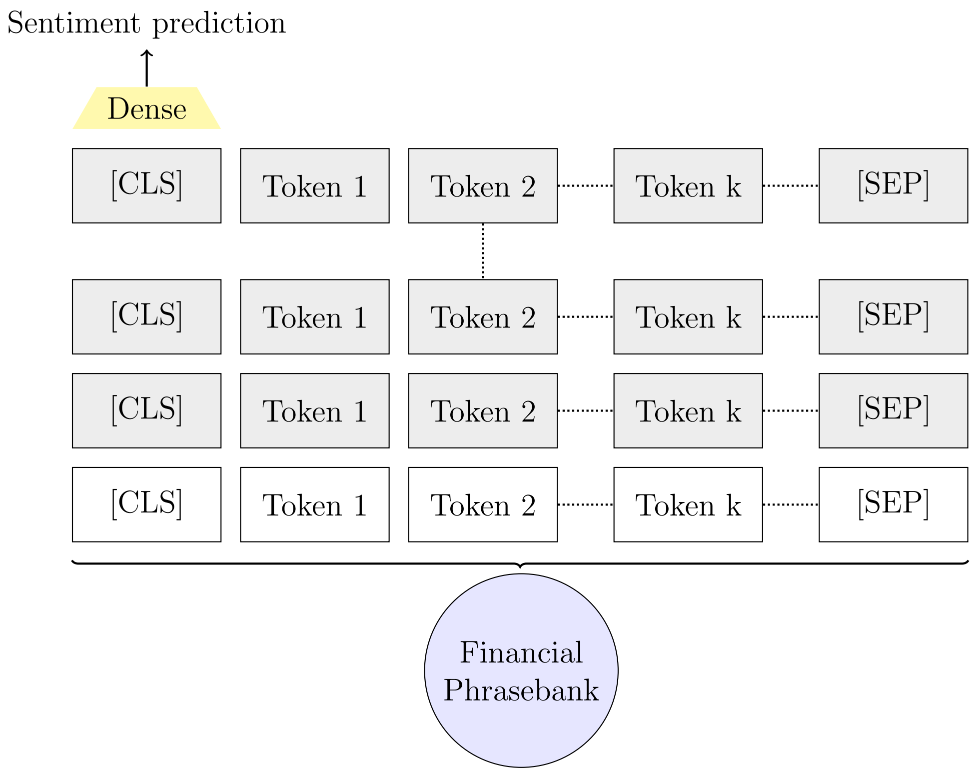

Fig. 3. FinBERT architecture developed by Araci (2019).
Notes: The Financial Phrasebank consists of 4840 sentences from financial news originally developed by Malo et al. (2014); [CLS] stands for Classification and represents the token at the beginning of each sequence, which contains one or two sentences; Token 1 to k refer to the tokens created by the model. Each token represents a word that is known to the model's vocabulary (or subwords and characters if the entire word is not known to the model); [SEP] stands for separation and represents the token at the end of each sequence; Dense refers to the dense layer, a neural network layer in which the final classification takes place; Sentiment prediction represents the output of the FinBERT model, which is a positive, negative, or neutral sentiment value.

indicates that the use of domain-adapted BERT models is appropriate for sentiment analysis, we follow this approach and apply the FinBERT model of Araci (2019) that has been widely used and discussed before (e.g., Koshiyama et al., 2021; Kumar and Sachdeva, 2020; Sinha et al., 2022). It should be noted that we do not use an extra-parametrization and refer to Araci (2019), who also reports other measures, like F1 values.

FinBERT relies on the same mathematical approach as BERT but implements additional computations to apply NLP tasks in the financial domain. The architecture of the FinBERT model is displayed in Fig. 3. As an extension to BERT, the model was further pre-trained on the Financial Phrasebank developed by Malo et al. (2014), which consists of 4840 sentences from financial news. Like the original BERT model, FinBERT consists of a multi-layer bidirectional transformer encoder. The sets of transformer encoders are located one above the other. At the beginning and end of the sequence there are two tokens named [CLS] and [SEP]. Each token represents a word that is known to the model's vocabulary. If an entire word is unknown, tokens are generated by subwords or characters. After the last [CLS] token, a dense layer (i.e., the neural network layer) performs the final classification, after which the sentiment prediction takes place. Araci (2019) shows that FinBERT outperforms the original BERT model in the financial domain regarding accuracy and F1 score. Various studies confirm FinBERT's eligibility for financial market research.

In the context of our study, research in the field of financial market prediction is especially relevant. Xing et al. (2020) use FinBERT to predict Forex price movements. Kabbani and Duman (2022) focus on share prices and include 10 stocks in the analysis, such as Apple, Microsoft, and IBM. Sinha et al. (2022) investigate FinBERT's eligibility for aggregate stock market prediction by studying the influence of sentiment on 900 Indian companies. Fazlija and Harder (2022) show that FinBERT can be used to predict the S&P 500 movement. The latter study is especially relevant since we investigate the influence of COVID-19 sentiment on the S&P 500.

The split into training, test, and validation set was 64%, 20% and 16% respectively. The model classifies all tokenized sentences as positive, negative, neutral, along with their respective probabilities (logits). We calculate sentiment as follows:

$$Sentiment = Logit_{Positive} - Logit_{negative} \tag{2}$$

Sentiment is the probability that the sentence is positive minus the probability that the sentence is negative. We therefore follow similar approaches in the literature (e.g., Andriotis et al., 2014). As an example, we provide the following token and illustrate the calculation of the sentiment:

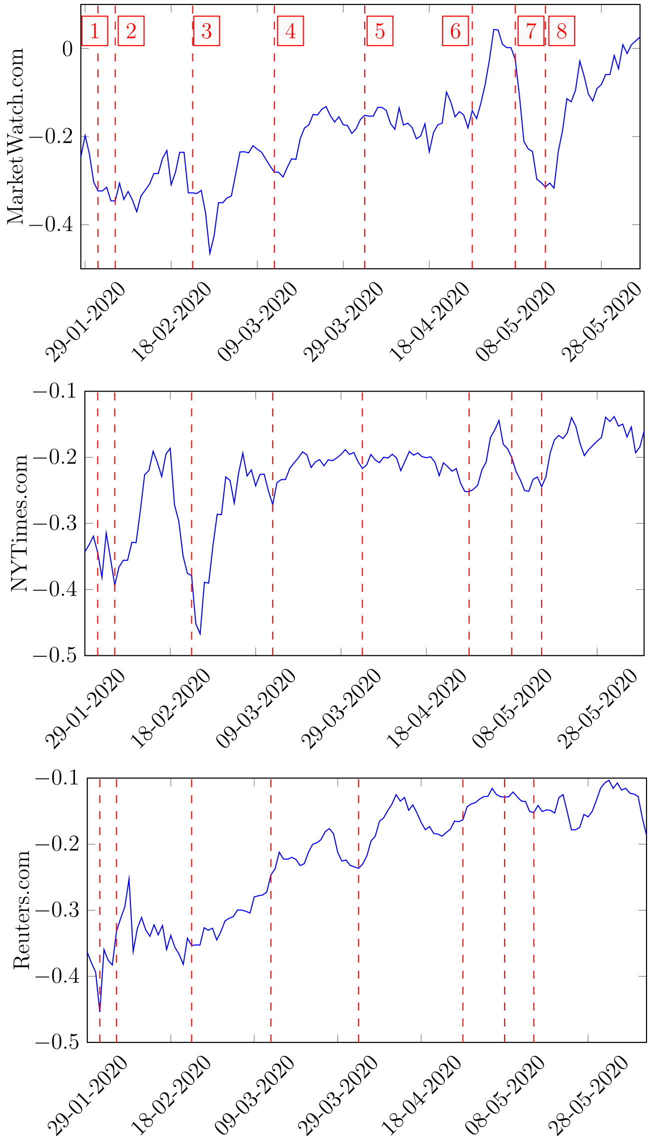

Fig. 4. Rolling seven-day average of the sentiment indicators over time according to the three news sources: MarketWatch.com, NYTimes.com, and Reuters.com. Dashed red lines indicate numbered significant events related to the COVID-19, as reported in Table 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

> The latest tranche of $60 million was raised a month ago from existing investors who wanted to provide more capital as they saw the business booming with the pandemic, said Gustavo Sapoznik, founder and chief executive of ASAPP.

The BERT model classifies this message into a positive value of 0.843, a negative value of 0.007841, and a neutral value of 0.1484. This results in a sentiment of 0.835, which represents a positive prediction of the model. We obtain the sentiment for an entire article by averaging the individual sentiment of its sentences. Fig. 4 depicts the rolling seven-day average of the COVID-19 sentiment extracted over the considered period. As expected, that COVID-19 sentiment was mostly negative across all platforms. However, we observe an upward trend during the course of the pandemic, suggesting that the negativity of the sentiment was greater in magnitude during the initial outbreak and has decreased over time. The figure indicates a sharp decline of the sentiment in late February, when the Western world started to realize that COVID-19 could be a global threat to the economy due to generalized lockdowns and travel limitations.

Table 3
Dating of significant COVID-19 events according to the Centers for Disease Control and Prevention (cdc.gov) and FRASER (fraser.stlouisfed.org). The episodes are reported in Fig. 4 (dashed red lines).

#	Date	Event
1	02 February 2020	The WHO declares the COVID-19 outbreak a public health emergency of international concern.
2	05 February 2020	The United States announces the first case of the new coronavirus in Wisconsin.
3	23 February 2020	Italy is the first Western country to introduce lockdown measures.
4	13 March 2020	President Trump declares COVID-19 disease a nationwide emergency.
5	03 April 2020	Second largest rise in US unemployment insurance claims in history.
6	28 April 2020	The United States becomes the first country with one million confirmed cases of COVID-19.
7	08 May 2020	The US Bureau of Labor Statistics reports historic levels of unemployment and job losses.
8	15 May 2020	Historic decline in US advance monthly retail and food service sales.

Table 4
Weekly correlations of sentiment scores based on the three media sources: MarketWatch, NYTimes, and Reuters.

	MarketWatch.com	NYTimes.com	Reuters.com
MarketWatch.com	1		
NYTimes.com	0.081	1	
Reuters.com	0.338	0.876	1

We have included in the figure red dashed lines that indicate significant eight milestones related to COVID-19 and its economic consequences, as described in Table 3. Specifically, we have identified a set of events that lies in the proximity of local minima or turning points within the three sentiments.[2]

The first event corresponds to the declaration of COVID-19 disease as a public health emergency of international concern (2 February 2020). That announcement has signaled that the severity of the disease was no longer limited to China and a few other countries. The second event is associated with the announcement of the first U.S. case in Wisconsin (5 February 2020). The third event is related to the introduction of lockdown measures by Italy (23 February 2020). Italy has been the first western nation to implement lockdown measures since World War II. This event helped create to form uncertainty in the markets, given the severity of the economic consequences of these policy measures, which were later applied by other countries. The fourth is President Trump declaring COVID-19 a nationwide emergency (13 March 2020). The fifth event represents the second largest rise in US unemployment insurance claims in US history (3 April 2020). This economic variable represents an early warning signal of economic recession since it by definition involves unemployment. The sixth event corresponds to the millionth US COVID-19 case (28 April 2020). The seventh event relates to the historic levels of unemployment and job losses reported by the US Bureau of Labor Statistics (8 May 2020). After the lockdown measures were implemented in the United States, about 20.5 million jobs were lost, and the unemployment rate reached 14.7%, the highest rate since the Great Depression. The eighth and final event was a historic decline in advance monthly retail and food service sales in the United States (15 May 2020). Once again, this economic indicator highlights the severity of the slowdown in economic activity.

It is worth recalling that the three sentiments are obtained from the cumulative news published each day on the corresponding platforms. Consequently, the different directions we observe regarding the impact of the listed events are attributable to the heterogeneity in the reported news as documented in the correlations included in Table 4.[3]

If we compare the weekly correlations of sentiments between platforms, as shown in Table 4, Reuters and NYTimes sentiments are most closely correlated. This might be explained by the nature of the websites. In contrast to MarketWatch, which focuses almost exclusively on financial news, NYTimes and Reuters cover a broader range of topics (politics, economics, science, etc.).

In addition to sentiment, we also include daily variance in sentiment and the volume of articles as additional measures of COVID-19 news sentiment. The variance in sentiment in the articles is the average of the squared differences from the mean value on day t, while the volume simply represents the number of COVID-19-related news articles on day t.

3. Stock market and COVID-19 news sentiment

In this section, we investigate the relationship between the extracted COVID-19 news sentiment and the financial markets. Given that all three news platforms report heavily on the United States, we focus on the S&P 500 market and consider stock index returns ($\mathrm{ret}_{S\&P500}$), the realized volatility of the stock index ($\mathrm{rv}_{S\&P500}$), and changes in the trading (log) volume ($\Delta \log Volume_{S\&P500}$).[4]

[2] We are indebted to an anonymous referee for the suggestion.

[3] Note that in order to have a direct correspondence between news and a considered sentiment, one should map the magnitude of the tokens belonging to each news that contribute to the logit components for that sentiment in a considered day. For sake of simplicity and conciseness, we have preferred to provide a homogeneous identification.

[4] Data for the stock index returns and trading volume were downloaded from Bloomberg. Data for the realized volatility (five-minute sub-sampled) were downloaded from the Oxford-Man Realized Library at the University of Oxford.

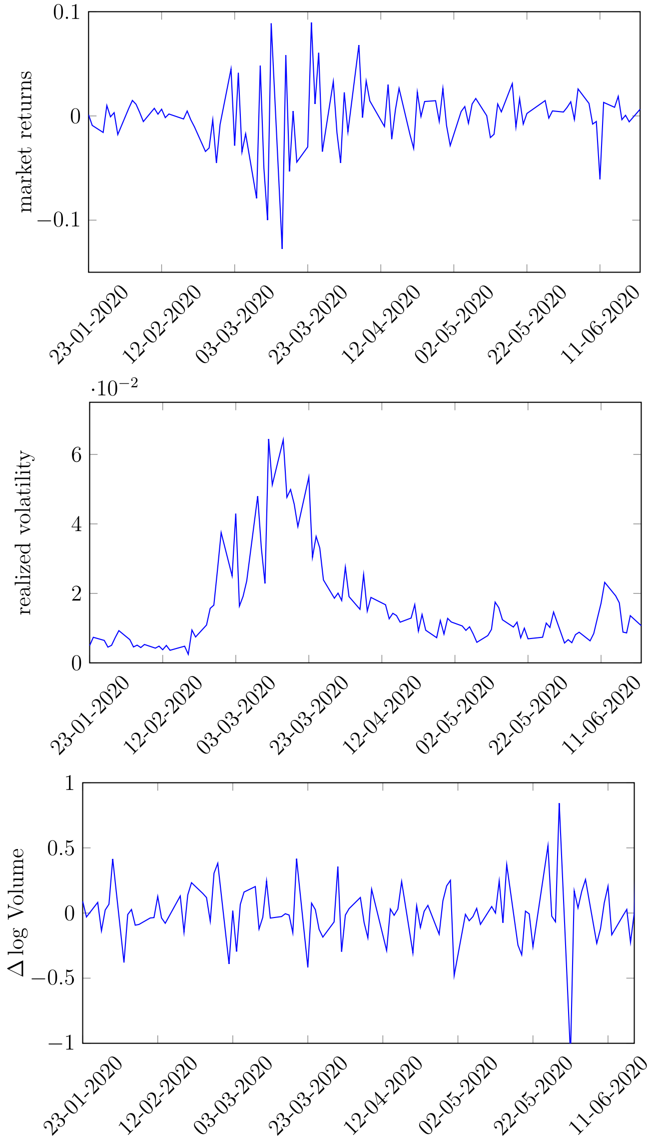

Fig. 5. Market returns, realized volatility, and $\Delta \log$ trading volumes of the S&P 500 over the period from 23 January 2020 to 22 June 2020.

The period considered covers the first stages of the COVID-19 pandemic: from 23 January 2020 to 22 June 2020 at a daily frequency. Fig. 5 shows the financial returns, daily realized volatility, and changes in the trading log-volume of the S&P 500 market index over time. In this respect, we consider three daily components for each COVID-19 news sentiment: (i) the sentiment (levels); (ii) the daily variance in sentiment in the articles from the three news sources (σ^2); and (iii) the volume of the articles for each news source (Volume). Furthermore, we add a set of control variables that could exert an impact on the stock market returns: (i) the VIX; (ii) the OFR Financial Stress Index,[5] which measures the stress in global financial markets; (iii) the growth rate in global COVID-19 cases (grCOVID-19)[6]; and (iv) Google searches matching the "coronavirus" topic on Google Trends to proxy for the general public attention as in Costola et al. (2020). Therefore, we estimate the following models:

$$\boldsymbol{y}_{S\&P500} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \epsilon, \tag{3}$$

where $\boldsymbol{y}_{S\&P500} = \left(ret_{S\&P500}, rv_{S\&P500}, \Delta \log Volume_{S\&P500}\right)$ is a $T \times 1$ vector, $\boldsymbol{X}$ is a $T \times K_1$ matrix for the sentiment variable(s), and $\boldsymbol{Z}$ is a $T \times K_2$ matrix for the control group previously defined.

[5] The index is proposed by the US Office of Financial Research and is available at https://www.financialresearch.gov/financial-stress-index/.

[6] Data were downloaded by Bloomberg using the mnemonic NCOVCNCA.

Table 5
Estimates of S&P 500 returns using COVID-19 news sentiments (MarketWatch, NYTimes, and Reuters), plus a set of control variables (the VIX, the OFR index, the growth rate in COVID-19 world cases, and Google searches for "coronavirus").

	(1)	(2)	(3)	(4)	(5)
Constant	0.0736***	0.0747***	0.0736***	0.0807***	0.0843***
	(0.0183)	(0.0183)	(0.0218)	(0.0198)	(0.0182)
MarketWatch	0.0373***			0.0387***	0.0385***
	(0.0108)			(0.0112)	(0.0113)
MarketWatch (σ^2)	−0.0273				0.0091
	(0.0566)				(0.0603)
MarketWatch (Volume)	−0.0018*				−0.0016
	(0.0011)				(0.001)
NYTimes		0.0599***		0.0434**	0.047**
		(0.0223)		(0.0165)	(0.0184)
NYTimes (σ^2)		0.0247			0.0403
		(0.0563)			(0.0759)
NYTimes (Volume)		0.0000			0.0000
		(0.0009)			(0.0009)
Reuters			0.0668***	0.042	0.0288
			(0.0233)	(0.0278)	(0.027)
Reuters (σ^2)			−0.0943		−0.1017*
			(0.066)		(0.0567)
Reuters (Volume)			0.0001		0.0002
			(0.0001)		(0.0001)
VIX	−0.0029***	−0.0031***	−0.0029***	−0.0031***	−0.0032***
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0007)
OFR	0.0045***	0.0039	0.0016	0.0017	0.0013
	(0.0016)	(0.0032)	(0.0018)	(0.0016)	(0.003)
grCOVID-19	−0.0091	−0.0051	−0.0124	0.0008	−0.0009
	(0.0163)	(0.0106)	(0.0157)	(0.0127)	(0.0132)
Google Trends	0.0011**	0.0011**	0.0013***	0.0015***	0.0016***
	(0.0005)	(0.0004)	(0.0004)	(0.0005)	(0.0004)
Observations	105	105	105	105	105
Adjusted R^2	0.2273	0.208	0.198	0.2556	0.2611
F-stat	5.3701	4.9019	4.6671	6.1006	3.8266
p-value	3.1964e−05	9.0607e−05	0.00015346	6.4611e−06	6.6351e−05

Notes: HAC standard errors are in parentheses; *, **, and *** denote statistical significance at 10%, 5%, and 1%, respectively.

Table 5 presents the estimates of the returns on the S&P 500. The estimates in Models 1, 2, and 3 contain the daily market sentiment, the variance in sentiment, the volume of articles for each news source, and the set of control variables as previously defined. All the sentiment indicators are statistically significant at a 1% level and positively related to the returns on the S&P500. This indicates that an increase (decrease) in sentiment implies a rise in positive (negative) news and corresponds to positive (negative) market returns. This highlights that there is a relationship between sentiment on COVID-19 and the stock market after controlling for the VIX (statistically significant and negatively related), the OFR Financial Stress Index (statistically significant and positively related only for MarketWatch), the growth rate of COVID-19 cases (not statistically significant for any of the three sources), and global Google searches for "coronavirus" (statistically significant and positively related for all three sources).

Regarding the variance and volume of the news sources, the only statistically significant variable is the volume for MarketWatch (at a 10% level), which is negatively associated with market returns, indicating that an increase in the arrival of news is mostly related to negative market sentiment during the considered period. If we consider the three sentiments jointly (Model 4), the results confirm the previous findings for MarketWatch and NYTimes at 1% and 10% significance level, respectively. Reuters is no longer statistically significant. Most likely, the correlation between Reuters and NYTimes leads to the result that the variable for NYTimes captures this effect alone. Model 5 includes all sentiments with their associated variances and volumes and shows that the results for sentiments remain unchanged with respect to Model 4.

Regarding the variance in the sentiment of the articles related to Reuters, we found that it is statistically significant at the 10% level and negatively related to market returns. This behavior is similar to the leverage effect between returns and volatility observed in asset pricing. As shown in Web Appendix A, the results for score, variance, and volume remain substantially unchanged if we consider them individually.

Table 6 shows the results for the realized volatility on the S&P 500 with the S&P 500 returns included as an additional control variable. NYTimes is the only news source that has a statistically significant result in the estimates. Model 2 shows that the NYTimes sentiment score is negatively related to the realized market volatility at the 5% level. This is an expected result due to the asymmetry in volatility, which is higher during turbulent periods and lower in tranquil ones.

Regarding the variance and volume of the news sources, we find that only the NYTimes volume is statistically significant at 10% level; it is negatively associated with market volatility. As the New York Times is a general media outlet, this relationship could simply reflect a different time scale impact of the exogenous shock represented by COVID-19. As a matter of fact, the market has immediately and negatively reacted after the outbreak of the pandemic due to changes in investors' expectations, while the

Table 6
Estimates of S&P 500 realized volatility using COVID-19 news sentiments (MarketWatch, NYTimes, and Reuters), plus a set of control variables (the returns on the S&P 500, the VIX, the OFR index, the growth rate in COVID-19 world cases, and Google searches for "coronavirus").

	(1)	(2)	(3)	(4)	(5)
Constant	−0.0204***	−0.0205***	−0.0207***	−0.023***	−0.0198***
	(0.0032)	(0.0038)	(0.0045)	(0.0041)	(0.0045)
MarketWatch	0.001			0.0013	0.0005
	(0.0034)			(0.0032)	(0.0037)
MarketWatch (σ^2)	−0.0129				−0.0152
	(0.0143)				(0.0138)
MarketWatch (Volume)	0.0001				0.0001
	(0.0003)				(0.0003)
NYTimes		−0.0095**		−0.0087**	−0.0122**
		(0.0042)		(0.0041)	(0.0051)
NYTimes (σ^2)		0.0164			0.0195
		(0.0126)			(0.014)
NYTimes (Volume)		−0.0003***			−0.0002***
		(0.0001)			(0.0000)
Reuters			0.0009	0.0055	0.0036
			(0.0073)	(0.0068)	(0.0062)
Reuters (σ^2)			0.0199		0.0185
			(0.016)		(0.0166)
Reuters (Volume)			0.0000		0.0000
			(0.0000)		(0.0000)
$\text{ret}_{S\&P500}$	−0.0233	−0.0115	−0.0174	−0.0151	−0.0092
	(0.0211)	(0.0231)	(0.0204)	(0.0244)	(0.0218)
VIX	0.001***	0.001***	0.001***	0.0011***	0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
OFR	−0.0027***	−0.0019***	−0.0024***	−0.0029***	−0.0017**
	(0.0005)	(0.0005)	(0.0006)	(0.0005)	(0.0007)
grCOVID-19	−0.0009	−0.0041	−0.0006	−0.0021	−0.0036
	(0.0032)	(0.0038)	(0.0035)	(0.0041)	(0.0042)
Google Trends	0.0002**	0.0002**	0.0002*	0.0002*	0.0002*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Observations	105	105	105	105	105
Adjusted R^2	0.8207	0.8397	0.8233	0.825	0.8381
F-stat	60.4995	69.115	61.556	62.306	39.4478
p-value	3.9152e−34	1.8988e−36	1.9713e−34	1.2184e−34	2.9176e−32

Notes: HAC standard errors are in parentheses; *, **, and *** denote statistical significance at 10%, 5%, and 1%, respectively.

volumes of news articles on the topic naturally increased in number as time went by. These results are confirmed when we consider the three sentiments jointly (Model 4) and the sentiments with their associated variances and volumes (Model 5). If we consider each sentiment individually (see Web Appendix A), none of the scores or variance provides statistically significant results. Interestingly, all article volumes for all three sources are statistically significantly and positively related to volatility. Given that these models are missing relevant market features, it could be that they implicitly capture the public interest in the COVID-19 pandemic (Costola et al., 2020).

Finally, Table 7 reports the estimates of changes in the log trading volumes of the S&P 500. First, the adjusted-R^2 is close to zero and negative in several cases, indicating a null explanatory power of the covariates. This is also confirmed in the F-test, where none of the estimated models rejects the null hypothesis that the coefficients are jointly equal to zero. During the pandemic, Chiah and Zhong (2020) provided evidence of strengthening trading activities in the financial markets, especially those markets with better corporate and sovereign governance. In our case, we found that the *New York Times* is a news source that provides a statistically significant result in the estimates. The sentiment score is statistically significant at the 10% level and negatively associated with positive variations in volumes. This could indicate that over the considered period an increase (decrease) of positive (negative) COVID-19 news mitigates the number of trades with respect to the previous trading day.

Interestingly, the variance in the sentiment of the articles from the *New York Times* is statistically significant at the 5% level and negatively related to changes in the log trading volumes of the S&P 500. This suggests that an increase in the variance of the sentiment news on a given day is associated with a decrease in trading volume with respect to the previous trading day.

The result remains unchanged if we consider the volume individually, as reported in Web Appendix A. Once again, these findings are confirmed when the three sentiments are jointly considered in Model 4 and with their associated variances and volumes in Model 5.

To account for the common variation among the three news sources, we have estimated the model using principal component analysis (PCA) and have included in the regression the first principal component on the sentiment scores, variances, and article volumes. The results are presented in Web Appendix B and confirm a statistically significant and positive relationship between the first component on the sentiment scores and market returns. We do not find such evidence on realized volatility and changes in the log trading volumes of the S&P 500. The only statistically significant result is provided by the first component on sentiment volume for realized volatility.

Table 7
Estimates of $\Delta \log$ trading volumes for the S&P 500 using COVID-19 news sentiments (MarketWatch, NYTimes, and Reuters), plus a set of control variables (the VIX, the OFR index, the growth rate in COVID-19 world cases, and Google searches for "coronavirus").

	(1)	(2)	(3)	(4)	(5)
const	0.0512	0.005	−0.131	−0.1091	−0.0749
	(0.0963)	(0.1244)	(0.1226)	(0.1271)	(0.1545)
MarketWatch	−0.1225			−0.0986	−0.1018
	(0.2958)			(0.2903)	(0.2943)
MarketWatch (σ^2)	−0.3115				−0.6323
	(0.5889)				(0.5472)
MarketWatch (Volume)	−0.0052				−0.0024
	(0.0063)				(0.0068)
NYTimes		−0.357*		−0.331*	−0.3532*
		(0.2088)		(0.1947)	(0.21)
NYTimes (σ^2)		−1.1744**			−1.5859***
		(0.5153)			(0.5463)
NYTimes (Volume)		−0.0023			−0.0037
		(0.0043)			(0.0037)
Reuters			−0.1049	0.0468	−0.0539
			(0.2049)	(0.1767)	(0.1886)
Reuters (σ^2)			−0.0333		0.122
			(0.4874)		(0.4501)
NYTimes (Volume)			0.0021		0.0027
			(0.0018)		(0.002)
$\text{ret}_{S\&P500}$	−1.1694	−0.6269	−1.0574*	−0.5823	−0.7009
	(0.7119)	(0.5577)	(0.5766)	(0.7549)	(0.7284)
VIX	−0.0013	0.0007	0.0023	0.0031	0.002
	(0.0039)	(0.0045)	(0.0044)	(0.005)	(0.0056)
OFR	−0.0066	−0.0012	−0.0293	−0.0123	−0.0108
	(0.0091)	(0.0146)	(0.0216)	(0.0102)	(0.0198)
grCOVID-19	−0.2618*	−0.311**	−0.242	−0.3352**	−0.2287**
	(0.1549)	(0.1188)	(0.1838)	(0.1432)	(0.1132)
Google Trends	0.0017	0.0005	0.0007	−0.0004	−0.0002
	(0.0029)	(0.0024)	(0.0023)	(0.0035)	(0.0038)
Observations	105	105	105	105	105
Adjusted R^2	−0.0331	0.0024	−0.0262	−0.0195	−0.0153
F-stat	0.583	1.0306	0.6683	0.7508	0.8881
p-value	0.78963	0.41862	0.71814	0.64663	0.5739

Notes: HAC standard errors are in parentheses; *, **, and *** denote statistical significance at 10%, 5%, and 1%, respectively.

3.1. Disentangling the impact on the sentiment components and news categories

The findings above show that there is a relationship between COVID-19 news sentiment and stock market returns across all three news platforms. The structure of the NYTimes.com data allows us to further study the specific nature of the news since the website is divided into different categories that can be retrieved through the API.[7]

Below, we aim to first disentangle the sentiment component by distinguishing between positive and negative sentiment values. Additionally, we also consider the component of the sentiment, which is neutral regarding COVID-19. Second, we focus on the type of news released by the *New York Times* to investigate whether this result is driven by specific news categories.[8] The news sentiments are based on the following categories: (i) business, (ii) science, (iii) health, (iv) culture, (v) opinion, (vi) United States, (vii) world, and (viii) other.

Table 8 presents the overall news sentiments, but this time without news categorization. The sentiments we consider involve (i) the positive component of news sentiment (NYTimes^{+}), (ii) the negative component of news sentiment (NYTimes^{-}), (iii) news sentiment (NYTimes) as defined in Eq. (2) and previously considered, and (iv) sentiment based on neutral news not related to COVID-19 ($\text{NYTimes}^{\text{Neutral}}$).

Interestingly, it is the negative component (NYTimes^{-}) that explains the results regarding *New York Times* sentiment. It is significant at a 5% level and negatively related to S&P 500 returns. Given that the COVID-19 pandemic represents an adverse event, it confirms that bad news is the main driver in forming market expectations. Conversely, the positive component (NYTimes^{+}) is not statistically significant. This result also has further implications for the previous findings, at least for the *New York Times*. It is ultimately a reduction in bad news that is statistically significant and positively related to financial returns, rather than an increase in good news. Finally, the neutral component of the sentiment ($\text{NYTimes}^{\text{Neutral}}$) is not significantly related to COVID-19 news.

[7] Reuters and MarketWatch data were collected using a crawler from unstructured web pages. The categorization of those news items was therefore daunting.

[8] The authors are indebted to Guest Editor Sabri Boubaker for providing insightful comments and directions for additional work, which has resulted in this section.

Table 8
Estimates of S&P 500 returns using COVID-19 news sentiments in the *New York Times* for all categories, plus a set of control variables (the VIX, the OFR index, the growth rate in COVID-19 world cases, and Google searches for "coronavirus").

	(1)	(2)	(3)	(4)
Constant	0.059**	0.0847***	0.077***	0.063***
	(0.0225)	(0.0225)	(0.0193)	(0.0217)
$NYTimes^{+}$	0.0161			
	(0.0594)			
$NYTimes^{-}$		−0.0564**		
		(0.0219)		
NYTimes			0.0592***	
			(0.0189)	
$NYTimes^{Neutral}$				−0.0023
				(0.014)
VIX	−0.0027***	−0.0032***	−0.0031***	−0.0028***
	(0.001)	(0.0009)	(0.0009)	(0.0009)
OFR	0.0047***	0.0044***	0.0042***	0.0048***
	(0.0016)	(0.0015)	(0.0014)	(0.0016)
grCOVID-19	−0.0168	−0.0069	−0.0046	−0.0177
	(0.0161)	(0.0126)	(0.0128)	(0.0157)
Google Trends	0.0008*	0.0012**	0.0011**	0.0009*
	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Observations	105	105	105	105
Adjusted R^2	0.1588	0.2177	0.2231	0.1582
F-stat	4.9254	6.7881	6.9746	4.9099
p-value	0.00045557	1.758e−05	1.2792e−05	0.00046824

Notes: The news sentiments included are the positive component of the news sentiment ($NYTimes^{+}$), the negative component of the news sentiment ($NYTimes^{-}$), the news sentiment (NYTimes), and the sentiment based on neutral news ($NYTimes^{Neutral}$). HAC standard errors are in parentheses; *, **, and *** denote statistical significance at 10%, 5%, and 1%, respectively.

Regarding type of news, we have estimated the model for all sentiment built on the distinct categories. The findings show that the business news category represents the key source that provides statistically significant results. This insight serves as evidence that the information flow that shaped the market expectations was related to COVID-19 business news. Table 9 presents the estimated results.[9] The positive component ($NYTimesBusiness^{+}$) is statistically significant at the 1% level and positively related to S&P 500 returns. $NYTimesBusiness^{-}$ is negative and statistically significant at the 5% level. As expected, if negative news goes up, market returns go down. The *New York Times* sentiment resulting from the difference between positive and negative news (NYTimesBusiness) is statistically significant at the 1% level and positively related to S&P 500 returns. Interestingly, the magnitude of the coefficient is similar to that of MarketWatch reported in Table 5, which is almost exclusively a financial news source. In addition, in this case, the neutral component ($NYTimesBusiness^{Neutral}$) does not provide any statistically significant results.

The other statistically significant result is provided by the science news category. In this case, *New York Times* sentiment is statistically significant at the 10% level and positively related to stock market returns (see Web Appendix C). As this category is focused on disseminating science news, it is not surprising that articles responding to the COVID-19 pandemic could have contributed to the perception of the ongoing pandemic in the financial domain.

4. Conclusion

Natural language processing represents a useful tool for researchers and policymakers interested in extracting sentiment from news sources. The present study investigates the relationship between the stock market and news sentiment related to COVID-19. We have applied the Natural Language Toolkit, which uses machine learning models to extract COVID-19-related news sentiment in 203,886 online articles published on MarketWatch.com, Reuters.com, and NYtimes.com between January and June 2020. We applied a financial market-adapted BERT model to carry out our COVID-19 sentiment analysis. Our findings reveal a statistically significant relationship between COVID-19 sentiments and the stock market. This highlights that the flow of news on COVID-19 had an impact on the financial market since it contributed to forming market participants' expectations about the evolution of the pandemic, the real economy, and the stock market.

The results show that there is a statistically significant and positive relationship between sentiment and market returns. The sentiment scores and volumes for *the New York Times* exhibit a statistically significant and negative relationship with realized volatility. As to changes in S&P 500 trading volumes, we found that sentiments and control variables had almost a null explanatory power. Further, we considered positive, negative, and neutral sentiment components and showed that a decrease in negative news exerts a statistically significant impact on financial returns. Finally, we have provided evidence that the business news category is the

[9] For space reasons, we having estimates for all other news categories in Web Appendix C. Please note that the number of news changes among news categories and is due to their different availability over time.

Table 9
Estimates of S&P 500 returns using COVID-19 news sentiments in the *New York Times* for the business category, plus a set of control variables (the VIX, the OFR index, the growth rate in COVID-19 world cases, and Google searches for "coronavirus").

	(1)	(2)	(3)	(4)
Constant	0.0626***	0.0781***	0.0751***	0.0642***
	(0.0217)	(0.0198)	(0.0202)	(0.021)
$NYTimesBusiness^{+}$	0.0429***			
	(0.0112)			
$NYTimesBusiness^{-}$		−0.0227**		
		(0.0087)		
NYTimesBusiness			0.0311***	
			(0.0076)	
$NYTimesBusiness^{Neutral}$				0.0199
				(0.0131)
VIX	−0.0036***	−0.0036***	−0.0036***	−0.0035***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)
OFR	0.0044**	0.0057***	0.0054***	0.0044**
	(0.0017)	(0.0016)	(0.0016)	(0.0019)
grCOVID-19	−0.0082	−0.0111	−0.0069	−0.0158
	(0.0241)	(0.0263)	(0.0253)	(0.0258)
Google Trends	0.0014***	0.0013**	0.0014**	0.0013**
	(0.0005)	(0.0006)	(0.0006)	(0.0006)
Observations	92	92	92	92
Adjusted R^2	0.2366	0.2153	0.2552	0.2043
F-stat	6.6397	5.9929	7.2352	5.6742
p-value	2.8444e−05	8.3727e−05	1.0717e−05	0.00014353

Notes: The news sentiments included are the positive component of the news sentiment ($NYTimesBusiness^{+}$), the negative component of the news sentiment ($NYTimesBusiness^{-}$), the news sentiment (NYTimesBusiness), and the sentiment based on neutral news ($NYTimesBusiness^{Neutral}$).
HAC standard errors are in parentheses; *, **, and *** denote statistical significance at 10%, 5%, and 1%, respectively.

main sentiment driver for S&P 500 returns. The other news category from the *New York Times* that provides statistically significant outcomes is science.

Our findings complement recent studies on COVID-19 news and the financial market. Differently from the branch that emerged with the use of Google search volume as a proxy of public attention (Lyócsa et al., 2020; Costola et al., 2020; Salisu and Vo, 2020; Dey et al., 2022), we contribute to the stream of literature on news sentiment analysis. In this regard, Haroon and Rizvi (2020) investigate the relationship between coronavirus-related news and volatility of equity markets and find that panic news is associated with volatility in the stock markets. We provide evidence that also the volume of the news exerts an impact on the market realized volatility after controlling for several drivers such as implied volatility, financial stress, google search query volume on COVID-19, and the growth rate of the global COVID-19 cases.

Biktimirov et al. (2021) analyze COVID-19 news in the printed edition of the Wall Street Journal and find that the intensity of coverage for the "debt market" and "financial markets" are positively associated with the market performance. We disentangle the impact on the sentiment components and news categories for NYTimes.com and show that the business category represents the main driver in forming market expectations. Consistently with the previous studies, we find that this link remains the most relevant also during the outbreak of COVID-19 due to its economic and financial implications. For instance, Garcia (2013) shows that financial news from the New York Times provides stock returns predictability in recession times.

The present study provides several implications for investors, policymakers and financial institutions. Online news platforms contain valuable information that can be analyzed by machine learning models to predict stock market developments. Especially during the early phase of any future pandemic, investors might be able to assess the risks and manage their assets under high uncertainty. Sentiment analysis revealing the severity of a given instance of financial turmoil can also help policymakers evaluate the benefits and harms of their interventions. For instance, governments might restrict business activity and mobility depending on the expected intensity of the resulting economic downturn. Central banks and other financial institutions can also use stock market predictions to prepare accurate and timely responses, such as an increase in the money supply.

Our study does have certain limitations. We investigate a relatively short period between January and June 2020. Thus, we follow previous research in the field of stock market crash prediction, which also limited its analysis to the few months directly before or after an episode of financial turmoil (e.g., Butler et al., 2005; Jiang et al., 2010). Our period includes some medical and non-medical interventions by governments (e.g., lockdowns, mask mandates).

It would be interesting for a future research agenda to investigate the impact of COVID-19 news related to the government response on stock markets. Specifically, medical and non-medical interventions such as vaccination could be analyzed in more detail (e.g., lifting of lockdowns, vaccinations, and central bank responses). Once COVID-19 is no longer officially classified as a pandemic, it could be measured how events and actions affected the sentiment scores over time (e.g., different COVID-19 waves over time; countries using different news sources; key events such as vaccination rounds and initial lockdowns).

Another avenue for future research might be the inclusion of additional news sources. In this paper, we have focused on three platforms that predominantly publish economic and political news. Thus, adding opinions from social networks or discussion boards could contribute to a richer assessment of news sentiment.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ribaf.2023.101881.

References

Ahmad, W., Kutan, A.M., Gupta, S., 2021. Black swan events and COVID-19 outbreak: Sector level evidence from the US, UK, and European stock markets. Int. Rev. Econ. Finance 75, 546–557. http://dx.doi.org/10.1016/j.iref.2021.04.007.

Ahmed, S., Alshater, M.M., El Ammari, A., Hammami, H., 2022. Artificial intelligence and machine learning in finance: A bibliometric review. Res. Int. Bus. Finance 61, 101646. http://dx.doi.org/10.1016/j.ribaf.2022.101646.

Akhtaruzzaman, M., Boubaker, S., Sensoy, A., 2021. Financial contagion during COVID-19 crisis. Finance Res. Lett. 38, 101604. http://dx.doi.org/10.1016/j.frl.2020.101604.

Akhtaruzzaman, M., Boubaker, S., Umar, Z., 2022. COVID–19 media coverage and ESG leader indices. Finance Res. Lett. 45, 102170. http://dx.doi.org/10.1016/j.frl.2021.102170.

Andriotis, P., Takasu, A., Tryfonas, T., 2014. Smartphone message sentiment analysis. In: Peterson, G., Shenoi, S. (Eds.), Advances in Digital Forensics. Springer, Cham, pp. 253–265. http://dx.doi.org/10.1007/978-3-662-44952-3_17.

Antweiler, W., Frank, M.Z., 2004. Is all that talk just noise? The information content of internet stock message boards. J. Finance 59 (3), 1259–1294. http://dx.doi.org/10.1111/j.1540-6261.2004.00662.x.

Araci, D., 2019. FinBERT: Financial sentiment analysis with pre-trained language models. http://dx.doi.org/10.48550/arXiv.1908.10063, arXiv.

Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. Annu. Rev. Econ. 11, 685–725. http://dx.doi.org/10.1146/annurev-economics-080217-053433.

Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M., Viratyosin, T., 2020. The unprecedented stock market reaction to COVID-19. Rev. Asset Pric. Stud. 10 (4), 742–758. http://dx.doi.org/10.1093/rapstu/raaa008.

Baker, S.B., Bloom, N., Davis, S.J., Sammon, M., 2018. Coding large daily financial market moves: Data construction guide. Tech. Rep. 9, 2020.

Bianchi, D., Büchner, M., Tamoni, A., 2021. Bond risk premiums with machine learning. Rev. Financ. Stud. 34 (2), 1046–1089. http://dx.doi.org/10.1093/rfs/hhaa062.

Biktimirov, E.N., Sokolyk, T., Ayanso, A., 2021. Sentiment and hype of business media topics and stock market returns during the COVID-19 pandemic. J. Behavioral and Experimental Finance 31, 100542. http://dx.doi.org/10.1016/j.jbef.2021.100542.

Butler, A.W., Grullon, G., Weston, J.P., 2005. Can managers forecast aggregate market returns? J. Finance 60 (2), 963–986. http://dx.doi.org/10.1111/j.1540-6261.2005.00752.x.

Caporin, M., Corazzini, L., Costola, M., 2019. Measuring the behavioural component of the S&P 500 and its relationship to financial stress and aggregated earnings surprises. Br. J. Manage. 30 (3), 712–729. http://dx.doi.org/10.1111/1467-8551.12285.

Carlini, F., Cucinelli, D., Previtali, D., Soana, M.G., 2020. Don't talk too bad! stock market reactions to bank corporate governance news. J. Bank. Financ. 121, 105962. http://dx.doi.org/10.1016/j.jbankfin.2020.105962.

Casarin, R., Squazzoni, F., 2013. Being on the field when the game is still under way: The financial press and stock markets in times of crisis. PLoS One 8 (7), e67721. http://dx.doi.org/10.1371/journal.pone.0067721.

Chen, K., Zhou, Y., Dai, F., 2015. A LSTM-based method for stock returns prediction: A case study of China stock market. In: IEEE International Conference on Big Data. IEEE, pp. 2823–2824. http://dx.doi.org/10.1109/BigData.2015.7364089.

Chiah, M., Zhong, A., 2020. Trading from home: The impact of COVID-19 on trading volume around the world. Finance Res. Lett. 37, 101784. http://dx.doi.org/10.1016/j.frl.2020.101784.

Costola, M., Iacopini, M., Santagiustina, C.R., 2020. Google search volumes and the financial markets during the COVID-19 outbreak. Finance Res. Lett. 101884. http://dx.doi.org/10.1016/j.frl.2020.101884.

Das, S.R., Chen, M.Y., 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. Manage. Sci. 53 (9), 1375–1388. http://dx.doi.org/10.1287/mnsc.1070.0704.

De Spiegeleer, J., Madan, D.B., Reyners, S., Schoutens, W., 2018. Machine learning for quantitative finance: Fast derivative pricing, hedging and fitting. Quant. Finance 18 (10), 1635–1643. http://dx.doi.org/10.1080/14697688.2018.1495335.

DeLong, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1987. The Economic Consequences of Noise Traders. National Bureau of Economic Research Cambridge, Mass., USA, http://dx.doi.org/10.3386/w2395.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1 (Long and Short Papers). Association for Computational Linguistics, pp. 4171–4186. http://dx.doi.org/10.18653/v1/N19-1423.

Dey, A.K., Hoque, G.T., Das, K.P., Panovska, I., 2022. Impacts of COVID-19 local spread and Google search trend on the US stock market. Physica A 589, 126423. http://dx.doi.org/10.1016/j.physa.2021.126423.

Fazlija, B., Harder, P., 2022. Using financial news sentiment for stock price direction prediction. Mathematics 10 (13), 2156. http://dx.doi.org/10.3390/math10132156.

Garcia, D., 2013. Sentiment during recessions. J. Finance 68 (3), 1267–1300. http://dx.doi.org/10.1111/jofi.12027.

Ghoddusi, H., Creamer, G.G., Rafizadeh, N., 2019. Machine learning in energy economics and finance: A review. Energy Econ. 81, 709–727. http://dx.doi.org/10.1016/j.eneco.2019.05.006.

Gormsen, N.J., Koijen, R.S., 2020. Coronavirus: Impact on stock prices and growth expectations. Rev. Asset Pric. Stud. 10 (4), 574–597. http://dx.doi.org/10.1093/rapstu/raaa013.

Hamilton, J.T., 1995. Pollution as news: Media and stock market reactions to the toxics release inventory data. J. Environ. Econ. Manag. 28 (1), 98–113. http://dx.doi.org/10.1006/jeem.1995.1007.

Haroon, O., Rizvi, S.A.R., 2020. COVID-19: Media coverage and financial markets behavior—A sectoral inquiry. J. Behav. Experimental Finance 27, 100343. http://dx.doi.org/10.1016/j.jbef.2020.100343.

Hu, M., Liu, B., 2004. Mining Opinion Features in Customer Reviews. In: Proceedings of the 19th National Conference on Artifical Intelligence. AAAI '04, AAAI Press, pp. 755–760. http://dx.doi.org/10.5555/1597148.1597269.

Huang, X., 2018. Macroeconomic news announcements, systemic risk, financial market volatility, and jumps. J. Futures Mark. 38 (5), 513–534. http://dx.doi.org/10.1002/fut.21898.

Huynh, T.L.D., Foglia, M., Nasir, M.A., Angelini, E., 2021. Feverish sentiment and global equity markets during the COVID-19 pandemic. J. Econ. Behav. Organ. 188, 1088–1108. http://dx.doi.org/10.1016/j.jebo.2021.06.016.

Jiang, Z.-Q., Zhou, W.-X., Sornette, D., Woodard, R., Bastiaensen, K., Cauwels, P., 2010. Bubble diagnosis and prediction of the 2005–2007 and 2008–2009 Chinese stock market bubbles. J. Econ. Behav. Organ. 74 (3), 149–162. http://dx.doi.org/10.1016/j.jebo.2010.02.007.

Kabbani, T., Duman, E., 2022. Deep Reinforcement Learning Approach for Trading Automation in The Stock Market. MLWA-D-22-00192.

Koch, A., Huynh, T.L.D., Wang, M., 2022. News sentiment and international equity markets during BREXIT period: A textual and connectedness analysis. Int. J. Finance Econ. http://dx.doi.org/10.1002/ijfe.2635.

Koshiyama, A., Blumberg, S.B., Firoozye, N., Treleaven, P., Flennerhag, S., 2021. QuantNet: Transferring learning across trading strategies. Quant. Finance http://dx.doi.org/10.1080/14697688.2021.1999487.

Kumar, A., Sachdeva, N., 2020. Multi-input integrative learning using deep neural networks and transfer learning for cyberbullying detection in real-time code-mix data. Multimedia Syst. http://dx.doi.org/10.1007/s00530-020-00672-7.

Le, L.H.V., Huynh, T.L.D., Weber, B.S., Nguyen, B.K.Q., 2021. Different firm responses to the COVID-19 pandemic shocks: Machine-learning evidence on the Vietnamese labor market. Int. J. Emerg. Mark. http://dx.doi.org/10.1108/IJOEM-02-2021-0292.

Liu, Z., Dai, P.-F., Huynh, T.L., Zhang, T., Zhang, G., 2022. Industries' heterogeneous reactions during the COVID-19 outbreak: Evidence from Chinese stock markets. J. Int. Financ. Manag. Account. http://dx.doi.org/10.1111/jifm.12166.

Liu, Z., Huang, D., Huang, K., Li, Z., Zhao, J., 2021. Finbert: A pre-trained financial language representation model for financial text mining. In: Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence. pp. 4513–4519. http://dx.doi.org/10.24963/ijcai.2020/622.

Lyócsa, Š., Baumöhl, E., Vỳrost, T., Molnár, P., 2020. Fear of the coronavirus and the stock markets. Finance Res. Lett. 36, 101735. http://dx.doi.org/10.1016/j.frl.2020.101735.

Malo, P., Sinha, A., Korhonen, P., Wallenius, J., Takala, P., 2014. Good debt or bad debt: Detecting semantic orientations in economic texts. J. Assoc. Inform. Sci. Technol. 65 (4), 782–796. http://dx.doi.org/10.1002/asi.23062.

Marty, T., Vanstone, B., Hahn, T., 2020. News media analytics in finance: A survey. Account. Finance 60 (2), 1385–1434. http://dx.doi.org/10.1111/acfi.12466.

Masini, R.P., Medeiros, M.C., Mendes, E.F., 2021. Machine learning advances for time series forecasting. J. Econ. Surv. http://dx.doi.org/10.1111/joes.12429.

Mitchell, M.L., Mulherin, J.H., 1994. The impact of public information on the stock market. J. Finance 49 (3), 923–950. http://dx.doi.org/10.1111/j.1540-6261.1994.tb00083.x.

Nofer, M., Hinz, O., 2015. Using Twitter to predict the stock market. Bus. Inform. Syst. Eng. 57 (4), 229–242. http://dx.doi.org/10.1007/s12599-015-0390-4.

Peng, B., Chersoni, E., Hsu, Y.-Y., Huang, C.-R., 2021. Is domain adaptation worth your investment? Comparing BERT and FinBERT on financial tasks. In: Proceedings of the Third Workshop on Economics and Natural Language Processing. pp. 37–44. http://dx.doi.org/10.18653/v1/2021.econlp-1.5.

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1. http://dx.doi.org/10.18653/v1/N18-1202.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 2018. Improving Language Understanding by Generative Pre-Training. Technical Report, OpenAI.

Rebucci, A., Hartley, J.S., Jiménez, D., 2022. An event study of COVID-19 central bank quantitative easing in advanced and emerging economies. In: Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling. http://dx.doi.org/10.1108/S0731-90532021000043A014.

Salisu, A.A., Vo, X.V., 2020. Predicting stock returns in the presence of COVID-19 pandemic: The role of health news. Int. Rev. Financ. Anal. 71, 101546. http://dx.doi.org/10.1016/j.irfa.2020.101546.

Sinha, A., Kedas, S., Kumar, R., Malo, P., 2022. SEntFiN 1.0: Entity-aware sentiment analysis for financial news. J. Assoc. Inform. Sci. Technol. 1–22. http://dx.doi.org/10.1002/asi.24634.

Smales, L.A., 2021. Investor attention and global market returns during the COVID-19 crisis. Int. Rev. Financ. Anal. 73, 101616. http://dx.doi.org/10.1016/j.irfa.2020.101616.

Tetlock, P.C., 2007. Giving content to investor sentiment: The role of media in the stock market. J. Finance 62 (3), 1139–1168. http://dx.doi.org/10.1111/j.1540-6261.2007.01232.x.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30, 6000–6010.

Wan, X., Yang, J., Marinov, S., Calliess, J.-P., Zohren, S., Dong, X., 2021. Sentiment correlation in financial news networks and associated market movements. Sci. Rep. 11 (1), 1–12. http://dx.doi.org/10.1038/s41598-021-82338-6.

Xing, F.Z., Cambria, E., Welsch, R.E., 2018. Natural language based financial forecasting: A survey. Artif. Intell. Rev. 50 (1), 49–73. http://dx.doi.org/10.1007/s10462-017-9588-9.

Xing, F., Hoang, D.H., Vo, D.-V., 2020. High-frequency news sentiment and its application to Forex market prediction. In: Proceedings of the 54th Hawaii International Conference on System Sciences. HICSS, http://dx.doi.org/10.24251/HICSS.2021.191.

Xu, Y., Liang, C., Li, Y., Huynh, T.L., 2022. News sentiment and stock return: Evidence from managers' news coverages. Finance Res. Lett. 48, 102959. http://dx.doi.org/10.1016/j.frl.2022.102959.

Yarovaya, L., Matkovskyy, R., Jalan, A., 2021. The effects of a "black swan" event (COVID-19) on herding behavior in cryptocurrency markets. J. Int. Financ. Mark. Inst. Money 75, 101321. http://dx.doi.org/10.1016/j.intfin.2021.10132.

Zhang, Q., Cai, C.X., Keasey, K., 2013. Market reaction to earnings news: A unified test of information risk and transaction costs. J. Account. Econ. 56 (2–3), 251–266. http://dx.doi.org/10.1016/j.jacceco.2013.08.002.