Association for Information Systems

AIS Electronic Library (AISeL)

ICIS 2019 Proceedings

Crowds, Social Media and Digital Collaborations

Information or noise: How Twitter facilitates stock market information aggregation

Florian Kiesel
Grenoble Ecole de Management, florian.kiesel@grenoble-em.com

Thomas Pöppe
Technische Universität Darmstadt, poeppe@bwl.tu-darmstadt.de

Sascha Kolaric
Technische Universität Darmstadt, kolaric@bwl.tu-darmstadt.de

Dirk Schiereck
Technische Universität Darmstadt, schiereck@bwl.tu-darmstadt.de

Information or Noise: How Twitter Facilitates Stock Market Information Aggregation*

Completed Research Paper

Thomas Pöppe
Technische Universität Darmstadt
Dept. of Economics and Law
thomas@poeppe.com

Florian Kiesel
Grenoble Ecole de Management,
Univ Grenoble Alpes ComUE
florian.kiesel@grenoble-em.com

Sascha Kolaric
Technische Universität Darmstadt
Dept. of Economics and Law
kolaric@bwl.tu-darmstadt.de

Dirk Schiereck
Technische Universität Darmstadt
Dept. of Economics and Law
schiereck@bwl.tu-darmstadt.de

Abstract

We assess the relevance of Twitter for stock-relevant information dissemination in financial markets on the single stock level. We use a unique dataset including more than 12 million Twitter feeds linked to specific firms. Using intraday data for the computation of advanced trading metrics, such as effective spreads, intraday volatility, and a daily version of the microstructure variable probability of informed trading (PIN), we measure the impact of Twitter activity on trading and information dissemination. The PIN model indicates that more uninformed than informed traders rush to the market along with rising Twitter activity. These results indicate that Twitter serves as an excellent indicator of news that is relevant for the stock market. However, we show that Twitter does not lead traditional news channels. In contrast, Twitter activity follows the market and has no predictive power with regard to future stock trading volume or volatility on the single stock level.

Keywords: Social media, Twitter, microblogging, probability of informed trading (PIN), information dissemination, information asymmetry

Introduction

Since Twitter's foundation in 2006, activity and content created on it grew exponentially and firms increasingly use Twitter as a channel to directly communicate with their customers (Culnan et al. 2010). This makes it an important channel to publish news without first being filtered by an editorial office and Twitter may thereby provide relevant information for financial markets almost instantaneously. There is evidence that the information content of microblogs, such as Twitter, has significant impact on financial markets (e.g., Li et al. 2018; Nisar and Yeung 2018; Zhang et al. 2011) and that social media networks help to reduce market inefficiency and reduce the cost of capital as information dissemination is easier using

---

* This paper is partially based on the fifth chapter of the PhD thesis "Information Asymmetry and Information Dissemination in High-Frequency Capital Markets" of Thomas Pöppe.

networks and social connections (Han and Yang 2013). Information sharing through social interaction has important consequences on the cost of information and the trading behaviour of market participants (e.g., Hong et al. 2004, 2005; Shive 2010). In particular, using information technology to aggregate social media activity allows for the creation of sentiment indicators, which have some forecasting power on future stock price movements (e.g., Deng et al. 2018; Li et al. 2018; Leitch and Sherif 2017, Sul et al. 2017). Bollen et al. (2011) and Zhang et al. (2011) provide first evidence that the sentiment of Twitter feeds has some predictive power with regard to overall future market movements.[1]

The present study distinguishes itself from majority of the existing literature among two dimensions: First, while Deng et al. (2018), Sul et al. (2017) and Li et al. (2018) document that Twitter sentiment has some predictive power on overall stock market movements, we directly connect Twitter activity and stock market data on the single stock level as opposed to the overall market level. This allows us to draw important conclusions whether Twitter is a relevant information source in stock markets for trading activity of the respective firm. To this end, we use a comprehensive and complete dataset consisting of more than 12 million Twitter feeds related to specific German stock-listed companies.

Second, while prior studies rely on simpler metrics retrieved from daily trading data, we use intraday stock data and compute advanced metrics for information processing in trading, such as effective spreads and intraday volatility. Additionally, we compute a modified (daily) version of the microstructure variable probability of informed trading (PIN), originally introduced by Easley et al. (1996) and extended by Easley et al. (1997), and determine whether Twitter activity attracts more uninformed traders to invest or divest in a company's stock than informed ones. We thereby extend the findings of Li et al. (2018) who show that an increase in message volume may induce trading from side-lined investors or uninformed investors.

Our results provide evidence that Twitter serves as an excellent indicator of news arrivals that affect stock trading activity (e.g., Deng et al. 2018; Li et al. 2018; Nisar and Yeung 2018; Sul et al. 2017). In particular, trading volume, volatility, and spread rise contemporaneously with Twitter activity. Additionally, the PIN model indicates that more uninformed than informed traders rush to the market along with rising Twitter activity, a finding which contributes to the discussion on microblogs and their information content. However, our results also document that Twitter does not lead traditional news channels. On the contrary, on an intraday basis it appears that Twitter activity follows the market and has no predictive power with regard to future stock trading volume or volatility.

The remainder of this paper is organised as follows: The next section defines the key variables, while the following section describes the data collection process and provides descriptive sample statistics. The following two section provide the empirical results and additional robustness checks, while the final sections conclude the paper with a discussion of the main findings, its managerial implications and provides inherent limitations.

Research Design

Metrics of Trading Activity

To characterize trading activity, we analyze stock return, trading volume, spread, volatility, and trade size. The variable *lnVolume* is defined as the product of the number of traded shares and the share price in Euro (EUR). The bid-ask spread is one of the most widely employed trading metrics to measure transaction costs, liquidity, information risk, or the degree of information asymmetry. We employ two measures: First, the percent quoted spread (*qspread*), which is defined as the value of the absolute spread relative to the midpoint price for stock i at time t. For the variable *qspread*, each quote i over time t is weighted. We also report the results for an equal-weighted quoted spread using the variable *qspreadeqw*. We use the logarithm of the variables for the empirical analysis. Second, we compute the realized or effective spread on actual trades as the effective spread better captures the actual spread for investors. Following Holden and Jacobsen (2014), the effective spread for a trade k is defined as:

[1] Evidence can also be extracted from earlier studies about information dissemination through the internet (e.g., Antweiler and Frank 2004) or analysing Google Search Volume (e.g., Andrei and Hasler 2015; Da et al. 2015).

$$effspread_k = \frac{2D_k(D_k - M_k)}{M_k} \tag{1}$$

where D_k is an indicator variable equal to *+1* if the k^{th} trade is a buy and *-1* if the trade is a sell, while M_k is the midpoint price of the quote at the time of the k^{th} trade. The effective spread is volume-weighted and analysed in its log transformation.

In addition, we construct several variables to indicate which traders react to certain news. In analogy to Antweiler and Frank (2004), we categorize trades with thresholds of 100,000 EUR and 1 million EUR as small, medium and large trades (variables *slt100*, *slt1mio*, *slt1miop*), respectively, to distinguish small from large traders in an attempt to differentiate between retail and institutional investors. In addition, we use 50,000 EUR and 500,000 EUR as an additional set of smaller thresholds (variables *slt50*, *sl500*, *slt500p*).

We further account for potential differences in the typical trade size of each stock within our sample by constructing a historical trade size distribution for each stock. The trade size distribution allows us to determine what constitutes a "large" and "small" trade for each stock. The variable $xt3share_{i,t}$ measures the share of trades relative to all trades on day t for stock i that are smaller than the tercile of the stock's historical trade size distribution. Variable $xt3lnratio_{i,t}$ is the log of the ratio of the number of trades in the lowest tercile relative to the number of trades in the top tercile for stock i on day t. The variables $xt5share_{i,t}$ and $xt5lnratio_{i,t}$ are computed in a similar manner for quintiles instead of terciles.

The Probability of Informed Trading

PIN is a composite variable based on a microstructure trading model (e.g., Easley et al. 1996, 1997; Easley and O'Hara 1987). We use a modification of the model's estimation procedure to allow for a daily estimation of the model's parameters instead of one estimation per 30 trading days (see Pöppe et al. 2016). The model's assumption is as follows: There are two types of traders, informed and uninformed, who arrive sequentially to trade a risky asset with a competitive and risk-neutral market maker. The asset's value is determined by information events, which happen with probability α, and contain bad news with probability δ (good news with probability 1-δ, respectively). Only the informed traders can observe the existence and direction of a signal and consequently only trade if there is a signal. The uninformed traders trade independently of the arrival of a signal purely for liquidity reasons. The probabilities μ and ϵ describe the ability of informed and uninformed investors, respectively, to actually trade, once they decide to trade. The market maker knows the structure of the trading process and must update his believes of the realizations of underlying parameters after every trade or, in the absence of a willing trade partner, to adjust his quotes. The key inputs into the market maker's thought process are buys, sells, and no-trades, and the outcomes are quoted spreads and trades, which in turn allow for a derivation of the underlying trade process through a maximum likelihood estimation. The described parameters can also be estimated from tick data with the only input being the number of buys and sells and the number of times no trade happened for a certain amount of time during a given day (Easley et al. 1996). The estimation of the four parameters allows to calculate the probability of informed trading:

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon(1 - \alpha\mu)} \tag{2}$$

Sample Composition and Descriptive Statistics

Company Selection and Twitter Data

We examine the stocks in the major German indices DAX, MDAX, and TecDAX. This sample combines the largest 100 German companies. An advantage of the German market is the low level of fragmentation compared to the U.S. market, where the share of stocks traded in their home venues can be as low as 25% for NYSE and 30% for NASDAQ (Holden and Jacobsen 2014). We collected a usable and complete sample starting on October 20, 2013 and ending on April 30, 2014.[2] This time period follows the German federal

[2] The public application programming interface (API) of Twitter allows to retrieve at most 1% of the total message stream on Twitter. If a query is specific enough that it contains less than 1% of the total message

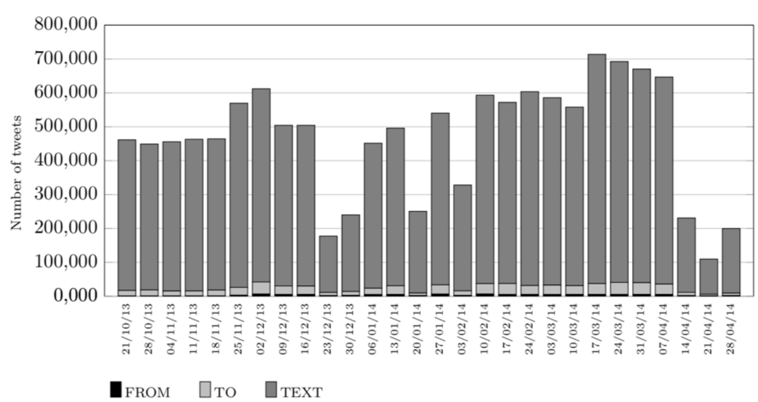
election at the end of September 2013, thereby avoiding potential distortion of our data set by political campaigning. In addition, the time period also covers firms' year end results and first quarter results, which should create increased Twitter activity and also allow for a differentiated analysis of Twitter activity around earnings announcements (and potential information leakage prior to these announcements).

We distinguish three types of tweets: (i) tweets containing the company's name or associated stock tickers in the tweet message (TEXT-tweets), (ii) tweets sent from the firm's Twitter account to the public if the company has a Twitter account (FROM-tweets), and (iii) tweets sent from Twitter users to the company account (TO-tweets). In our sample 72% of the potential firms had their own Twitter account. For most of the analysis, we will focus on the largest group, the TEXT-tweets. We do not collect the number of followers of a given Twitter account, but we include re-tweets in our sample[3]. We therefore have an implicit weighting by re-tweets in our sample as re-tweets are not filtered out but instead captured as an additional new tweet with a new timestamp.

The correct identification and assignment of tweets to a specific firm in our sample is crucial. The majority of the literature on Twitter and stocks avoids this issue by analysing overall Twitter activity. We exclude a number of tweets from the initial Twitter text search where the synonym usage is obvious and difficult to circumvent. The raw data collected amounts to more than 50 million tweets. After the necessary pre-processing steps and joining the Twitter data with the trading data of the sample firms, 12 million tweets are left for our analyses.

Descriptive Statistics

Figure 1 presents the weekly activity on Twitter during the observation period. The holiday breaks around Christmas and Easter are clearly visible. In addition, a significant difference between TEXT-tweets, TO-tweets and FROM-tweets can be observed.



This figure illustrates the tweet count per week during the observation period. TEXT-tweets are tweets that contain the company's name or associated stock tickers, FROM-tweets are tweets sent from the company's Twitter account to the public, and TO-tweets are tweets send from users to the company's Twitter account.

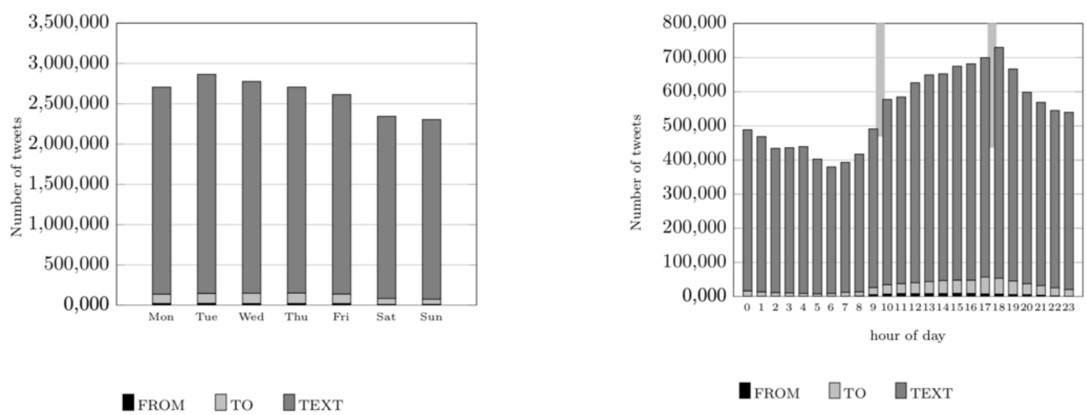
Figure 1. Weekly Twitter activity over the observation period

volume, the full result set is returned. Our study therefore uses all relevant data for the sample and not a random subsample.

[3] The literature has conflicting views on the importance of followership. Some scholars produce "better" results with follower-weighed tweets (e.g., Nofer and Hinz 2014) and argue that more followership implies tweets of higher relevance or even quality. On the other hand, a number of articles conclude that re-tweets, which we cover, are a valid indicator of the quality of information (Kwak et al. 2010; Sprenger et al. 2014), whereas the number of followers is not informative (Kwak et al. 2010; Zhang et al. 2011).

Figure 2 Panel A shows the distribution of Twitter tweets over the week. The distribution of tweets is relatively uniform during the weekdays Monday through Friday. On weekends, however, Twitter activity slightly decreases. In line with Bollen et al. (2011) and Das and Chen (2007), we exclude the Tweet activity on the weekend for our analysis, but we shift the Twitter volume occurring after the close of trading to the next trading day. Figure 2 Panel B plots the number of tweets per hour during a trading day. Tweet activity is increasing during the trading hours from 9:00 a.m. to 5:30 p.m. and is lower during the German night-time hours between midnight and 7 a.m. Almost half of the Twitter activity occurs during trading hours.

Panel A: Tweets by weekday *Panel B: Tweets by hour*



This figure illustrates the distribution of Twitter tweets by the average number of tweets per day (Panel A) and by the average number of tweets per hour during the investigation period (Panel B). TEXT-tweets are tweets that contain the company's name or associated stock tickers, FROM-tweets are tweets sent from the company's Twitter account to the public, and TO-tweets are tweets send from users to the company's Twitter account.

Figure 2. Daily and hourly distribution of tweets

Standardization of Twitter Tweet Volumes

We standardize Twitter tweet volumes and Twitter activity for a given firm to avoid a size effect[4]. We normalize tweet activity by subtracting the median number of tweets for a company on a particular weekday during the observation period from the number of tweets for that company on that weekday (Da et al. 2011)[5]:

$$atwwk_{i,t} = ln\left(\frac{1 + tw_{i,t}}{1 + median_{k\in\{k,k+7,\dots,n-mod(n,k)\}}(tw_{i,k})}\right) \tag{3}$$

This definition is advantageous as it controls for weekday effects.

Trading Data

We obtain intraday trading data from Thomson Reuters Tick History[6]. The data contains all updates to the best bid and best ask as well as all trades for the covered stocks, both with their respective price and volume

[4] We also use a normal standardization. For company i on day t with tweet count $tw_{i,t}$, sample mean $\overline{tw_{i,t}}$ and sample standard deviation $\widehat{\sigma_i}$ the variable $twbyz_{i,t}$, which is calculated as $twbyz_{i,t} = \frac{tw_{i,t}-\overline{tw_i}}{\widehat{\sigma_i}}$. The results are qualitatively similar, but more robust using the *atwwk* variable (see also Figure 2 Panel A). For reasons of brevity, we forego to show the results, but they are available upon request.
[5] Given our relatively short sample period, we do not use a rolling standardization.
[6] We thank the Capital Markets Corporative Research Centre for their support in the provision of access to trading data.

on offer or cleared. Table 1 provides the descriptive statistics of the trading activity. While the differences in trading activity are large even in a sample of stocks composed from the leading indices, the stocks in the lowest decile trade sufficiently frequent to warrant a daily or intraday analysis.

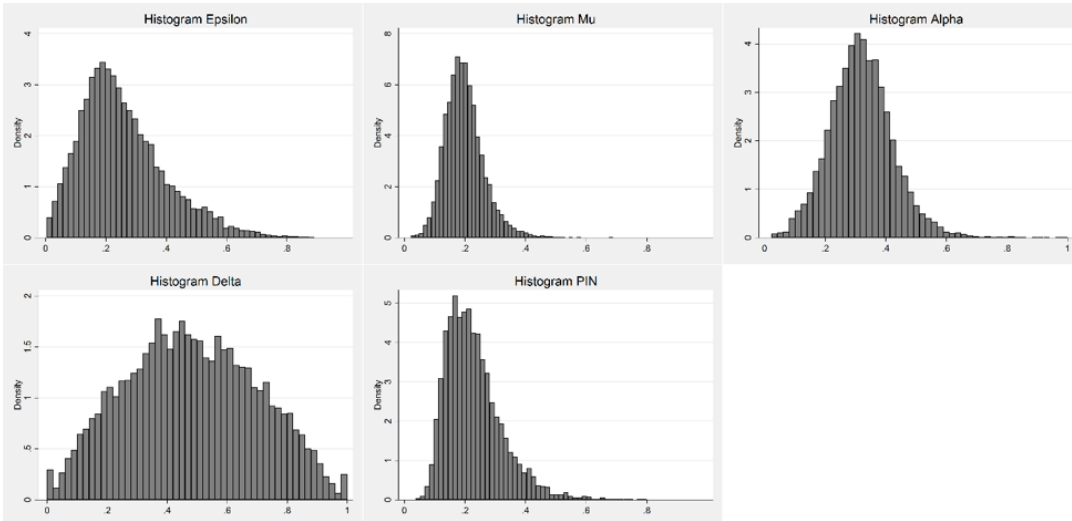
PIN Estimation from Trading Data

We estimate one PIN per day and stock for different combinations of no-trade intervals (5, 10, 20 and 30 seconds) and bucket lengths (8, 12 and 15 minutes). Figure 3 depicts the histograms for the four parameters of the model from which PIN is composed, ε, μ, α, δ, and of PIN itself. All parameters approximately resemble a normal to log-normal distribution without visible clustering around extreme solutions. PIN is skewed to the right and we can confirm that PIN and α increase from high to low volume stocks. The average of δ is close to 50%. The arrival probability for the uninformed traders, ε, increases in line with the length of the no-trade interval. The arrival probability for the informed trades, μ, is also increasing but does so less strongly than ε, resulting in a decreasing PIN from high to low volume stocks.

	Totals			Daily traded value per stock (mio. EUR)		Daily trade count per stock		Mean time between trades (sec.)		Default PIN est. param.	
	Trade count (mio.)	Value (mio. EUR)	Volume (mio.)	Mean	Max	Mean	Max	based on mean	based on max	no-trad e	bucket length
Panel A: by Index											
DAX	18.99	464,801	12,526	107	924	4,365	28,299	7	1	-	-
MDAX	6.65	54,129	1,902	9	238	1,147	17,598	27	2	-	-
TecDAX	3.20	19,928	1,326	5	75	736	11,293	42	3	-	-
Pabel B: by Decile (trade volume)											
1	9.21	271,431	5,182	187	924	6,348	28,299	5	1	5	8
2	5.78	130,785	5,875	90	452	3,985	26,728	8	1	5	8
3	4.16	63,092	1,565	44	313	2,866	16,745	11	2	5	8
4	2.59	26,005	1,138	18	230	1,788	17,598	17	2	10	8
5	1.90	15,839	491	11	238	1,311	7,244	23	4	10	8
6	1.81	12,252	382	8	46	1,245	6,035	25	5	10	8
7	1.47	9,337	637	6	34	1,012	6,562	30	5	20	12
8	1.01	5,449	261	4	31	697	3,310	44	9	20	12
9	0.64	3,642	146	3	35	441	4,111	69	7	30	15
10	0.29	1,026	77	1	9	197	2,417	155	13	30	15
Total	28.84	538,859	15,754	37	924	1,989	28,299	15	1	-	-

This table shows the aggregated trading activity of all stocks in the sample, aggregated by index membership in Panel A and deciles by trade volume in Panel B. Columns two to four show the total of the number of trades, the traded value and traded volume aggregated over the observation period. The first column of the sections daily traded value and daily trade count is the mean of the daily trade value or trade count. "Max" is the maximum of all daily values. The following two columns approximate the mean time in seconds between two trades by dividing the mean and maximum number of trades per day by the number of seconds per trading day (30,600 seconds). The last two columns show the default assignment of the two parameters required for estimating PIN, the length of the no-trade interval in seconds and the length of an intraday bucket in minutes.

Table 1. Trading activity on XETRA and default choice of PIN estimation parameters



This figure shows the histograms of the four parameters of the PIN model and PIN itself. Per stock and trading day one estimation was run.

Figure 3. PIN histograms

Results

Univariate Analysis

Table 2 displays the results for the univariate correlations of the full sample for the variables *atwwk*. Table 3 aggregates the results for univariate correlations by a single company by averaging the correlation coefficients and counting for each variable pair how many are statistically different from zero. Table 4 provides the results of the non-parametric Mann-Whitney U-test on differences in means and the Moods-Median-test on differences in medians between days with Twitter activity in the bottom tercile and days with Twitter activity in the top tercile. Table 5 repeats the analysis of Table 4, but on a single company basis with aggregated data.

		Same day *atwwk*	Lag 1 day *atwwk*	Future 1 day *atwwk*
Volume	*lnVolume*	0.007	-0.016	-0.011
Volatility	*hlVola*	0.068**	0.015	0.049**
	15minVola	0.073**	0.10	0.048**
	15minVolaMax	0.079**	0.018	0.053**
Spread	*effspread*	0.078**	0.054**	0.074**
	qspread	0.025	0.011	0.030*
	qspreadeqw	0.049**	0.034**	0.050**
Trades	*xt3share*	-0.088**	-0.065**	-0.054**
	xt5share	-0.075**	-0.052**	-0.045**
	xt3lnratio	-0.095**	-0.074**	-0.063**
	xt5lnratio	-0.098**	-0.078**	-0.064**
	slt50	0.008	-0.020	-0.008
	slt500	0.016	-0.10	-0.005
	slt500p	0.017	-0.016	-0.008
	slt100	0.008	-0.020	-0.008
	slt1mio	0.014	-0.015	-0.006
	slt1miop	0.009	-0.013	-0.019
Return	*retln*	0.036**	-0.006	0.031
	retNeg	-0.030*	-0.014	-0.020
	retPos	0.089**	0.004	0.070**
PIN	*PIN*	-0.010	0.004	0.005
	μ	0.067**	0.041**	0.054**

ε	0.030*	0.007	0.005
δ	-0.038**	-0.019	-0.029
α	-0.003	-0.001	-0.007

This table shows pairwise correlations coefficients between Twitter activity and indicators of trading activity for the full sample using the metric *atwwk* as a proxy for Twitter activity. * and ** indicate statistically significant correlations at the 1% and 0.1% level, respectively.

Table 2. Pairwise correlations for daily Twitter activity and trading on XETRA

		Same day			Lag 1 day			Future 1 day		
		Avg ρ	*#sig. 1%*	*#sig. 0.1%*	*Avg ρ*	*#sig. 1%*	*#sig. 0.1%*	*Avg ρ*	*#sig. 1%*	*#sig. 0.1%*
Volume	*lnVolume*	0.169	54	35	0.083	38	16	0.115	35	19
Volatility	*hlVola*	0.089	36	15	0.019	22	6	0.064	26	10
	15minVola	0.102	37	18	0.013	20	9	0.062	26	10
	15minVolaMax	0.100	26	14	0.022	16	2	0.066	20	9
Spread	*effspread*	0.092	26	7	0.040	16	4	0.081	28	7
	qspread	-0.013	15	3	-0.029	13	3	-0.005	15	4
	qspreadeqw	0.025	32	10	-0.057	30	7	0.011	30	10
Small-	*xt3share*	-0.097	32	18	-0.075	25	11	-0.061	21	11
Large	*xt5share*	-0.079	25	16	-0.058	21	8	-0.048	20	8
Trades	*xt3lnratio*	-0.110	32	20	-0.090	31	14	-0.074	28	10
	xt5lnratio	-0.111	32	18	-0.092	30	14	-0.072	29	9
	slt50	0.141	50	33	0.056	34	16	0.097	32	15
	slt500	0.160	54	26	0.080	25	9	0.106	32	11
	slt500p	0.077	18	7	0.031	10	2	0.037	12	4
	slt100	0.142	49	34	0.057	34	15	0.098	32	15
	slt1mio	0.142	43	18	0.065	20	7	0.096	26	8
	slt1miop	0.051	12	3	0.026	8	1	0.017	8	2
Return	*retln*	0.044	20	4	-0.002	8	0	0.038	13	2
	retNeg	-0.022	23	9	-0.007	10	3	-0.011	14	1
	retPos	0.096	31	8	0.003	8	0	0.074	21	3
PIN	*PIN*	-0.055	19	4	-0.031	16	2	-0.029	15	5
	μ	0.074	26	8	0.043	14	2	0.059	23	5
	ε	0.106	43	26	0.049	32	8	0.064	33	9
	δ	-0.038	15	1	-0.014	12	4	-0.029	11	0
	α	-0.004	7	1	0.001	7	1	-0.005	9	1

This table shows the number of correlations calculated per company that are statistically significant at the 1% and 0.1% level. Correlations are calculated between standardized daily Twitter activity (*atwwk*) and indicators of trading activity on XETRA. The total number of companies and hence the maximum possible count is 83. "Avg ρ" displays the average of the 83 correlation coefficients.

Table 3. Pairwise correlations per single firm for daily Twitter activity and trading on XETRA (atwwk)

		Same day			Lag 1 day			Future 1 day		
		sig. of difference			sig. of difference			sig. of difference		
Variable		MW	MM	Δ Avg	MW	MM	Δ Avg	MW	MM	Δ Avg
Volume	*lnVolume*	0.000	0.000	63.02%	0.000	0.000	49.39%	0.000	0.000	53.24%
Volatility	*hlVola*	0.000	0.000	12.01%	0.045	0.091	2.90%	0.000	0.000	9.65%
	15minVola	0.000	0.000	9.73%	0.005	0.022	3.05%	0.000	0.000	7.79%
	15minVolaMax	0.000	0.000	11.44%	0.005	0.100	3.89%	0.000	0.000	9.86%
Spread	*effspread*	0.806	0.749	-0.06%	0.007	0.235	-3.93%	0.265	0.438	-1.23%
	qspread	0.000	0.000	-11.28%	0.000	0.000	-11.66%	0.000	0.000	-9.35%
	qspreadeqw	0.000	0.000	-10.46%	0.000	0.000	-12.27%	0.000	0.000	-10.17%
Trades	*xt3share*	0.000	0.000	-9.02%	0.000	0.000	-7.01%	0.000	0.000	-7.07%
	xt5share	0.000	0.000	-4.38%	0.000	0.000	-3.10%	0.000	0.000	-3.29%

	xt3lnratio	0.000	0.000	-10.79%	0.000	0.000	-8.56%	0.000	0.000	-8.23%
	xt5lnratio	0.000	0.000	-4.59%	0.000	0.000	-2.97%	0.000	0.000	-3.45%
	slt50	0.000	0.000	41.33%	0.000	0.000	31.43%	0.000	0.000	35.67%
	slt500	0.000	0.000	72.49%	0.000	0.000	52.95%	0.000	0.000	57.69%
	slt500p	0.000	0.000	7.93%	0.000	0.000	4.31%	0.000	0.000	5.05%
	slt100	0.000	0.000	41.77%	0.000	0.000	31.79%	0.000	0.000	35.99%
	slt1mio	0.000	0.000	52.19%	0.000	0.000	36.04%	0.000	0.000	41.50%
	slt1miop	0.000	0.000	3.50%	0.004	0.004	1.34%	0.114	0.122	0.90%
Return	*retln*	0.004	0.049	0.13%	0.092	0.115	0.13%	0.076	0.227	0.13%
	retNeg	0.965	0.000	-0.08%	0.028	0.000	-0.08%	0.754	0.000	-0.08%
	retPos	0.000	0.020	0.22%	0.222	0.162	0.22%	0.003	0.130	0.22%
PIN	*PIN*	0.000	0.000	-7.50%	0.000	0.000	-6.82%	0.000	0.000	-6.13%
	μ	0.000	0.000	5.42%	0.000	0.000	2.74%	0.000	0.000	4.52%
	ε	0.000	0.000	16.31%	0.000	0.000	12.77%	0.000	0.000	13.68%
	δ	0.001	0.007	-3.26%	0.158	0.443	-1.73%	0.022	0.009	-2.38%
	α	0.045	0.021	1.23%	0.061	0.066	1.32%	0.064	0.108	0.89%

This table compares indicators of trading activity on days in the top tercile of Twitter activity to days in the bottom tercile of Twitter activity. Terciles are calculated per single stock over the whole observation period based on the variable *atwwk*. In each section "MW" is the p-value of a Mann-Whitney mean comparison, "MM" the p-value of a Moods-Median comparison and "Δ Avg" is the difference of the averages of the bottom vs the top tercile.

Table 4. Differences in trading indicators - top vs bottom tercile of Twitter activity

		Same day		*Lag 1 day*		*Future 1 day*	
		atwwk		*atwwk*		*atwwk*	
Variables		10%	1%	10%	1%	10%	1%
Volume	*lnVolume*	49	30	30	9	34	13
Volatility	*hlVola*	25	11	17	7	21	6
	15minVola	25	10	18	8	20	8
	15minVolaMax	22	5	17	3	16	7
Spread	*effspread*	28	8	14	2	20	6
	qspread	13	3	11	2	13	3
	qspreadeqw	28	5	22	5	27	7
Trades	*xt3share*	31	17	24	9	26	13
	xt5share	32	15	21	9	24	11
	xt3lnratio	38	19	28	12	28	13
	xt5lnratio	34	19	28	11	32	12
	slt50	44	21	23	6	32	11
	slt500	48	23	28	8	32	12
	slt500p	13	2	9	2	11	1
	slt100	44	21	23	6	32	11
	slt1mio	33	19	17	1	22	7
	slt1miop	10	1	5	1	4	1
Return	*retln*	12	2	10	0	14	1
	retNeg	12	4	11	1	11	1
	retPos	12	3	8	0	13	1
PIN	*PIN*	19	6	18	4	16	4
	μ	18	4	11	3	17	6
	ε	34	13	23	6	27	6
	δ	16	2	11	3	10	1
	α	5	1	8	0	6	0

This table shows the number of stocks for which the difference in the mean between the top and bottom tercile trading days is significant at the 10% or 1% level, respectively. The tercile split is based on each stock's individual Twitter activity.

Table 5. Quantile comparison for daily Twitter activity and trading on XETRA (company level)

A positive and statistically significant correlation of Twitter activity and trading volume is present in all analyses. Only the full sample linear correlation for the *atwwk* is not significantly correlated with trading volume on the 0.1% level of significance. Volatility results are the strongest of all evaluated variables in the full sample analysis (see Table 2 and 4). Regardless of which of the three measures for volatility is considered, Twitter activity is always positively correlated with volatility.

The results for the spread are not as consistent. The full sample correlation with the effective spread is positive and significant. An elevated effective spread indicates a higher degree of information asymmetry or the presence of information being processed by market participants, which would suggest that Twitter helps, or is at least an indication, for information dissemination. Contrary to prior expectations, the split into terciles reveals no significance, whereas all other variables show the strongest results in this analysis. This discrepancy may be driven by a size effect, as spreads depend on liquidity.

Examining the results of the PIN variable and its parameters provides a differentiated view on the types of traders. The arrival rates of informed and uninformed traders, μ and ε, represent total trading volume in the underlying model. Therefore, it is not surprising that the correlation is positive and significant, as Twitter activity and trading volume are positively correlated. The quantile comparison, however, strongly indicates a much larger increase in ε, i.e. uninformed traders. The parameter α reveals one of the most noteworthy results so far. Twitter activity is not correlated with the probability of information arrival. This is surprising given that every other variable is at least weakly correlated to Twitter activity in the full sample. Therefore, it appears as if Twitter volume does not necessarily convey new information, but that it does attract uninformed traders, who push up volume and volatility.

Multivariate Analysis

To investigate the causal link between Twitter activity and trading activity on firm level, we use a panel regression. The panel dataset consists of the cross section of the 83 firms and the time-series of 131 trading days controlled for company fixed-effects.[7]

Table 6 presents the results for the panel regressions with trading activity as the dependent variables. Twitter activity is the key independent variable, but we include several control variables from the univariate analysis. The first model looks purely at the contemporaneous relationship whereas the second model includes five lags of both the Twitter variable and the dependent variable.

Group	Variable	*lnVolume*		*15minVola*		*effspread*		*retln*	
Twitter	*atwwk*	0.030 ***	0.031 ***	-0.001	0.008	0.022	0.016 **	0.001	0.001 **
	L.atwwk		-0.010		-0.017 **		-0.008 *		0.000
	L2.atwwk		0.007		-0.009		-0.007		-0.001
	L3.atwwk		0.006		0.003		-0.002		0.000
	L4.atwwk		0.000		-0.004		0.003		0.000
	L5.atwwk		0.005		-0.001		-0.004		0.000
Control	*lnVolume*			0.225 ***	0.239 ***	0.009	0.005	0.000	0.001
	15minVola	0.241 ***	0.244 ***			0.239 ***	0.205 ***	-0.002	-0.002 *
	effspread	0.007	0.013	0.166 ***	0.155 ***			-0.003	-0.003
	retln	0.028	-0.017	-0.479	-0.813 *	-1.421 *	-1.662 ***		
PIN	*PIN*	-1.257 ***	-1.100 ***	0.270 *	0.333 **	0.330 **	0.216 *	-0.001	-0.002
	μ	2.521 ***	2.304 ***	1.021 ***	0.689 ***	-0.344 *	-0.178 *	-0.011	-0.016 *
	ε	2.912 ***	2.777 ***	0.339 **	0.326 **	0.002	-0.107	0.008	0.010
	δ	1.488 ***	1.365 ***	-0.032	-0.075	-0.251 **	-0.252 ***	0.004	0.005
	α	-0.023	-0.022	0.002	-0.001	-0.068 **	-0.043 **	-0.025 ***	-0.025 ***
Lags dependent variable	L1 dep. var.		0.062 ***		0.155 ***		0.254 ***		-0.056
	L2 dep. var.		0.039 ***		0.080 ***		0.168 ***		-0.047
	L3 dep. var.		0.012		0.038 **		0.110 ***		-0.063 *
	L4 dep. var.		0.023 *		0.036 **		0.092 ***		-0.030
	L5 dep. var.		0.018		0.015		0.100 ***		0.019
	Constant	15.970 ***	13.700 ***	-8.900 ***	-7.130 ***	-5.670 ***	-0.680 *	-0.024	-0.032

[7] Running a random effects model yields very similar results in terms of direction and significance of coefficients.

Observations	10,467	10,065	10,467	10,065	10,467	10,004	10,467	10,065
Adj. R^2	0.969	0.970	0.622	0.646	0.753	0.821	0.099	0.114
Clustered std. errors	date	date	date	date	date	date	date	date
Firm FE	yes	yes	yes	yes	yes	yes	yes	yes

This table shows the results of a fixed-effects panel regression with *lnVolume*, *15minVola*, *effspread*, and *retln* as dependent variables. For each of the dependent variables the first regression is run on contemporaneous data without any lags. The second regression is run including 5 lags of the dependent variable and 5 lags of the Twitter variable. Fixed effects are per company and standard errors clustered by date. *, ** and *** indicate statistical significance at the 5%, 1% and 0.1% level, respectively.

Table 6. Regression on trade metrics

In the regression on trading volume, all control variables, including PIN and its parameters, react significantly in the expected direction. PIN declines with increasing volume, the coefficients of both arrival probabilities are significant and positive, the coefficient of the variable α is positive as the arrival of new information pulls investors to the market and δ is insignificant as positive or negative news trigger volume. Volatility also increases along with trading volume. These effects change only marginally when lagged values of Twitter and volume are included. Twitter's contemporaneous coefficient (*atwwk*) is significant and positive even when lagged volume is added to the regression. However, no predictive power remains: None of the lagged Twitter variables shows significant influence, whereas the autocorrelation of volume is clearly visible.

Volatility is associated with increased volume, higher spreads, higher PIN, and higher arrival rates of both informed and uninformed traders, all of which are known relations and support our modelling. We also find a significant negative correlation of volatility and Twitter activity lagged by one day, although the standardized coefficients (not shown for brevity) are very small compared to the control variables.

The log of each stock's daily raw return is at least partially explained by the dependent variable in our regression models. The coefficient of Twitter activity is again only significant for the same trading day. The effect is slightly positive, indicating a higher chance of positive news being discussed on Twitter. The probability of bad news, δ, has a negative and highly significant coefficient.

A similar regression using PIN and its components as dependent variables is provided in Table 7. Apart from a few significant coefficients at the 5% level of significance in higher lags, Twitter activity does not add any explanatory power for PIN or its components.

Overall, in the multivariate setting, the explanatory power of lagged Twitter activity completely disappears. In addition, the contemporaneous relationships are weaker than expected. The relation holds for volume and the effective spread. The rejection Twitter's relevance for information processing is even stronger when PIN variables are analysed.

		PIN		ε		μ		α	
Regular coefficients									
Twitter	*atwwk*	-0.001	-0.001	-0.001	0.000	-0.001	0.000	-0.001	-0.001
	L.atwwk		-0.001		-0.001		0.001		-0.001
	L2.atwwk		0.001		-0.002*		0.000		0.001
	L3.atwwk		0.001		-0.002*		0.000		0.000
	L4.atwwk		-0.002*		0.001		0.000		-0.001
	L5.atwwk		0.000		-0.002*		-0.001		0.000
Control	*lnVolume*	-0.037***	-0.032***	0.119***	0.106***	0.052***	0.051***	0.017***	0.015***
	15minVola	-0.006	-0.006*	0.037***	0.029***	0.016***	0.017***	-0.005	-0.007
	effspread	0.006	0.004	-0.012***	-0.008**	0.004*	0.004*	-0.010*	-0.009*
	retln	0.094*	0.093*	-0.089	-0.131*	-0.008	0.008	0.102	0.092
Lags dep. var.	*L1 dep. Var.*		0.165***		0.222***		0.102***		0.113***
	L2 dep. var.		0.030*		0.025*		-0.004		-0.008

	L3 dep. var.		0.017		0.019		0.007		0.023
	L4 dep. var.		-0.003		0.010		0.004		-0.016
	L5 dep. var.		0.019		0.024 **		0.026 *		0.001
Standardized coefficients									
Twitter	*atwwk*	-0.010	-0.008	-0.010	0.002	-0.011	-0.006	-0.014	-0.014
	L.atwwk		-0.011		-0.008		0.010		-0.009
	L2.atwwk		0.012		-0.015 *		-0.002		0.011
	L3.atwwk		0.012		-0.017 *		0.000		0.003
	L4.atwwk		-0.020 *		0.006		0.000		-0.012
	L5.atwwk		0.004		-0.013 *		-0.010		-0.004
Control	*lnVolume*	-0.659 ***	-0.567 ***	1.403 ***	1.249 ***	1.359 ***	1.224 ***	0.266 ***	0.232 ***
	15minVola	-0.028	-0.032 *	0.120 ***	0.102 ***	0.114 ***	0.124 ***	-0.024	-0.031
	effspread	0.045	0.034	-0.058 ***	-0.038 **	0.038 *	0.041 *	-0.066 *	-0.059 *
	retln	0.018 *	0.019 *	-0.011	-0.018 *	-0.002	0.002	0.017	0.016
Lags dep. var.	*L1 dep. var.*		0.165 ***		0.223 ***		0.102 ***		0.113 ***
	L2 dep. var.		0.030 *		0.025 *		-0.004		-0.008
	L3 dep. var.		0.017		0.020		0.007		0.023
	L4 dep. var.		-0.003		0.010		0.004		-0.016
	L5 dep. var.		0.018		0.024 **		0.027 *		0.001
	Constant	0.823 ***	0.686 ***	-1.494 ***	-1.384 ***	-0.511 ***	-0.497 ***	-0.047	-0.068
Observations		10,467	8,708	10,467	8,708	10,467	8,708	10,467	8,708
Adj. R^2		0.549	0.545	0.846	0.858	0.405	0.415	0.054	0.055
Clustered std. errors		date	date	date	date	date	date	date	date
Firm FE		yes	yes	yes	yes	yes	yes	yes	yes

This table shows the results of several fixed-effects panel regressions with PIN and its components ε, μ and α as dependent variables. Twitter activity is measured by the variable *atwwk*. For each of the dependent variables the first regression is run on contemporaneous data without any lags. The second regression includes five lags of the dependent variable and five lags of the Twitter variable. Fixed effects are per company and standard errors clustered by date. *, ** and *** indicate statistical significance at the 5%, 1% and 0.1% level, respectively.

Table 7. Regression on PIN

Further Empirical Checks

Interday Event Study

We run an interday event study to draw further conclusions on the timing of stock market-relevant information dissemination and the impact of Twitter. The event study covers two major types of announcements: ad-hoc and quarterly earnings announcements. During the sample period, we document 208 events, 110 ad-hoc announcements and 98 earnings announcements. The requirement of an event window from day t-5 up to day t+5 being covered by our data and non-overlapping windows reduces the joined sample of ad-hoc and earnings announcements to 140 events. The analysis of PIN further reduces the sample to 98 events, as the maximum likelihood estimation does not converge on every trading day.

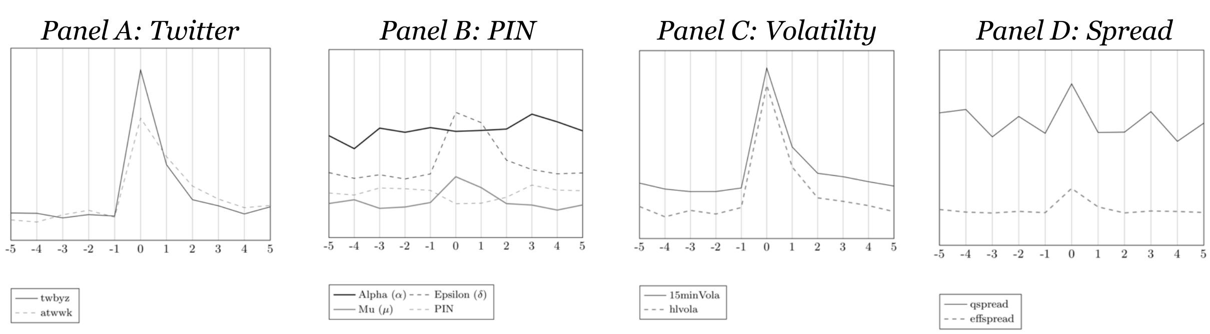
The results are illustrated in Figure 4. Table 8 presents the corresponding *t*-test statistics and Kruskal-Wallis test statistics. The reaction on the announcement day is clearly visible for Twitter activity, volatility, spread, and arrival probabilities. PIN declines due to the release of previously private information. Activity on Twitter is strongest on the announcement day indicating a strong link of Twitter to the processing of information that is not just chatter but also relevant for stock markets. However, no variable, including Twitter activity, shows a reaction on the day prior to the announcement. Moreover, the reversal to a normal level of Twitter activity starts with two days of statistically significant downward jumps and another two days of a slower decline until Twitter activity returns to its pre-event average. In contrast, volume and spread take just one day for their reversal. Volatility also takes longer to again decrease to pre-event levels.

The results provide evidence that information processing on Twitter is related to trading and offers information, which may move prices. Yet, Twitter appears to be a post-processing platform for the dissemination and interpretation of news and not a platform for news or information generation.

		t-test on successive event days								Kruskal-Wallis (-5, 5)		Kruskal-Wallis (-1, 1)	
Variable		[-4; -3]	[-3;-2]	[-2; -1]	[-1; 0]	[0; 1]	[1; 2]	[2; 3]	[3; 4]	χ^2	P	χ^2	p
Trading and Twitter data for every day in the event window: n = 140													
Twitter	*atwwk*	-0.83	-0.57	0.93	-10.56 ***	3.92 ***	3.46 ***	1.47	0.92	210.3	0.00	112.5	0.00
Volume	*lnVolume*	-0.15	0.36	-0.40	-4.02 ***	1.80 *	1.31	0.42	-0.14	37.5	0.00	18.5	0.00
Volatility	*hlVola*	-0.71	0.42	-0.75	-8.95 ***	5.13 ***	2.76 ***	0.35	0.45	139.5	0.00	72.0	0.00
	15minVola	0.28	-0.01	-0.41	-8.77 ***	5.09 ***	2.30 **	0.33	0.52	123.6	0.00	70.6	0.00
	15minVola Max	0.39	0.65	-0.19	-7.72 ***	4.39 ***	1.98 **	0.21	0.38	94.4	0.00	53.9	0.00
Spread	*effspread*	1.56	-1.11	0.92	-2.61 ***	2.54 **	-0.02	-1.19	1.78 *	16.6	0.08	9.8	0.01
	qspread	0.17	-0.18	0.15	-1.96 *	1.94 *	0.06	-0.04	0.05	6.9	0.74	6.0	0.05
PIN estimation for every day in the event window: n = 98													
PIN	*PIN*	-1.10	0.12	0.28	2.46 **	-0.11	-1.01	-1.75 *	0.73	17.1	.072	12.3	0.00
	μ	1.51	-0.26	-0.72	-3.73 ***	1.75 *	3.01 ***	0.25	0.89	55.1	0.00	29.7	0.00
	ε	-0.31	0.36	-0.48	-4.91 ***	0.72	3.06 ***	0.90	0.42	64.4	0.00	30.3	0.00
	δ	-1.90 *	-0.57	1.42	1.45	0.30	-1.58	-0.11	-0.51	19.9	0.03	1.90	0.38
	α	-2.32 **	0.49	-0.54	.42	-0.11	-0.17	-1.58	0.85	16.8	0.08	1.20	0.56

This table provides the results of an event study around ad-hoc and earnings announcements. Columns three to ten display *t*-statistics to test difference in means on successive days in an event window from t – 4 to t + 4. The final four columns show the χ^2 statistic and the resulting p-value from a Kruskal-Wallis test on difference across all days in an event window, once for a t–5 to t+5 window and the t–1 to t+1 event window. *, ** and *** indicate statistical significance at the 5%, 1% and 0.1% level, respectively.

Table 8. Event study with ad-hoc and earnings announcements



This figure depicts the evolution of relevant metrics in the event study around ad-hoc and earnings announcements. Panel A shows the two metrics to measure Twitter activity, atwwk and twbyz, Panel B shows PIN and its components ε, μ and α, Panel C shows two measures for volatility, 15minVola and hlvola, and Panel D shows two measures for the spread, qspread and effspread.

Figure 4. Event study results

Intraday Event Study

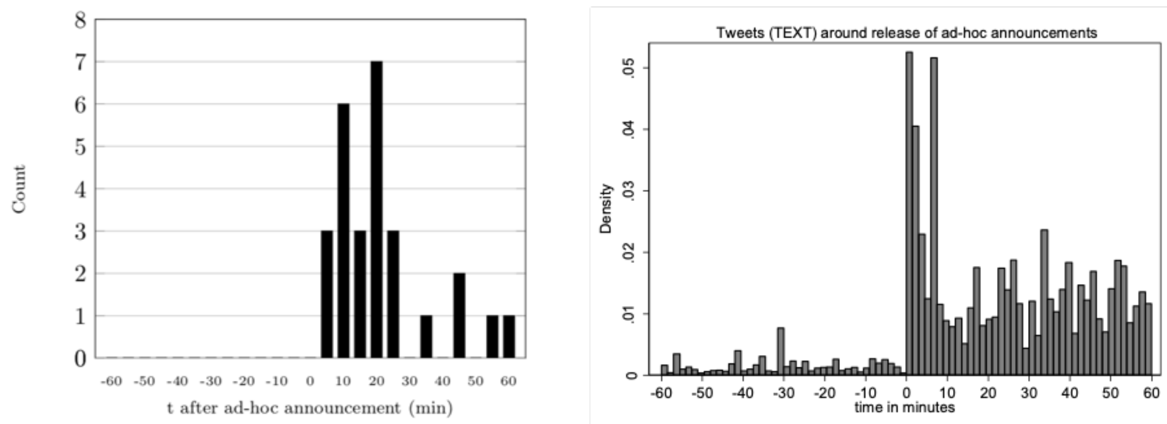
For a deeper analysis of the intraday Twitter activity around ad-hoc announcements, we analyse FROM-tweets and TEXT-tweets and run a descriptive intraday event study around the ad-hoc announcements.

Panel A of Figure 5 illustrates the Twitter tweets from the company around ad-hoc releases. We find that no tweet is released prior to an announcement. This is expected, as firms are legally obliged to first use the ad-hoc channel. However, all tweets replicating or referring to the ad-hoc announcement are released with a considerable delay of several minutes. Given the known relevance and price impact of ad-hoc announcements (e.g., Muntermann and Güttler 2007), relying on Twitter alone, notwithstanding the share of companies who do not use Twitter as a secondary release channel for ad-hoc announcements, is not a dominate strategy.

Finally, we use the TEXT-tweets to analyse the overall tweet volume around ad-hoc announcements. Panel B of Figure 5 presents the results. Twitter activity jumps significantly and immediately during the first minute after the release of an ad-hoc announcement. The activity in the hour prior to the announcement shows no pre-event run-up. Twitter activity stays elevated during the one-hour observation period. This provides further evidence that Twitter acts as post-processing and information dissemination platform and that it does not predict or forecast news events.

Panel A: FROM-tweets around ad-hoc announcements

Panel B: Panel B: Overall tweet activity around ad-hoc announcements



This figure illustrates the intraday Twitter activity surrounding the sample of ad-hoc news announcements. Panel A shows a histogram that depicts the time it takes until the first tweet covering the content of an ad-hoc announcement is released from the company Twitter account for a sample of 27 announcements. Panel B shows a histogram that depicts the number of TEXT-tweets published within one hour before and after the release of an ad-hoc announcement by the respective company for a sample of 99 announcements.

Figure 5. ***Intraday Twitter activity around ad-hoc announcements***

Discussion

Findings

This study relates a comprehensive sample of Twitter tweets to a large and diverse sample of companies in a major equity market to test the influence of Twitter on financial markets and to estimate the composition of informed and uninformed market participants. We evaluate the relevance of Twitter as a tool for information dissemination in financial markets on the single stock level. Using a unique dataset including more than 12 million Twitter feeds linked to specific firms and calculating advanced trading metrics, such as effective spreads, intraday volatility, and a daily version of PIN, we test the impact of Twitter activity on trading and information dissemination.

Our approach and our results differ from previous studies on Twitter but relate to and echo the literature on stock message boards and general short-term information processing. We show that Twitter can serve as an excellent indicator of news, which are also relevant for the stock market. In line with previous research (e.g., Deng et al. 2018; Li et al. 2018), we document that trading volume, volatility, and spread rise contemporaneously with Twitter activity, all of which indicate that relevant information is processed or at least disseminated with the help of Twitter. The PIN model indicates that more uninformed than informed traders rush to the market along with rising Twitter activity. This provides further evidence to the notion of Li et al. (2018) that increased Twitter activity may induce less informed investors to move from the side-line and participate in trading once they see the signals from other investors.

Additionally, using a multivariate analysis with lagged variables, the results show that all initial indications of a forward-looking predictive power of Twitter disappears. An interday event study around ad-hoc and earnings announcements again confirms that Twitter activity is related to trading and to information moving stock prices, but also reveals that Twitter is a post-processing platform for the dissemination and interpretation of news. Therefore, information procession on Twitter may be close to real-time, but it does not lead traditional news channels. The long reversal after a news spike further supports the notion that Twitter primarily supports information dissemination rather than serving as a primary means for information generation.

Theoretical Implications

Our study provides evidence on the ability of Twitter to serve as an information dissemination channel for single stocks. However, we cannot find robust evidence that Twitter activity is actually the main platform for information creation. Considering the increasing importance of social media for information dissemination on a firm level in the presence of an ever-increasing connected community, this may change in the future and does merit further research, as there is a lack of comprehensive research so far.

A number of questions remain for further investigation: First, we did not analyze the smaller subset of FROM- and TO-tweets in more detail, which can potentially aid in the understanding how companies and investors use Twitter as an additional communication channel between them. Second, repeating the study with more recent data may offer valuable insights given the current environment on Twitter. On the one hand, Twitter is now an established information channel and may now contribute to news generation, while on the other hand the purportedly large number of bots may contribute to an increase in noise. Third, it would also be possible to split the data based on industry sectors and to thereby distinguish between companies like Adidas, who are consumer-focused, from other purely business-to-business oriented companies, such as SAP. Finally, the interaction between different information channels, such as Twitter and Facebook, could provide a deeper understanding how price-sensitive information travels between various social media platforms and their differential effect on financial markets.

Managerial Implications

The questions remain to whom and how it may be useful to observe Twitter activity. In principle, one could differentiate between the management of a firm and current or potential investors. For firms who are not yet active on Twitter, the results of the present study may provide yet another argument to start utilizing Twitter as an additional channel. Given the near instant information dissemination, managers may also find that Twitter can potentially be a useful tool to direct and shape the information flow on their company and to effectively communicate with customers, suppliers and investors alike.

From an investor perspective, our results document that Twitter can be used for information gathering purposes. The information contained in single tweets as well as overall Twitter activity can, when aggregated, support investors' decision investment decision processes. Advanced algorithms or prediction models are not necessarily required in this context, but Twitter activity may simply guide investor interest to those stocks where new information requires further attention.

From another viewpoint, given the absence of investor-aimed filtering or sentiment-processing, our results are also unexpected. We did not retrieve data from a platform aimed at investors, but rather one of the most generic and general platforms available. We did not just filter stock tickers but considered every tweet that mentioned the name of a firm. Still, we observed a strong link between financial markets and firm-linked activity on Twitter. Yet, all things considered the more traditional and perhaps less broadly popular stock discussion forums might be a preferable place to data mine for relevant information as discussions are more continuous and focused than the (infrequent) outbursts on Twitter.

References

Andrei, D., and Hasler, M. 2015. "Investor attention and stock market volatility," *Review of Financial Studies* (28:1), pp. 33–72.

Antweiler, W., and Frank, M.Z. 2004. "Is all that talk just noise? The information content of internet stock message boards," *Journal of Finance* (59:3), pp. 1259–1294.

Bollen, J., Mao, H., and Zeng, X. 2011. "Twitter mood predicts the stock market," *Journal of Computational Science* (2:1), pp. 1–8.

Culnan, M.J., McHugh, P.J., and Zubillaga, J.I. 2010. "How large U.S. companies can use Twitter and other social media to gain business value," *MIS Quarterly Executive* (9:4), pp. 243–259.

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

Da, Z., Engelberg, J., and Gao, P. 2015."The sum of all FEARS investor sentiment and asset prices," *Review of Financial Studies* (28:1), pp. 1–32.

Das, S.R., and Chen, M.Y. 2007. "Yahoo! for Amazon: Sentiment extraction from small talk on the web," *Management Science* (53:9), pp. 1375–1388.

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

Easley, D., and O'Hara, M. 1987. "Price, trade size, and information in securities markets," *Journal of Financial Economics* (19:1), pp. 69–90.

Easley, D., Kiefer, N.M., and O'Hara, M. 1997. "One day in the life of a very common stock," *Review of Financial Studies* (10:3), pp. 805–835.

Easley, D., Kiefer, N.M., O'Hara, M., and Paperman, J.B. 1996. "Liquidity, information, and infrequently traded stocks," *Journal of Finance* (51:4), pp. 1405–1436.

Han, B., and Yang, L. 2013. "Social networks, information acquisition, and asset prices," *Management Science* (59:6), pp. 1444–1457.

Holden, C.W., and Jacobsen, S. 2014. "Liquidity measurement problems in fast, competitive markets: expensive and cheap solutions," *Journal of Finance* (69:4), pp. 1747–1785.

Hong, H., Kubik, J.D., and Stein, J.C. 2004. "Social interaction and stock-market participation," *Journal of Finance* (59:1), pp. 137–163.

Hong, H., Kubik, J.D., and Stein, J.C. 2005. "Thy neighbor's portfolio: Word-of-mouth effects in the holdings and trades of money managers," *Journal of Finance* (60:6), pp. 2801–2824.

Kwak, H., Lee, C., Park, H., and Moon, S. 2010. "What is Twitter, a social network or a news media?" *Proceedings of the 19th International Conference on World Wide Web,* pp. 591-600.

Leitch, D., and Sherif, M. 2017. "Twitter mood, CEO succession announcements and stock returns," *Journal of Computational Science* (21), pp. 1–10.

Li, T, van Dalen, J, and van Rees, P.J., 2018. "More than just noise? Examining the information content of stock microblogs on financial markets," *Journal of Information Technology* 33, pp. 50-69.

Muntermann, J., and Güttler, A. 2007. "Intraday stock price effects of ad hoc disclosures: The German case," *Journal of International Financial Markets, Institutions and Money* (17:1), pp. 1–24.

Nisar, T.M., and Yeung M., 2018. "Twitter as a tool for forecasting stock market movements: A short-window event study", *Journal of Finance and Data Science* 4, pp. 101-119.

Nofer, M., and Hinz, O. 2014. "Using Twitter to Predict the Stock Market: Where is the Mood Effect?," *Business & Information Systems Engineering*, (57:4) pp. 229-242.

Pöppe, T., Aitken, M., Schiereck, D., and Wiegand, I. 2016. „A PIN per day shows what news convey: the intraday probability of informed trading," *Review of Quantitative Finance and Accounting* (47:4), pp. 1187–1220.

Shive, S. 2010. "An epidemic model of investor behaviour," *Journal of Financial and Quantitative Analysis* (45:1), pp. 169–198.

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

Sul, H.K., Dennis, A.R., and Yuan, L. (2017). "Trading on Twitter: Using social media sentiment to predict stock returns," *Decision Sciences* (48:3), pp. 454–488.

Zhang, X., Fuehres, H., and 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 - The 2nd Collaborative Innovation Networks Conference - COINs2010* (26), pp. 55–62.