Received: 5 August 2021 | Accepted: 19 December 2023

DOI: 10.1111/jfir.12380

ORIGINAL ARTICLE

The Journal of Financial Research

Tweets versus broadsheets: Sentiment impact on stock markets around the world

Baoqing Gan[1] | Vitali Alexeev[1,2] | Danny Yeung[1]

[1]Finance Department, UTS Business School, University of Technology Sydney, New South Wales, Australia

[2]Department of Economics and Finance, University of Guelph, Ontario, Canada

Correspondence
Baoqing Gan, Finance Department, UTS Business School, University of Technology Sydney, Bldg 8, L7, 14-28 Ultimo Rd, Ultimo, New South Wales, 2007, Australia.
Email: baoqing.gan@alumni.uts.edu.au

Abstract

We contrast sentiment derived from social and news media to investigate its impact across 14 international markets. We find that heightened media sentiment during nontrading periods significantly affects the next day's opening returns even after accounting for the previous-day activity. Markedly, only the US market exhibits strong reactions to social media, whereas other markets are more responsive to the news. We find that most variability in overnight returns is explained by sentiment aggregated 3 h before markets open. Our findings suggest that the overnight sentiment does not simply subsume previous-day market activity but contains additional information that helps improve predictability in return forecasting models.

JEL CLASSIFICATION

C58, G17, G41

"It isn't events that move markets, but reactions to them, so long as they are shared by a big enough bunch of traders."

---- The Economist, July 15, 2010

1 | INTRODUCTION

Recent literature has demonstrated that accurately measured social media sentiment can capture useful aggregate opinions on economic factors that sway the financial markets' movement (Azar & Lo, 2016; Sprenger, Sandner, et al., 2014; Yang et al., 2015). Indeed, Gan et al. (2020) show that social media sentiment has now eclipsed that of the news media as the dominant information channel for US stock markets. This growing and pervasive influence of social media is of interest to market participants but should elicit concerns for regulators. The recent frenzied rise of GameStop is a perfect example of how social media sentiment can be harnessed to generate price spikes and excessive volatility in the market ("Hero to Villain," 2021). The rising influence of social media is perhaps unsurprising as it now permeates every facet of our daily lives, shaping the consensus opinion inducing herd-like behavior. Although the literature has focused on US evidence in contrasting the effects of social and news media sentiment (e.g., Alexeev et al., 2022; Chen et al., 2014; Jiao et al., 2020), the question is whether a similar trend can be observed in other financial markets.

In the United States, the impact of sentiment in the market is shown to be statistically and economically significant at an intraday level (Deng et al., 2018; Renault, 2017; Sun et al., 2016). These US studies, however, concentrate on analyzing the sentiment and stock return interactions during trading hours, leading to a mutual sentiment–return causality loop—a tricky endogeneity issue to address. Instead, from an international perspective, we seek answers to the following questions: How does overnight sentiment affect the opening price? Is this impact different for social media compared to traditional news? Which markets are predominantly driven by social media? Does the effect of media pessimism exceed that of optimism?

To tackle these questions, we employ 1-min sentiment scores on 14 global stock markets from the Thomson Reuters MarketPsych Indices (TRMI) database.[1] To our knowledge, these entity-specific sentiment measures are the most granular among currently available data sets and cover both well-known developed markets and some affluent emerging markets. The availability of high-frequency sentiment scores allows us to sever the sentiment–return feedback loop and form a unique perspective to analyze the impact of media sentiment during nontrading sessions on market behavior at the opening. Controlling for variables known to determine the return rate, namely, previous-day return, volume, realized volatility, and the Chicago Board Options Exchange (CBOE) Volatility Index (VIX), we evaluate the impact of sentiment on overnight returns and contrast the effects of social media sentiment with that of the news. We account for asymmetries in the market behavior in response to optimistic and pessimistic attitudes in the media.

We confirm a positive relation between social (news) media sentiment and overnight returns for all the countries in our sample.[2] The more optimistic (pessimistic) the overnight sentiment is, the higher (lower) the next opening price (relative to previous close). We confirm the robustness of this result to different sentiment aggregation periods ranging from 30 min before markets open to amassing sentiment from the previous day's close. We find that most variability in overnight returns is explained when sentiment is aggregated over the 3 h before the market opens.

Our analyses reveal that the United States is the only market more sensitive to social media sentiment than news. A 1 SD increase in social media sentiment leads to a 1.17% increase in the Dow Jones Industrial Average (DJIA) overnight returns, which is statistically significant at the 1% level. Conversely, a 1 SD rise in news media sentiment causes only a 0.8% increase in the DJIA overnight returns, which is not statistically significant. In other markets, especially Hong Kong, Japan, India, and Singapore, news sentiment asserts a stronger impact on opening values than social media sentiment. For example, in Japan, overnight returns on the Nikkei 225 increase by 15.59% on a 1 SD increase in social media sentiment, whereas it rises remarkably by 28.16% on a 1 SD increase in news media sentiment. Japan, Hong Kong, India, and France appear to be swayed by news media sentiment much more strongly than Canada, Singapore, and the United Kingdom. The responses to news media sentiment in Australia and the United States appear to have little consequence.

[1]Although TRMI provides data for additional markets, the sparsity of these data prevents a meaningful cross-country comparison. For a summary of this database, see https://www.marketpsych.com/app

[2]For all the markets in our sample, the coefficients on sentiment variables are positive with at least one media type (news media, social media, or both) statistically significant at the 5% level.

We show that information is absorbed faster in markets with higher liquidity and greater media attention. For instance, examining sentiment and the returns of 11 Asian markets, Chen et al. (2013) argue that because of arbitrage restrictions, Asian markets react stronger to investor sentiment than do their developed market counterparts. Consistent with this, we find that the magnitude of market reactions to sentiment in the United States is smaller than in other markets. This finding is intuitive, given that US financial markets are among the most efficient and liquid markets.

It is important to distinguish large sentiment swings in the market. Prior studies have shown that the predictive power of sentiment is concentrated in high-sentiment periods (e.g., Agrawal et al., 2018; Alexeev et al., 2022; Berkman et al., 2012; Stambaugh et al., 2012). Accordingly, we explicitly account for extreme sentiment readings, represented by the top (most positive) and bottom (most negative) deciles, and uncover substantial differences in return responses at these extreme sentiment magnitudes. These highly polarized media tones tend to be more influential than moderate sentiment readings within interior deciles. We find that the markets in Australia, Canada, India, Japan, United Kingdom, and United States are more sensitive to elevated negative tones than positive ones. Hong Kong, in contrast, is the only market that exhibits higher responses to heightened positive sentiment. Our results reconfirm that negative sentiment generally has a stronger effect on stock markets than positive sentiment. This finding is in line with the recent works revealing the asymmetric effects between positive and negative sentiment (Agrawal et al., 2018; Akhtar et al., 2012; Deng et al., 2018; Smales, 2014).

One may argue that these results could be driven by the interval choice in which sentiment is accumulated. To verify the robustness of our results, we test a range of durations for sentiment aggregation and confirm the robustness of our findings. Specifically, we find that windows spanning from 30 min to 6 h before the market opening contain consistent predictive power. Interval durations of less than 30 min weaken the return predictability and do not provide consistent results; sentiment accumulated over such short intervals tends to be more volatile than sentiment accumulated in more extended periods. This is primarily due to the sparsity of available observations and the small sample size in shorter intervals. On the contrary, amassing sentiment from the previous close to the next open may dilute the predictive ability of overnight sentiment.

Two main implications emerge from this study. First, our finding of cross-country differences in return response to media tones indicates that one should be cautious in extrapolating evidence from one market to another. In this respect, we side with Xiong et al. (2020), who arrive at a similar conclusion, suggesting it would be unwise to hastily adopt the US evidence in other markets. Moreover, Griffin et al. (2011) find that the differences in the quality of news dissemination mechanisms can explain cross-country differences in stock price reactions. Similarly, Calomiris and Mamaysky (2019) find that news text flows about emerging markets contain more incremental information for predicting returns compared with the news about developed markets. Their topic-based sentiment analysis confirms that the nature of news in emerging and developed markets tends to be different. Whereas Calomiris and Mamaysky (2019) focus on 1-month- and 1-year-ahead returns, our investigation centers on intraday patterns contrasting the effects of social versus news media on overnight returns.

Second, our results paint a more complete picture of the propagation of overnight moods in media and their impact on market behavior. Important announcements are increasingly scheduled outside normal trading sessions (Bagnoli et al., 2005; Bradley et al., 2014; Jiang et al., 2012; Michaely et al., 2013). von Beschwitz et al. (2013) finds that coverage in news analytics, which affect the market in a separate and distinct way from the underlying informational content of the event, speeds up the market reaction. Heightened media coverage increases both stock price updates and trading volume in the first few seconds after an event. Using novel textual-based sentiment measures, we show that overnight moods in social and news media affect overnight returns and, subsequently, intraday return patterns. These findings are of great importance for investors and regulators in a highly connected world where social media platforms such as Twitter, Facebook, and StockTwits are gaining prevalence.

Using a novel data set of sentiment measures that are entity specific and covering international markets, we show the direct impact of overnight sentiment on market returns at the open. The importance of overnight episodes for markets has been the subject of lively debate in the recent behavioral finance literature. Berkman et al. (2012) shows that following high-attention days, irrational retail investors who prefer to post orders overnight create price pressure at the opening of the next trading day, leading to mispricing at the opening and a subsequent intraday reversal effect. Aboody et al. (2018) and Weißofner and Wessels (2019) examine and prove the suitability of overnight returns as a proxy for firm-specific sentiment in the US and other global markets, respectively. As pointed out by Baker et al. (2012), most studies investigating the return predictability from investor sentiment (e.g., Heston & Sinha, 2017; Renault, 2017; Sun et al., 2016) are US centered because of limited data availability for other markets. Benefiting from the availability of highly granular overnight sentiment data, we contribute to this line of literature by focusing on the causal relation between overnight sentiment and returns, mitigating the need for a proxy as in Aboody et al. (2018) and Weißofner and Wessels (2019). Furthermore, our data allow us to differentiate social media influences from that of traditional news and perform a comprehensive comparison across 14 of the most affluent financial markets. Whereas studies often apply sentiment measures that account for only a single source of sentiment,[3] studies that directly contrast the rising social media with the traditional financial news remain rare. In this respect, our research extends the works of Ahoniemi et al. (2015), Chen et al. (2018), and Boudoukh et al. (2018), which demonstrate the crucial role of textual analytics and sentiment in the current digital world.

The debate on whether sentiment is a momentum or contrarian return predictor remains highly contentious. Using weekly Google search index, Gao et al. (2019) find evidence that sentiment is a contrarian predictor in stock markets worldwide. In 36 of their 38 country samples, the authors find a negative relation between sentiment and the following week's market returns. In contrast, Han and Li (2017) find that investor sentiment is a reliable momentum predictor at a monthly frequency in China; that is, there is a positive relation between sentiment and subsequent returns. Our results provide additional evidence on this undetermined topic. What distinguishes our article is that we use alternative sentiment measures that are entity specific and available in a more instant fashion. This allows us to gauge the emotion swings for a specific market in a more timely manner and, therefore, conduct almost anatomical investigations across international markets. To the best of our knowledge, we are the first to employ textual-based 1-min sentiment scores to examine a sentiment–return relation across global markets. Hence, by examining return predictability in several international markets, we add value to the literature on sentiment contagion and its return predictability in international markets (e.g., Bai, 2014; Baker et al., 2012; Chen et al., 2013; Feldman & Liu, 2017; Hudson & Green, 2015).

2 | DATA AND METHODOLOGY

2.1 | Data source

Our sentiment data are from TRMI. Using its natural-language-processing algorithm, TRMI analyzes news and social media in real time to convert the quantity and variety of professional news and Internet messages into manageable information flows. The most granular indicators in TRMI are updated at a 1-min frequency. Specifically, we investigate the 14 international market indices listed in Table 1. For each index, TRMI scans and analyzes English language articles and posts referring to that index. Because there is a vast distinction in communication styles between social and news media, TRMI uses differentiated text analytic models to improve sentiment scoring accuracy for social and news media sources (Peterson, 2016). The TRMI social media source comprises the top 30% of over 2 million blogs, stock message boards, and other social media sites, for example, SeekingAlpha, Yahoo! Finance, and StockTwits. TRMI's social media analytics date back to 1998, when it began to analyze Internet forum

[3]For example, Tetlock (2007) considers sentiment contained in news columns from the *Wall Street Journal*, and Garcia (2013) examines effects from the *New York Times*. Chen et al. (2014) focus on the stock message board seekingalph. com, Sprenger, Tumasjan, et al. (2014) analyze short-period Twitter content, Siganos et al. (2014) extract sentiment from Facebook, and Da et al. (2011) derive their sentiment measure from the Google Search Volume index. Fang and Peress (2009) synthesize several popular US newspapers (*USA Today*, *Wall Street Journal*, *New York Times*, and *Washington Post*).

TABLE 1 Data sources.

TRMI	Resembling indices	Oxford-Man	Time zone	Trading hours
MPTRXUS30	DJIA 30	DJIA	UTC-5(−4)	9:30–16:00
MPTRXCA250	S&P/TSX Composite	S&P/TSX Composite	UTC-5(−4)	9:30–16:00
MPTRXBR50	IBRX 50	Bovespa	UTC-3(−2)	10:00–17:00
MPTRXGB100	FTSE 100	FTSE 100	UTC (+1)	8:00–16:30
MPTRXCH20	Swiss Market	Swiss Market	UTC + 1(+2)	9:30–17:00
MPTRXDE30	Deutsche Borse DAX 30	DAX 30	UTC + 1(+2)	9:00–17:30
MPTRXES35	IBEX 35	IBEX 35	UTC + 1(+2)	9:00–17:30
MPTRXEU50	EURO STOXX 50	EURO STOXX 50	UTC + 1(+2)	9:00–17:30
MPTRXFR40	CAC 40	CAC	UTC + 1(+2)	9:00–17:30
MPTRXIN50	Nifty 50	S&P CNX Nifty	UTC + 5:30	9:15–15:30
MPTRXHK50	Hang Seng	Hang Seng	UTC + 8	9:30–12:00; 13:00–16:00
MPTRXSG30	FTSE Straits Times	FTSE Straits Times	UTC + 8	9:00–12:00; 13:00–17:00
MPTRXJP225	Nikkei 225	Nikkei 225	UTC + 9	9:00–11:30; 12:30–15:00
MPTRXAU500	ASX All Ordinaries	ASX Ordinaries	UTC + 10(+11)	10:00–16:00

Note: Sentiment variables are obtained from Thomson Reuters MarketPsych Indices (TRMI); overnight returns, daily returns, and daily average realized volatility are from the Oxford-Man Institute of Quantitative Finance. The TRMI source data are dated using coordinated universal time (UTC), which we adjust to local exchange time to match the trading hours. Changes in daylight saving times across countries have been taken into account. Descriptive statistics for the Oxford-Man variables are reported in Appendix C.

and message board content. In 2008, TRMI added Twitter content, and in 2009, LexisNexis (Moreover Technologies then) social media content was incorporated. TRMI news sentiment includes Reuters news and a host of mainstream news sources in its total historical data set, such as *Wall Street Journal*, *New York Times*, and *Financial Times*. In 2005, the archive began including Internet news content from Moreover Technologies.[4]

The TRMI algorithm then scores the index's specific sentiment on social and news media based on the Loughran and McDonald (2014) financial dictionary.[5] We use 1-min sentiment scores from January 1, 2011 to November 30, 2017. These aggregated sentiment scores mimic and target specific entities via representative equity indices. We supplement the TRMI sentiment scores with daily stock market data from the Oxford-Man Institute of Quantitative Finance (hereafter, Oxford-Man) country index archives.[6]

One could argue that there might be substantial linguistic differences among social and news media worldwide. However, we do not examine non-English media sources, such as Japanese, Spanish, and Cantonese, for reasons of data comparability. Moreover, TRMI sentiment scores are based on a financial dictionary constructed from English words. Although financial dictionaries could be easily obtained for other languages, the sentiment scoring algorithm could not guarantee comparability. At the time of our analysis, TRMI released Japanese-based sentiment measures, but we chose English-based sentiment to keep the measurement consistent across all markets in our sample. Employing local language sentiment data from other providers may jeopardize our results' comparability.

[4]For more details on this data set, see Gan et al. (2020, Section 3.1, and supplementary appendix B).

[5]TRMI covers a plethora of securities and markets, including more than 12,000 companies, 36 commodities and energy subjects, 187 countries, 62 sovereign markets, 45 currencies, and since 2009, and more than 150 cryptocurrencies. For more details, see Thomson Reuters MarketPsych Indices 2.2 User Guide, March 23, 2016, Document Version 1.0.

[6]A summary of descriptive statistics for the Oxford-Man variables is in Appendix C.

Our sentiment variables range from −1 (maximally negative tone) to 1 (maximally positive tone), with a sentiment score of 0 representing neutral tonality. Heatmaps, day-of-the-week, and time-of-day groupings enable visualization of the vast high-frequency sentiment and stock return data to help identify patterns and irregularities in our data set. In Figures 1 and 2 (left-hand-side panels), using the DJIA, Financial Times Stock Exchange (FTSE) 100, and Nikkei 225 indices as the most prominent examples, we allot all available 1-min sentiment observations into pixelated heatmaps by time of the day (horizontal axis) on each day of our sample (vertical axis). The horizontal axis spans 12:00 AM to 11:59 PM with 1440 min in total, and the vertical axis covers the entirety of our sample period, totaling 2526 days. Each pixel represents a single 1-min observation. Positive values are shown in red, negative values in blue, and missing values as blank. A mixture of positive and negative sentiment scores brings out an overall purple hue, attesting to the frequent reversal in sentiment polarity at high frequencies. A strong tendency of DJIA social media to coincide with the exchange trading hours can be observed by contrasting the saturation of Figure 1a and Figure 2a data in the heatmaps. Coincidentally, such a pattern in the news media is less obvious but has more pronounced threads weaving through each morning "on the hour" (i.e., prominent ridges at the 6:00, 7:00, 8:00, and 9:00 AM marks in Figure 1a). In striking contrast, the flow of sentiment data for the FTSE 100 and Nikkei 225 indices exhibit a more pronounced activeness in the news media segment, coinciding with the exchange trading hours much closer than its social media counterpart (Figures 1c, 1e, 2c, 2e). This indicates substantial dissimilarities in information flows in the UK and Japanese markets compared to the US market. Panels on the right-hand side display the proportions of nonmissing observations in variables on the left-hand side and capture intraday and day-of-the-week patterns in these variables, including nontrading days (e.g., weekends and public holidays). Figures 1 and 2 highlight the nontrivial nature of sentiment analysis due to the irregularity of the data, especially in light of the asynchronicity with the returns.

2.2 | Model specifications

To evaluate the impact of cumulative sentiment from social or news media during nontrading hours on overnight returns, we employ the framework from Fraiberger et al. (2018). Specifically, we estimate the following set of baseline and hypothesis-specific regressions to capture various media sentiment effects and control for potential confounding factors:

$$Ro_t = \alpha + \beta_2 Rc_{t-1} + \beta_3 VLM_{t-1} + \beta_4 RV_{t-1} + \beta_5 VIX_{t-1} + \epsilon_t, \tag{1}$$

$$Ro_t = \alpha + \beta_1 Sent_t^k + \epsilon_t, \tag{2}$$

$$Ro_t = \alpha + \beta_1 Sent_t^k + \beta_2 Rc_{t-1} + \beta_3 VLM_{t-1} + \beta_4 RV_{t-1} + \beta_5 VIX_{t-1} + \epsilon_t, \tag{3}$$

$$\begin{aligned} Ro_t = {} & \alpha + \beta_1 Sent_t^k + \gamma_1 D_t^{-k} + \gamma_2 D_t^{+k} + \gamma_3 D_t^{-k} \times Sent_t^k + \gamma_4 D_t^{+k} \times Sent_t^k \\ & + \beta_2 Rc_{t-1} + \beta_3 VLM_{t-1} + \beta_4 RV_{t-1} + \beta_5 VIX_{t-1} + \epsilon_t, \end{aligned} \tag{4}$$

where Ro_t indicates the overnight return of an aggregate country index on day t (in percent) and is computed as $Ro_t = \log(Po_t/Pc_{t-1}) \times 100$, with Po_t and Pc_t denoting open and close index values on day t.[7] We incorporate control variables known to influence the market opening indices. Specifically, Rc_{t-1} is the close-to-close return of the country index on day $t-1$ (in percent), calculated as $Rc_{t-1} = \log(Pc_{t-1}/Pc_{t-2}) \times 100$. The lagged close-to-close return

[7]Equations (1)–(4) are estimated for each country separately. All variables, coefficients, and error terms are therefore country specific. We omit the country subscript for simplicity. That is, instead of $Ro_{t,i}$, we use Ro_t without loss of generality. A summary of country-specific equity indices, including the list of data sources, time zones, and trading hours, is provided in Table 1.

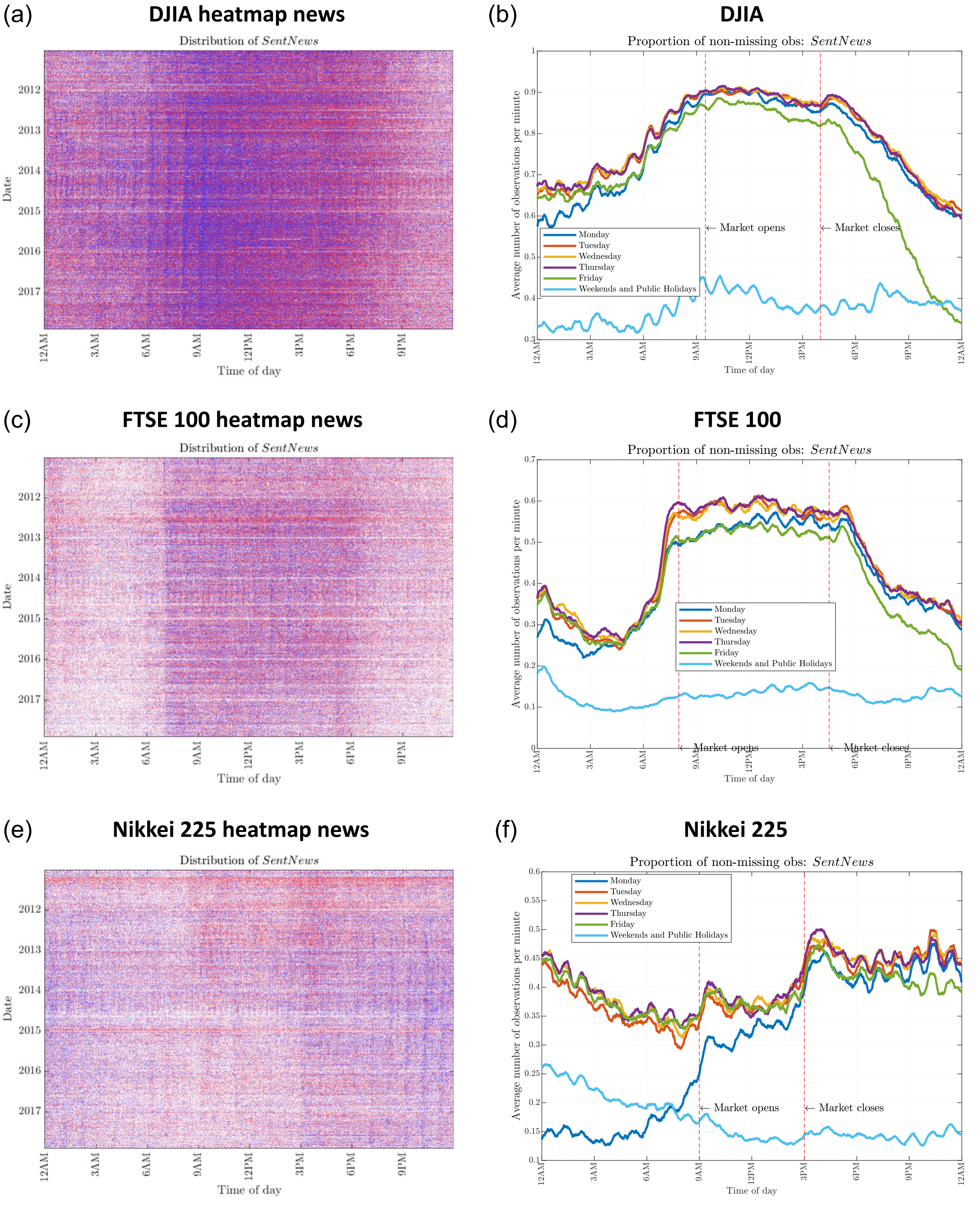

FIGURE 1 Heatmap of news media sentiment. This figure offers a visualization of the representative Thomson Reuters MarketPsych Indices (TRMI) news media indices and the proportion of nonmissing data by day of the week and minute of the day (horizontal axis). Three major markets are presented in the figure. By far, the DJIA is the most sentiment rich, as evidenced by the heatmaps' color saturation in the panels on the left. The peaks of news media activity fall within the trading times of the markets for the DJIA and FTSE 100. Only transient spikes corresponding to postopening and postclosing market times are observed for the Nikkei 225. Japan's financial market reliance on the US and UK markets is further evidenced by substantially low news media activity on Mondays, with activity picking up toward the market open times in Tokyo. (a) DJIA heatmap news, (b) DJIA, (c) FTSE 100 heatmap news, (d) FTSE 100, (e) Nikkei 225 heatmap news, (f) Nikkei 225. [Color figure can be viewed at wileyonlinelibrary.com]

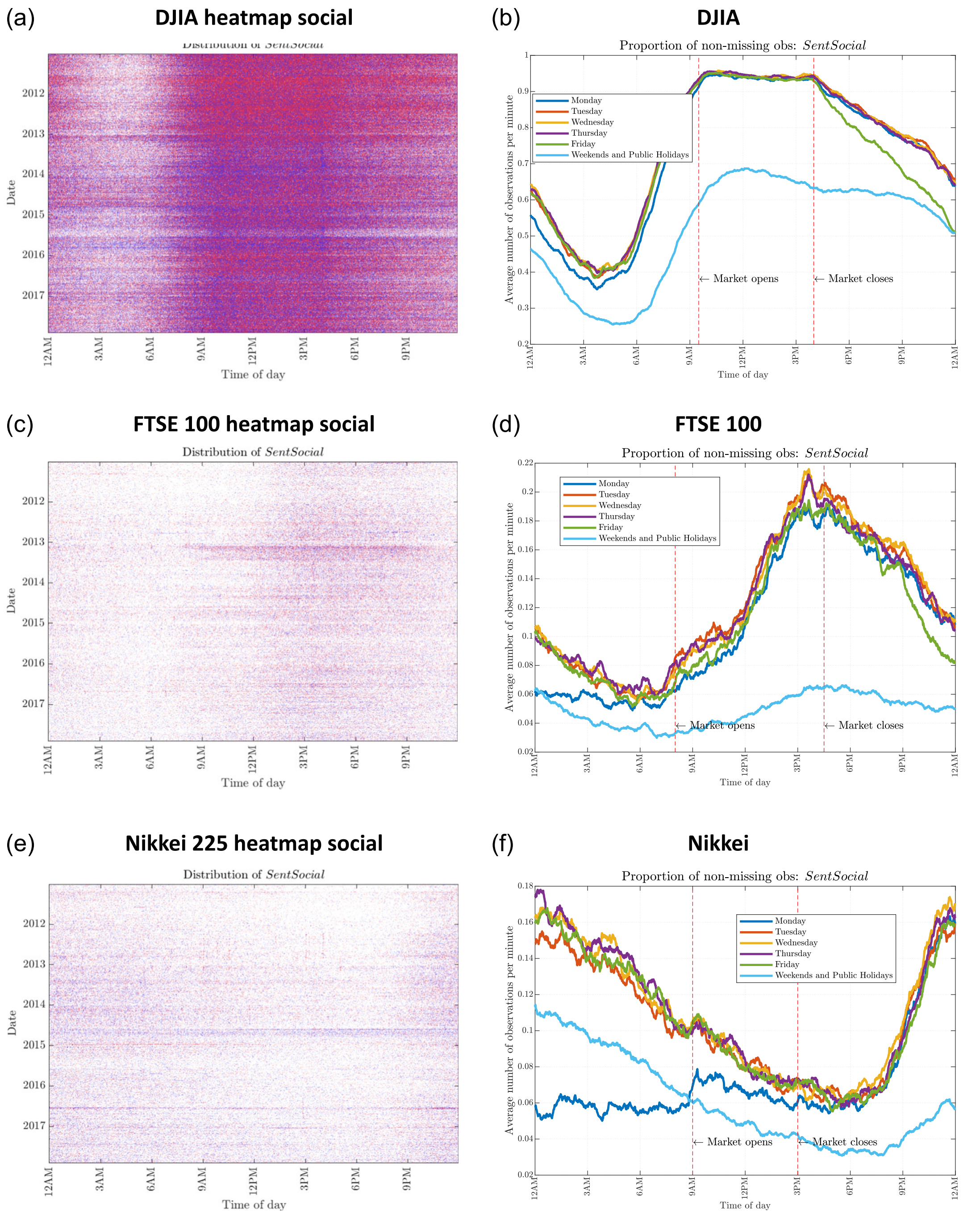

FIGURE 2 Heatmap of social media sentiment. This figure offers a visualization of the representative Thomson Reuters MarketPsych Indices (TRMI) social media indices and the proportion of nonmissing data by day of the week and minute of the day (the horizontal axis). By far, the DJIA is the most sentiment rich, as evidenced by the heatmaps' color saturation in the panels on the left. The peaks of social media activity fall within the trading times of the markets for the DJIA. Only transient spikes corresponding to postopening and postclosing market times are observed for the FTSE 100 and the Nikkei 225. Japan's financial market reliance on the US and UK markets is further evidenced by substantially low social media activity on Mondays, with activity picking up toward the trading hours in New York. (a) DJIA heatmap social, (b) DJIA, (c) FTSE 100 heatmap social, (d) FTSE 100, (e) Nikkei 225 heatmap social, (f) Nikkei. [Color figure can be viewed at wileyonlinelibrary.com]

is included to account for the autocorrelation resulting from possible market microstructure phenomena, such as nonsynchronized trading, bid–ask bounce, and trading costs. *VLM* denotes the demeaned log number of trades and is used to proxy for the changes in market liquidity. *RV* is the demeaned daily average realized volatility (in percent) used to account for changes in short-run market frictions other than the return autocorrelation. *VIX* is the demeaned log VIX, which controls for the general sentiment swings in the global market.[8] *Sent* stands for media sentiment, the superscript k denotes the source of sentiment data and can be either S or N depending on whether the source of the sentiment is social or news media, respectively. Although the effect of $Sent_t^k$ on Ro_t may appear contemporaneous in Equations (1)–(4), our construction of overnight sentiment variables ensures accumulation of 1-min sentiment scores after the previous day close and before the market opening time on day t.

Equation (1) is a baseline model that includes a set of controls but omits sentiment. Comparing models that include sentiment to this benchmark allows for an evaluation of changes in the model predictability due to the inclusion of sentiment variables. An alternative benchmark model in Equation (2) that includes only the sentiment variable assists in determining whether overnight sentiment subsumes information from the previous day's trading activity captured by the controls in Equation (1). In Equation (3), we focus on the single sentiment variable while accounting for the set of controls. Finally, our most flexible model in Equation (4) allows for asymmetric effects and pockets of heightened sentiment via binary regressors and their interactions with the sentiment variables.

Our aim is to investigate the effect of sentiment on overnight returns in Equations (1)–(4) while allowing for several alternative control variables commonly used in the literature.[9] There are several reasons why this causal relation deserves attention. Measuring media sentiment over nontrading periods helps avoid any endogeneity concerns by breaking the return–sentiment loop; using overnight sentiment offers a more accurate signal-to-noise proxy for what can be a very noisy measure. Nowadays, overnight information flows and the emotions expressed within them are of greater importance than they used to be (Ahoniemi et al., 2015). As the world becomes increasingly interconnected, regional and global events can trigger investor reactions across multiple markets. It stands to reason that one needs to consider the similarities and differences between the US and other global markets.

In our setup, $Sent_t^k$ is the focal independent variable, the standardized[10] average cumulative sentiment before the market opens on day t, from media type k, where $k = N$ for news media and $k = S$ for social media. For our main discussion, we define $Sent_t^k$ over the window at which we aggregate sentiment from the previous day close to the current day open; for example, for the United States, from 4:00 PM on day $t - 1$ to 9:29 AM on day t. For robustness, we probe different window lengths over the nontrading hours to gauge emotions between 2 consecutive trading days. The results of close-to-open sentiment analysis are reported here, but we explore various sentiment windows and discuss robustness checks in Section 4.

Our key independent variable, $Sent_t^k$, is constructed from the intraday sentiment scores provided by TRMI. On each day, we use unequally distanced data 1-min frequencies to compute average cumulative sentiment. To maintain comparability with different levels of media coverage volume across markets, we avoid prefilling the missing observations.[11] Specifically, if $x_{t,j}^k$ denotes raw sentiment from media type k on day t at time j, the close-to-open average cumulative sentiment on day t is the cumulative sum of all available 1-min sentiment scores (positive and negative) from the market closing time (τ_c) on day $t - 1$ to the market open time (τ_o) on day t, divided by the number of

[8]Different from Fraiberger et al. (2018), who use detrended log trading volume and detrended volatility, we use the demeaned variables of log trades, realized volatility, and the VIX to improve the comparability of regression constants. The demeaned results are computed as the daily observations in excess of sample averages.

[9]Appendix A provides a detailed list of variable names and definitions.

[10]Unlike Fraiberger et al. (2018), who normalize the news-based sentiment index, we use the term *standardize* instead. Although several studies use the terms *normalization* and *standardization* interchangeably, we differentiate them. Standardization is the process of demeaning and unifying variance, in other words, obtaining the z-score, whereas normalization is the process of rescaling variables between 0 and 1.

[11]The TRMI sentiment data we use span 24 h at the highest possible frequency of 1 min. However, when there are no postings on social media or no articles in the news about a specific entity, sentiment scores are represented as missing values or "not a number."

available observations over the same duration.[12] In other words, the average cumulative overnight sentiment is computed as follows:

$$X_t^k = \frac{1}{n_t^k}\sum_{j=\tau_c}^{\tau_o} x_{t,j}^k,$$

where n_t^k is the cardinality (i.e., number of nonmissing elements) of a set of overnight sentiment scores from media type k terminating on day t. This definition implicitly assumes that $\tau_c \in t-1$ and $\tau_o \in t$. If the entire overnight period contains no observations, we assign a 0 value to the cumulative sentiment for that period to maintain a neutral emotion. Given the diversity of the markets in our sample and their sentiment score variability, it is important to standardize the average cumulative sentiment. We compute it as follows:

$$Sent_t^k = \frac{X_t^k - \overline{X^k}}{\sigma_{X^k}},$$

where $\bar{X}^k$ is the mean score of X_t^k averaged across days, and σ_{X^k} is the sample standard deviation.[13]

To account for the impact from polarized (strong positive or negative) sentiment, we further include two binary regressors, D_t^{+k} and D_t^{-k}, in Equation (4) to indicate the top and bottom decile days of $\{Sent_t^k\}_{\forall t}$. Furthermore, we add interaction terms between these dummy variables and sentiment to capture the magnitude of the effect from the highly polarized sentiment.[14]

2.3 | Descriptive statistics

We rescale and transform the variables to generate comparable regression coefficients across the markets. For example, as shown in the raw data columns in Table 2, the number of trades (*VLM*) and the VIX (*VIX*) are at much higher magnitudes, whereas realized volatility (*RV*) is at a much lower level compared to the other variables. This scale difference presents difficulty in interpreting regression coefficients.

We standardize $Sent^S$ and $Sent^N$ to improve the comparability of these key independent variables across multiple markets. Because standardization results in zero means and unit standard deviations in sentiment variables, the interpretation of the coefficient estimates and their economic significance is simplified. This rescaling procedure ensures we are quantifying the causality we set out to measure at comparable levels. Hence, our subsequent discussions are based on the transformed variables unless stated otherwise.

2.4 | Model validity

Before estimating the models in Equations (1)–(4), we assess the pairwise correlations between all continuous variables to alleviate concerns over possible collinearity or omitted variable bias. The implications of collinearity and omitted variable bias could be dire, potentially resulting in biased estimates, high standard errors of the regression estimates, large changes in the coefficients when adding predictors, and opposite signs of the coefficients from

[12] Market closing and opening times are listed in Table 1.

[13] That is, $\bar{X}^k = \frac{\sum_{t=1}^{T} X_t^k}{T}$ and $\sigma_{X^k} = \sqrt{\frac{\sum_{t=1}^{T}\left(X_t^k - \bar{X}^k\right)^2}{T-1}}$.

[14] In unreported results, we tested for different magnitudes of "polarized" average cumulative sentiment, constructing dummy variables based on the top and bottom quintiles instead of deciles, and found consistent results.

TABLE 2 Descriptive statistics.

Variable	Raw data Mean	SD	Min.	Max.	Rescaled data Mean	SD	Min.	Max.
Panel A: Dow Jones Industrial Average (DJIA)								
Ro	0.0000	0.0010	−0.0063	0.0084	0.0039	0.0979	−0.6307	0.8366
$Sent^S$	−0.0055	0.0450	−0.2872	0.1934	0.0000	1.0000	−6.2605	4.4206
$Sent^N$	0.0120	0.0449	−0.1838	0.1760	0.0000	1.0000	−4.3593	3.6532
Rc	0.0004	0.0084	−0.0556	0.0402	0.0411	0.8383	−5.5624	4.0179
VLM	16,009.9	6353.6	4,899.0	23,412.0	0.0000	0.4046	−1.1030	0.4612
RV	0.000064	0.000178	0.000002	0.005946	0.0000	0.0178	−0.0062	0.5882
VIX	16.36	5.60	9.14	48.00	0.0000	0.2873	−0.5366	1.1219
Panel B: Financial Times Stock Exchange (FTSE) 100								
Ro	0.0002	0.0062	−0.0666	0.0317	0.0227	0.6243	−6.6640	3.1714
$Sent^S$	0.0042	0.0777	−0.2668	0.2662	0.0000	1.0000	−3.4886	3.3728
$Sent^N$	−0.0248	0.0624	−0.2881	0.2000	0.0000	1.0000	−4.2180	3.5995
Rc	0.0001	0.0094	−0.0495	0.0373	0.0126	0.9352	−4.9549	3.7257
VLM	68,291.7	21,294.6	11,899.0	314,308.0	0.0000	0.2940	−1.7043	1.5697
RV	0.000047	0.000073	0.000004	0.001596	0.0000	0.0073	−0.0043	0.1549
VIX	16.36	5.60	9.14	48.00	0.0000	0.2873	−0.5366	1.1219

Note: This table presents summary statistics for the data used in our analysis. The left-hand side columns present descriptive statistics for original (raw) data before applying any transformations. The right-hand side columns present descriptive statistics for each regression variable after we made the following transformations: ***Ro*** and ***Rc*** are expressed in percentage; $\boldsymbol{Sent^S}$ and $\boldsymbol{Sent^N}$ are standardized to have zero means and unit standard deviations; ***VLM*** and ***VIX*** take logarithm formats first and are then demeaned; ***RV*** is demeaned and transformed to percent. The variables are defined in Appendix A. The daily sentiment data are constructed by aggregating 1-min sentiment data from the previous day close to the next day open. For example, for the US market, the overnight sentiment on day $\boldsymbol{t}$ is constructed by aggregating 1-min sentiment data from 4:00 PM on day $\boldsymbol{t-1}$ to 9:29 AM on day $\boldsymbol{t}$. Similarly, for the UK market, the aggregation window is from 4:30 PM to 7:59 AM. Summary statistics for the other 12 markets are omitted for brevity. Time subscripts in the variables are omitted because the distinction between $\boldsymbol{t}$ and $\boldsymbol{t-1}$ is of no consequence to the univariate descriptive statistics.

what is anticipated in theory. We demonstrate our preestimation assessment procedure in Table 3, using the DJIA and FTSE 100 as examples. It is unsurprising that the highest correlation in the DJIA data is 0.44 between VIX_{t-1} and RV_{t-1}, followed by 0.38 between $Sent_t^S$ and VLM_{t-1} and 0.37 between $Sent_t^S$ and $Sent_t^N$. Similarly, the highest correlation in the FTSE 100 data is 0.60 between VIX_{t-1} and RV_{t-1}, followed by 0.35 between VLM_{t-1} and RV_{t-1} and 0.30 between $Sent_t^S$ and $Sent_t^N$. In the US market, $Sent_t^S$ shows stronger correlations with other variables than does $Sent_t^N$. In the UK market, however, the difference in correlation with other variables between $Sent_t^S$ and $Sent_t^N$ is less prominent, implying an important distinction between the FTSE 100 and DJIA in terms of information transmission from the two media sources. The magnitudes of correlation coefficients in Table 3 indicate no evidence of collinearity and a low possibility of omitted variable bias in the model.[15]

[15]In Appendix D, we report the results of variance inflation factors (VIFs) to assess the severity of multicollinearity. Although proper collinearity diagnostic tests such as in Farrar and Glauber (1967) and Belsley et al. (1980) may be performed, the magnitudes of the correlation coefficients and the VIFs do not warrant these tests. The reported results include DJIA and FTSE 100 only; the results for other markets are similar and available upon request from the authors.

TABLE 3 Pairwise correlation coefficients: DJIA and FTSE 100 close to open.

	Ro_t	$Sent_t^S$	$Sent_t^N$	Rc_{t-1}	VLM_{t-1}	RV_{t-1}	VIX_{t-1}
Panel A: Dow Jones Industrial Average (DJIA) pairwise correlations							
Ro_t	1.00						
$Sent_t^S$	0.05**	1.00					
$Sent_t^N$	0.04*	0.37***	1.00				
Rc_{t-1}	0.05**	0.03	0.04	1.00			
VLM_{t-1}	0.11***	0.38***	−0.01	−0.01	1.00		
RV_{t-1}	−0.01	−0.11***	−0.05**	−0.16***	−0.04	1.00	
VIX_{t-1}	−0.05*	−0.33***	−0.07***	−0.15***	−0.33***	0.44***	1.00
Panel B: Financial Times Stock Exchange (FTSE) 100 pairwise correlations							
Ro_t	1.00						
$Sent_t^S$	0.06***	1.00					
$Sent_t^N$	0.23***	0.30***	1.00				
Rc_{t-1}	−0.01	0.03	0.01	1.00			
VLM_{t-1}	0.06**	0.00	−0.05**	−0.08***	1.00		
RV_{t-1}	0.07***	−0.09***	−0.14***	−0.13***	0.35***	1.00	
VIX_{t-1}	−0.03	−0.16***	−0.26***	−0.11***	0.20***	0.60***	1.00

Note: To assist in checking possible omitted variable bias, this table lists the pairwise Pearson correlation coefficients and their respective significance levels between variables in Equation (3). The variables are defined in Appendix A. Sentiment is aggregated from the previous day's close to the next day's open, from 4:00 PM on day $t - 1$ to 9:29 AM on day t.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

As shown in Table 3, the correlation between overnight social and news media sentiment is around 0.3; that is, the correlation between $Sent^S$ and $Sent^N$ is 0.37 for social media in Panel A and 0.30 for news media in Panel B. As a result, the top and bottom decile dummy indicators we generate based on these overnight sentiments may contain the same tail-event days, involving a possible double-counting problem if the strong overlapping rate of these dummy variables existed and we incorporated them together. Thus, another important procedure before estimating regression coefficients is to check the coincidence rates in the dummy variables to identify the heightened sentiment levels. We make this evaluation for every country and report the results of the DJIA and the FTSE 100 in Table 4 for illustration. Both panels of Table 4 show that the coincidence rates between D_t^{-S} and D_t^{+S}, and between D_t^{-N} and D_t^{+N}, are 0, which is intuitively correct by construction. We discover a low coincidence rate in polarized emotions across social and news media sources in both the DJIA and FTSE 100 data sets. In particular, the coincidence rates of the most negative sentiment (bottom decile) social and news media days (D_t^{-S} and D_t^{-N}) are only 3.39% for the DJIA and 2.65% for the FTSE 100. Similarly, the coincidence rates between the most positive (top decile) social and news media days (D_t^{+S} and D_t^{+N}) are 2.76% for the DJIA and 1.90% for the FTSE 100. Our analysis of the other 12 markets reveals a similar pattern: The coincidence rates of highly negative sentiment from social and news media are higher than the coincidence rates of highly positive sentiment from these two sources. We also find that the coincidence rates between the dummy variables that indicate the opposite polarities across different media sources (i.e., D_t^{+S} and D_t^{-N}, D_t^{+N} and D_t^{-S}) are all lower than 1%. This evidence further corroborates our assertion that only a diminutive possibility of coincidence exists in the extremely positive (negative) sentiment from one media source to another.

TABLE 4 Coincidence rates in extreme sentiment across media platforms.

	D_t^{-S}	D_t^{+S}	D_t^{-N}	D_t^{+N}
Panel A: DJIA sentiment coincidence rates				
D_t^{-S}	100%			
D_t^{+S}	0.00%	100%		
D_t^{-N}	3.39%	0.35%	100%	
D_t^{+N}	0.12%	2.76%	0.00%	100%
Panel B: FTSE 100 sentiment coincidence rates				
D_t^{-S}	100%			
D_t^{+S}	0.00%	100%		
D_t^{-N}	2.65%	0.29%	100%	
D_t^{+N}	0.23%	1.90%	0.00%	100%

Note: This tables presents close-to-open overnight sentiment for the Dow Jones Industrial Average (DJIA) and Financial Times Stock Exchange (FTSE) 100 indices data sets. We check the coincidence rates (in percent) among the four dummy variables representing strong positive or negative sentiment from social and news media. The dummy variables are based on sentiment aggregated from the previous day close to the next day open, from 4:00 PM on day $\boldsymbol{t-1}$ to 9:29 AM on day $\boldsymbol{t}$. Similarly, for the UK market, the aggregation window is from 4:30 PM to 7:59 AM. The coincidence rates for the other 12 markets are similar and omitted for brevity. $\boldsymbol{D_t^{\pm S} = 1}$ if the average cumulative social media sentiment (S) belongs to the top **(+)** or bottom **(−)** decile, $\boldsymbol{D_t^{\pm N} = 1}$ if the average cumulative news media sentiment (N) is ranked in the top **(+)** or bottom **(−)** decile.

3 | SENSITIVITY TO OVERNIGHT SENTIMENT IN GLOBAL MARKETS

In this section, we formally analyze the global market sensitivity patterns based on overnight (close-to-open) social media and news sentiment data. For each country in the sample, we estimate Equations (1)–(4) using an ordinary least squares (OLS) regression. For illustration purposes, Tables 5 and 6 report the estimates using DJIA and FTSE 100 data, respectively.[16] We summarize the results for the 14 markets in Table 7 to contrast the similarities and differences in market responses to changes in social and news sentiment.

We report the estimates of the baseline model in Column 1 of Tables 5 and 6. This model does not include the key independent variable, $Sent_t^k$, but does incorporate four variables controlling for the impact of the previous-day trading activities on overnight returns (Rc_{t-1}, VLM_{t-1}, RV_{t-1}, and VIX_{t-1}). In the US market, we observe in Table 5 that the previous-day return and trading volume significantly affect the DJIA's overnight return the following day. The estimated positive coefficients suggest that the higher the previous-day return and volume, the higher the index values at the next day's opening. In other words, daily return autocorrelation and changes in the market liquidity significantly affect the DJIA's overnight return. In contrast, for the UK market, the change in daily realized volatility and the variation in the VIX strongly influence the FTSE 100's overnight returns (Column 1 of Table 6). A heightened level of realized volatility on the previous day leads to a statistically significant increase in the overnight return of the FTSE 100. However, the increased global "fear" index (VIX) on the previous day triggers a decline in the index at the market opening the following morning.

[16]The detailed estimates for other countries are obtained similarly but omitted here for brevity. For each market and model specification, we use the maximum available number of observations, which leads to variations in sample sizes, for example, because of the inclusion of lagged variables or differences in nontrading days across countries.

TABLE 5 Sentiment sensitivity of the DJIA.

Dependent variable = Overnight DJIA return on day t (in percent)							
		Social media (k = S)			News media (k = N)		
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0037	0.0039*	0.0037	0.0039	0.0039*	0.0037	0.0030
	(1.39)	(1.68)	(1.41)	(1.32)	(1.68)	(1.40)	(1.02)
$Sent_t^k$		0.0053**	0.0014	0.0117***	0.0040*	0.0040*	0.0008
		(2.24)	(0.53)	(2.71)	(1.72)	(1.71)	(0.19)
D_t^{-k}				−0.05**			0.0203
				(−2.32)			(0.76)
D_t^{+k}				0.0032			−0.0199
				(0.12)			(−0.75)
$D_t^{-k} \times Sent_t^k$				−0.0433***			0.0131
				(−3.41)			(0.90)
$D_t^{+k} \times Sent_t^k$				−0.0130			0.0178
				(−0.90)			(1.15)
Rc_{t-1}	0.0056**		0.0056**	0.0062**		0.0055*	0.0056**
	(1.98)		(1.98)	(2.20)		(1.95)	(1.98)
VLM_{t-1}	0.0256***		0.0245***	0.0204***		0.0259***	0.0259***
	(4.13)		(3.79)	(3.11)		(4.19)	(4.17)
RV_{t-1}	−0.0046		−0.0055	−0.0187		−0.0016	0.0037
	(−0.03)		(−0.04)	(−0.13)		(−0.01)	(0.02)
VIX_{t-1}	−0.0012		−0.0001	0.0044		−0.0001	−0.0005
	(−0.12)		(−0.002)	−0.44		(−0.01)	(−0.05)
Adj. R^2	0.0114	0.0023	0.0110	0.0168	0.0013	0.0125	0.0116
F-statistic	6.00	5.02	4.85	4.29	2.96	5.39	3.27
p-value	0.00	0.03	0.00	0.00	0.09	0.00	0.00
Obs.	1738	1739	1738	1738	1739	1738	1738

Note: This table summarizes the regression results based on Equations (1)–(4) for the Dow Jones Industrial Average (DJIA) index. The variables are defined in Appendix A. The t-statistics are provided in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

We proceed by gradually introducing sentiment variables based on model variations in Equations (2)–(4) and report our results in Tables 5 and 6. Columns 2–4 present results based on social media, and Columns 5–7 are based on news media. For the US market, we observe that although the impact of overall sentiment from social media in Column 2 is significant and the coefficient is of the expected positive sign, the effect is marginal. Moreover, this effect is insignificant after controlling for the previous-day market activity (see Column 3). However, when we consider the asymmetry and effects of extreme sentiment, the evidence is consistent with our expectations: The impact of extreme negative sentiment is much more pronounced, whereas the effect of extreme positive sentiment is insignificant. This evidence is in line with Alexeev et al. (2022), who show that

TABLE 6 Sentiment sensitivity of the FTSE 100.

FTSE 100 dependent variable = Overnight return on day *t* (in percent)							
		Social media (k = S)			**News media (k = N)**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
Constant	0.0227	0.0227	0.0227	0.0325*	0.0227	0.0227	0.0487***
	(1.53)	(1.52)	(1.53)	(1.95)	(1.56)	(1.57)	(3.04)
$Sent_t^k$		0.0391***	0.0355**	0.0186	0.1447***	0.1491***	0.1227***
		(2.62)	(2.36)	(0.71)	(9.93)	(9.95)	(4.82)
D_t^{-k}				−0.0889			0.5100***
				(−0.44)			(2.97)
D_t^{+k}				−0.0009			−0.0197
				(−0.00)			(−0.11)
$D_t^{-k} \times Sent_t^k$				0.0044			0.3833***
				(0.04)			(4.03)
$D_t^{+k} \times Sent_t^k$				−0.0002			−0.0351
				(−0.00)			(−0.33)
Rc_{t-1}	0.0008		0.0003	0.0004		0.0045	0.0057
	(0.05)		(0.02)	(0.03)		(0.28)	(0.37)
VLM_{t-1}	0.0825		0.0787	0.0808		0.0874*	0.1017*
	(1.52)		(1.45)	(1.49)		(1.66)	(1.94)
RV_{t-1}	10.834***		10.808***	10.697***		10.461***	11.083***
	(4.04)		(4.04)	(3.99)		(4.01)	(4.27)
VIX_{t-1}	−0.2453***		−0.2246***	−0.2285***		−0.1062*	−0.1078*
	(−3.79)		(−3.45)	(−3.49)		(−1.65)	(1.68)
Adj. R^2	0.0126	0.0034	0.0152	0.0144	0.0532	0.0655	0.0791
F-statistic	6.56	6.84	6.37	3.82	98.7	25.3	17.6
p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Obs.	1738	1739	1738	1738	1739	1738	1738

Note: This table summarizes the regression results of Equations (1)–(4) of the Financial Times Stock Exchange (FTSE) 100 index. The variables are defined in Appendix A. The ***t***-statistics are provided in parentheses.
*p < 0.10; **p < 0.05; ***p < 0.01.

sentiment from extreme deciles often possesses a higher signal-to-noise ratio than does sentiment from moderate deciles. This finding is consistent with the large body of empirical literature demonstrating that the influence of negative investor sentiment prevails over positive sentiment (e.g., Akhtar et al., 2012; Sprenger, Sandner, et al., 2014; Stambaugh et al., 2012, 2014).

Our results for the US markets indicate that accumulated positive overnight social media sentiment leads to significant increases in the overnight market return. Without controlling for the previous-day trading activities, a 1 SD increase (decrease) in cumulative social media sentiment is associated with 0.53% higher (lower) opening returns the next day (Column 2 of Table 5). Similarly, a 1 SD increase (decrease) in cumulative social media

TABLE 7 Global market sensitivity to overnight sentiment.

	US	UK	Canada	Brazil	Australia	Switzerland	Germany	Spain	EU	France	HK	Japan	India	Singapore
	DJIA	FTSE100	TSX	Bovespa	AORD	Swiss	DAX30	IBEX35	STOXX	CAC40	HangSeng	Nikkei225	Nifty	FTStraits
Panel A: Social media														
Constant	0.0039	0.0325*	0.0214***	0.0545***	0.0079	0.0243	0.0371*	0.0685***	0.0154	0.0276	0.0810***	0.0649***	0.0379**	0.0601***
	(1.32)	(1.95)	(3.04)	(2.84)	(1.05)	(1.57)	(1.89)	(3.20)	(1.32)	(1.45)	(3.46)	(2.89)	(2.48)	(5.21)
$Sent_t^S$	0.0117***	0.0186	0.0105	0.0329	0.0257**	0.0215	0.1230***	−0.0074	0.0424**	0.0766**	0.1191***	0.1559***	0.0721***	0.0480**
	(2.71)	(0.71)	(0.95)	(0.99)	(2.10)	(0.87)	(4.06)	(−0.21)	(2.26)	(2.47)	(3.20)	(4.48)	(3.018)	(2.25)
D_t^{-S}	−0.05**	−0.0889	−0.1046	−0.2972*	−0.0618	−0.0011	−0.0821	−0.2307	−0.0784	−0.2354	−0.0939	0.5516**	−0.0840	−0.0335
	(−2.32)	(−0.44)	(−1.25)	(−1.66)	(−0.90)	(−0.01)	(−0.35)	(−1.01)	(−0.68)	(−1.15)	(−0.36)	(2.34)	(−0.52)	(−0.39)
D_t^{+S}	0.0032	−0.0009	0.1066	0.0194	−0.0334	0.056	0.0442	0.0217	0.1576	−0.0953	0.4420*	0.3680	−0.1702	−0.0126
	(0.12)	(−0.00)	(1.22)	(0.66)	(−0.57)	(0.33)	(0.21)	(0.11)	(1.50)	(−0.44)	(1.71)	(1.32)	(−1.17)	(−0.12)
$D_t^{-S} \times Sent_t^S$	−0.0433***	0.0044	−0.0472	−0.0801	−0.0588	0.0130	−0.1431	−0.0396	−0.0693	−0.1535	−0.1016	0.2861**	−0.0446	−0.0414
	(−3.41)	(0.04)	(−0.96)	(−0.73)	(−1.51)	(0.12)	(−1.07)	(−0.30)	(1.06)	(−1.35)	(−0.70)	(2.18)	(−0.50)	(−0.83)
$D_t^{+S} \times Sent_t^S$	−0.0130	−0.0002	−0.0394	−0.0721	−0.0047	−0.0457	−0.1179	−0.0379	−0.1066*	0.0079	−0.2506*	−0.2355	0.0540	−0.0364
	(−0.90)	(−0.00)	(−0.84)	(−0.84)	(−0.14)	(−0.49)	(−0.98)	(−0.36)	(−1.77)	(0.07)	(−1.74)	(−1.44)	(0.64)	(−0.64)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
adj. R^2	0.0168	0.0144	0.0354	0.0078	0.0092	−0.0001	0.017	0.0155	0.0146	0.0118	0.031	0.0556	0.0257	0.0543
F-stat	4.29	3.82	8.08	2.51	2.8	0.871	4.35	4.04	3.86	3.31	7.18	12.4	6.10	12.1
Panel B: News media														
Constant	0.0030	0.0487***	0.022***	0.0514***	0.0158**	0.0266*	0.0465**	0.0524**	0.0219*	0.0403**	0.0891***	0.0601***	0.0281*	0.0619***
	(1.02)	(3.04)	(3.15)	(2.69)	(2.12)	(1.72)	(2.39)	(2.46)	(1.89)	(2.17)	(3.87)	(2.73)	(1.91)	(5.54)
$Sent_t^N$	0.0008	0.1227***	0.0615***	0.0647**	0.1702	0.0585**	0.1478***	0.0716**	0.0745***	0.1720***	0.1996***	0.2816***	0.1675***	0.1014***
	(0.19)	(4.82)	(5.49)	(2.16)	(1.46)	(2.37)	(5.02)	(2.16)	(4.22)	(6.04)	(5.53)	(7.82)	(7.69)	(5.87)
D_t^{-N}	0.0203	0.5100***	−0.1250*	0.0973	0.1702**	−0.1022	0.1534	0.2972	0.0209	−0.3483*	0.1766	0.4961**	0.4597**	−0.1119

TABLE 7 (Continued)

	US	UK	Canada	Brazil	Australia	Switzerland	Germany	Spain	EU	France	HK	Japan	India	Singapore
	DJIA	FTSE100	TSX	Bovespa	AORD	Swiss	DAX30	IBEX35	STOXX	CAC40	HangSeng	Nikkei225	Nifty	FTStraits
	(0.76)	(2.97)	(−1.85)	(0.36)	(2.36)	(−0.60)	(0.60)	(1.18)	(0.16)	(−1.85)	(0.65)	(2.25)	(2.25)	(−0.94)
D_t^{+N}	−0.0199	−0.0197	−0.1803**	0.1602	0.0945	0.1919	−0.1957	0.1634	0.0876	0.0480	0.8705***	−0.0638	−0.1476	0.0486
	(−0.75)	(−0.11)	(−2.12)	(0.90)	(1.30)	(1.20)	(−0.81)	(0.65)	(0.63)	(0.18)	(2.77)	(−0.20)	(−0.71)	(0.35)
$D_t^{-N} \times Sent_t^N$	0.0131	0.3833***	−0.1087***	0.0685	0.1336***	−0.0031	0.0885	0.1959	−0.0029	−0.1325	0.1025	0.1746	0.2424**	−0.0480
	(0.90)	(4.03)	(−2.96)	(0.41)	(3.37)	(−0.03)	(0.62)	(1.35)	(−0.04)	(−1.30)	(0.74)	(1.57)	(2.15)	(−0.73)
$D_t^{+N} \times Sent_t^N$	0.0178	−0.0351	0.0631	−0.1673*	−0.0582	−0.0973	0.0700	−0.0945	−0.0998	−0.0675	−0.5384***	−0.0141	0.0870	−0.0473
	(1.15)	(−0.33)	(1.28)	(−1.78)	(−1.41)	(−1.05)	(0.50)	(−0.69)	(−1.19)	(−0.42)	(−2.72)	(−0.08)	(0.70)	(−0.59)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.0116	0.0791	0.0526	0.0042	0.0302	0.0165	0.0418	0.0229	0.0268	0.0652	0.0714	0.103	0.0984	0.101
F-statistic	3.27	17.6	11.7	1.81	7.01	4.24	9.43	5.52	6.31	14.5	15.8	23.2	22.1	22.8

Note: This table summarizes the regression results of all sample indices based on Equation (4). Panel A (Panel B) reports global market sensitivity to social media (news media). The dependent and independent variables are defined in Appendix A. The t-statistics are provided in parentheses. Control variables include the close-to-close return on the previous day (Rc_{t-1}, in percent), demeaned log number of trades on the previous day (VLM_{t-1}), demeaned realized volatility on the previous day (RV_{t-1}, in percent), and demeaned $\log(VIX)$ on the previous day (VIX_{t-1}). The estimated coefficients for control variables are omitted for brevity. The adjusted R^2 and F-statistics for each country regression are provided at the bottom of each panel. For the F-test, the unrestricted and restricted models are Equations (4) and (1), respectively. The F critical values with $df_1 = 5$, $df_2 = 1738 - 5 - 1$ are 1.85, 2.22, and 3.03 for the 10%, 5% and 1% significance levels, respectively.

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

sentiment in the FTSE 100 is related to 3.91% upswings (drops) in the next day's opening returns (Column 2 of Table 6). When we control for the impact of the previous-day trading activities on overnight returns in Column 3 of Tables 5 and 6, we observe that the economic magnitudes of influence from social media sentiment shrink for both the DJIA and FTSE 100. Moreover, in Table 5, the *t*-statistics of $Sent^S$ for the DJIA become insignificant (0.53), whereas the significance and magnitude of the news media sentiment remain largely intact. Although the coefficient of social media sentiment for the FTSE 100 subsides from 0.0391 to 0.0355, it remains significant at the 5% level. Compared with Column 2, the overall model fit for both markets improves, with adjusted R^2 increasing from 0.23% to 1.1% for the DJIA (Table 5) and from 0.34% to 1.52% for the FTSE 100 (Table 6).

Da et al. (2014, p. 12) posit that a central prediction of theories of investor sentiment is reversal: *"When sentiment is high, prices are temporarily high but later become low."* To look for evidence of return reversals, in addition to a contemporaneous relation, we consider the effects of sentiment on lead overnight return up to 1 month ahead. We find that much of the direct sentiment effect on day 0 reported earlier is temporary: The consistently positive coefficients for *Sent* in Tables 5 and 6 are followed by a series of negative coefficients on sentiment in the ensuing days, suggesting gradual reversal of the effect.[17]

The model in Equation (4) controls for the most optimistic and pessimistic days, using dummy variables indicating the top (D_t^{+S}) and bottom (D_t^{-S}) deciles of $Sent^S$, as well as their interaction terms with $Sent^S$. The model estimated in Column 4 of Tables 5 and 6 highlights the asymmetry in sentiment impact. For the DJIA index (Column 4 of Table 5), where the social media sentiment coefficient is significant, only the large negative swings appear to affect the index returns at the opening. We find no such evidence for the news media sentiment. In contrast, for the FTSE 100 (Column 4 of Table 6), the results reveal the opposite: It is the news media sentiment rather than social media that exhibits a significant impact, and consistent with the DJIA results, it is the large negative sentiment that has the most pronounced influence. All else equal, a 3 SD increase in pessimistic sentiment (i.e., decrease in $Sent^S$) causes a 21.5% decrease in overnight returns.[18] We focus on a 3 SD change rather than the ad hoc 1 SD in interpretating the results of Equation (4) because only under such a magnitude are the dummy variables representing the top and bottom deciles of sentiment switched. Milder magnitudes of change, such as 1 *SD*, do not constitute the top and bottom decile events.

Unlike in the US market, news media sentiment in the UK market displays a more profound role in determining overnight returns. Controlling for other market variables, a 1 SD increase in news media sentiment of the FTSE 100 generates a 14.91% increase in the overnight returns (Column 6 of Table 6) at the 1% significance level. The same effect from social media, however, results in only a 3.55% increase in overnight returns (Column 3). Similar to the social media patterns in the US market, negative news sentiment in the UK exerts a greater impact than does positive news sentiment (Column 7 of Table 6).

In Table 7, following the same analytical approach, we report the results for all markets in our sample. To contrast the effects of social and news media across all 14 stock markets, we refrain from reporting results individually for each market as we do in Tables 5 and 6 for illustrative purposes. Instead, we detail the results of country-level regressions based on Equation (4) in Table 7, where Panels A and B contain the results based on social and news media, respectively. Overall, we find that with the exception of the US and Brazilian markets, news media sentiment displays a more pronounced influence on overnight returns.

The coefficients of $Sent^S$ and $Sent^N$ in Table 7 offer several insights. First, except for the IBEX35 index for social media, all the coefficients of $Sent^S$ and $Sent^N$ are positive, suggesting a direct effect of overnight sentiment on overnight returns. That is, positive (negative) sentiment over the nontrading hours is associated with an increase (decrease) in the next-day opening prices relative to the previous-day closing prices. Second, because we employed standardized sentiment variables, we can directly contrast the coefficient magnitudes

[17]Refer to the Online Appendix for estimation details and results.

[18]For example, the effect from a 3 SD increase in negative sentiment can be estimated with $Sent^S = -3$, $D_t^{-S} = 1$, and $D_t^{+S} = 0$ as: $0.0117 \times (-3) + (-0.05) \times 1 + (-0.0433) \times (-3) = -0.2150$ based on values in Column 4 of Table 5.

between social and news media effects and compare sentiment effects across countries. We find that only the US market exhibits a stronger reaction to social media sentiment than to news, as evidenced by the magnitude of the $Sent^S$ coefficient relative to $Sent^N$. In contrast, the rest of the countries in our sample display greater responses to news media. Moreover, social media sentiment has little effect on the UK, Canadian, Swiss, and Spanish stock markets. Although the estimated coefficient of $Sent^S$ for Brazil is insignificant, the heightened negative social media sentiment, D_t^{-S}, has a hefty detrimental impact on Bovespa's overnight returns. Other markets, for instance, France, Germany, Hong Kong, India, Japan, and Singapore, are sensitive to both social and news media, given the significance of both $Sent^S$ and $Sent^N$ coefficients. However, these markets appear more responsive to news media sentiment than to social media. Because Germany and France primarily drive the performance of the European Union's STOXX index, the estimates for the STOXX index are consistent with these two markets. Last, the only market similar to the United States in its response to media sentiment is Australia, where social media exerts significant influences on opening returns, whereas news media sentiment is largely muffled. This finding is consistent with Bertram (2004), who finds that the general performance of the ASX is mainly affected by overnight sessions when any material market-moving information arrives from larger trading venues (e.g., United States and Europe). It is worth noting, however, that the magnitude of the social media impact on the US market is smaller compared with Australia. Given that the US financial markets are among the most efficient, this result is not surprising.

A closer look at the binary regressors and interaction terms in Table 7 further illustrates the asymmetric effect of media sentiment on overnight returns. The results in Panel A suggest that the US, Brazilian, and Japanese markets are prone to excessive negative swings in social media sentiment. At the same time, Hong Kong is the only market in our sample that is more sensitive to extreme positive social media sentiment. This asymmetry is more pronounced for news media in Panel B. We find that overnight returns in the United Kingdom, Australia, France, Japan, and India are highly sensitive to excessive negative swings in news media sentiment. The Canadian market is highly susceptible to negative and positive news sentiment variations but displays no association with social media. Thus, the Canadian market is a striking contrast to the United States, where social media sentiment is dominant. On the contrary, Hong Kong offers an interesting case; it is the only market that shows a statistically significant reaction to positive but not negative sentiment spikes across social and news media.

In Figure 3, we evaluate increments in the proportion of overnight return variation explained by adding sentiment-based variables. Specifically, we contrast the adjusted R^2 in Table 7 based on Equation (4) for social and news media, respectively, with the adjusted R^2 from a baseline model in Equation (1) that contains no sentiment data. This allows for a value-added assessment of the sentiment signal.[19] From the figure, we observe that in India, France, and the United Kingdom, the signal contained in the news media sentiment data more than doubles the proportion of the explained overnight return variation. In contrast, the United States and Brazil are the only countries in our sample where social media sentiment offers relatively larger explanatory power compared to news media. The addition of social media variables in Switzerland and news media variables in Brazil shows no significant improvements, resulting in low R^2s and F-statistics below conventional critical values (see Table 7). The R^2 levels in Figure 3 may appear low, but they are nonetheless consistent with the R^2 bounds in the empirical stock return forecasting literature. For instance, Fama and French (1988), Zhou (2010), and Kan and Zhou (2006) present close to or less than 1% monthly return forecasting R^2 statistics. According to Rapach and Zhou (2013), monthly R^2 statistics below 1% can still be economically relevant. Contrasting R^2 in models with and without sentiment variables allows us to assess the importance of news versus social media in

[19]Our aim is not to compare information environments or market efficiencies across countries, although it could be an interesting topic question for future research. As noted in Bartram et al. (2009), there are large variations across countries in uncertainty about country fundamentals, financial development, and information environment, all of which determine the cross-country variation in R^2.

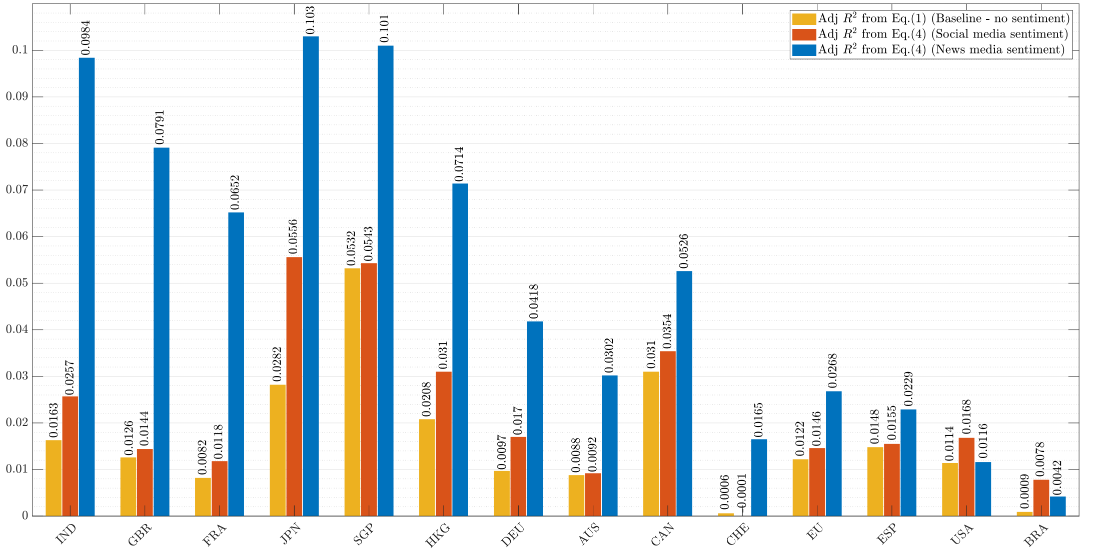

FIGURE 3 Proportion of overnight return variation explained. This figure contrasts adjusted R^2 s from the baseline regression model in Equation (1) (absent sentiment data) with adjusted R^2 s from Equation (4) based on social and news media sentiment. The country plots are ordered based on the difference between the adjusted R^2 s from Equation (4) for news and social media. [Color figure can be viewed at wileyonlinelibrary.com]

the global markets, with their specific information-processing limitations, market inefficiencies, and any cultural or psychological influences.

Given the complexity of the model in Equation (4), we provide visual summaries of our findings in Figures 4 and 5 where we assess the overall impact of 1 SD and 3 SD changes in overnight sentiment on the index return. In these figures, we contrast the sensitivity of index returns to changes in news versus social media sentiment. If social and news media impacts are equivalent, the country would lie on the 45-degree line. Except for the United States, our results in Figure 4 convey that a 1 SD change in news media sentiment has a greater impact on index returns than does social media.

To assist with the interpretation of the results in Table 7 and to contrast the effects of news versus social media, we plot the estimated coefficients from Panel A against the estimated coefficients from Panel B along with a 45-degree reference line. To that end, Figure 4 contrasts market sensitivities to overnight news sentiment ($Sent^N$) and social media sentiment ($Sent^S$). Countries plotted above (below) the 45-degree line exhibit stronger sensitivity to news (social) media sentiment. Considering the standardization of the country-specific sentiment scores and the inclusion of controls, the coefficients of $Sent^N$ and $Sent^S$ from Equation (4) represent the magnitude of change in country index returns (in percent) in response to a 1 SD change in sentiment after controlling for other factors that are known to determine the market index return rate. Each point on the graph represents an intersection of the estimated coefficients for $Sent^N$ and $Sent^S$ from Equation (4). The sizes of the plotted points are scaled to represent the absolute values of the larger t-statistic of the corresponding coefficients. For example, the United States is the only country whose broad market index exhibits a dependency on daily social media sentiment fluctuations. Although the magnitude of the coefficient is small, the US market is among the most efficient markets in the world, and the coefficient (however small) is still significant after accounting for other factors known to determine index returns.[20] Unlike the United States,

[20]Specifically, we account for the previous-day close-to-close returns (Rc_{t-1}), volume (VLM_{t-1}), realized volatility (RV_{t-1}), and global fear index (VIX_{t-1}).

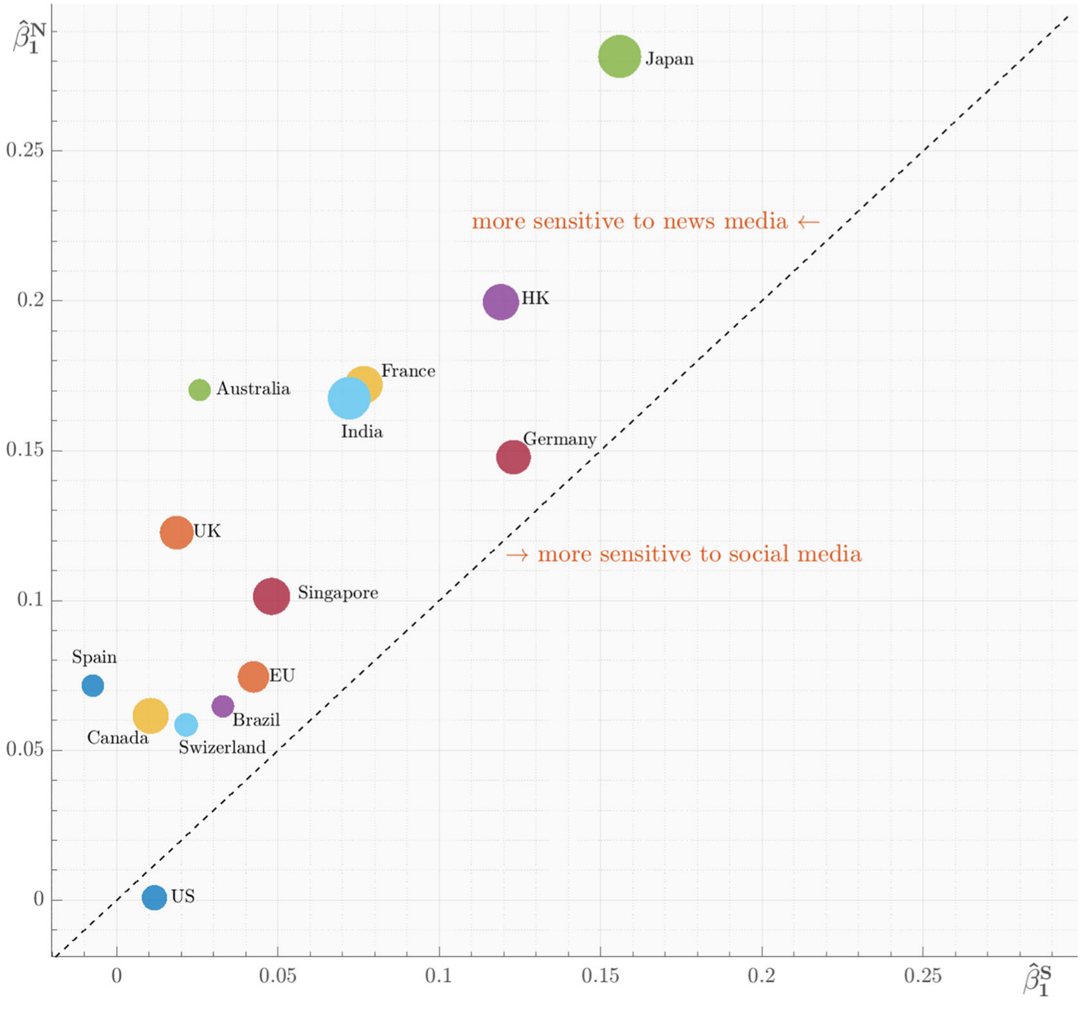

FIGURE 4 Market sensitivities to overnight sentiment. The figure contrasts the market sensitivities to the overnight news and social media sentiment. Considering the standardization of country sentiment scores and the inclusion of controls, the coefficients on $Sent^N$ and $Sent^S$ from Equation (4), $\hat{\beta}_1^N$ and $\hat{\beta}_1^S$, represent the magnitude of change in returns (in percent) in response to a 1 SD change in sentiment after controlling for other factors that are known to determine the return rate. Each point on the graph represents an intersection of the estimated $Sent^N$ and $Sent^S$ from Equation (4), and the sizes of plotted points are scaled to represent the absolute values of the larger t-statistic of the corresponding coefficients. [Color figure can be viewed at wileyonlinelibrary.com]

markets in other countries are more sensitive to sentiment from news versus social media. France, Hong Kong, India, and Japan appear to be swayed by news media sentiment much more strongly than Canada, Singapore, and the United Kingdom. Response to daily news media sentiment in Australia and the United States appears to be of little consequence judging by the t-statistics on $Sent^N$ from Table 7.

Figure 5 contrasts the total effects of large swings in news and social media sentiment. Given a 3 SD change in the relevant sentiment type, the response in the market index return is calculated based on estimated coefficients from Table 7. This allows us to consider the full complexity of sentiment variables in Equation (4), including overall sensitivity to sentiment and polarized emotions via binary regressors and interaction terms.

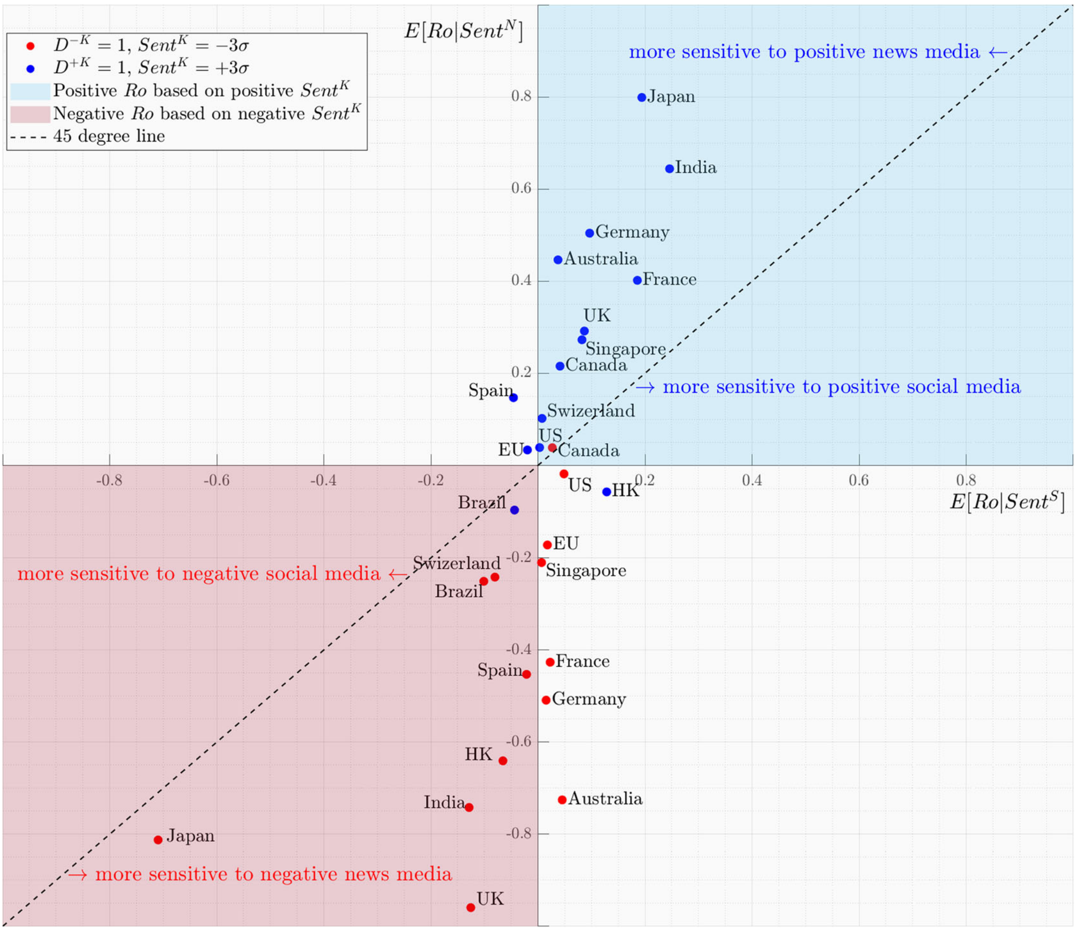

FIGURE 5 Market reactions to positive and negative sentiment. This figure contrasts market reactions to optimistic and pessimistic overnight news and social media sentiment. Considering the standardization of country sentiment scores and the inclusion of controls, the horizontal axis represents expected overnight returns conditional on strong social media sentiment ($Sent^S = \pm 3\sigma$), and the vertical axis indicates expected overnight returns conditional on strong news media sentiment ($Sent^N = \pm 3\sigma$). Each point on the graph represents an intersection of the estimated responses from the model in Equation (4) using 3σ change in sentiment based on social media (x-axis) and news media (y-axis). The blue dots indicate responses to the positive sentiment, that is, $E[Ro|D^{+S} = 1, D^{+N} = 0, Sent^S = 3\sigma]$ for the x-coordinate, and $E[Ro|D^{+S} = 0, D^{+N} = 1, Sent^N = 3\sigma]$ for the y-coordinate. The red dots indicate similar responses but for the negative sentiment, that is, $E[Ro|D^{-S} = 1, D^{-N} = 0, Sent^S = -3\sigma]$ and $E[Ro|D^{-S} = 0, D^{-N} = 1, Sent^N = -3\sigma]$. The variables are defined in Appendix A. [Color figure can be viewed at wileyonlinelibrary.com]

The horizontal axis in Figure 5 represents expected overnight returns conditional on strong social media sentiment ($Sent^S = \pm 3\sigma$), and the vertical axis indicates expected overnight returns conditional on strong news media sentiment ($Sent^N = \pm 3\sigma$). Here, $\sigma = 1$ for all countries and sentiment media types because of the standardization of sentiment scores that facilitates the comparison. Each blue mark in the figure represents the intersection of the predicted index return from large positive changes in social media sentiment (x-axis) and news media sentiment (y-axis). Similarly, each red mark represents the intersection of the predicted index return from large negative changes in sentiment.[21]

In Figure 5, the location of scatter points along the horizontal axis indicates the total effect from social media sentiment on market returns. The deviation of scatter points along the vertical axis and away from the origin attests to the magnitude of the total effect from news media sentiment. If market returns are positive in response to positive sentiment from social and news media, the scatter points would be located in the shaded blue area in Quadrant I. Similarly, if market returns are negative in response to negative sentiment from social and news media, the scatter points would be located in the shaded red area in Quadrant III. If the effects of social and news media are equivalent, the scatter points would be plotted along the 45-degree line. Points concentrated around the origin display smaller magnitudes of sensitivity to sentiment. Contrasting the red and blue scatter points reveals that negative sentiment effects from news media appear to be stronger than positive effects. For example, a − 3 SD change in news media sentiment results in a 96 basis points (bps) decrease in FTSE 100 opening returns, whereas +3 SD change in news media sentiment accounts for only a 31 bps increase. Another interesting observation is that the Hong Kong market reacts negatively to both positive and negative extreme news media sentiment. Points placed in Quadrants II and IV indicate sensitivity to only one type of media source. For example, the red dots at the bottom right (namely, Australia, France, Germany, and Singapore) indicate sensitivity to negative news sentiment but not to negative social media sentiment. We perform a similar analysis with 2 SD change in the relevant sentiment variables and find qualitatively similar results.

Together, Figures 4 and 5 aid in our understanding of the results in Table 7. Figure 4 indicates that except in the United States, markets are more easily swayed by variations in news sentiment rather than in social media. Consequently, scholars should and must refrain from adopting US evidence naively in the context of other markets. Figure 5 contrasts strong polarized (positive and negative) sentiment effects across different media types, offering new insights into the literature on sentiment impact on stock returns.

3.1 | Evolution of the media landscape

The media landscape is changing, and the way investors absorb new information has changed dramatically as social media has become more accessible and widespread. Our analysis of media activity trends reveals disproportionate changes in the proliferation of social media among countries in our sample. A notable case is India's explosive social media activity growth since 2015. Other markets—Singapore, Brazil, and Hong Kong—have also shown a rapid increase in social media activity over time. This has a material impact on the sentiment effects we report in Table 7. For example, based on our total sample period, the estimated effects of sentiment on Nifty index returns are 0.0721 (social media) and 0.1675 (news media). We find substantive differences when splitting our sample into two subperiods: January 2011–June 2015 and July 2015–November 2017. The former is characterized by the stable activity of social media relative to news, and the latter is marked with the uptake in social media activity. The effect of social media on Nifty increases almost two-fold from 0.056 in the earlier period to 0.096 in the later period, whereas the effect of news media drops from 0.181 to 0.151.[22] An increase in social media effect and a drop in news media effect is also observed in Germany, France, and the

[21]More precisely, the x-coordinates for the blue marks in Figure 5 are calculated as $E[Ro|Sent^S = +3, D^{+S} = 1, D^{+N} = 0, D^{-S} = 0, D^{-N} = 0]$ with estimates from Panel A of Table 7 for social media sentiment. The y-coordinates are $E[Ro|Sent^N = 3, D^{+S} = 0, D^{+N} = 1, D^{-S} = 0, D^{-N} = 0]$ with estimates from Panel B for news media sentiment. Therefore, blue marks in Figure 5 represent the relative sensitivities of market index returns to extreme positive sentiment shifts. Similarly, the coordinates for the red marks are calculated by passing the negative 3 SD change in the relevant sentiment type through model estimates in Table 7. That is, the (x, y) coordinates of the red marks are represented by the values $E[Ro|Sent^S = -3, D^{+S} = 0, D^{+N} = 0, D^{-S} = 1, D^{-N} = 0]$ and $E[Ro|Sent^N = -3, D^{+S} = 0, D^{+N} = 0, D^{-S} = 0, D^{-N} = 1]$.

[22]For brevity, we omit the complete set of results here and refer to the Online Appendix for analysis of news and social media trend dynamics and subperiod estimation results.

United States. In contrast, Brazil and Switzerland show increased social media effects along with increased news media effects. An interesting topic for future research could focus on social media content quality instead of the quantity proxied by media buzz.

4 | ROBUSTNESS ANALYSIS

In the previous section, we report the results of accumulating sentiment of the prior-day closing to the next-day market opening. We check the robustness of our findings by varying the lengths of sentiment aggregation windows to explore the effectiveness of return predictability from sentiment signals. Following the same format, we regress overnight returns on standardized cumulative sentiment using Equations (1)–(4) and report the results when social and news media sentiment are aggregated 3 h before the market opens. Our analysis of the optimal window for gauging the emotional scores indicates that windows from 30 min to 6 h before markets open form the most effective signals in terms of overall model fit and significance of estimated coefficients on sentiment variables. A duration of fewer than 30 min (e.g., 15 min) before the opening suffers from a data sparsity problem that is detrimental to signal performance. Another negative aspect of shorter aggregation windows comes from the high volatility of sentiment scores. In contrast, aggregating sentiment over more extended periods allows for reduced noise and a more persistent trend. The overall model fit for each country in our sample varies slightly when the aggregation window is altered, but the best fit is achieved between 30 min and 6 h before the market opening. Furthermore, we observe an inverse relation between data availability in each market (proxied by the number of available 1-min news and social media scores reported in Appendix B) and the optimal length of the aggregation window. For consistency, we use 3-h windows in our robustness tests in Tables 8 and 9. Results for alternative lengths of aggregation windows are available upon request from the authors.

Tables 8 and 9 report sensitivities of the DJIA and FTSE 100 overnight returns, respectively, to social and news media sentiment 3 h before the market opens. The results in these two tables are consistent with results in Section 3. Similarly, Columns 2–4 measure social media impacts and Columns 5–7 measure news media impacts.

When aggregating sentiment just 3 h before the market opens instead of using an entire overnight period, social media sentiment of the DJIA retains significant predictability, although at relatively lower economic magnitudes. A comparison between the restricted model (Column 2) in Tables 5 and 8 shows that the coefficient shrinks from 0.0053 to 0.0039 if we consider only the morning preopening sentiment (e.g., from 6:29 AM to 9:29 AM in the US market, and from 4:59 AM to 7:59 AM in the UK market). After including all control variables and dummy variable terms (unrestricted model), the coefficients on $Sent^S$ improve in both Tables 5 and 8 (Column 4), relative to their restricted models (Column 2). However, the statistical and economic significance are reduced in the shorter period sentiment tests. Compared with Table 5, an 1 SD increase in social media sentiment 3 h before the opening leads to only a 0.76% increase in the opening prices (Column 4 of Table 8), a reduction of more than 35% from 1.17% reported in Column 4 of Table 5. Moreover, negative tonality continues to display a stronger impact than does positive tonality, as both D_t^{-S} and its interaction term maintain similar magnitudes at statistically significant levels. In contrast, D_t^{+S} and its interaction term remain insignificant. Together, in Table 8, a 1 SD spike in negative DJIA social media sentiment (decline in sentiment) 3 h before the opening leads to a 2.23% decrease in opening returns $(0.0076 \times (-1) + (-0.049) + (-0.0343) \times (-1) = 0.0223)$. By contrast, this negative DJIA social media sentiment for the entire overnight period (Table 5) gives rise to only a 1.84% reduction in opening returns $(0.0117 \times (-1) + (-0.05) + (-0.0433) \times (-1) = -0.0184)$. The coefficients based on news sentiment data, however, are now insignificant in all three models (Columns 5–7 of Table 8).

Table 9 shows that the FTSE 100 social media sentiment causes similar effects whether sentiment is aggregated at a shorter or full overnight period, whereas the FTSE 100 news sentiment effect diminishes remarkably if aggregated only 3 h before the opening. The estimated coefficients from models based on social media in Columns 2 and 3 are at

TABLE 8 DJIA 3-h cumulative sentiment regressions.

DJIA dependent variable = Overnight return on day t (in percent)							
		Social media (k = S)			News media (k = N)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0037	0.0039*	0.0037	0.0055*	0.0039*	0.0037	0.0043
	(1.39)	(1.68)	(1.39)	(1.96)	(1.68)	(1.38)	(1.48)
$Sent_t^k$		0.0039*	0.0004	0.0076*	−0.0009	−0.001	0.0006
		(1.65)	(0.18)	(1.80)	(−0.37)	(−0.42)	(0.15)
D_t^{-k}				−0.049*			−0.0091
				(−1.89)			(−0.42)
D_t^{+k}				−0.0021			−0.0346
				(−0.07)			(−1.16)
$D_t^{-k} \times Sent_t^k$				−0.0343**			−0.0058
				(−2.21)			(−0.5)
$D_t^{+k} \times Sent_t^k$				−0.0119			0.0163
				(−0.78)			(0.9)
Rc_{t-1}	0.0056**		0.0056**	0.0056**		0.0057**	0.0058**
	(1.98)		(1.98)	(1.98)		(1.99)	(2.02)
VLM_{t-1}	0.0256***		0.0253***	0.0251***		0.0256***	0.0254***
	(4.13)		(3.98)	(3.94)		(4.12)	(4.08)
RV_{t-1}	−0.0046		−0.0057	−0.0257		−0.0041	−0.0058
	(−0.03)		(−0.04)	(−0.17)		(−0.0275)	(−0.04)
VIX_{t-1}	−0.0012		−0.0008	0.0026		−0.0013	−0.012
	(−0.12)		(−0.08)	(0.26)		(−0.13)	(−0.13)
Adj. R^2	0.0114	0.0010	0.0108	0.0133	−0.0005	0.0102	0.0097
F-statistic	6.00	2.73	4.80	3.61	0.14	4.83	2.89
p-value	0.00	0.099	0.0002	0.0002	0.711	0.0002	0.0022
Obs.	1738	1739	1738	1738	1739	1738	1738

Note: This table summarizes the regression results of Equations (1)–(4) on the Dow Jones Industrial Average (DJIA). The variables are defined in Appendix A. The t-statistics are provided in parentheses.
*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$.

magnitudes similar to those in Table 6 (marginal decrease from 0.0391 to 0.0334, and from 0.0355 to 0.0327, respectively), with their t-statistics lower but still significant. An astonishing result is observed in Column 4, where after considering tonality, the explanatory power of $Sent^S$ is largely improved from 0.0186 to 0.0518, with t-statistics significant at the 10% level. The result suggests that for the FTSE 100, aggregating social media sentiment from the previous-day close to the next-day open might have generated an obscurely wider window that has dampened the precision of the signal. However, if we focus on just 3 h before opening, the signal from social media might be more helpful in predicting opening prices. In contrast, for the FTSE 100 news sentiment (Columns 5–7 of Table 6), focusing

TABLE 9 FTSE100 3-h cumulative sentiment regressions.

FTSE 100 dependent variable = Overnight return on day *t* (in percent)							
		Social media ($k = S$)			News media ($k = N$)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	0.0227	0.2267	0.0227	0.0209	0.0227	0.0227	0.0317*
	(1.53)	(1.52)	(1.53)	(1.25)	(1.52)	(1.53)	(1.90)
$Sent_t^k$		0.0334**	0.0327**	0.0518*	0.0335**	0.0301**	0.0336
		(2.24)	(2.18)	(1.90)	(2.24)	(1.98)	(1.28)
D_t^{-k}				0.0263			0.0096
				(0.19)			(0.06)
D_t^{+k}				−0.0933			−0.14
				(−0.61)			(−0.82)
$D_t^{-k} \times Sent_t^k$				−0.0203			0.0216
				(−0.26)			(0.23)
$D_t^{+k} \times Sent_t^k$				0.0280			0.0433
				(0.32)			(0.46)
Rc_{t-1}	0.0008		0.0008	0.0010		0.0023	0.0027
	(0.05)		(0.05)	(0.06)		(0.14)	(0.17)
VLM_{t-1}	0.0825		0.0767	0.0796		0.0791	0.0787
	(1.52)		(1.42)	(1.46)		(1.46)	(1.45)
RV_{t-1}	10.834***		11.056***	10.941***		10.884***	10.884***
	(4.04)		(4.13)	(4.07)		(4.06)	(4.06)
VIX_{t-1}	−0.2453***		−0.2367***	−0.2383***		−0.2241***	−0.2229***
	(−3.79)		(−3.66)	(−3.66)		(−3.42)	(−3.40)
Adj. R^2	0.0126	0.0023	0.0148	0.0132	0.0023	0.0143	0.0131
F-statistic	6.56	5.00	6.21	3.58	5.02	6.04	3.56
p-value	0.00	0.03	0.00	0.00	0.03	0.00	0.00
Obs.	1738	1739	1738	1738	1739	1738	1738

Note: This table summarizes the regression results of Equations (1)–(4) on the Financial Times Stock Exchange (FTSE) 100. The variables are defined in Appendix A. The ***t***-statistics are provided in parentheses.
$^*p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

just on the preopening period might be less ideal compared with watching the entire overnight period. This distinction between short-period social and news sentiment predictability might result from the more dynamic nature of social media than news media. The coefficient of $Sent^N$ drops sharply from 0.1447 to 0.0335 in Column 5, and it dwindles from 0.1491 to 0.0301 after further controlling for market variables in Column 6. The t-statistics are also considerably smaller compared with the overnight sentiment group, whereas key variable coefficients of the most flexible model in Column 7 are now all insignificant. For the FTSE 100, these results suggest that news sentiment should be aggregated at longer windows.

5 | CONCLUSION

We investigate the influence of overnight social and news media sentiment on the returns of 14 major stock markets at the opening. Controlling for factors known to affect the return rate, we find that changes in the overnight return, that is, the opening price relative to the previous-trading-day close, can be attributed to the build-up of emotions in news and social media. This effect is significant both statistically and economically. Our results suggest that the more optimistic (pessimistic) the media tone is, the higher (lower) the next-day opening price. Highly polarized positive or negative emotions, such as top and bottom decile sentiment, tend to be more influential than moderate sentiment (interior deciles). Our analysis shows that overnight sentiment does not simply subsume previous-day market activity but, in fact, contains additional information that helps improve the explanatory power of return forecasting models.

We also find that only in the United States does social media exert a greater effect in the market, whereas news sentiment remains the strongest effect on stock markets worldwide. This finding cautiously highlights the issue of hastily applying US-based evidence to other markets. Furthermore, the economic magnitudes of return predictability induced from overnight social and news media sentiment in the United States are much smaller than in other countries, reaffirming that the US market is among one of the most liquid and efficient markets in the world.

By incorporating the direction of tonality and allowing for asymmetry in our modeling framework, we find that negative news sentiment plays a greater role than positive sentiment in most international markets. Among others, Australia, India, and Japan tend to be easily affected by both positive and negative news sentiment. In contrast, Hong Kong is the only market that is highly prone to positive sentiment based on both social and news media. These mixed results pose difficulty in drawing a unified conclusion but nonetheless offer new directions for future research. With the availability of extended trading sessions as well as futures and options trading across markets, the potential overlap between sentiment and return could be explored in more detail. One possible direction is to consider weekend media tones when most spot and futures markets are closed. Such an approach, albeit with lower data frequency, allows for better isolation between sentiment and returns in search for the cause and effect.

Overall, this article contributes to the behavioral finance literature on investor sentiment and its impact on stock markets. It assists in understanding the price-discovery process in markets other than the United States, with a novel data set of high-frequency, textual-based sentiment and an approach that helps disentangle the return–sentiment feedback loop.

ACKNOWLEDGMENTS

We thank Thomson Reuters Financial and Risk for offering MarketPsych Indices (TRMI) as part of our research data; we are grateful to Ron Bird for assisting with the TRMI data set acquisition. We appreciate insightful comments and suggestions from Bart Frijns, Tony He, and Stefan Trueck. This research is supported by the Australian Government Research Training Program scholarship. Open access publishing facilitated by University of Technology Sydney, as part of the Wiley - University of Technology Sydney agreement via the Council of Australian University Librarians.

ORCID

Baoqing Gan http://orcid.org/0000-0001-6903-0433
Vitali Alexeev https://orcid.org/0000-0003-3092-2500
Danny Yeung https://orcid.org/0000-0002-3032-0449

REFERENCES

Aboody, D., Even-Tov, O., Lehavy, R., & Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, *53*(2), 485–505.

Agrawal, S., Azar, P. D., Lo, A. W., & Singh, T. (2018). Momentum, mean-reversion, and social media: evidence from StockTwits and twitter. *Journal of Portfolio Management*, *44*(7), 85–95.

Ahoniemi, K., Fuertes, A.-M., & Olmo, J. (2015). Overnight news and daily equity trading risk limits. *Journal of Financial Econometrics*, *14*(3), 525–551.

Akhtar, S., Faff, R., Oliver, B., & Subrahmanyam, A. (2012). Stock salience and the asymmetric market effect of consumer sentiment news. *Journal of Banking & Finance*, *36*(12), 3289–3301.

Alexeev, V., Gan, B., & Yeung, D. (2022). *Moods on the move: Overnight sentiment and the intraday return dynamics* (Technical Report). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4184707

Azar, P. D., & Lo, A. W. (2016). The wisdom of twitter crowds: predicting stock market reactions to FOMC meetings via twitter feeds. *Journal of Portfolio Management*, *42*(5), 123–134.

Bagnoli, M., Clement, M. B., & Watts, S. G. (2005). *Around-the-clock media coverage and the timing of earnings announcements* (Working Paper). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=570247

Bai, Y. (2014). Cross-border sentiment: an empirical analysis on EU stock markets. *Applied Financial Economics*, *24*(4), 259–290.

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

Bartram, S., Brown, G., & Stulz, R. (2009). *Why do foreign firms have less idiosyncratic risk than U.S. firms?* (Working Paper 14931). National Bureau of Economic Research.

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. Wiley.

Berkman, H., Koch, P. D., Tuttle, L., & Zhang, Y. J. (2012). Paying attention: overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, *47*(4), 715–741.

Bertram, W. K. (2004). An empirical investigation of Australian stock exchange data. *Physica A: Statistical Mechanics and Its Applications*, *341*, 533–546.

Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2018). Information, trading, and volatility: evidence from firm-specific news. *Review of Financial Studies*, *32*(3), 992–1033.

Bradley, D., Clarke, J., Lee, S., & Ornthanalai, C. (2014). Are analysts' recommendations informative? intraday evidence on the impact of time stamp delays. *Journal of Finance*, *69*(2), 645–673.

Calomiris, C. W., & Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*, *133*(2), 299–336.

Chen, C., Fengler, M. R., Härdle, W. K., & Liu, Y. (2018). *Textual sentiment, option characteristics, and stock return predictability* (IRTG 1792 Discussion Paper 2018-023). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3210585

Chen, H., De, P., Hu, Y. J., & Hwang, B.-H. (2014). Wisdom of crowds: the value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367–1403.

Chen, M.-P., Chen, P.-F., & Lee, C.-C. (2013). Asymmetric effects of investor sentiment on industry stock returns: panel data evidence. *Emerging Markets Review*, *14*, 35–54.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*(5), 1461–1499.

Da, Z., Engelberg, J., & Gao, P. (2014). The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, *28*(1), 1–32.

Deng, S., Huang, Z. J., Sinha, A. P., & Zhao, H. (2018). The interaction between microblog sentiment and stock returns: an empirical examination. *MIS Quarterly*, *42*(3), 895–918.

The Economist. (2010, July 15). The difference engine: The wisdom of crowds. *The Economist*. https://www.economist.com/babbage/2010/07/15/the-difference-engine-the-wisdom-of-crowds

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, *22*(1), 3–25.

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

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *Review of Economics and Statistics*, *49*(1), 92–107.

Feldman, T., & Liu, S. (2017). Contagious investor sentiment and international markets. *Journal of Portfolio Management*, *43*(4), 125–136.

Fraiberger, S., Lee, D., Puy, D., & Ranciere, R. (2018). *Media sentiment and international asset prices* (Policy Research Working Papers, 18(274)). World Bank.

Gan, B., Alexeev, V., Bird, R., & Yeung, D. (2020). Sensitivity to sentiment: news vs social media. *International Review of Financial Analysis*, *67*, Article 101390.

Gao, Z., Ren, H., & Zhang, B. (2019). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis*, *55*(2), 549–580.

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

Griffin, J. M., Hirschey, N. H., & Kelly, P. J. (2011). How important is the financial media in global markets? *Review of Financial Studies*, *24*(12), 3941–3992.

Han, X., & Li, Y. (2017). Can investor sentiment be a momentum time-series predictor? evidence from China. *Journal of Empirical Finance*, *42*, 212–239.

Hero to villain: How the GameStop saga threatens Robinhood. (2021, February 1). *The Economist*. https://www.economist.com/finance-and-economics/2021/02/01/how-the-gamestop-saga-threatens-robinhood

Heston, S. L., & Sinha, N. R. (2017). News vs sentiment: predicting stock returns from news stories. *Financial Analysts Journal*, *73*(3), 67–83.

Hudson, Y., & Green, C. J. (2015). Is investor sentiment contagious? international sentiment and UK equity returns. *Journal of Behavioral and Experimental Finance*, *5*, 46–59.

Jiang, C. X., Likitapiwat, T., & McInish, T. H. (2012). Information content of earnings announcements: evidence from after-hours trading. *Journal of Financial and Quantitative Analysis*, *47*(6), 1303–1330.

Jiao, P., Veiga, A., & Walther, A. (2020). Social media, news media and the stock market. *Journal of Economic Behavior & Organization*, *176*, 63–90.

Kan, R., & Zhou, G. (2006). A new variance bound on the stochastic discount factor. *Journal of Business*, *79*(2), 941–961.

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

Michaely, R., Rubin, A., & Vedrashko, A. (2013). Corporate governance and the timing of earnings announcements. *Review of Finance*, *18*(6), 2003–2044.

Peterson, R. L. (2016). *Trading on sentiment: The power of minds over markets*. Wiley.

Rapach, D., & Zhou, G. (2013). Forecasting stock returns. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (2A, pp. 328–383). Elsevier.

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

Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*, *107*, 730–743.

Smales, L. A. (2014). News sentiment and the investor fear gauge. *Finance Research Letters*, *11*(2), 122–130.

Sprenger, T. O., Sandner, P. G., Tumasjan, A., & Welpe, I. M. (2014). News or noise? using twitter to identify and understand company-specific news flow. *Journal of Business Finance & Accounting*, *41*(7–8), 791–830.

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

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: investor sentiment and anomalies. *Journal of Financial Economics*, *104*(2), 288–302.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2014). The long of it: odds that investor sentiment spuriously predicts anomaly returns. *Journal of Financial Economics*, *114*(3), 613–619.

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, *73*, 147–164.

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

von Beschwitz, B., Keim, D. B., & Massa, M. (2013). *First to 'read' the news: News analytics and high frequency trading*. Working Paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2356547

Weißofner, F., & Wessels, U. (2019). Overnight returns: an international sentiment measure. *Journal of Behavioral Finance*, *21*(2), 205–217.

Xiong, X., Meng, Y., Li, X., & Shen, D. (2020). Can overnight return really serve as a proxy for firm-specific investor sentiment? cross-country evidence. *Journal of International Financial Markets, Institutions and Money*, *64*, Article 101173.

Yang, S. Y., Mo, S. Y. K., & Liu, A. (2015). Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance*, *15*(10), 1637–1656.

Zhou, G. (2010). How much stock return predictability can we expect from an asset pricing model? *Economics Letters*, *108*(2), 184–186.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gan, B., Alexeev, V., & Yeung, D. (2024). Tweets versus broadsheets: Sentiment impact on stock markets around the world. *Journal of Financial Research*, *47*, 601–633. https://doi.org/10.1111/jfir.12380

APPENDIX A: VARIABLE DEFINITIONS

Variable	Definition
k	Media type, where k take values S and N denoting social and news media, respectively
n_t^k	Number of nonmissing observations of sentiment of media k across the overnight period
Ro	Overnight return, or close-to-open return
Rc	Daily return, or close-to-close return
RV	Demeaned daily average realized volatility
$Sent^k$	Standardized cumulative sentiment from media type k
VLM	Demeaned log daily trades
VIX	Demeaned log daily Chicago Board Options Exchange (CBOE) Volatility Index
$x_{t,j}^k$	Raw sentiment data from the k media type on day t at time j
X_t^k	Cumulative sentiment of media k averaged across the overnight period (divided by the cardinality)
$\bar{X}^k$	Mean score of the cumulative sentiment of media k across the sample period (days)
σ_{X^k}	Standard deviation of the cumulative sentiment of media k across the sample period (days)
D_t^{+k} and D_t^{-k}	Binary regressors that indicate the top and bottom decile days of $\{Sent_t^k\}_{\forall t}$

Note: Sentiment variables are obtained from Thomson Reuters MarketPsych Indices (TRMI). Overnight returns, daily returns, and daily average realized volatility are from Oxford-Man Institute of Quantitative Finance.

APPENDIX B: TRMI DESCRIPTIVE STATISTICS

(See Table B1).

TABLE B1 TRMI descriptive statistics.

Index		Mean	SD	Min.	Max.	Obs.
ASX	x^S	0.01	0.49	−1	1	618,017
	x^N	-0.01	0.44	−1	1	1,033,660
Bovespa	x^S	−0.03	0.58	−1	1	121,920
	x^N	−0.08	0.47	−1	1	221,731
CAC	x^S	0.03	0.53	−1	1	167,340
	x^N	0.01	0.52	−1	1	557,285
DAX	x^S	0.02	0.54	−1	1	223,247
	x^N	0.01	0.47	−1	1	841,920
DJIA	x^S	−0.01	0.35	−1	1	2,753,605
	x^N	0.02	0.33	−1	1	2,536,911
EUSTOXX	x^S	0.03	0.51	−1	1	557,034
	x^N	0.01	0.44	−1	1	1,446,764

Index		Mean	SD	Min.	Max.	Obs.
FTSE100	x^S	0.00	0.52	−1	1	450,888
	x^N	−0.02	0.43	−1	1	1,410,028
FTStraits	x^S	0.05	0.55	−1	1	83,053
	x^N	0.05	0.53	−1	1	150,955
HangSeng	x^S	0.06	0.53	−1	1	387,350
	x^N	0.05	0.53	−1	1	468,960
IBEX	x^S	−0.01	0.57	−1	1	46,791
	x^N	0.00	0.51	−1	1	214,968
Nifty	x^S	0.11	0.56	−1	1	332,530
	x^N	0.05	0.47	−1	1	566,467
Nikkei225	x^S	0.06	0.51	−1	1	412,601
	x^N	0.01	0.44	−1	1	1,311,914
Swiss	x^S	0.03	0.54	−1	1	213,296
	x^N	−0.01	0.50	−1	1	650,666
TSX	x^S	0.03	0.47	−1	1	1,038,727
	x^N	0.05	0.46	−1	1	859,937

Note: This table summarizes the descriptive statistics of all the Thomson Reuters MarketPsych Indices (TRMI) sample company group indices. $\boldsymbol{x^S}$ and $\boldsymbol{x^N}$ are the net positive and negative emotions of a specific entity (the representative index) on social and news media, respectively. The sample period is January 1, 2011 to November 30, 2017 at a 1-min frequency.

In this appendix, we provide descriptive statistics for the 1-min Thomson Reuters MarketPsych Indices (TRMI) data from January 1, 2011 to November 30, 2017 for all country indices in our analysis. Based on the average sentiment scores over the entire sample period, Brazil is the only country that nets overall negative tonality from both social and news media sources. Asian markets (Singapore, Hong Kong, India, and Japan) exhibit overall positive tonality from both media sources. The magnitudes of overall sentiment tonality for Australia, Spain, Switzerland, the United States, and the United Kingdom only marginally deviate from zero. Overall sentiment tonalities for France, Germany, and, as a consequence, the European Union's STOXX indices are positive but have lower magnitudes compared to Asian markets. The volatility of social media sentiment is higher than the sentiment volatility based on news media sources (except for Singapore, where the two volatilities are equal). The United States and Canada are the only markets where the number of 1-min sentiment observations based on social media is greater than traditional news media sources, pointing to heightened activity in the social media domain in these two markets.

APPENDIX C: OXFORD-MAN DESCRIPTIVE STATISTICS

Index	Oxford-Man variables	Mean	SD	Min.	Max.	Obs.
ASX	*Ro*	0.0000	0.0077	−0.0291	0.0370	1807
	Rc	0.0002	0.0083	−0.0365	0.0356	1806
	RV	0.0000	0.0000	0.0000	0.0008	1807
	Trades	719	19	514	969	1807
Bovespa	*Ro*	0.0004	0.0129	−0.0482	0.0831	1807
	Rc	0.0001	0.0145	−0.0880	0.0660	1806
	RV	0.0001	0.0002	0.0000	0.0034	1807
	Trades	817	45	391	845	1807

(Continues)

Index	Oxford-Man variables	Mean	SD	Min.	Max.	Obs.
CAC	*Ro*	0.0001	0.0102	−0.0516	0.0750	1807
	Rc	0.0003	0.0125	−0.0785	0.0630	1806
	RV	0.0001	0.0001	0.0000	0.0023	1807
	Trades	2,040	51	1164	3677	1807
DAX	*Ro*	0.0001	0.0104	−0.0560	0.0772	1807
	Rc	0.0005	0.0125	−0.0669	0.0560	1806
	RV	0.0001	0.0002	0.0000	0.0024	1807
	Trades	29,637	1045	16,926	30,601	1807
DJIA	*Ro*	−0.0003	0.0082	−0.0393	0.0563	1807
	Rc	0.0004	0.0084	−0.0541	0.0410	1806
	RV	0.0001	0.0002	0.0000	0.0059	1807
	Trades	16,018	6355	4899	23,412	1807
EUSTOXX	*Ro*	0.0001	0.0112	−0.0458	0.0904	1807
	Rc	0.0002	0.0126	−0.0840	0.0601	1806
	RV	0.0001	0.0002	0.0000	0.0054	1807
	Trades	2,040	4	1949	2041	1807
FTSE100	*Ro*	0.0001	0.0069	−0.0384	0.0462	1807
	Rc	0.0002	0.0092	−0.0483	0.0380	1806
	RV	0.0000	0.0001	0.0000	0.0016	1807
	Trades	68,130	21,228	11,899	314,308	1807
FTStraits	*Ro*	0.0001	0.0050	−0.0338	0.0258	1807
	Rc	−0.0001	0.0076	−0.0412	0.0283	1806
	RV	0.0000	0.0000	0.0000	0.0005	1807
	Trades	15,618	7337	2941	46,690	1807
HangSeng	*Ro*	0.0008	0.0076	−0.0405	0.0492	1807
	Rc	0.0003	0.0103	−0.0654	0.0498	1806
	RV	0.0001	0.0001	0.0000	0.0010	1807
	Trades	6911	3914	602	9661	1807
IBEX	*Ro*	0.0005	0.0120	−0.0569	0.0788	1807
	Rc	0.0002	0.0141	−0.1194	0.0559	1806
	RV	0.0002	0.0002	0.0000	0.0055	1807
	Trades	6022	255	2540	6214	1807
Nifty	*Ro*	0.0003	0.0053	−0.0524	0.0396	1807
	Rc	0.0002	0.0098	−0.0592	0.0437	1806
	RV	0.0001	0.0001	0.0000	0.0015	1807
	Trades	16,788	1889	4311	21,322	1807
Nikkei225	*Ro*	0.0001	0.0100	−0.0539	0.0972	1807
	Rc	0.0006	0.0135	−0.1055	0.0771	1806
	RV	0.0001	0.0002	0.0000	0.0030	1807
	Trades	1318	552	1071	3602	1807
Swiss	*Ro*	0.0000	0.0080	−0.0412	0.1022	1807
	Rc	0.0003	0.0097	−0.0867	0.0502	1806
	RV	0.0001	0.0001	0.0000	0.0042	1807

Index	Oxford-Man variables	Mean	SD	Min.	Max.	Obs.
	Trades	12,318	2615	6183	27,156	1807
TSX	*Ro*	0.0002	0.0071	−0.0329	0.0455	1807
	Rc	0.0001	0.0079	−0.0431	0.0395	1806
	RV	0.0000	0.0001	0.0000	0.0032	1807
	Trades	1574	65	721	2855	1807

Note: This table lists descriptive statistics for the key market variables for all sample markets. Trades represents the total number of daily trades in each market. The variables are defined in Appendix A. The sample period is January 1, 2011 to November 30, 2017 at a daily frequency. The data source is Oxford-Man Institute of Quantitative Finance.

APPENDIX D: VARIANCE INFLATION FACTORS

	DJIA		FTSE100	
	Social	News	Social	News
$Sent_t^k$	1.1605	1.0055	1.0385	1.1918
Rc_{t-1}	1.0377	1.0390	1.0313	1.1124
VLM_{t-1}	1.1671	1.1051	1.2183	1.2206
RV_{t-1}	1.3629	1.3570	2.0364	2.0358
VIX_{t-1}	1.5501	1.4787	1.8056	1.8522

Note: This table summarizes the five variance inflation factors (VIFs) for each independent variable in Equation (3) for the Dow Jones Industrial Average (DJIA) and Financial Times Stock Exchange (FTSE) 100 indices and social media versus news media, respectively. The variables are defined in Appendix A. To help quantify the severity of multicollinearity in an ordinary least squares (OLS) regression analysis, VIF provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity. If $\boldsymbol{VIF(\hat{\beta}_i) > 10}$, multicollinearity is high, and if $\boldsymbol{5 < VIF(\hat{\beta}_i) \leq 10}$, multicollinearity has certain influences on the model. The square root of the VIF indicates how much larger the standard error increases compared to when the variable had zero correlation with other predictor variables in the model. For example, suppose the VIF of a predictor variable is 5.27 ($\sqrt{\mathbf{5.27}}$ = 2.3). In this case, the standard error for the coefficient of the predictor variable is 2.3 times larger than if the predictor variable had zero correlation with the other predictor variables. This table shows that all of the DJIA and FTSE 100 variables' VIFs are between 1 and 2 (VIF less than 5 is the commonly used cutoff), indicating our models do not suffer from multicollinearity.