International Journal of Information Management Data Insights 4 (2024) 100273

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

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

Headlines or Hashtags? The battle in social media for investor sentiment in the stock market

Yudhvir Seetharam, Kingstone Nyakurukwa *

University of the Witwatersrand, School of Economics and Finance, Johannesburg, South Africa

ARTICLE INFO

Keywords:
Behavioural finance
Transfer entropy
Online investor sentiment

ABSTRACT

This study tackles the complex task of measuring investor sentiment, a latent variable often measured through various proxies. The focus here is on textual sentiment extracted from online sources, specifically news media and social media sentiment. The central inquiry is whether these proxies are equivalent indicators of investor sentiment. Employing firm-level daily sentiment scores for DJIA stocks and leveraging Granger causality and transfer entropy, the research investigates the dynamics of information flow between these proxies. The findings show a prevailing pattern: information predominantly flows from social media to news for the majority of stocks while a reverse relationship is established for some stocks. The variations across stocks suggest that these proxies do not uniformly capture the same underlying phenomena. The study shows the significant role of social media in shaping news media sentiment and prompts considerations about regulating social media platforms in the context of their impact on financial markets.

1. Introduction

Research in information systems and related fields has explored how information and communication technologies, particularly social media and online news media, can empower communities by facilitating communication, sharing reliable information, and identifying opportunities for collective action (Eismann et al., 2021). This has permeated different domains including financial markets where the collective action of investors has been associated with stock market outcomes like returns, trading volume and volatility (Chen et al., 2021; Maqsood et al., 2020; Peng et al., 2022). Information from online media has been used in various studies in attempts to understand how sentiment inferred from these sources of information can be used for market timing purposes. Advances in big data have further amplified this by using data-driven and theory-driven approaches to understand how investor sentiment extracted from news and social media can be used to gauge the pulse of financial markets (Maass et al., 2018).

Though the literature is unanimous on the role of investor sentiment in financial markets, various proxies exist. The challenge with measuring investor sentiment is that it is not observable. The question that remains is whether these different proxies of investor sentiment measure the same phenomena. Smales (2016) outlines the significance of exploring the relationship between various proxies of investor sentiment. Intuitively if certain market sentiment measures lack correlation, it raises doubts about the validity of at least one aspect of the evolving literature on investor sentiment.

According to Chan et al. (2017), if sentiment proxies measure the same phenomenon, they should at least be correlated with each other. In a test of this hypothesis, Chan et al. (2017) constructed text-based sentiment proxies and compared them with the traditionally used sentiment proxies formalised by Baker and Wurgler (2006) and Baker and Wurgler (2007). The authors reported that these proxies of investor sentiment are not correlated and concluded that at least one or perhaps all the sentiment proxies considered were not valid proxies of investor sentiment. Bu (2023) compares indirect proxies of investor sentiment and direct proxies of investor sentiment and reports that they are weakly correlated, which implies that they might be measuring different phenomena.

Though investor sentiment is not directly observable, studies have used various latent variables to represent it. These include *inter-alia* market variables (e.g. Baker & Wurgler, 2006), surveys (e.g. Solanki & Seetharam, 2014) and textual sentiment (Hong et al., 2020; Nyakurukwa & Seetharam, 2023a; Wilksch & Abramova, 2023). Textual sentiment is sentiment that is extracted from text posted on online platforms like news and social media platforms. According to Kushwaha et al. (2021), the past decade has witnessed the integration of text-based

* Corresponding author.
E-mail address: knyakurukwa@gmail.com (K. Nyakurukwa).

https://doi.org/10.1016/j.jjimei.2024.100273

Available online 16 August 2024
2667-0968/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

and multimedia (image and video)-based information retrieval as essential components in various research domains within the analytics field. News and social media platforms have emerged as some of the most used sources of information in empirical finance. However, there is increasing convergence between these two sources of information as there is a constant flow of information between them.

This intertwining of news and social media is indicative of a symbiotic relationship, where news outlets harness the immediacy and virality of social platforms to broaden their audience reach. Simultaneously, social media platforms benefit from the credibility and journalistic rigour associated with traditional news sources, enhancing the overall reliability of the information disseminated through these channels. A typical individual is no longer only a consumer of information but a *"prosumer"* capable of producing and consuming information (Grover et al., 2022). This convergence has redefined the landscape of information dissemination, blurring the lines between traditional journalism and user-generated content on social media. As a consequence, individuals are exposed to a diverse range of news content, with the collaborative efforts of news media and social platforms creating a more interconnected and accessible information ecosystem (Kim et al., 2023). This evolution reflects the need for an understanding of the interconnected dynamics shaping contemporary media landscapes.

Ren et al. (2021) contend that previous studies on the correlation between sentiment from news media and social media have neglected the growing interdependence between these two sources of information. They propose a model that illustrates how information diffuses between the two sources while acknowledging their increased interdependence. This mechanism suggests that media outlets are compelled to cater to the demands of their readership. This is rooted in the theoretical rationale of "demand-driven media bias" (Mullainathan & Shleifer, 2005). According to confirmation bias, people tend to experience psychological stress when confronted with information that contradicts their pre-existing beliefs, compared to information that supports their beliefs (Festinger, 1957). Consequently, media outlets may design news articles that align with readers' beliefs to attract their attention and enhance audience engagement, leading to higher advertising revenues. Accordingly, mass media platforms might feel inclined to align their news coverage with the sentiments expressed on social media, especially when reinforced by widespread agreement within social media circles. The demand-driven media bias therefore hypothesises that there is a dominant flow of information from social media to news media sentiment.

The *"echo chamber effect"* model can also be used as a theoretical model to explain the flow of information from traditional news to social media. According to this model, social media platforms often reiterate signals from traditional news sources. This repetition creates an "echo chamber" where information is continuously reinforced without the introduction of new perspectives. The model posits that within these echo chambers, a subset of "behavioural" traders misinterpret repeated information as new. This misinterpretation occurs because these traders neglect the correlation between the original news signals and their repetitions on social media. Consequently, the echo chamber effect amplifies the impact of traditional news on social media sentiment, as the same piece of information is perceived multiple times as fresh and relevant. This model provides a framework for understanding how information flows from traditional news sources to social media, influencing market behaviour and sentiment through repeated exposure and reinforcement. Thus, the echo chamber effect suggests a dominant flow of information from traditional news to social media. Fig. 1 visualises the two models that explain the potential flow of information between traditional news media and social media:

Our findings indicate that, in most cases, information primarily flows from social media sentiment to news media sentiment in the majority of stocks. However, there are some stocks in which information dominantly flows from news to social media. Thus, our findings mostly support the demand-driven bias model. Also, our results show that these two proxies of investor sentiment cannot be regarded as equivalent as they are likely to be measuring different phenomena.

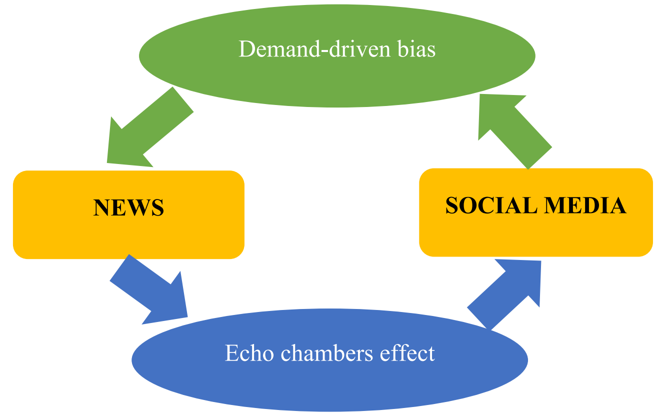

Fig. 1. Models explaining the potential flow of information between news media and social media.

We proceed as follows; in Section 2, we look at the literature review, Section 3 describes the data and methods, Section 4 reports the findings of the study, Section 5 discusses the results and Section 6 concludes.

2. Literature review

2.1. Proxies of investor sentiment

Within the domain of behavioural finance investor sentiment has been measured using different proxies because it is not directly observable. This is even more significant in this age of big data where individuals can easily depend on online sources to gauge sentiment (e.g. online news platforms). This is because different sources of information can be used to gauge the pulse of the market (e.g. news platforms, social media platforms, online search engines etc.). According to Shleifer and Vishny (1997), such information gathering using online sources is associated with unsophisticated investors, who are generally regarded as being at an information disadvantage compared to institutional investors. The problem is that different online sources of information measure different phenomena and rather than trying to isolate which one is better, it is ideal to understand how these sources are related. This can be established by trying to identify which one, within a network, is more dominant in transmitting information.

2.2. Theoretical framework

Theoretically, Milas et al., 2021 employ the agenda-setting theory of media in political contexts (see Neuman et al., 2014), which posits that the media holds the authority to shape the salient topics of the day. This principle applies to both Twitter[1] and conventional media outlets. The authors investigate the agenda-setting hypothesis by empirically testing its applicability to a subject concerning the sovereign bond markets. A unique dataset focusing on news related to the "Grexit" topic was created and included news from both Twitter and traditional media sources. A bivariate system was used to establish the causal relationship between news and social media using Dufour et al., (2006) causality framework that employs multiple horizon VAR specifications. The study reveals a two-way exchange of information between Twitter and traditional news outlets. This implies that both forms of news play a crucial role in predicting the behaviour of the sovereign bond market, indicating a connection between the "old" (traditional news) and the "new"

[1] The social media platform is now called X. In this study, we use Twitter since the data were collected before the platform was acquired by X.

(Twitter) media. Moreover, the influence of Twitter on traditional news is found to be more enduring, influential, and statistically significant, highlighting the dominance of Twitter as the primary agenda-setting news source in the European sovereign bonds market. Furthermore, even after removing the influence of traditional news (by orthogonalising the Twitter variable with respect to the traditional news variable), Twitter continues to exhibit significant predictive power. The demand-driven media bias also complies with the agenda-setting theory explained above, as both theories elevate the role of social media in affecting public opinions used by traditional media companies in setting public agenda. Thus, sentiment can flow from social media to traditional news media.

2.3. Information flow between different proxies of investor sentiment

One question that has dominated scholarly discourse is whether different proxies of investor sentiment can be regarded as equivalent. Behrendt and Prange (2021) use two online searches as proxies for investor sentiment; Google searches and company-specific Wikipedia queries to establish if they could be considered equivalent proxies. According to Behrendt and Prange (2021), studies examining the equivalence of online investor sentiment are associated with ongoing research that seeks to establish the preferred proxy. Behrendt and Prange (2021) use a dataset of Google stock ticker searches and company-specific Wikipedia queries for DJIA stocks to establish the equivalence of the two proxies using an information-theoretic framework. Linear Granger causality was also used to compare the results and establish if nonlinear dynamics are important in modelling the relationship between the variables.

The results by Behrendt and Prange (2021) reveal no particular pattern in terms of the dominant direction of information transfer between Google and Wikipedia searches. Some companies exhibit a dominant direction of information flow from Google to Wikipedia while others exhibit a reverse relationship. Based on these findings, the authors conclude that the two proxies do not measure the same online phenomena, and cannot, therefore, be regarded as equivalent. Behrendt and Prange (2021) define an unambiguous flow of information flow between the variables as being detected if information transfer is statistically significant in one direction. Also, the study establishes nonlinear information transfer between the variables as revealed by the significant effective transfer entropy values for the residual series that are purged of any auto and cross-correlations. Thus, modelling the dependencies between Google searches and Wikipedia queries for the sampled companies using linear Granger causality could not capture the nonlinear dependencies.

While there are studies that have sought to understand information flow between different proxies of investor sentiment, the most researched aspect in this strand of literature is the potential diffusion of information between social media and news media. Most naïve investors tend to repeat on social media information what they would already have encountered on traditional news platforms. On the other hand, in the contemporary world, most traditional media companies have also adopted social media handles which they use to give a snapshot of their news to increase traffic to their news websites. Thus, there is information flow between these two sources of information, but the question remains; what is the dominant direction in the flow of information between these important sources of financial information? An interesting angle on understanding the phenomenon can be seen from the proposal by the European Union Parliament to impose a *"link tax"*[2] on social media platforms that use links to news articles on the presumption that this hurts mass media by "stealing" news. Thus, this proposal by the EU parliament acknowledges the diffusion of information between news media and social media. Various studies have examined the dynamics of information flow between social media and news media using textual analysis.

Beckers (2018) reasons that though several studies have been done on the role played by social media and news media sentiment in financial markets, these studies have failed to allow inference of whether these two proxies of investor sentiment are competitive or complementary. To fill this gap, Beckers (2018) utilises a universe of data gathered from Thomas Reuters MarketPsych indices (TRMI) that provide an inclusive source of sentiment extracted from text. News sentiment is extracted from media articles originating from >2000 international news publications. Social media content on the other hand is gathered from stock forums, *tweets*, comment streams as well as blogs originating from >700 primary sources. In a direct comparison of social media and news media sentiment, Beckers (2018) reports that these two proxies for investor sentiment are moderately correlated. Notably, the study reported that news trumps social media as a sentiment source for market timing.

In their study, Gan et al. (2020) sought to find the dynamics of information flow between social media and traditional news media using sentiment scores extracted from TRMI. Their main aim was to capture the dependencies between the two types of online textual sentiment proxies using a vector autoregressive framework to avoid explicit exogeneity assumptions. The study used two proxies for investor sentiment; *buzz* – which was used as a measure of the volume of the information flow and *sentiment* – which was created from a principal component analysis of the various emotional states existing in the TRMI database. A rolling window approach was utilised to establish the dynamic relationship between the two proxies of online investor sentiment for companies listed on the SP500. The findings from the empirical model based on VAR revealed a waning of the news media in terms of imparting sentiment between 2011 and 2017. Notably, the study reports that by 2016, social media had become the dominant source of information in generating media activity.

Jiao et al. (2020) used news media and social media buzz for 1848 stocks between 2009 and 2014. The ultimate panel dataset excluded regulated utilities, depository institutions as well as holding and investment companies. The authors also excluded stocks that were delisted within the sample period. The study specifically investigated the mechanisms at play between social media sentiment and news media sentiment using a stock-level panel VAR including the two sentiment indexes as well as volatility and turnover as endogenous variables. The empirical findings indicate that a rise in news media correlates with a subsequent increase in social media activity, as demonstrated by Granger causality. The relationship is unidirectional as there is no evidence of a statistically significant reverse relationship between the variables. This is consistent with the view that content posted on social media platforms is generated by repeating and retweeting content that has already appeared on traditional news platforms.

Though various studies have used Twitter and StockTwits as proxies for social media, alternative proxies for social media have also been used in other studies. In China, western social media platforms like Twitter

[2] The link tax is a component of a law that would provide publishers ownership over anything that is shared online via Facebook or YouTube as well as other social media platforms and would also apply to services that aggregate news, like Google News. Anyone who intends to use online journalistic content excerpts must first obtain permission from the publisher. This new publisher privilege would be valid for 20 years following publication (https://felixreda.eu/eu-copyright-reform/extra-copyright-for-news-sites/).

and StockTwits are not allowed. However, there are alternative proxies for social media used in China such as Baidu and Sina Weibo. Dong and Gil-Bazo (2020) used Sina Weibo (A Twitter-like platform as a proxy for social media) and Sina Finance (For the extraction of news sentiment) to examine how mass media affects social media for finance-related posts. The study included 2501 stocks in a two-year-period between 2013 and 2014 and found that social media amplifies the persistence of news media over time.

3. Research methodology

3.1. Data

The study uses a population of firms constituting the Dow Jones Industrial Average (DJIA) between 1 January 2016 and 30 April 2023. DJIA stocks have been chosen because of the high quality of textual sentiment data extracted from the Bloomberg terminal. We targeted all DJIA constituent stocks because they attract considerable media attention because of their sizes. Moreover, several studies have used DJIA stocks in studies investigating the role of textual sentiment in stock markets (e.g. Nyakurukwa & Seetharam, 2023b). Sentiment scores from both social media and news media are sourced from the Bloomberg terminal. While Bloomberg began including social media sentiment data on January 1, 2015, the quality of these sentiment scores notably improved in 2016, marked by a decrease in the percentage of missing values. We, therefore, restrict the study period to the period after 31 December 2015. This leads us to an ultimate dataset of 29 DJIA constituent stocks. Ultimately, the dataset contains 1906 daily observations of news sentiment and social media scores, together constituting 55,274 firm-day observations for each variable.

In Tables A1 and A2, most of the stocks have a positive news media and social media score average. This means that on average, the listed stocks were associated with positive sentiment on the two media platforms (News and Twitter). In terms of stationarity, using the bootUR (Smeekes & Wilms, 2022) test under the null hypothesis that each series has a unit root, the results across all the markets and all the variables for each stock sampled show statistically significant results at the 1 % level of significance. This means that all the series do not have a unit root and are therefore stationary. Using the Fat-tailed normality test because of its robustness for financial variables (Jelito & Pitera, 2021), we observe that for all stocks the test statistic is statistically significant at the 1 % level of significance. As expected, we conclude that our firm-level variables are all not normally distributed at the 1 % level of significance.

3.2. Variables

For this study, average sentiment from Twitter and news was taken from Bloomberg Inc. Bloomberg Inc. uses human specialists to manually analyse massive datasets of tweets and news items to determine the average sentiment. Each tweet or news article is then given a label, which can be divided into positive, neutral and negative categories. Following manual classification, the feeds are fed into machine-learning models that are trained to mimic the analysis of text communications by language specialists. Once the machine learning models are finished, they are applied to new tweets and news stories that have been tagged with tickers. Each tweet or news story is then given a sentiment score, which ranges from −1 to +1, in real-time.

3.3. Methods

3.3.1. Linear causality

The Granger causality test (Granger, 1980) is a traditional method used to establish if one variable causes another. Specifically, it helps figure out if a variable (X) has an impact on predicting another variable (Y). We create two models to predict Y. Model 1 looks at Y's past values to make predictions. Model 2 considers both the past values of Y and X to predict Y. The models essentially help us see if adding X improves our ability to predict Y. Using a VAR model that includes firm-level news sentiment and social media sentiment, we calculate a Granger causality index (GCI)[3] to measure how much better Model 2 predicts Y compared to Model 1. This index is based on the difference in errors between the two models and is computed as follows:

$$GCI = log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) \tag{1}$$

where σ_1^2 and σ_2^2 are the variances of the errors in Models 1 and 2, respectively. We use the Fisher test to check if the improvement in predictions is statistically significant. This involves comparing the error differences from both models against a critical value. We set up hypotheses to test if X indeed affects Y. The null hypothesis assumes X does not Granger cause Y, and the alternative hypothesis assumes significant Granger causality from X to Y. To determine the dominant direction of information flow, we use the GCI. If GCI values show that news Granger causes Twitter more than the other way around for a company, we conclude that information flows mainly from news to social media. If, for all sampled stocks, the results consistently show that information flows dominantly from news (social media) to social media (news), we conclude that the two are similar equivalent proxies of sentiment.

3.3.2. Transfer entropy

The Granger causality test described above has several limitations. It typically assumes that the variables involved are normally distributed and have linear relationships. This can be problematic when dealing with real-world data, which often exhibits non-normal distributions and nonlinearities. Transfer entropy offers an alternative approach. It is a non-parametric measure, meaning it does not rely on specific assumptions about the underlying data-generating process. This makes it more robust in the presence of non-normal distributions and nonlinearities, which are common in financial data. Transfer entropy focuses on information transfer between two time series. Here, Shannon entropy (Shannon, 1948) plays a key role. Shannon entropy essentially quantifies the uncertainty associated with a random variable. In simpler terms, it tells us how much information, on average, is needed to represent the variable. The formula for Shannon entropy for a discrete random variable X with probability distribution $p(x)$:

$$H_J = -\sum p(j)\ .\ \log(p(j)) \tag{2}$$

Here, Σ represents the sum across all possible outcomes (x) that the variable X can take, and log denotes the logarithm to base 2. The summation calculates the average number of bits needed to optimally encode the information about X based on its probability distribution. Transfer entropy builds upon Shannon entropy and quantifies the directed information flow between two time series, X and Y. It utilizes the concept of a Markov process, where the probability of a future state depends only on the current state and not on the entire history. The transfer entropy, TE(X->Y), measures the information gained about the future of Y by knowing the past and present of X, beyond what is already predictable from the past of Y alone.

The Effective transfer entropy (ETE) addresses the small sample bias by comparing the actual transfer entropy with a shuffled version of the time series where dependencies are shuffled. The statistical significance of transfer entropy estimates is evaluated through Markov block bootstrap, a method designed to maintain the temporal dependencies within the time series. The presented approach assumes discrete data. When working with continuous data, such as sentiment scores, symbolic coding is utilized. This involves dividing the data into a limited number of bins determined by the quantiles of the empirical distribution. Each data

[3] For more details on the details please read Nyakurukwa and Seetharam (2023).

point is subsequently assigned a symbol associated with the bin to which it belongs.

If the bounds specified for n bins are represented by $q_1, q_2, \ldots, q_n$ where $q_1 < q_2 \ldots < q_n$ and y_t represents the time series under investigation, the data points are recoded as follows:

$$\begin{cases} 1 & \text{for } y_t \leq q_1 \\ 2 & \text{for } q_1 < y_t \leq q_2 \\ \vdots & \vdots \\ n-1 & \text{for } q_{n-1} < y_t \leq q_2 \\ n & \text{for } y_t \geq q_n \end{cases}$$

Each value in the time series y_t is substituted with an integer $(1, 2, \ldots, n)$ based on how S_t corresponds to the interval defined by the lower and upper bounds q_1 *to* q_n. The choice of bins was justified by the empirical distribution of the data. The dominant direction of information flow is determined by comparing the effective transfer entropies in both directions (ETE(X->Y) and ETE(Y->X)). The difference between these values indicates a stronger information flow direction. If the sign (positive or negative) of the difference in ETEs is consistent across all sampled firms, then we conclude that there is evidence for equivalence between the two investor sentiment proxies. However, we only consider a significant flow of information if at least one of the directional information flows (ETE(X->Y) or ETE(Y->X)) is statistically significant at the 5 % level of significance. If this condition is not met, we conclude that for that stock, there is no statistically significant flow of information between news media sentiment and social media sentiment.

4. Results

In this section, we report the findings on the extent of information flow between news media sentiment and social media sentiment at the firm level using Granger causality. We start by recalling the definition of equivalence of sentiment proxies. The two proxies of online investor sentiment are regarded as equivalent if information unambiguously flows from one sentiment proxy to the other. Thus, if for some stocks information flows dominantly from news to social media while for others information dominantly flows from social media to news, then we conclude that these proxies are not equivalent.

4.1. Granger causality

In the Granger causality models used for each stock level data, the lag length was chosen using the Bayesian information criterion (BIC). The results on the information flow between news media sentiment and social media sentiment at the firm level are presented in Table 1. Statistically significant flow of information for each stock is considered if one of the causal relationships in the bidirectional tests is significant[4] at the 5 % level.

The findings from Table 1 show that the stocks that exhibit a significant flow of information from news to social media and from social media to news are 6 and 15 respectively, while 8 stocks show no statistically significant flow of information at the 5 % level. What we can conclude from the results of the Granger causality test using the USA market is that information flow does not follow a predictable trend. There are stocks where information dominantly flows from news (social media), while for others, no statistically significant flow of information can be observed. Thus, using the definition of equivalence of investor sentiment proxies used by Behrendt and Prange (2021) where there should be an unambiguous dominant flow of information from one proxy to another for all sampled stocks, we can conclude that these two proxies of online investor sentiment cannot be regarded as equivalent proxies.

[4] i.e if one of the tests of Granger causality to news (twitter) is statistically significant at the 5% level.

Table 1
Information flow using Granger causality.

Ticker	News → Twitter		Twitter → News		Dominance
	F-test	GCI	F-test	GCI	
AAPL	6.2885***	0.0132	13.1210***	0.0273	Twitter
AMGN	0.0573	0.0000	2.2759	0.0012	Insignificant
AXP	0.6646	0.0011	1.2177	0.0019	Insignificant
BA	0.9996	0.0021	12.9356***	0.0270	Twitter
CAT	10.3660***	0.0054	38.4574***	0.0200	Twitter
CRM	7.5386***	0.0040	12.4471***	0.0065	Twitter
CSCO	6.4606**	0.0034	10.8079***	0.0057	Twitter
CVX	4.0938**	0.0021	9.1360***	0.0048	Twitter
DIS	17.3254***	0.0091	3.5625*	0.0019	News
GS	0.0018	0.0000	41.2100***	0.0214	Twitter
HD	15.2637***	0.0080	18.2317***	0.0095	Twitter
HON	6.6585***	0.0070	3.9298**	0.0041	News
IBM	1.5768	0.0008	20.1405***	0.0105	Twitter
INTC	5.5270***	0.0116	2.1924*	0.0046	News
JNJ	17.0585***	0.0089	3.7807*	0.0020	News
JPM	2.0384*	0.0043	6.1273***	0.0129	Twitter
KO	10.4326***	0.0055	10.9040***	0.0057	Twitter
MCD	4.6289**	0.0024	2.5342	0.0013	News
MMM	2.6443**	0.0042	2.7050*	0.0043	Twitter
MRK	3.9576**	0.0021	5.2770**	0.0028	Twitter
MSFT	0.9814	0.0005	3.0937*	0.0016	Insignificant
NKE	2.5696	0.0013	12.4621***	0.0065	Twitter
PG	9.1663***	0.0048	21.1092***	0.0110	Twitter
TRV	2.4673	0.0013	0.3933	0.0002	Insignificant
UNH	22.3083***	0.0117	16.1702***	0.0085	News
V	1.4051	0.0015	0.7730	0.0008	Insignificant
VZ	0.8175	0.0004	0.1532	0.0001	Insignificant
WBA	0.6681	0.0004	3.5912*	0.0019	Insignificant
WMT	0.6804	0.0004	1.8776	0.0010	Insignificant

4.2. Findings from transfer entropy

Transfer entropy uses the probability theory to determine the flow of information between two time series and is effective even with nonlinear variables. In this section, we present the findings of the flow of information between the two proxies of investor sentiment. Like the criteria used in the Granger causality models, we only consider the existence of a significant flow of information between variables if at least one of the causal directions (news to Twitter or Twitter to news) is statistically significant at the 5 % level. To determine the dominant flow of information, we use the ETE scores for each stock. News (social media) is presumed to be dominant if the ETE score for transfer entropy from the news (social media) is higher than the ETE score for transfer entropy from social media (news) for each stock. According to Behrendt et al. (2019), choosing the lag length in transfer entropy is not a trivial task as it depends on the time series being considered. In this study, for the lag length, we follow Behrendt and Prange (2021) by selecting an appropriate lag length using the BIC recovered from a bivariate VAR model that includes time series vectors of social media sentiment and news media sentiment scores for each stock. We use the same lag length for investigating the bidirectional flow of information for each stock in line with Schreiber (2000) who argues that the interpretation of results is facilitated by setting the same number of lags. Table 2 shows the results of the dominant flow of information between the two proxies of investor sentiment.

Slight differences can be seen in the results reported in terms of percentage dominance in information flow using transfer entropy compared to the results obtained using Granger causality. Eight stocks (27.58 %) are associated with an insignificant flow of information while in 6 (20.68 %) stocks, news sentiment is dominant and in 15 stocks (51.29 %) social media is dominant.

4.3. Robustness

To guarantee that our findings are not influenced by specific model

Table 2
Information flow using transfer entropy.

	News → Twitter			Twitter → News			Dominance
	TE	ETE	SE	TE	ETE	SE	
AAPL	0.0783	0.0068***	0.0091	0.0723	0.0233**	0.0090	Twitter
AMGN	0.0032	0.0000	0.0017	0.0066	0.0017	0.0016	Insignificant
AXP	0.0516	0.0118**	0.0065	0.0659	0.0205***	0.0072	Twitter
BA	0.0553	0.0000	0.0080	0.0573	0.0115	0.0090	Insignificant
CAT	0.0101	0.0056***	0.0017	0.0084	0.0041**	0.0016	News
CRM	0.0044	0.0000	0.0018	0.0098	0.0053***	0.0017	Twitter
CSCO	0.0110	0.0066***	0.0020	0.0164	0.0122***	0.0016	Twitter
CVX	0.0072	0.0028*	0.0018	0.0093	0.0043**	0.0018	Twitter
DIS	0.0137	0.0088***	0.0017	0.0089	0.0037**	0.0017	News
GS	0.0096	0.0052**	0.0019	0.0180	0.0134***	0.0017	Twitter
HD	0.0056	0.0013	0.0016	0.0089	0.0044**	0.0016	Twitter
HON	0.0313	0.0100**	0.0047	0.0295	0.0078**	0.0044	News
IBM	0.0082	0.0043**	0.0017	0.0129	0.0078***	0.0017	Twitter
INTC	0.0642	0.0000	0.0096	0.0780	0.0222**	0.0090	Twitter
JNJ	0.0069	0.0026*	0.0017	0.0049	0.0004	0.0018	Insignificant
JPM	0.0720	0.0041	0.0101	0.0757	0.0064*	0.0099	Insignificant
KO	0.0119	0.0074***	0.0018	0.0146	0.0098***	0.0019	Twitter
MCD	0.0070	0.0025	0.0017	0.0035	0.0000	0.0017	Insignificant
MMM	0.0491	0.0114**	0.0071	0.0571	0.0171***	0.0070	Twitter
MRK	0.0070	0.0025*	0.0016	0.0107	0.0061***	0.0017	Twitter
MSFT	0.0074	0.0028*	0.0017	0.0116	0.0070***	0.0018	Twitter
NKE	0.0079	0.0032*	0.0018	0.0077	0.0029*	0.0016	Insignificant
PG	0.0092	0.0045**	0.0016	0.0070	0.0018	0.0017	News
TRV	0.0123	0.0079***	0.0015	0.0034	0.0000	0.0018	News
UNH	0.0125	0.0079***	0.0018	0.0034	0.0000	0.0018	News
V	0.0233	0.0040	0.0042	0.0185	0.0004	0.0038	Insignificant
VZ	0.0074	0.0025*	0.0017	0.0141	0.0089***	0.0019	Twitter
WBA	0.0043	0.0000	0.0017	0.0032	0.0000	0.0018	Insignificant
WMT	0.0144	0.0100***	0.0018	0.0169	0.0117***	0.0019	Twitter

Notes: TE, ETE, and SE stand for Transfer Entropy, Effective Transfer Entropy, and standard errors, respectively. ***, **, and * indicate statistical significance at the 1 %, 5 %, and 10 % levels, respectively.

specifications, we implement a series of robustness checks to determine if our results are responsive to changes in model parameters. First, as explained earlier, setting the lag length in transfer entropy models highly depends on the nature of the time series being analysed. In the results reported in this study, we used a commonly utilised approach of determining the lag length using a bivariate VAR model that includes the two proxies of investor sentiment and choosing the lag length that minimises the BIC. In further robustness checks, we choose a lag length of one for all transfer entropy models. The decision to use a lag length of one is based on literature indicating that financial time series can be effectively modelled using Markov chains of order one (Mcqueen & Thorley, 1991; Turner et al., 1989). Second, instead of discretisation based on the 5 % and 95 % quantiles, we follow Behrendt and Prange (2021) by also utilising discretisation based on the 10 % and 90 % quantiles. Using both robustness checks, our results do not qualitatively change. Information dominantly flows from social media.

The dominance of social media over news media in the transmission of sentiment shows that firms should actively manage their social media presence to influence public sentiment and media coverage. It is important to develop strategies to engage with followers, address potential issues, and promote positive narratives. This is particularly important as studies have shown that one of the challenges of big data is the existence of misinformation, spam messages as well as the deployment of bots to drive a particular narrative (Kar et al., 2023). Also, in the long run, management of social media presence reduces the operational cost of handling complaints and improves customer satisfaction, thereby increasing overall firm profitability.

5. Discussion

5.1. Contributions to literature

The results of the study show that generally, DJIA stocks are associated with a significant information flow of information between news and social media. Considerable information flow between social media and news media sentiment observed can be explained by several factors. The real-time nature of Twitter enables users to quickly share and disseminate information, including sentiments and opinions about the stock market. This rapid transmission of sentiment can catch the attention of news media outlets, who may then pick up and amplify the sentiment in their coverage. This phenomenon is consistent with the concept of Twitter acting as an "early warning system" for news media, as discussed by McGregor and Molyneux (2020).

Second, the inherent characteristics of Twitter, such as brevity and conciseness, make it a popular platform for succinctly expressing opinions and emotions (Tandon et al., 2021). This allows for a large volume of sentiment-related data to be generated on Twitter, providing news media outlets with a vast pool of potential news leads or angles to cover. As a result, news media may rely on Twitter as a source of information and gauge public sentiment to shape their own narratives and coverage of stocks. This result is also consistent with the demand-driven media bias hypothesis of Mullainathan and Shleifer (2005). This hypothesis suggests that media outlets may be compelled to cater to the demands of their readership and thereby report news that aligns with the narratives prevalent on social media platforms to increase advertising revenue. These results are also in line with the agenda-setting theory where social media has become an important source of public opinion which can then be picked up by traditional news platforms. The results concur with Milas et al. (2021) who report findings highlighting Twitter's significance as a primary source for setting the agenda regarding discussions related to Grexit. For a few of the stocks, the findings showed a significant flow of information from news sentiment to social media sentiment. This supports the echo chamber effect where information previously appearing on traditional news is repeated on social media platforms and interpreted as new information by investors. This is in line with Tetlock (2011) who suggested that this “stale information” can affect stock market outcomes like stock returns and volatility.

The fact that information dominantly flows from social media to news is not surprising given the regulations surrounding the disclosure of corporate and strategic information *via* social media. In 2013, the Securities and Exchange Commission (SEC) issued a report that made it clear that firms can utilise social media platforms like Twitter to disclose key corporate information in line with Regulation Fair Disclosure as long as investors are sufficiently alerted about which social media is used to disseminate the information (Alexander & Gentry, 2014). Thus, it is possible for key corporate information to appear first on social media before it is picked up by traditional news media later during the day or sometimes the following day. This could explain why for most of the stocks, information predominantly flows from social media to news media.

5.2. Practical implications

The study raises questions about the regulation of social media platforms in the context of their influence on news media sentiment and, subsequently, financial markets. Discussions may emerge on how to strike a balance between ensuring freedom of expression on these platforms and preventing potential market manipulation. Regulatory bodies may need to assess the role of social media in disseminating information that can impact financial markets. The findings highlight social media as a dominant source influencing investor sentiment, with information predominantly flowing from social media to news for most stocks. This indicates that social media platforms have a pivotal role in influencing market sentiment.

5.3. Future scope

Future studies could exploit the relationship between social media and news media sentiment reported in this study in various directions. First, fake accounts, often referred to as social bots or Sybil accounts, are generated in large numbers by software programs to artificially boost the spread of false information by propagating it across social media platforms (Haustein et al., 2016; Michail et al., 2022). Future studies could leverage the findings found from this study to understand whether information generated by bots could end up on mainline news platforms as real news. Future studies could therefore specifically pursue this using primary data and not third-party data sources which may be more restrictive. Future studies could also explore the flow of information between news sentiment and social media sentiment using sectoral data to establish whether the differences in the direction of the flow of information could be explained by sectoral characteristics. Finally, rather than depending on data provided by third-party data providers, future studies could use primary data and leverage the rise of generative AI to compute sentiment scores. Several recent studies have shown that generative AI is accurate in several business applications (Dwivedi et al., 2023; Ooi et al., 2023)

6. Conclusion

This study sought to address the ongoing challenge of identifying an effective proxy for the unobservable variable of investor sentiment, focusing specifically on the equivalence of news media and social media sentiments. Analysing daily sentiment scores for DJIA stocks through Granger causality and transfer entropy, the research shows distinct patterns in information flow among the sampled stocks. The prevailing trend demonstrates a significant flow from social media to news, though variations exist, with certain stocks indicating information dominance from news to social media or showing no significant flow. We conclude that these two proxies of investor sentiment are not equivalent and financial market participants need to understand these complexities in their investing decisions. The major limitation of the study is the exclusion of smaller stocks because of poor quality sentiment scores generated because of the less attention these stocks attract on online platforms.

CRediT authorship contribution statement

Yudhvir Seetharam: Writing – review & editing, Methodology, Investigation, Formal analysis, Conceptualization. **Kingstone Nyakurukwa:** Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare no potential conflict of interest in the process of writing this manuscript.

Appendix. Sampled stocks

Table A1
Descriptive statistics and tests for stationarity (News sentiment).

Ticker	Mean	Median	SD	Min	Max	bootUR	FT.Normality
AAPL	−0.0010	−0.0018	0.0862	−0.8542	0.7784	−3.3673***	86.8593***
AMGN	0.0608	0.0026	0.3772	−0.9991	0.9997	−2.0706***	20.1674***
AXP	−0.0111	0.0000	0.2000	−0.9869	0.9750	−3.4615***	63.8244***
BA	−0.0060	−0.0010	0.1483	−0.9815	0.8527	−3.0867***	82.6210***
CAT	0.0255	0.0000	0.3494	−0.9975	0.9933	−3.6323***	20.9695***
CRM	0.0590	0.0000	0.2042	−0.8527	0.9534	−4.9416***	37.9915***
CSCO	0.0162	0.0000	0.1743	−0.8997	0.9653	−3.9268***	73.5489***
CVX	−0.0022	0.0000	0.1604	−0.7995	0.8419	−3.9069***	59.4271***
DIS	−0.0003	−0.0002	0.0547	−0.5863	0.7938	−3.5180***	97.8415***
GS	−0.0258	−0.0023	0.1370	−0.9466	0.9566	−3.4599***	72.8500***
HD	0.0520	0.0037	0.2368	−0.9537	0.9860	−4.9925***	45.7016***
HON	0.0073	0.0000	0.1932	−0.9933	0.9117	−2.0637***	61.2458***
IBM	−0.0193	0.0000	0.2409	−0.9924	0.9893	−3.0957***	60.0605***
INTC	0.0027	0.0000	0.1063	−0.7041	0.8329	−3.9318***	85.6996***
JNJ	−0.0406	−0.0005	0.2247	−0.9976	0.9102	−2.9027***	45.9378***
JPM	−0.0155	−0.0026	0.0931	−0.8418	0.5378	−4.9483***	76.0772***
KO	0.0149	0.0000	0.1478	−0.9427	0.8802	−6.5666***	70.6448***
MCD	0.0007	0.0000	0.1550	−0.9359	0.9502	−3.5653***	74.1031***
MMM	−0.0337	0.0000	0.3643	−1.0000	1.0000	−2.8823***	19.9884***

(*continued on next page*)

Table A1 (*continued*)

Ticker	Mean	Median	SD	Min	Max	bootUR	FT.Normality
MRK	0.0601	0.0000	0.2690	−0.9933	0.9933	−2.8266***	38.9688***
MSFT	0.0066	0.0000	0.0492	−0.2681	0.7345	−2.5118***	100.735***
NKE	0.0003	0.0000	0.1635	−0.9667	0.9449	−4.1083***	70.0664***
PG	0.0383	0.0000	0.2759	−0.9991	0.9913	−3.9426***	39.4504***
TRV	−0.0392	0.0000	0.3637	−0.9999	0.9997	−2.8061***	16.8877***
UNH	0.0563	0.0000	0.3268	−0.9975	0.9933	−3.6730***	27.8993***
V	0.0168	0.0000	0.1602	−0.9292	0.8760	−3.7121***	67.8804***
VZ	0.0044	0.0000	0.1283	−0.9560	0.9806	−3.9246***	85.8540***
WBA	−0.0046	0.0000	0.3420	−0.9975	0.9933	−3.1106***	28.3487***
WMT	0.0050	−0.0002	0.1301	−0.9534	0.9048	−4.0691***	88.3278***

Notes: The table shows the mean, median, standard deviation (SD), minimum value (Min), maximum value (Max), bootUR statistics (Smeekes & Wilms, 2022) and Fat tail normality test (Jelito & Pitera, 2021) statistics. ***, **, * represents statical significance at the 1 %, 5 % and 10 % levels of significance.

Table A2

Descriptive statistics and tests for stationarity (Social media sentiment).

Ticker	Mean	Median	SD	Min	Max	bootUR	FT.Normality
AAPL	−0.0057	−0.0021	0.0377	−0.3241	0.3218	−2.7123***	78.4273***
AMGN	0.0292	0.0165	0.1162	−0.9516	0.8902	−2.7139***	54.4354***
AXP	0.0182	0.0065	0.1239	−0.9795	0.9153	−2.7413***	65.0485***
BA	−0.0133	−0.0033	0.1032	−0.9119	0.8893	−3.4204***	72.8338***
CAT	0.0225	0.0123	0.1776	−0.9904	0.9510	−3.6737***	51.6769***
CRM	0.0124	0.0053	0.0688	−0.9725	0.5887	−2.5379***	87.4591***
CSCO	0.0202	0.0103	0.0927	−0.8672	0.7404	−4.2951***	66.1571***
CVX	0.0063	0.0079	0.1267	−0.9711	0.8630	−2.9634***	54.9377***
DIS	0.0003	−0.0004	0.0410	−0.9757	0.5366	−4.4327***	95.9119***
GS	−0.0185	−0.0021	0.1280	−0.9448	0.7441	−3.3964***	60.2197***
HD	0.0384	0.0155	0.1661	−0.9577	0.8680	−3.5293***	41.5436***
HON	0.0345	0.0090	0.1299	−0.9747	0.9108	−2.3582***	52.7578***
IBM	0.0033	0.0013	0.0400	−0.4410	0.6598	−7.5433***	100.7231***
INTC	0.0024	0.0030	0.0638	−0.8226	0.9712	−2.6161***	79.0261***
JNJ	−0.0060	0.0011	0.1443	−0.9925	0.9650	−4.1858***	62.9547***
JPM	0.0032	0.0047	0.1665	−0.9797	0.7775	−2.8586***	46.1880***
KO	0.0030	0.0000	0.1376	−0.9618	0.9762	−2.7781***	90.6095***
MCD	0.0028	0.0005	0.1065	−0.9789	0.8647	−3.7006***	80.0208***
MMM	0.0142	0.0094	0.1654	−0.9010	0.9658	−2.7063***	44.5803***
MRK	0.0207	0.0122	0.0922	−0.8487	0.7864	−3.0613***	64.9895***
MSFT	0.0014	0.0014	0.0458	−0.6997	0.5960	−2.6051***	100.9573***
NKE	0.0157	0.0048	0.1618	−0.9608	0.9897	−2.4254***	59.6669***
PG	0.0254	0.0057	0.1145	−0.9158	0.7772	−3.1611***	57.3445***
TRV	0.0421	0.0045	0.2278	−0.9600	0.9502	−8.7739***	34.8702***
UNH	0.0540	0.0172	0.1623	−0.8793	0.9204	−4.0062***	44.6801***
V	0.0159	0.0092	0.0972	−0.9882	0.7453	−3.4651***	74.5340***
VZ	−0.0006	0.0006	0.0689	−0.6866	0.6781	−4.0161***	88.4192***
WBA	0.0027	0.0000	0.1116	−0.9872	0.8177	−3.1318***	61.1199***
WMT	−0.0209	−0.0021	0.1310	−0.9682	0.7823	−3.0622***	79.2956***

Notes: The table shows the mean, median, standard deviation (SD), minimum value (Min), maximum value (Max), bootUR statistics (Smeekes & Wilms, 2022) and Fat tail normality test (Jelito & Pitera, 2021) statistics. ***, **, * represents statical significance at the 1 %, 5 % and 10 % levels of significance.

References

Alexander, R. M., & Gentry, J. K. (2014). Using social media to report financial results. *Business Horizons, 57*(2), 161–167. https://doi.org/10.1016/j.bushor.2013.10.009

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

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *The Journal of Economic Perspectives, 21*(2), 129–151.

Beckers, S. (2018). Do social media trump news? The relative importance of social media and news based sentiment for market timing. *The Journal of Portfolio Management, 45*(2), 58–67. https://doi.org/10.3905/jpm.2018.45.2.058

Behrendt, S., Dimpfl, T., Peter, F. J., & Zimmermann, D. J. (2019). RTransferEntropy—Quantifying information flow between different time series using effective transfer entropy. *SoftwareX, 10*. https://doi.org/10.1016/j.softx.2019.100265

Behrendt, S., & Prange, P. (2021). What are you searching for? On the equivalence of proxies for online investor attention. *Finance Research Letters, 38*, Article 101401. https://doi.org/10.1016/j.frl.2019.101401

Bu, Q. (2023). Are all the sentiment measures the same? *Journal of Behavioral Finance, 24*(2), 161–170. https://doi.org/10.1080/15427560.2021.1949718

Chan, F., Durand, R. B., Khuu, J., & Smales, L. A. (2017). The validity of investor sentiment proxies: Sentiment proxies. *International Review of Finance, 17*(3), 473–477. https://doi.org/10.1111/irfi.12102

Chen, K., Li, X., Luo, P., & Zhao, J. L. (2021). News-induced dynamic networks for market signaling: Understanding the impact of news on firm equity value. *Information Systems Research, 32*(2), 356–377. https://doi.org/10.1287/isre.2020.0969

Dong, H., & Gil-Bazo, J. (2020). Sentiment stocks. *International Review Of Financial Analysis, 72*. https://doi.org/10.1016/j.irfa.2020.101573

Dufour, J. M., Pelletier, D., & Renault, E. (2006). Short run and long run causality in time series: Inference. *Journal of Econometrics, 132*(2), 337–362. https://econpapers.repec.org/article/eeeeconom/v_3a132_3ay_3a2006_3ai_3a2_3ap_3a337-362.htm.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., et al. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eismann, K., Posegga, O., & Fischbach, K. (2021). Opening organizational learning in crisis management: On the affordances of social media. *The Journal of Strategic Information Systems, 30*(4), Article 101692. https://doi.org/10.1016/j.jsis.2021.101692

Festinger, L. (1957). *A theory of cognitive dissonance (pp. xi, 291)*. Stanford University Press.

Gan, B., Alexeev, V., Bird, R., & Yeung, D. (2020). Sensitivity to sentiment: News vs social media. *International Review of Financial Analysis, 67*, Article 101390. https://doi.org/10.1016/j.irfa.2019.101390

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control, 2*, 329–352. https://doi.org/10.1016/0165-1889(80)90069-X

Grover, P., Kar, A. K., & Dwivedi, Y. (2022). The evolution of social media influence—A literature review and research agenda. *International Journal of Information*

Management Data Insights, 2(2), Article 100116. https://doi.org/10.1016/j.jjimei.2022.100116

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology, 67*(1), 232–238. https://doi.org/10.1002/asi.23456

Hong, H., Ye, Q., Du, Q., Wang, G. A., & Fan, W. (2020). Crowd characteristics and crowd wisdom: Evidence from an online investment community. *Journal of the Association for Information Science and Technology, 71*(4), 423–435. https://doi.org/10.1002/asi.24255

Jelito, D., & Pitera, M. (2021). New fat-tail normality test based on conditional second moments with applications to finance. *Statistical Papers, 62*(5), 2083–2108. https://doi.org/10.1007/s00362-020-01176-2

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization, 176*, 63–90. https://doi.org/10.1016/j.jebo.2020.03.002

Kar, A. K., Angelopoulos, S., & Rao, H. R. (2023). Guest Editorial: Big data-driven theory building: Philosophies, guiding principles, and common traps. *International Journal of Information Management, 71*, Article 102661. https://doi.org/10.1016/j.ijinfomgt.2023.102661

Kim, A., Moravec, P. L., & Dennis, A. R. (2023). *When do details matter?* News source evaluation summaries and details against misinformation on social media. *International Journal of Information Management, 72*, Article 102666. https://doi.org/10.1016/j.ijinfomgt.2023.102666

Kushwaha, A. K., Kar, A. K., & Dwivedi, Y. K. (2021). Applications of big data in emerging management disciplines: A literature review using text mining. *International Journal of Information Management Data Insights, 1*(2), Article 100017. https://doi.org/10.1016/j.jjimei.2021.100017

Maass, W., Parsons, J., Purao, S., Storey, V., & Woo, C. (2018). Data-driven meets theory-driven research in the Era of Big data: Opportunities and challenges for information systems research. *Journal of the Association for Information Systems, 19*(12). https://doi.org/10.17705/1jais.00526

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., et al. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management, 50*, 432–451. https://doi.org/10.1016/j.ijinfomgt.2019.07.011

McGregor, S. C., & Molyneux, L. (2020). Twitter's influence on news judgment: An experiment among journalists. *Journalism, 21*(5), 597–613. https://doi.org/10.1177/1464884918802975

Mcqueen, G., & Thorley, S. (1991). Are stock returns predictable? A test using markov chains. *The Journal of Finance, 46*(1), 239–263. https://doi.org/10.1111/j.1540-6261.1991.tb03751.x

Michail, D., Kanakaris, N., & Varlamis, I. (2022). Detection of fake news campaigns using graph convolutional networks. *International Journal of Information Management Data Insights, 2*(2), Article 100104. https://doi.org/10.1016/j.jjimei.2022.100104

Milas, C., Panagiotidis, T., & Dergiades, T. (2021). Does it matter where you search? Twitter versus traditional news media. *Journal of Money, Credit and Banking, 53*(7), 1757–1795. https://doi.org/10.1111/jmcb.12805

Mullainathan, S., & Shleifer, A. (2005). The Market for News. *The American Economic Review, 95*(4), 1031–1053.

Neuman, W., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication, 64*(2), 193–214. https://doi.org/10.1111/jcom.12088

Nyakurukwa, K., & Seetharam, Y. (2023a). Can textual sentiment partially explain differences in the prices of dual-listed stocks? *Finance Research Letters, 104529*. https://doi.org/10.1016/j.frl.2023.104529

Nyakurukwa, K., & Seetharam, Y. (2023b). From Shanghai to wall street: The influence of Chinese news sentiment on US stocks. *Journal of Behavioral Finance, 0*(0), 1–14. https://doi.org/10.1080/15427560.2023.2270100

Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., et al. (2023). The potential of generative artificial intelligence across disciplines: perspectives and future directions. *Journal of Computer Information Systems, 0*(0), 1–32. https://doi.org/10.1080/08874417.2023.2261010

Peng, J., Zhang, J., & Gopal, R. (2022). The good, the bad, and the social media: financial implications of social media reactions to firm-related news. *Journal of Management Information Systems, 39*(3), 706–732. https://doi.org/10.1080/07421222.2022.2096547

Ren, J., Dong, H., Padmanabhan, B., & Nickerson, J. V. (2021). How does social media sentiment impact mass media sentiment? A study of news in the financial markets. *Journal of the Association for Information Science & Technology, 72*(9), 1183–1197.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters, 85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of Finance, 52*(1), 35–55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x

Smales, L. A. (2016). Time-varying relationship of news sentiment, implied volatility and stock returns. *Applied Economics, 48*(51), 4942–4960. https://doi.org/10.1080/00036846.2016.1167830

Smeekes, S., & Wilms, I. (2022). *bootUR: An r package for bootstrap unit root tests* (arXiv: 2007.12249). arXiv. https://doi.org/10.48550/arXiv.2007.12249.

Solanki, K., & Seetharam, Y. (2014). *Is consumer confidence an indicator of JSE performance?* (SSRN Scholarly Paper 2504414). https://papers.ssrn.com/abstract=2504414.

Tandon, C., Revankar, S., Palivela, H., & Parihar, S. S. (2021). How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. *International Journal of Information Management Data Insights, 1*(2), Article 100035. https://doi.org/10.1016/j.jjimei.2021.100035

Tetlock, P. C. (2011). All the news that's fit to reprint: do investors react to stale information? *The Review of Financial Studies, 24*(5), 1481–1512. https://doi.org/10.1093/rfs/hhq141

Turner, C. M., Startz, R., & Nelson, C. R. (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics, 25*(1), 3–22. https://doi.org/10.1016/0304-405X(89)90094-9

Wilksch, M., & Abramova, O. (2023). PyFin-sentiment: Towards a machine-learning-based model for deriving sentiment from financial tweets. *International Journal of Information Management Data Insights, 3*(1), Article 100171. https://doi.org/10.1016/j.jjimei.2023.100171