Does Twitter Activity Proxy for Idiosyncratic Information ? Evidence from the Indian Stock Market

Sayantan Kundu [1]
Aditya Banerjee [2]

Abstract

This study analyzed the impact of stock-specific information, proxied by Twitter activity, on the stock returns of listed Indian firms. Twitter activities at different levels of information assimilation were considered through tweet-count, retweet-count, and favorite-count as proxies for information supply, information propagation, and information affirmation/validation, respectively. Day-to-day price movement and price movements in two sub-periods: the off-market-hours and the market-hours were considered separately for the analysis. It was also tested whether Twitter activities had an asymmetric impact on stock returns on days with positive and negative sentiments. The study was carried out with data of over 2.4 million tweets about 437 Indian firms listed on the Bombay Stock Exchange for 124 trading days. Panel data analysis with random and fixed effects was employed to test whether the Twitter activity is a significant price mover. The results showed that all three measures of Twitter activity significantly impacted stock returns on a day-to-day basis, especially during the overnight period. However, during market hours, only tweet-count had a significant impact on stock price movements. Further, the results revealed that Twitter activity had the most significant impact on the price of a stock when the market sentiment about the stock was negative. This study is the first large-scale study in the Indian context and opens up the possibilities of further research on these lines.

Keywords : Twitter activity, information supply, asymmetric impact, Indian stocks, returns

JEL Classification Codes : G12, G14, G41

Paper Submission Date : March 10, 2020 ; Paper sent back for Revision : December 23, 2020 ; Paper Acceptance Date : February 20, 2021 ; Paper Published Online : August 30, 2021

In recent times, mass and social media channels have become important mediums for propagating news and speculations related to the financial performance of firms (Renault, 2017 ; Ryan & Taffler, 2004). Twitter is one of the most popular social media platforms for open discussions based on news and speculations about firms. Twitter allows developers access to their data, thus providing researchers with invaluable insight into tweets as a proxy for investor sentiment. In this paper, the impact of the volume of Twitter activities on stock returns is analyzed for tweets arriving during different periods of the day.

Analysis of 'sentiments' in tweets has been the norm in most studies so far. Researchers usually segregated tweets in words and classified those tweets as 'positive' and 'negative' either manually (Aggarwal et al., 2012) or using various software tools (Deng et al., 2018 ; Renault, 2017 ; Sprenger et al., 2014a, 2014b). Then, the segregated tweets were counted, and the count of positive and negative tweets was taken as the variables which are compared against stock returns.

However, as Deng et al. (2018) pointed out, the techniques of segregating tweets based on sentiments were not

[1] *Assistant Professor, Finance (Corresponding Author),* Xavier Business School, St. Xavier's University Kolkata, Action Area IIIB, New Town, Kolkata - 700 160. (Email : sayantan.kundu@sxuk.edu.in) ; ORCID iD : 0000-0002-6047-7156
[2] *Assistant Professor,* School of Business, University of Petroleum and Energy Studies, Dehradun - 248 007, Uttarakhand. (Email : aditya.banerjee@ddn.upes.ac.in) ; ORCID iD : 0000-0002-6776-1643

DOI : https://doi.org/10.17010/ijf/2021/v15i8/165815

yet perfect. Their study used the opinion of human observers to confirm the accuracy of tweet segregation and found it to be less than perfect. Researchers also pointed out that tweets themselves might not affect the movement of stock prices (Bollen et al., 2011). Therefore, this paper aims to analyze whether stock-specific information proxied by Twitter activity can explain the movement of stock prices. Further, it is done on the volume of tweets, without segregating them into positive and negative categories based on sentiments contained in the tweets.

The current paper starts with testing the hypothesis that Twitter activity is a significant 'proxy' for disseminating information about individual stocks. Most of the extant literature used only Twitter count (Bollen et al., 2011 ; Deng et al., 2018 ; Sprenger et al., 2014a, 2014b). However, the affirmation and further propagation of those tweets are neglected. Thus, this study uses three measures of Twitter activity, namely, tweet-count (a proxy for the supply of new information), favorite-count (a proxy for affirmation/validation of information), and retweet-count (a proxy for further propagation of information). Literature is also sparse about whether tweets arriving during the market and off-market hours impact the movement of stock prices differently. Information arriving during market hours is readily absorbed in the price discovery process (Deng et al., 2018). Thus, tweet-count should be the most influential variable during trading hours.

On the other hand, information arriving in off-market-hours is not immediately reflected in prices. Instead, information is discussed, accepted, and gets further propagated before investors act on them. Hence, along with tweet-count, favorite-count and retweet-count should also be significant determinants of overnight stock returns. Thus, the sample has been divided into off-market-hour and market - hours to analyze the impact of Twitter activity volume on overnight and market-hour price movements, respectively.

Further, there is enough evidence in favour of the asymmetric impact of positive and negative information (Sprenger et al., 2014b ; Yang et al., 2015). Thus, another objective of this paper is to test whether Twitter activity influences stock returns differently on the days when the market sentiment about a stock is positive vs. the days when the market sentiment about a stock is negative.

This paper employs panel data methodology on a large sample of over 2.4 million tweets about 437 Indian stocks over 124 trading days. It reveals that the stock-specific Twitter activity is a significant proxy of information propagated through mass and social media, even after controlling for the movement of the market index. The impact of information supply, propagation, and affirmation through Twitter activity is highly significant during off-market hours and is observable in overnight price jumps. However, during market-hours, only the supply of information as proxied by tweet-count is significant. The empirical evidence in this paper proves that information proxied by Twitter activity about individual stocks has a more significant impact on the days when stock returns are negative and has a minimal impact on the days when stock returns are positive. It can be inferred from the results that idiosyncratic information spread through Twitter grabs more attention when market sentiment about a particular stock is negative. Thus, Twitter spreads fear and panic more effectively than enthusiasm.

Review of Literature

Publicly available information is long known to be the key driver of stock prices because news affects the sentiment of investors, which in turn influences their buy/sell decision in the stock markets. Researchers (Berry & Howe, 1994 ; Carretta et al., 2011) studied the impact of the firm-related news published in newspapers, journals, and magazines on stock price movement and found that information supplied through news and media had a measurable and significant impact on stock price movement. Ryan and Taffler (2004) found that corporate news announcements were the most important stock price mover, followed by other information such as analyst opinions. In the Indian context, Tewari and Pathak (2015) found that news about India in the *New York Times* affected the volume of foreign institutional investor (FIIs) investments in the Indian financial markets.

The possible explanation for the impact of news on stock prices is that stocks are traded by people susceptible to their behavioral dispositions. Investor biases play a significant role in their decision-making processes (Mangala & Sharma, 2014) and often prevent investors from making rational decisions (Dangi & Kohli, 2018). Some crucial aspects of investor behavior are their tendency towards herd behavior, which is often the outcome of emotional contagion in response to a piece of information (Raut & Das, 2015). With the availability of the world-wide-web to the masses, propagation and discussion of news have become pervasive, resulting in an increased likelihood of contagions. Therefore, knowledge of how public sentiment towards a firm is affected by its social media activities has become essential in understanding how sentiments affect stock prices.

The advent of the internet has enabled a large number of people to access and propagate a vast amount of information at a low cost. Rubin and Rubin (2010) explained through their study involving Wikipedia editing frequency and accuracy of analyst reports that the internet is the most inexpensive medium for information gathering. They suggested that if the public is more involved in gathering information about a firm, then the analyst forecasts for that firm are also more accurate than usual. There are numerous channels of gathering information on the internet, such as Google search. Researchers (Da et al., 2011 ; Vlastakis & Markellos, 2012) used Google search-volume data on many firms to analyze the impact of demand for such information on stock price levels and found that higher demand for data on Google explained stock price variations.

Social media has emerged as an important medium of communication within the internet. Social media provides a platform for bringing together groups of people and enables them to share their ideas (Yang et al., 2015 ; Zhang et al., 2011). Twitter has become a gold mine for researchers since it came into existence. It is a major social media platform that allows developers access to its users, enabling researchers to gather data about investor sentiment on social media.

Zhang et al. (2011) found that the flurry of emotions expressed by Twitter users regarding stock markets had a significant effect on various stock market indices in the United States on the following day. Another contemporary study by Bollen et al. (2011) revealed that by analyzing the 'public mood' through Twitter activities, one could increase the accuracy of predicting the movement of the stock index (in this case, the DJIA).

Twitter brings together large groups of people who share common interests. Yang et al. (2015) showed that by analyzing the tweets of financial communities within Twitter, it was possible to predict stock market movements more robustly. Twitter microblogs are devoted to discussing stock markets events that signal investors' sentiments towards stocks (Sprenger et al., 2014a). Similarly, Renault (2017) found that investor sentiment conveyed by investors' Twitter activities helped predict intraday index returns (S&P 500) to some extent.

In the Indian context, few papers applied machine-learning-based sentiment analysis with data from Twitter and some stock market blogs to predict the movement of stock prices of banks (Ranjan et al., 2018) and stock indices like NIFTY and Sensex (Bhardwaj et al., 2015) as well as very few chosen companies from different sectors (Nayak et al., 2016). However, for a comprehensive cross-section of stocks, the dependence of stock returns on Twitter activity is still not established in the Indian context.

As evident from extant literature, though studies on the impact of social media activities on stock returns have evolved over the years, some research questions have not been adequately addressed. The first problem is that trading hours are limited each day, but Twitter (and other social media) activities continue throughout the day. Thus, the impact of Twitter activities happening during trading hours and non-trading hours should be separately analyzed as the time available to assimilate Twitter activities differs.

Further, several studies have evidenced an asymmetric impact of good and bad news on stock returns and volatility (Malik, 2011), with negative shocks increasing volatility. Therefore, how the volume of social media activities, which convey the strength of emotions, affects stock price movement differently when stock returns are positive and negative needs to be analyzed separately.

This study uses available social media information in the form of Twitter activity to answer the two problems discussed above that have not been adequately addressed in the literature. Further, this study explores the problems with a large sample of stocks in the Indian context.

Theoretical Background and Hypothesis Formation

If a piece of information is vital for a firm's stock price, it is likely to be tweeted, liked, and retweeted more often. The discussion about a firm on Twitter and the sentiments conveyed through the tweets may act as a mirror for the market-wide sentiment about the firm. The following theoretical derivation helps in explaining the role of information in the movement of stock prices :

$$\ln P_{i,t} = \ln P_{i,t-1} + \varepsilon_{i,t} \qquad (1)$$

Equation 1 is a simple random walk model where the log of the stock price of a firm i for the period t, ($\ln P_{i,t}$) is determined by the stock price in the last period ($\ln P_{i,t-1}$) and an innovation component ($\varepsilon_{i,t}$). The innovation component is the shock, which in turn is determined by the information arriving at the market on that day. Information can be classified into two types : the pervasive market-wide factors ($I_{m,t}$) and the idiosyncratic stock-specific factors ($I_{i,t}$). The total innovation ($\varepsilon_{i,t}$) can thus be partially explained by equation 2.

$$\varepsilon_{i,t} = f(I_{m,t}) + g(I_{i,t}) + u_{i,t} \qquad (2)$$

where $u_{i,t}$ in equation 2 is the random error component. Thus, equation 1 can be expanded as follows :

$$SR_{i,t} = \ln P_{i,t} - \ln P_{i,t-1} = f(I_{m,t}) + g(I_{i,t}) + u_{i,t} \qquad (3)$$

where, $SR_{i,t}$ is the log-return of stock i on day t.

The market-wide factors are proxied by the returns of a well-diversified stock index (MR_t). The idiosyncratic stock-related factors are the information that affects the movement of individual stock prices. If measures of Twitter activity ($A_{i,t}$) are efficient proxies for stock-specific information, then equation 3 can be re-written after incorporating the measurable proxies of the market and the idiosyncratic factors :

$$SR_{i,t} = \alpha + \beta\, MR_t + \gamma A_{i,t} + v_{i,t} \qquad (4)$$

where $v_{i,t}$ is the error term of the equation that can be used in regression analysis. α, β, and γ are the parameters of the regression equation 4.

In equation 4, the vector $A_{i,t}$ represents some measure of Twitter activity. Three measures of Twitter activity are used in this study. The first is the count of tweets ($T_{i,t}$), second is the count of retweets ($R_{i,t}$), and third is the count of favorites ($F_{i,t}$) for each firm on each day. Researchers have bifurcated $T_{i,t}$ into positive and negative tweet-count. The number of tweets per day is a proxy for information supply. This generation and supply of information through social media reduces information asymmetry and moves the prices closer to a new equilibrium. When the existing tweets are retweeted, it leads to the propagation of information. Lastly, the number of favorites is considered as the affirmation/validation of the supplied information. Combined, they reduce adverse selection and promote trade that facilitates better price discovery. Thus, these three variables are taken as explanatory variables, which may have a measurable impact on the stock price movement.

The common thread in the literature suggests that positive tweets positively impact stock prices and negative tweets impact negatively (Deng et al., 2018 ; Yu et al., 2013). Another common finding in the literature is that negative emotions conveyed through Twitter are associated with a more significant impact on stock returns and volatility (Deng et al., 2018 ; Yang et al., 2015).

If absolute returns are considered, then the need to segregate tweets disappears since the results can demonstrate whether the tweets indeed act as a proxy for idiosyncratic stock information that moves stock prices. Equation 5 formulates this notion :

$$|SR_{i,t}| = \alpha + \beta_0 |MR_t| + \gamma_0 A_{i,t} + v_{i,t} \tag{5}$$

Thus, the first set of testable hypotheses (only alternate hypotheses are presented, while the null hypotheses represent the status quo in each case) are :

➯ **H_a1 :** Twitter activity significantly impacts absolute daily stock returns.

➯ **H_a2 :** Twitter activity during off-market hours significantly impacts absolute stock returns during off-market hours (overnight price jump).

➯ **H_a3 :** Twitter activity during market hours significantly impacts absolute stock returns during the market hours.

These hypotheses are tested by validating the significance of γ_0 for three distinct periods (day-today, off-market-hour, market-hour) using equation 5 presented above.

Further, if the stock returns themselves are divided into two exclusive and exhaustive samples of days with negative ($SR^-_{i,t}$) and positive ($SR^+_{i,t}$) returns, then it is possible to explain which one of them is affected more by the information proxied by Twitter. This notion is formulated in equations 6 and 7 :

$$SR^+_{i,t} = \alpha + \beta_1 MR_t + \gamma_1 A_{i,t} + v_{i,t} \tag{6}$$

$$SR^-_{i,t} = \alpha + \beta_2 MR_t + \gamma_2 A_{i,t} + v_{i,t} \tag{7}$$

From equations 6 and 7, the following (alternate) hypotheses can be tested by validating the significance of γ_1 and γ_2 :

➯ **H_a4 :** Twitter activity significantly impacts stock returns on days when the market sentiment for the stock is positive.

➯ **H_a5 :** Twitter activity significantly impacts stock returns on days when the market sentiment for the stock is negative.

➯ **H_a6 :** Overnight Twitter activity significantly impacts overnight stock returns on days when the market sentiment for the stock is positive.

➯ **H_a7 :** Overnight Twitter activity significantly impacts overnight stock returns on days when the market sentiment for the stock is negative.

➯ **H_a8 :** Market-hour Twitter activity significantly impacts market-hour stock returns on days when the market sentiment for the stock is positive.

➯ **H_a9 :** Market-hour Twitter activity significantly impacts market-hour stock returns on days when the market sentiment for the stock is negative.

All the hypotheses are tested for three measures of Twitter activity : Tweet-count, favorite-count, and retweet-count.

Data and Methodology

Twitter and stock price data were collected from March 6, 2019 to September 6, 2019. After removing non-trading days, the final sample consists of Twitter activity data of 124 days for 437 Indian firms listed on the Bombay Stock Exchange (BSE), providing a balanced panel of 54,188 firm-days. The market index considered is the S&P BSE 500 Index hosted by BSE. The daily open-high-low-close (OHLC) data for the firms in the sample and BSE 500 were collected from the PROWESS database hosted by the Centre for Monitoring Indian Economy (CMIE). Twitter data were collected from Twitter Server using the "*twitteR*" package in R. The total Twitter activity on a day-to-day basis (16:00 hours of the previous day to 16:00 hours of the current day) consists of the entire sample, which is bifurcated into two different periods for a more detailed analysis as follows :

(i) The first period consists of off-market-hour Twitter activities from the closing of the market on the previous day at 16:00 hours to the current day's market opening at 09:00 hours. The overnight price jump is computed from the adjusted closing price of the previous day and the adjusted opening price of the current day, as shown in equation 8. It also includes Twitter activities during the weekends and market holidays. The panel consists of 54,188 firm-years data.

$$Overnight\ Price\ Jump_t = ln\left(\frac{Adj.\ Opening\ Price_t}{Adj.\ Closing\ Price_{t-1}}\right) \qquad (8)$$

(ii) The second period consists of Twitter activities during the market hours, from the opening hour at 09:00 to the closing hour at 16:00 hours. The return for this period is computed from the adjusted opening price and closing price of the day, as shown in equation 9. The subsample consists of data on 429 firms for 124 days resulting in 53,196 firm-years panel data.

$$Market\ Hour\ Stock\ Returns_t = ln\left(\frac{Adj.\ Closing\ Price_t}{Adj.\ Opening\ Price_t}\right) \qquad (9)$$

For the day-to-day sample, the stock returns are computed from the adjusted closing prices of the previous day and the current day, as shown in equation 10.

$$Overnight\ Price\ Jump_t = ln\left(\frac{Adj.\ Closing\ Price_t}{Adj.\ Closing\ Price_{t-1}}\right) \qquad (10)$$

The panel data for each of the three periods (total, off-market, and market-hour) is further bifurcated into days where stock returns are positive and days where stock returns are negative. It resulted in two sets of unbalanced panels for each of the three periodic samples. Market returns for each period are considered as a control variable to account for the market-wide sentiments.

The regression models based on equations 5, 6, and 7 are fitted for each sample and their bifurcated subsamples. The regression parameters for each equation is estimated thrice with three different explanatory variables : $T_{i,t}$, $R_{i,t}$, and $F_{i,t}$. The panel regression for the entire sample has been controlled with random effects and fixed effects, as applicable.

Cross-section fixed effects are present in the unbalanced panels and are controlled. However, the bifurcated samples form unbalanced panels, and therefore, it is impossible to control for period random effects. The following section lays out the results of the empirical analysis.

Analysis and Results

Table 1 presents the summary statistics of the samples. Intuitively, it can be observed that though stock returns are negative on average, the average overnight price jump is positive for the sample period. The number of observations with negative returns is higher than that with positive returns for daily and market-hour samples. However, it is the opposite for the overnight sample, where there are more instances of positive price jumps.

Table 1. *Descriptive Statistics*

	Entire Day-to-Day Sample			Market Hour Sample			Overnight Sample		
Variables	Mean	Std. Deviation	Observations	Mean	Std. Deviation	Observations	Mean	Std. Deviation	Observations
Full Sample									
Tweet-Count ($T_{i,t}$)	44.63	203.80		19.47	107.18		25.52	123.77	
Retweet-Count ($R_{i,t}$)	10437.54	288317.5		3973.42	133677.0		6536.86	214182.4	
Favorite-Count ($F_{i,t}$)	57.64	497.21	54,188	27.76	370.19	53,196	30.39	266.53	54,188
Stock Return ($SR_{i,t}$)	–0.0015	0.0264		–0.0036	0.0239		0.0021	0.0158	
Market Return ($MR_{i,t}$)	–0.0002	0.0086		–0.0020	0.0076		0.0018	0.0030	
Positive Return Days									
Tweet-Count ($T_{i,t}$)	47.16	215.87		20.39	115.12		26.75	128.58	
Retweet-Count ($R_{i,t}$)	11745.34	337070.6		3871.354	104884.2		6828.55	206161.7	
Favorite-Count ($F_{i,t}$)	59.52	503.44	25,020	28.32	356.72	22,202	31.62	272.65	34,465
Positive Stock Return ($SR_{i,t}$)	0.0164	0.0184		0.0157	0.0177		0.0077	0.0098	
Negative Return Days									
Tweet-Count ($T_{i,t}$)	42.46	192.82		18.82	101.11		23.35	114.83	
Retweet-Count ($R_{i,t}$)	9315.725	238691.5		4046.531	150965.0		6027.15	227524.8	
Favorite-Count ($F_{i,t}$)	56.03	491.80	29,168	27.36	379.56	30,994	28.24	255.48	19,723
Negative Stock Return ($SR_{i,t}$)	–0.0169	0.0222		–0.0175	0.0172		–0.0077	0.0192	

Note. Table 1 shows the descriptive statistics of the sample under study. The entire day-to-day sample consists of all tweets from the previous day's market closing hours (16:00 hours). The market hour sample includes tweets occurring only during the trading hours (09:00 hours to 16:00 hours) of each day. The overnight sample includes tweets between 16:00 hours of the previous day to 09:00 hours of the current day.

Table 2 shows the results of regression equation 5 for the entire sample for day-to-day Twitter activities. The dependent variable, daily absolute stock return, is regressed on the three measures of Twitter activity, that is, tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$). Absolute market return ($|M_t|$) is held as the control variable. The results show that all measures of Twitter activities are significantly and positively related to absolute stock returns after controlling for absolute market returns. Thus, the daily stock returns reflect the number of tweets about the firms that have arrived in the market in the last 24-hours. The accepted information (proxied by favorite-count) and the propagated information (proxied by retweet-count) are also regarded as crucial by the market. Thus, it can be asserted that the measures of Twitter activity accurately proxy the relevant stock-specific

Table 2. *Impact of Twitter Activity on Day-to-Day Returns*

	Full Sample. Dependent Variable $\|SR_{i,t}\|$		
Variables	**Tweets**	**Retweets**	**Favorites**
Intercept (α)	0.0122***	0.0123***	0.0123***
Tweet-Count ($T_{i,t}$)	$2.49\times10^{-06***}$		
Retweet-Count ($R_{i,t}$)		$7.88\times10^{-10***}$	
Favorite-Count ($F_{i,t}$)			$5.96\times10^{-07***}$
Abs Market Return $\|MR_t\|$	0.6539***	0.6536***	0.6535***
Adj. R^2	0.0034	0.0031	0.0032
F-Statistic	94.6939***	85.7957***	88.1406***
Period Effects	RE	RE	RE
Cross -Section Effects	RE	RE	RE

Note. Table 2 shows the results of the regression of day-to-day absolute stock returns ($|SR_{i,t}|$) on three measures of Twitter activities : tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$) in the corresponding period. Absolute market return ($|MR_t|$) is the control variable. Panel data fixed effects (FE) and random effects (RE) are controlled for as applicable.

The symbol '***' refers to significance at 1% level ; '**' refers to significance at 5% level ; and '*' refers to significance at 10% level.

information. Absolute market returns have a positive and significant relationship with absolute stock returns in all instances, as expected. Thus, significant evidence in favour of the research hypothesis H_a1 is found.

Table 3 shows the results of regression equation 5 when only the overnight Twitter activity is considered. The dependent variable is the overnight price jump. The results are similar to those of Table 2. All three measures

Table 3. *Impact of Twitter Activity on Overnight Returns*

	Full Sample. Dependent Variable $\|SR_{i,t}\|$		
Variables	**Tweets**	**Retweets**	**Favorites**
Intercept (α)	0.0055***	0.0055***	0.0055***
Tweet-Count ($T_{i,t}$)	$3.45\times10^{-06***}$		
Retweet-Count ($R_{i,t}$)		$6.19\times10^{-10**}$	
Favorite-Count ($F_{i,t}$)			$1.64\times10^{-06***}$
Abs Market Return $\|MR_t\|$	0.9143***	0.0591***	0.9129***
Adj. R^2	0.0596	0.0591	0.0599
F-Statistic	8.8402***	8.7679***	8.8865***
Period Effects	RE	RE	RE
Cross - Section Effects	FE	FE	FE

Note. Table 3 shows the results of the regression of overnight absolute stock returns ($|SR_{i,t}|$) on three measures of Twitter activities : Tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$) in the corresponding period. Absolute market return ($|MR_t|$) is the control variable. Panel data fixed effects (FE) and random effects (RE) are controlled for as applicable.

The symbol '***' refers to significance at 1% level ; '**' refers to significance at 5% level ; and '*' refers to significance at 10% level.

Table 4. *Impact of Twitter Activity on Market-Hour Returns*

	Full Sample. Dependent Variable $\lvert SR_{i,t} \rvert$		
Variables	**Tweets**	**Retweets**	**Favorites**
Intercept (α)	0.0131***	0.0132***	0.0132***
Tweet-Count ($T_{i,t}$)	$2.38\times10^{-06***}$		
Retweet-Count ($R_{i,t}$)		-9.70×10^{-11}	
Favorite-Count ($F_{i,t}$)			8.22×10^{-08}
Abs Market Return $\lvert MR_t \rvert$	0.5766***	0.5765***	0.5765***
Adj. R^2	0.0028	0.0026	0.0026
F-Statistic	75.6006***	70.4117***	70.4863***
Period Effects	RE	RE	RE
Cross-Section Effects	RE	RE	RE

Note. Table 4 shows the results of the regression of absolute stock returns ($\lvert SR_{i,t} \rvert$) during market-hour on three measures of Twitter activities : Tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$), in the corresponding period. Absolute market return ($\lvert MR_t \rvert$) is the control variable. Panel data fixed effects (FE) and random effects (RE) are controlled for as applicable.

The symbol '***' refers to significance at 1% level ; '**' refers to significance at 5% level, and '*' refers to significance at 10% level.

of Twitter activity are significant and positively related to the overnight jump, and thus, the hypothesis H_a2 is valid. It can be concluded that information arriving during the off-market hours gets sufficient attention from the investors, reflecting the overnight price jump.

The results of regression equation 5, when considering only market hour Twitter activity, are presented in Table 4. The results are different from those of the entire sample (day-to-day) and off-market hour Twitter activities. Only the tweet-count ($T_{i,t}$) is significantly and positively related to returns during the market hour, while the other two measures of Twitter activity are insignificant. It suggests that investors only pay attention to the volume of new information arriving in the market during market hours. The further dissemination of information (retweets) and its validation (favorites) are not considered during the market hours as there is not enough time to do so. Results from Table 4 provide partial evidence in favour of the hypothesis H_a3.

The daily stock returns are further divided into positive and negative returns. Table 5 shows the results of the regression equations 6 and 7 for the segregated negative and positive stock return days. There is a stark contrast in results between the days with positive stock returns and those with negative stock returns. All measures of Twitter activity have a significant and negative relationship with negative stock returns (validating hypothesis H_a5). However, they have no significant relationship with positive returns (no significant evidence in favour of hypothesis H_a4).

Regression equations 6 and 7 are fitted after classifying the overnight returns into positive and negative days. The results presented in Table 6 are similar to those depicted in Table 5. All three measures of Twitter activity are significantly and negatively related to the negative overnight price jumps. Thus, evidence strongly supports the hypothesis H_a7. However, only favorite-count ($F_{i,t}$) has a significant and positive relationship with a positive overnight jump, providing partial evidence in favour of hypothesis H_a6.

Results from Table 6 show that when the stock-specific sentiment is negative, the arrival, affirmation, and propagation of idiosyncratic information causes higher negative overnight price jumps. On the contrary, when the stock-specific sentiment is positive, only favorite-count, which is a proxy for affirmation/validation of the

Table 5. *Impact of Twitter Activity on Day-to-Day Positive and Negative Stock Returns*

	Panel A : Positive Stock Returns. Dependent Variable $SR^{+}_{i,t}$			Panel B : Negative Stock Returns. Dependent Variable $SR^{-}_{i,t}$		
Variables	**Tweets**	**Retweets**	**Favorites**	**Tweets**	**Retweets**	**Favorites**
Intercept (α)	0.0149***	0.0150***	0.0150***	–0.0079***	–0.0081***	–0.0080***
Tweet-Count ($T_{i,t}$)	6.57×10^{-07}			–9.35×10^{-06***}		
Retweet-Count ($R_{i,t}$)		–6.47×10^{-11}			–1.31×10^{-09***}	
Favorite-Count ($F_{i,t}$)			1.37×10^{-07}			–3.87×10^{-06***}
Abs Market Return $\vert MR_t \vert$	0.4847***	0.4846***	0.4846***	0.4579***	0.4580***	0.4548***
Adj. R^2	0.1402	0.1401	0.1401	0.0574	0.0557	0.0578
F-Statistic	10.3112***	10.3084***	10.3093***	3.7440***	3.6559***	3.7621***

Note. Table 5 is divided into two parts. Panel A shows the results of the regression of day-to-day positive stock returns ($SR^{+}_{i,t}$) on three measures of Twitter activities : Tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$) in the corresponding period. The market return for the period (MR_t) is the control variable. In Panel B, the dependent variable is the day-to-day negative stock returns ($SR^{-}_{i,t}$) ; the explanatory variables and the control variable remain the same. Cross-section fixed effects (FE) are controlled for in all regressions.

The symbol '***' refers to significance at 1% level ; '**' refers to significance at 5% level ; and '*' refers to significance at 10% level.

Table 6. *Impact of Twitter Activity on Overnight Positive and Negative Stock Returns*

	Panel A : Positive Stock Returns. Dependent Variable $SR^{+}_{i,t}$			Panel B : Negative Stock Returns. Dependent Variable $SR^{-}_{i,t}$		
Variables	**Tweets**	**Retweets**	**Favorites**	**Tweets**	**Retweets**	**Favorites**
Intercept (α)	0.0056***	0.0056***	0.0056***	–0.0078***	–0.0081***	–0.0080***
Tweet-Count ($T_{i,t}$)	3.68×10^{-07}			–9.35×10^{-06***}		
Retweet-Count ($R_{i,t}$)		5.88×10^{-11}			–1.31×10^{-09**}	
Favorite-Count ($F_{i,t}$)			4.31×10^{-07**}			–3.87×10^{-06***}
Abs Market Return $\vert MR_t \vert$	0.8911***	0.8910***	0.8911***	0.4579***	0.4580***	0.4548***
Adj. R^2	0.1575	0.1575	0.1576	0.0574	0.0557	0.0578
F-Statistic	15.7114***	15.7099***	15.7241***	3.7440***	3.6559***	3.7621***

Note. Table 6 is divided into two parts. Panel A shows the results of the regression of overnight positive stock returns ($SR^{+}_{i,t}$) on three measures of Twitter activities : Tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$) in the corresponding period. The market return for the period (MR_t) is the control variable. In Panel B, the dependent variable is the overnight negative stock returns ($SR^{-}_{i,t}$) ; the explanatory variables and the control variable remain the same. Cross-section fixed effects (FE) are controlled for in all regressions.

The symbol '***' refers to significance at 1% level ; '**' refers to significance at 5% level ; and '*' refers to significance at 10% level.

information, seems to impact stock returns significantly. Thus, the arrival or propagation of positive information through off-market-hour tweets is not crucial unless many people affirm the same.

Table 7 presents the results obtained from regression equations 6 and 7 when fitted to the positive and negative market-hour stock returns. The results further strengthen the findings in Table 4. Tweet-count (arrival of information) is significantly related to both positive and negative market-hour returns. Higher tweet-count accentuates returns in days irrespective of whether the stock-specific sentiment is positive or negative, while other measures of Twitter activity are insignificant. It shows that only new information arriving in the form of tweets during the stock market during trading hours is important for investors, and its impact is assimilated in the stock prices. The results suggest that the research hypotheses H_a8 and H_a9 are valid only for tweet-count.

Table 7. ***Impact of Twitter Activity on Market-Hour Positive and Negative Stock Returns***

	Panel A : Positive Stock Returns. Dependent Variable $SR^{+}_{i,t}$			Panel B : Negative Stock Returns. Dependent Variable $SR^{-}_{i,t}$		
Variables	**Tweets**	**Retweets**	**Favorites**	**Tweets**	**Retweets**	**Favorites**
Intercept (α)	0.0154***	0.0154***	0.0154***	–0.0155***	–0.0155***	–0.0155***
Tweet-Count ($T_{i,t}$)	$2.75\times10^{-06**}$			$-2.33\times10^{-06**}$		
Retweet-Count ($R_{i,t}$)		4.82×10^{-10}			1.79×10^{-10}	
Favorite-Count ($F_{i,t}$)			1.33×10^{-07}			4.36×10^{-08}
Abs Market Return $\lvert MR_t \rvert$	0.4103***	0.4101***	0.4102***	0.5049***	0.5049***	0.5050***
Adj. R^2	0.1118	0.1115	0.1115	0.1663	0.1662	0.1662
F-Statistic	7.4957***	7.4799***	7.4799***	15.3761***	15.3622***	15.3620***

Note. This table is divided into two parts. Panel A shows the results of the regression of market-hour positive stock returns ($SR^{+}_{i,t}$) on three measures of Twitter activities : Tweet-count ($T_{i,t}$), retweet-count ($R_{i,t}$), and favorite-count ($F_{i,t}$) in the corresponding period. The market return for the period (MR_t) is the control variable. In Panel B, the dependent variable is the market-hour negative stock returns ($SR^{-}_{i,t}$) ; the explanatory variables and the control variable remain the same. Cross-section fixed effects (FE) are controlled for in all regressions.

The symbol '***' refers to significance at 1% level ; '**' refers to significance at 5% level ; and '*' refers to significance at 10% level.

Discussion

It is evident from the discussion in the previous section that three different measures of Twitter activity impact stock returns differently during different periods of a day. Moreover, Twitter activity impacts stock returns asymmetrically when market sentiment for a stock is positive versus when it is negative. To contrast the impact of Twitter activities on stock returns across the different samples used in this study, a comprehensive summary of the results from Table 2 through Table 7 is provided in Table 8. The summary of the results indicates that the

Table 8. ***Summary of Results***

Samples	Explanatory Variable	Tweet-Count	Retweet-Count	Favorite-Count
Day-to-day	Absolute Return (full sample)	Y	Y	Y
	Positive Return days (sub-sample)	N	N	N
	Negative Return days (sub-sample)	Y	Y	Y
Overnight	Absolute Return (full sample)	Y	Y	Y
	Positive Return days (sub-sample)	N	N	Y
	Negative Return days (sub-sample)	Y	Y	Y
Market-hour	Absolute Return (full sample)	Y	N	N
	Positive Return days (sub-sample)	Y	N	N
	Negative Return days (sub-sample)	Y	N	N

Note. Table 8 summarizes the results reported in Table 2 – Table 7. The significance of the three measures of Twitter activity (tweet-count, retweet-count, and favorite-count) on stock price movements for different samples are summarized. The dependent variable is absolute stock returns in the full sample for each period (day-to-day, overnight, and market-hour). The sample is bifurcated in days where stock returns are positive and negative ; wherein, the dependent variables are the respective positive and negative stock returns. If the impact of a measure of Twitter activity is significant at 5% level on the corresponding measure of the stock price movement of the given period (as reported in Tables 2 through 7), then it is indicated as 'Y' ; else, it is indicated as 'N.'

information about individual stocks spread through Twitter activity is most effective when sentiment about those stocks is negative.

Implications

Theoretical Implications

The results point towards information transmission through social media in the form of information supply, information propagation, and information affirmation/validation. It can be conjectured by comparing the results from the market hour and overnight samples (Table 4) that new information arriving during the market hours is readily absorbed in the price discovery process. This finding is similar to the findings of Deng et al. (2018). While trading happens in the stock markets, the supply of information is sufficient in driving the stock prices, and there is little time for analyzing and acting on all aspects of stock-specific information. Thus, tweet-count becomes the most influential variable during market hours.

In contrast, information arriving during the off-market hours gets sufficient time to be analyzed. Thus, other aspects of information, such as the further propagation of information and the affirmation of information, get attention during off-market-hours and are factored significantly in the overnight price jumps when investors act on them in the market opening auction. Hence, along with tweet-count, favorite-count and retweet-count also become significant determinants of overnight price jumps. Therefore, it can be inferred that investors need more time to incorporate the various aspects of information in their trading decisions.

Tables 5 and 6 show that the investors pay more attention to the prevalent stock-specific information when the sentiment about a stock is negative. When the sentiment about a stock is negative, it is more likely that the negative sentiment is discussed and propagated through Twitter activities. It, in turn, brings the stock price further down. Parallels of this finding can be drawn with those of prior researchers (Deng et al., 2018 ; Sprenger et al., 2014b ; Yang et al., 2015) who suggested that negative information leads to higher volatility in stock prices and stronger market reaction than positive information does.

Managerial Implications

The implications of the results are substantial for firm managers. Since the maximization of firms' value is a principal goal of managers, managing social media handles efficiently may help achieve it. Firms that have active social media handles attract more social media activity from the public. Customers and other stakeholders often discuss their experience with the firm's products and services on its social media handle. For managers, paying attention to social media activities may provide them an early warning mechanism to steer their decision-making process.

Therefore, the volume of Twitter activities has a significant impact on stock returns and volatility, and the results have important implications for investors and fund managers. Fund managers may tune their short-term trading decisions based on volumes of Twitter discussion, favorite count, and retweet count. However, though coefficients are significant, their values are small, indicating low sensitivity. Hence, implementing such models should be exercised with caution, and other factors important in price formation should be factored in.

Conclusion

This study analyzes the impact of stock-specific information, proxied by Twitter activity, on the stock returns of listed Indian firms. This paper tests whether Twitter activity is a significant price mover. Day-to-day price movement and price movements in the two sub-periods : the off-market-hours and the market-hours are considered separately for the analysis. Whether Twitter activities have an asymmetric impact on days with positive and negative sentiments has also been tested. Twitter activities at different levels of information assimilation have been considered through tweet-count, retweet-count, and favorite-count as proxies for information supply, information propagation, and information affirmation/validation, respectively.

The findings suggest that the day-to-day stock returns react to stock-specific information proxied by the three chosen measures of Twitter activities. However, the impact differs across market-hours and off-market-hours. The number of tweets arriving is a relevant mover of stock prices in both off-market-hours and market-hours. However, the favorite-count and retweet-count numbers have a significant impact on stock returns only in the off-market-hours. During the market-hours, only new information is factored into prices as investors and traders do not have the time to pay attention to the propagation or affirmation aspects of the information. On the contrary, during off-market-hours, investors and traders have ample time in their hands to take cognizance of how much the information is propagated or affirmed.

On a day-to-day basis, when the sentiment about a stock is positive, the information proxied by Twitter activity is not significant. However, negative sentiment about stocks intensifies with the arrival, propagation, and affirmation of information proxied by Twitter activity. Only affirmed information has a significant impact on positive overnight sentiment. On the contrary, the magnitude of negative overnight price jumps increases with the arrival, propagation, and affirmation of information proxied by Twitter activity. When market hour price movements are considered, only tweet-count matters for both positive and negative returns, suggesting that the traders only consider new information during trading hours, and they do not wait for the information to propagate or be affirmed.

This study uses an extensive sample of over 2.4 million tweets for 437 listed stocks on the Bombay Stock Exchange for about half a year. Therefore, the results obtained are not affected by small sample bias, and therefore, can be considered robust. This study is the most comprehensive research using Twitter, so far, in the context of the Indian stock market that uses a broad cross-section of Indian stocks.

Limitations of the Study and Scope for Further Research

The tweets obtained were not segregated as negative and positive since the techniques for doing so may have errors, as pointed in the extant literature. Thus, the study avoids certain model risks, though the findings of this study may be corroborated in the future by segregating the tweets based on the sentiment they convey to gain more insights on the subject. Future studies may also look at the granular impact of Twitter activities on stock price movement with high-frequency intra-day data. Using high-frequency intra-day data may reveal the effect of Twitter activities during periods of volatility clustering and provide more in-depth insights into the problem. Additionally, this study has not analyzed the effect of Twitter volumes on long-term stock returns, which could be another future research scope.

Authors' Contribution

During a conversation, Prof. Sayantan Kundu and Aditya Banerjee conceived that social media discussions might impact the stock returns. Mr. Banerjee subsequently did the literature review and found research papers from reputed journals on the relevant topic. Both authors further discussed and conceptualized the study's objective, which enabled Prof. Kundu to form the theoretical framework that helped them build the hypotheses. After that, Prof. Kundu and Mr. Banerjee wrote R-codes to download, save, and analyze company-specific tweets. Mr. Banerjee then painstakingly ran the code every week to download Twitter data. Finally, they worked together to investigate the tweet volumes and developed the regression models to test the hypotheses. The manuscript was written, reviewed, and improved by both authors. Upon receiving the journal's reviewers' comments, they again worked together to answer the comments and improve the manuscript.

Conflict of Interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Funding Acknowledgement

The authors received no financial support for the research, authorship, and/or for the publication of this manuscript.

References

Aggarwal, R., Gopal, R., Gupta, A., & Singh, H. (2012). Putting money where the mouths are : The relation between venture financing and electronic word-of-mouth. *Information Systems Research, 23*(3–part–2), 976 – 992. https://doi.org/10.1287/isre.1110.0402

Berry, T. D., & Howe, K. M. (1994). Public information arrival. *The Journal of Finance, 49*(4), 1331–1346. https://doi.org/10.1111/j.1540-6261.1994.tb02456.x

Bhardwaj, A., Narayan, Y., Vanraj, Pawan, & Dutta, M. (2015). Sentiment analysis for Indian stock market prediction using Sensex and Nifty. *Procedia Computer Science, 70,* 85 – 91. https://doi.org/10.1016/j.procs.2015.10.043

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1 – 8. https://doi.org/10.1016/j.jocs.2010.12.007

Carretta, A., Farina, V., Martelli, D., Fiordelisi, F., & Schwizer, P. (2011). The impact of corporate governance press news on stock market returns. *European Financial Management, 17*(1), 100–119. https://doi.org/10.1111/j.1468-036X.2010.00548.x

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance, 66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Dangi, M., & Kohli, B. (2018). Role of behavioral biases in investment decisions : A factor analysis. *Indian Journal of Finance, 12*(3), 43 – 57. https://doi.org/10.17010/ijf/2018/v12i3/121997

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.

Malik, F. (2011). Estimating the impact of good news on stock market volatility. *Applied Financial Economics, 21*(8), 545 – 554. https://doi.org/10.1080/09603107.2010.534063

Mangala, D., & Sharma, M. (2014). A brief mapping of theory and evidence of investors' behavioural biases. *Indian Journal of Finance, 8*(8), 44 – 56. https://doi.org/10.17010/ijf/2014/v8i8/71855

Nayak, A., Pai, M. M., & Pai, R. M. (2016). Prediction models for Indian stock market. *Procedia Computer Science, 89,* 441– 449. https://doi.org/10.1016/j.procs.2016.06.096

Ranjan, S., Singh, I., Dua, S., & Sood, S. (2018). Sentiment analysis of stock blog network communities for prediction of stock price trends. *Indian Journal of Finance, 12*(12), 7–21. https://doi.org/10.17010/ijf/2018/v12i12/139888

Raut, R. K., & Das, N. (2015). Behavioral prospects of individual investor decision making process : A review. *Indian Journal of Finance, 9*(4), 44 – 55. https://doi.org/10.17010/ijf/2015/v9i4/71457

Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance, 84,* 25 – 40. https://doi.org/10.1016/j.jbankfin.2017.07.002

Rubin, A., & Rubin, E. (2010). Informed investors and the internet. *Journal of Business Finance & Accounting, 37*(7 – 8), 841 – 865. https://doi.org/10.1111/j.1468-5957.2010.02187.x

Ryan, P., & Taffler, R. J. (2004). Are economically significant stock returns and trading volumes driven by firm - specific news releases ? *Journal of Business Finance & Accounting, 31*(1–2), 49 – 82. https://doi.org/10.1111/j.0306-686X.2004.0002.x

Sprenger, T. O., Sandner, P. G., Tumasjan, A., & Welpe, I. M. (2014a.). News or noise ? Using Twitter to identify and understand company - specific news flow. *Journal of Business Finance & Accounting,41*(7 – 8), 791 – 830. https://doi.org/10.1111/jbfa.12086

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014b.). Tweets and trades : The information content of stock microblogs. *European Financial Management, 20*(5), 926 – 957. https://doi.org/10.1111/j.1468-036X.2013.12007.x

Tewari, R., & Pathak, T. (2015). A correlation between mass media communication and foreign investments in India. *Prabandhan : Indian Journal of Management, 8*(10), 32–42. https://doi.org/10.17010/pijom/2015/v8i10/79829

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance, 36*(6), 1808 – 1821. https://doi.org/10.1016/j.jbankfin.2012.02.007

Yang, S. Y., Mo, S. Y., & Liu, A. (2015). Twitter financial community sentiment and its predictive relationship to stock market movement. *Quantitative Finance,15*(10), 1637–1656. https://doi.org/10.1080/14697688.2015.1071078

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value : A sentiment analysis approach. *Decision Support Systems, 55*(4), 919–926. https://doi.org/10.1016/j.dss.2012.12.028

Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through Twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences, 26,* 55–62. https://doi.org/10.1016/j.sbspro.2011.10.562

About the Authors

Dr. Sayantan Kundu, Fellow of IIM Calcutta, is an Assistant Professor of Finance at the Xavier Business School under St. Xavier's University Kolkata. He was previously an Assistant Professor at IIM Ranchi. His research and teaching interests lie in capital markets, asset pricing, corporate finance, and financial institutions. Dr. Kundu has published papers in reputed indexed journals and is a reviewer for few reputed journals.

Aditya Banerjee, a PhD Scholar in accounting and finance at the Indian Institute of Management Ranchi, is presently an Assistant Professor at School of Business, University of Petroleum and Energy Studies, Dehradun, Uttarakhand. His research interests lie in asset pricing, market efficiency, financial institutions, and corporate finance, and he has published papers in reputed indexed journals.