Finance Research Letters 30 (2019) 116–123

Contents lists available at ScienceDirect

Finance Research Letters

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

Social-media and intraday stock returns: The pricing power of sentiment

David C. Broadstock[a,*], Dayong Zhang[b]

[a] Center for Economic Sustainability and Entrepreneurial Finance, Hong Kong Polytechnic University, Hong Kong
[b] Research Institute of Economics and Management, Southwestern University of Finance and Economics, China

ARTICLE INFO

Keywords:
Textual analysis
Social-media
Sentiment
Asset pricing
Intraday

ABSTRACT

This paper tests whether sentiment extracted from social-media (Twitter), has pricing power towards stock market. Specifically, we evaluate whether firms' intraday stock returns react to sentiment on the firm itself, and/or sentiment on the wider financial market. Using intraday high frequency stock returns for a sample of US companies, we show that price dynamics are susceptible to social-media sentiment pricing factors, with varying balances of importance for firm specific and market wide sentiment.

1. Introduction

> *"A variety of surveys indicate that between 34% and 70% of investors are using social-media content at least to some extent in their investment decision-making... "* (de Jong et al. (2017), pp. 63)

Analysts of financial markets are no stranger to data, however social-media and news data have very different foundations to stock market data. The information presented by financial markets is highly structured, with attributes that permit clear sequencing and consistent recording of data. In contrast, news stories and social-media posts are ostensibly presented in the form of text, and beyond time stamps, it is not immediately obvious what structured information can be 'teased out' and analyzed. Nonetheless, financial practitioners already trade on information contained in social-media, and interest among scholars of finance is also emerging (Bukovina, 2016).

There is an obvious tension in the question as to whether or not social-media platforms contain a rich enough base of information to be able to *systematically* inform investor activity. Sprenger et al. (2014) question whether the information conveyed in social-media reflects 'news' or 'noise'. Several studies, including Chen et al. (2014), and Zhang et al. (2017) among others, argue that steps can be taken to gauge the 'financial literacy' of a message sender, using signals contained in the messages themselves. While this may aid identification of messages that are likely to provide more reliable information, it does not guarantee it, nor does it reflect the ability of potential users to either observe the signal, or act upon the content of the message. It therefore remains a timely, and open empirical question as to whether and how sentiment from social-media messaging platforms stimulates (or reinforces) trade activity.

Offering support in favor of a positive role for social-media, Mazboudi and Khalil (2017) remark that *"Overall, our results suggest that Twitter has become an important investor relation channel for major corporate events such as acquisition announcements and that large acquirers can use this new channel to enhance stability in their stock prices."* Bollen et al. (2011) illustrate that the 'collective mood' state extracted from twitter feeds can help improve predictions of the Dow-Jones Industrial Average index. However tension exists here too, on the basis of the SEC charging Elon Musk with securities fraud as a direct consequence of *"Musks tweets [which] caused Tesla's*

* Corresponding author.
E-mail address: david.broadstock@polyu.edu.hk (D.C. Broadstock).

https://doi.org/10.1016/j.frl.2019.03.030
Received 8 November 2018; Received in revised form 30 January 2019; Accepted 20 March 2019
Available online 21 March 2019
1544-6123/

stock price to jump by over six percent on August 7, and led to significant market disruption”.[1]

The action by the SEC against Tesla and Elon Musk serves as tangible evidence that Twitter has transitioned from a ‘mere’ popular social messaging platform, to a legitimate source of information which requires the attention of financial market regulator. While Twitter has been around since 2006, arguably the past 2 years have seen it’s key emergence as a tool of communication among global leaders. The emergence of TicToc (by Bloomberg), the first 24-hour global news network streaming live on Twitter, adds further legitimacy to Twitter as a platform for reliable news of immediate relevance to investors.

The primary focus, and source of contribution of the present study, is to empirically consider the responsiveness of prices on a given stock, not only to social-media sentiment generated on the stock itself, but to additional sentiment stemming from other sources but which may nonetheless connect to the value proposition of a company. To achieve this, we make use of standard Capital Asset Pricing Model (CAPM), augmented with sentiment factors constructed from social-media (Twitter) messages on the stock itself, as well as sentiment concerning the overall market index (e.g. S&P500 market index).

Most existing studies of sentiment and its impact to stock markets follow Bollen et al. (2011) and use daily frequency data. Cohen-Charash et al. (2013) was one of the first to follow from Bollen, and who also build on the notion of collective sentiment (mood), though in their case from news articles, placing emphasis on the definition of word lexicons (or bags of words) that are used to classify mood, and focus on the degree of pleasantness in messages along with signals of the level of ‘activation’ of the message author. Some previous works explore the role of ‘polarity’ in sentiment, for example disagreement and divergence in sentiment as discussed by Siganos (2017). Other applications making use of daily frequency data and measures of sentiment include Mazboudi and Khalil (2017) and You et al. (2017). There are only a handful of studies exploring these relationships using high-frequency or intraday data, which is known to contain additional information on market microstructure (e.g. Ji and Zhang, 2018). Borovkova and Mahakena (2015) examined aspects of price volatility, price jumps and volatility in NYMEX natural gas prices observed at 5-minutely intervals between 2006 and 2010. Corea (2016) directly questions whether Twitter can be a reliable proxy for investors sentiment using a two month sample of data spanning the end of September to the end of November of 2014, observed at the minutely frequency, for Apple, Facebook and Google. Similar to Bollen et al. (2011), predictive regression modeling was also a focus of Renault (2017). Using 30-min interval data on the S&P500 exchange traded fund (ETF) they claim to show *“direct evidence of sentiment-driven noise trading at the intraday level”*.

Unambiguously, the most common sampling frequency when working with intraday data cover 5-minutely intervals, with in excess of 40% of studies using the level of time aggregation.[2] We therefore aggregate our tick-level data to 5-min returns. In the end we estimate models with returns in 1-, 5- and 30-min intervals which we believe to encompass the most common frequencies appearing in intraday asset pricing research.

The remainder of the letter is structured as follows: Section 2 describes the steps taken in obtaining and preparing the data for estimation; Section 3 sets out our empirical strategy; Section 4 presents the results; and Section 5 concludes.

2. Data

Fig. (1) outlines a general schematic of the steps involved in the preliminary data collection. In brief, we take advantage of high-frequency (tick-level) data on stocks obtained from Norwegian bank ‘Netfonds’, permitting an up-to-date access to ultra-high-frequency data. From this source we identify a sample of highly traded stocks, which it is assumed will also have higher representation on social-media networks.

We then, on the basis of the identified ‘liquid’ stocks, stream live Twitter feeds for approximately one month, and continue to collect tick-level trade data. The Tweets require textual processing and analysis, to convert the text into numerical scores of sentiment. Following several previous studies we adopt the Canadian National Research Council (NRC) definitions of tone and sentiment, allowing us to extract the number of terms related to emotions of anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, as well as (partially overlapping) notion of positive and negative sentiment as well as something known as the ‘SMOG’ index.[3] SMOG refers to a simple measure of gobbledygook, and reflects the underlying readability of a passage of text. We extract sentiment from Tweets which refer to a given company ticker. Tweets for a ticker symbol are preceded either by a ‘#’-tag or a ‘$’-tag. Hashtags are widely used and require no special introduction. Cashtags, e.g. $SPY (for the S&P500 ticker), are tweets by tweeters who are self-declaring an above normal awareness of finance and are intended to signal a tweet as being of direct salience to investors and financial professionals (see for example Bartov, 2017). These allow us to construct variables $Hashtag_{t,k}$ and $Cashtag_{t,k}$ which sum up the number of words embedded in tweets, over time interval t, that include sentiment/emotion type k. We denote the summary variables relating to the wider market index as $RM\_Hashtag_{t,k}$ and $RM\_Cashtag_{t,k}$ respectively.

With respect to the financial market data, in order to have equally spaced time intervals it is helpful to aggregate the data over fixed-period intervals as discussed in Lee et al. (2015) among others. Following the discussion above, tick level data are aggregated up to 1-, 5- and 30 minute returns. For estimation we omit periods where markets are closed, and exclude from analysis all sentiment during non-trading hours. We consider stock returns for Exxon Mobil (XOM), General Electric (GE), Chessapeake Energy (CHK), Ford

[1] Please refer to https://www.sec.gov/news/press-release/2018-219 for additional information.

[2] We conducted a general review of intraday asset pricing research to gauge this number. We note that among the many studies we considered, two studies are in fact theoretical econometric studies on high-frequency estimation, but which explicitly make reference to methods that might be applied to 5-minutely data.

[3] See https://www.nrc-cnrc.gc.ca/eng/solutions/advisory/emotion_lexicons.html for further information.

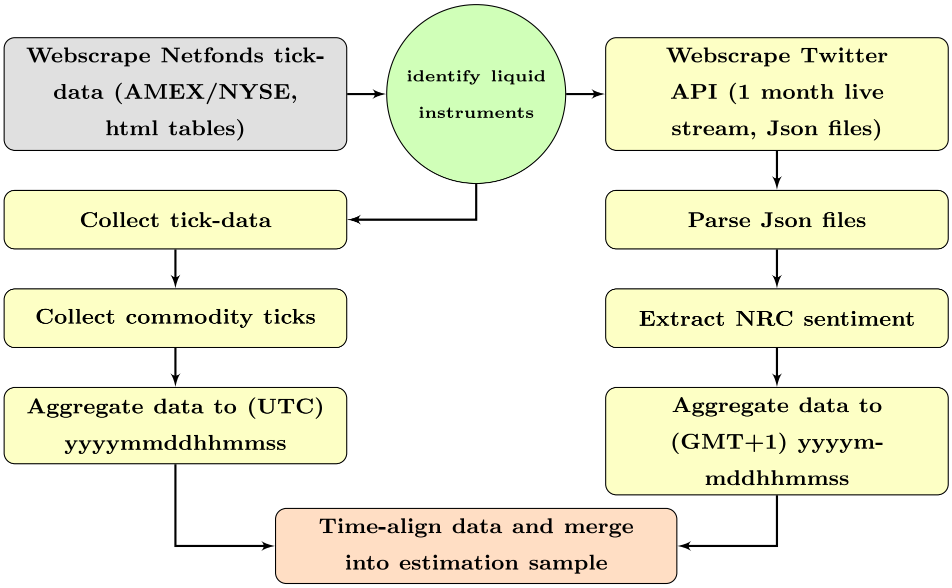

Fig. 1. Schematic for the steps involved in preparing the estimation dataset.

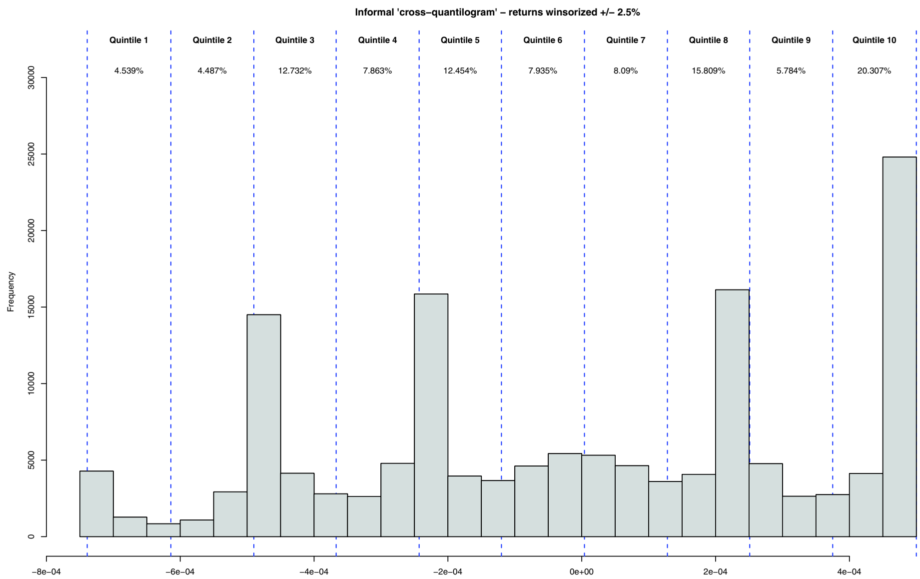

Fig. 2. This histogram shows the returns series for Ford (ticker 'F') using 1-second aggregated stock prices. The quantiles highlight along the top of the graph refer to the distribution of tweets that are being made within the corresponding returns quantiles. For example, quantile 10 refers to the upper 10% values of 1-second price returns for Ford, which also contains the highest peak in the returns series. The corresponding Tweet profile for this quantile shows that 20.507% of all tweets me occurred within this returns quantile. This graph therefore offers model-free verification of an association between social-media and stock returns.

Motor Company (F), Disney (DIS), and Walmart (WMT). The data cover trading for the month of August 2018.

To try and establish some salience of the research objective within the observed data, Fig. (2) plots a histogram the cross section of one-second returns (winsorized at the 2.5% upper and lower tails) on Ford Motor Company for a randomly sampled day. Along the top of this graph the quantiles of returns are highlighted by the dashed blue lines, and the percentage of tweets occurring at the corresponding price quantile is shown. Essentially it can be seen that Tweet numbers peak when price activity is active. This is verified by the large returns activity in quantiles 3, 5, 8 and 10, which have a corresponding Twitter volume accounting for close to 60% of Twitter volume. It is difficult to imagine such spikes are incidental.

Table 1

Main estimation results for sentiment connected to 'hashtag' messages. 'Full' models refer to Eq. (2) estimated by stepwise regression, with heteroskedasticity and autocorrelation corrected (HAC) standard errors used for inference. 'Bi' refers to regressions of the form in Eq. (3), again with HAC standard errors used for inference. '$\#^{-}$' denotes a significant negative reaction to sentiment emanating from a hashtag, and '$\#^{+}$' a significant positive reaction. Subscripts are used to denote whether significance is observed using 1-, 5- or 30-min returns e.g. '$\#^{+}_{(5,30)}$' would denote a positive significant coefficient on a hashtag when conducting regressions using 5- and 30-min returns, but not 1-min.

	Ticker					
Specification[#]	XOM full / individual	GE full / individual	CHK full / individual	F full / individual	DIS full / individual	WMT full / individual
Reaction to current period firm specific sentiment						
anger	$\#^{-}_{(30)}$/	$\#^{+}_{(30)}$/	$\#^{-}_{(1)}$/	/	/	/
anticipation	$\#^{+}_{(30)}$/$\#^{+}_{(5,30)}$	$\#^{-}_{(30)}$/	/$\#^{+}_{(1,30)}$, $\#^{-}_{(5)}$	$\#^{-}_{(30)}$/$\#^{+}_{(30)}$	/	$\#^{+}_{(30)}$/
disgust	$\#^{+}_{(1,30)}$/$\#^{+}_{(5)}$	$\#^{-}_{(5,30)}$/$\#^{-}_{(5)}$	/$\#^{-}_{(5)}$	$\#^{+}_{(1)}$/$\#^{+}_{(1)}$	$\#^{-}_{(30)}$/	/$\#^{-}_{(5,30)}$
fear	/	$\#^{+}_{(5)}$, $\#^{-}_{(30)}$/	$\#^{+}_{(1)}$, $\#^{-}_{(5)}$/$\#^{-}_{(5,30)}$	/	$\#^{+}_{(1)}$/$\#^{+}_{(30)}$	/$\#^{+}_{(1,5,30)}$
joy	/	$\#^{+}_{(5)}$/	/$\#^{+}_{(1,30)}$, $\#^{-}_{(5)}$	$\#^{-}_{(5)}$/$\#^{+}_{(30)}$	/$\#^{+}_{(1)}$	/
sadness	/	$\#^{+}_{(30)}$/	/$\#^{-}_{(5,30)}$	/	$\#^{-}_{(1)}$/	$\#^{-}_{(30)}$/$\#^{-}_{(30)}$
surprise	$\#^{-}_{(1)}$/$\#^{+}_{(5,30)}$	$\#^{-}_{(5)}$, $\#^{+}_{(30)}$/	/$\#^{+}_{(5)}$	/$\#^{+}_{(30)}$	/$\#^{+}_{(1)}$	/
trust	/	$\#^{+}_{(30)}$/	/$\#^{+}_{(1)}$, $\#^{-}_{(5)}$	$\#^{+}_{(30)}$/$\#^{+}_{(30)}$	/	/
negative	$\#^{+}_{(30)}$/	$\#^{-}_{(5,30)}$/	$\#^{+}_{(30)}$/	/	$\#^{+}_{(30)}$/$\#^{+}_{(30)}$	/
positive	/	$\#^{-}_{(30)}$/	/	$\#^{+}_{(5)}$/$\#^{+}_{(30)}$	/	/
SMOG	$\#^{+}_{(5)}$, $\#^{-}_{(30)}$/$\#^{+}_{(5)}$	$\#^{-}_{(30)}$/	$\#^{-}_{(30)}$/	/$\#^{+}_{(30)}$	/	$\#^{-}_{(30)}$/
Reaction to current period 'market' sentiment						
anger	/	/	$\#^{+}_{(5)}$/	/	/	/
anticipation	$\#^{+}_{(30)}$/	/	/	/	/	/
disgust	$\#^{+}_{(30)}$/	/	/	$\#^{+}_{(5)}$/$\#^{+}_{(5)}$	/	/
fear	$\#^{-}_{(30)}$/	/	/	/	/	/
joy	/	/	$\#^{+}_{(30)}$/	/$\#^{+}_{(30)}$	/	/
sadness	/	$\#^{-}_{(5)}$/	/	/	/	/
surprise	/	/	/	/$\#^{+}_{(5)}$	$\#^{-}_{(1,30)}$/	/
trust	$\#^{-}_{(30)}$/	/	/	$\#^{+}_{(30)}$/$\#^{+}_{(5,30)}$	$\#^{+}_{(30)}$/	/
negative	/	$\#^{+}_{(1,5)}$/$\#^{+}_{(1)}$	/	/	/	/
positive	/	/	$\#^{-}_{(30)}$/	/	/	$\#^{-}_{(5)}$/$\#^{-}_{(5)}$
SMOG	/	/	$\#^{+}_{(30)}$/	/	$\#^{+}_{(1)}$/	/$\#^{-}_{(5)}$
Reaction to lagged firm specific sentiment						
anger	/	$\#^{-}_{(30)}$/	/	/	/	/$\#^{-}_{(1)}$, $\#^{-}_{(5)}$, $\#^{-}_{(30)}$
anticipation	/$\#^{+}_{(5,30)}$	$\#^{+}_{(30)}$/	/$\#^{+}_{(1,30)}$, $\#^{-}_{(5)}$	/$\#^{+}_{(30)}$	$\#^{-}_{(30)}$/	/
disgust	/$\#^{+}_{(5)}$	$\#^{+}_{(30)}$/$\#^{-}_{(5)}$	/$\#^{-}_{(5)}$	/$\#^{+}_{(1)}$	/	/$\#^{-}_{(5,30)}$
fear	/	$\#^{+}_{(30)}$/	/$\#^{-}_{(5,30)}$	/	/$\#^{+}_{(30)}$	/$\#^{+}_{(1,5,30)}$
joy	/	/	/$\#^{+}_{(1,30)}$, $\#^{-}_{(5)}$	/$\#^{+}_{(30)}$	/$\#^{+}_{(1)}$	/
sadness	/	$\#^{-}_{(30)}$/	/$\#^{-}_{(5,30)}$	$\#^{-}_{(30)}$/	/	/$\#^{-}_{(30)}$
surprise	/$\#^{+}_{(5,30)}$	$\#^{-}_{(30)}$/	/$\#^{+}_{(5)}$	/$\#^{+}_{(30)}$	$\#^{+}_{(30)}$/$\#^{+}_{(1)}$	/
trust	/	$\#^{-}_{(30)}$/	/$\#^{+}_{(1)}$, $\#^{-}_{(5)}$	$\#^{-}_{(30)}$/$\#^{+}_{(30)}$	/	/
negative	$\#^{-}_{(30)}$/	$\#^{+}_{(30)}$/	$\#^{-}_{(30)}$/	/	/$\#^{+}_{(30)}$	/
positive	/	$\#^{+}_{(30)}$/	/	$\#^{+}_{(30)}$/$\#^{+}_{(30)}$	/	/
SMOG	/$\#^{+}_{(5)}$	$\#^{+}_{(30)}$/	$\#^{+}_{(30)}$/	$\#^{-}_{(30)}$/	/	/
Reaction to lagged 'market' sentiment						
anger	/	/	/	/	/	/
anticipation	/	/	/	/	/	/
disgust	/	/	/	/$\#^{+}_{(5)}$	/	/
fear	/	/	/	/	/	/
joy	/	/	/	/	/	/
sadness	/	/	/	/	/	/
surprise	/	/	/	/$\#^{+}_{(5)}$	/	/
trust	/	/	/	/$\#^{+}_{(5,30)}$	/	/
negative	/	/$\#^{+}_{(1)}$	/	/	/	/
positive	/	/	/	/	/	/$\#^{-}_{(5)}$
SMOG	/	/	/	/	/	/$\#^{-}_{(5)}$

(*continued on next page*)

Table 1 (*continued*)

Coefficients on market β(variable RM)						
1-min returns	0.750***	0.749***	1.452***	0.770***	0.777***	0.527***
5-min returns	0.778***	0.800***	1.903***	0.878***	0.660***	0.437***
30-min returns	0.690***	0.891***	2.396***	0.877***	0.943***	0.473***
*Diagnostic statistics***						
Observations						
1-min returns	7002	7002	7002	7002	7001	7002
5-min returns	1386	1386	1386	1386	1386	1386
30-min returns	216	216	216	216	216	216
Adjusted R^2						
1-min returns	0.14	0.05	0.03	0.06	0.09	0.07
5-min returns	0.14	0.08	0.07	0.09	0.07	0.06
30-min returns	0.25	0.13	0.12	0.25	0.28	0.17

Note: * 'full' refers to stepwise estimation of Eqn. (2), 'individual' refers to the partial sentiment regressions as in Eqn. (3).
** Diagnostic statistics reported for 'full' regressions only.

3. Methodology for analysis

The empirical analysis is developed around a simple CAPM model. Define the log-difference returns on an individual stock (R_t) as:

$$R_t = 100*(ln(P_t) - ln(P_{t-1})) \tag{1}$$

and apply a similar transformation to the S&P500 index data to obtain market returns (RM_t). Using these returns series we can specify our augmented CAPM structure as follows:[4]

$$\begin{aligned} R_t &= \alpha + \beta RM_t + \sum_{j=0}^{1}\sum_{k=1}^{K}\delta_{\#,j,k}Hashtag_{t-j,k} + \sum_{j=0}^{1}\sum_{k=1}^{K}\delta_{\$,j,k}Cashtag_{t-j,k} \\ &\quad + \sum_{j=0}^{1}\sum_{k=1}^{K}\gamma_{\#,j,k}RM\_Hashtag_{t-j,k} + \sum_{j=0}^{1}\sum_{k=1}^{K}\gamma_{\$,j,k}RM\_Cashtag_{t-j,k} + \varepsilon_t \end{aligned} \tag{2}$$

$$\begin{aligned} k &= anger, anticipation, disgust, fear, joy, sadness, \\ &\quad surprise, trust, positive, negative, SMOG \end{aligned}$$

Eq. (2) is estimated using ordinary least squares, with all inference being based upon heteroskedasticity and autocorrelation corrected (HAC) standard errors. General-to-specific model reduction is achieved through the use of stepwise regressions, which sequentially removes regressors on the basis of information criteria. GARCH specifications were also considered, but found to be impractical in early estimations, owing to the large number of variables in the most general specifications. While HAC standard errors do not contain controls for conditional heteroskedasticity, they should nonetheless eliminate the most severe concerns regarding such features.

There are patterns of high correlation between different sentiment types. Therefore, baseline regressions of (1) are complemented by a series of additional 'partial' regressions taking each of the sentiment variables into the CAPM specification one at a time. These regressions therefore take the form:

$$R_t = \alpha + \beta RM_t + \kappa_i Sentiment_{t,i} + \varepsilon_t \tag{3}$$

Where $Sentiment_{t,i}$ is a stacked vector of all variables in $Hashtag_{t-j,k}$, $Cashtag_{t-j,k}$, $RM\_Hashtag_{t-j,k}$ and $RM\_Cashtag_{t-j,k}$, containing $i = 8k$ elements (given that $j = 0, 1$). All analysis is done using the R software package, from data collection, to processing, and estimation.

4. Results

The main results are presented in Table (1) for hashtags, and Table (2) for cashtags. For each of the six stocks considered, we reveal significant intraday reactions to sentiment. In all cases we observe reactions to sentiment on the ticker in question, e.g. 'own' sentiment reactions, which under our preliminary assertions are a firm-specific sentiment factor. We additionally uncover reactions to market wide sentiment in all cases, to a greater or lesser degree. For each ticker there is at least some reaction to sentiment on the broader market index.

Lagged sentiment terms are significant across all stocks, suggesting that markets take some time to process the full value-implications of sentiment. These lags are mostly significant for firm-specific sentiment, especially for hashtags. For cashtags, 5/6 of the stocks show some systematic reaction to lagged market sentiment, but still the *firm-specific* lagged cashtags are more often significant

[4] We would like to thank an anonymous referee for encouraging us to consider specifications including lagged variables. Here, for pragmatic reasons we include only a first order lag. Future research might try to look for longer temporal dependence structures.

Table 2

Main estimation results for sentiment connected to 'cashtag' messages. 'Full' models refer to Eq. (2) estimated by stepwise regression, with heteroskedasticity and autocorrelation corrected (HAC) standard errors used for inference. 'Bi' refers to regressions of the form in Eq. (3), again with HAC standard errors used for inference. '$\$^{-}$' denotes a significant negative reaction to sentiment emanating from a cashtag, and '$\$^{+}$' a significant positive reaction. Subscripts are used to denote whether significance is observed using 1-, 5- or 30-min returns e.g. '$\$^{+}_{(5,30)}$' would denote a positive significant coefficient on a cashtag when conducting regressions using 5- and 30-min returns, but not 1-min.

	Ticker					
Specification[#]	XOM full / individual	GE full / individual	CHK full / individual	F full / individual	DIS full / individual	WMT full / individual
Reaction to current period firm specific sentiment						
anger	$\$^{+}_{(30)}/\$^{-}_{(30)}$	/	/	/	$\$^{+}_{(30)}/\$^{+}_{(30)}$	$/\$^{+}_{(5)}$
anticipation	$\$^{+}_{(30)}/\$^{-}_{(30)}$	$\$^{+}_{(30)}/\$^{-}_{(1)}$	$\$^{+}_{(30)}/$	$\$^{+}_{(1)}/\$^{+}_{(1)}$	/	/
disgust	$\$^{+}_{(30)}/$	/	$\$^{+}_{(1)}, \$^{+}_{(30)}/\$^{+}_{(1)}$	$/\$^{+}_{(30)}$	/	/
fear	$\$^{-}_{(30)}/\$^{-}_{(30)}$	/	$\$^{+}_{(30)}/\$^{+}_{(30)}$	$/\$^{+}_{(1)}$	/	$\$^{+}_{(30)}/\$^{+}_{(1)}$
joy	$\$^{+}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	/	$/\$^{+}_{(1)}$	/	$\$^{+}_{(1)}/\$^{+}_{(1)}$
sadness	$\$^{+}_{(30)}/$	/	$\$^{-}_{(30)}/$	$/\$^{+}_{(30)}$	/	$/\$^{+}_{(5)}$
surprise	$\$^{-}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	/	$\$^{-}_{(30)}/$	/	/
trust	$\$^{+}_{(1)}, \$^{-}_{(30)}/\$^{+}_{(1)}, \$^{-}_{(30)}$	$\$^{+}_{(5)}/\$^{-}_{(1)}$	/	$/\$^{+}_{(1)}$	/	$\$^{-}_{(30)}/$
negative	$\$^{+}_{(30)}/\$^{-}_{(30)}$	$\$^{+}_{(30)}/\$^{-}_{(1)}$	$/\$^{+}_{(30)}$	$\$^{+}_{(1,30)}/\$^{+}_{(1,5,30)}$	$\$^{-}_{(30)}/$	$\$^{+}_{(5)}/\$^{+}_{(1,5)}$
positive	$\$^{-}_{(5,30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	$\$^{-}_{(30)}/$	$/\$^{+}_{(30)}$	/	/
SMOG	$\$^{+}_{(30)}/$	$\$^{-}_{(5,30)}/\$^{-}_{(1)}$	/	$/\$^{+}_{(1,5,30)}$	/	$\$^{-}_{(5)}, \$^{+}_{(30)}/$
Reaction to current period 'market' sentiment						
anger	/	$\$^{-}_{(5)}/\$^{-}_{(5)}$	/	$\$^{-}_{(5)}/$	$\$^{-}_{(30)}/$	$\$^{+}_{(30)}/$
anticipation	$\$^{+}_{(5)}/$	/	$\$^{-}_{(1,30)}/$	$\$^{-}_{(30)}/$	/	$\$^{-}_{(1,30)}/$
disgust	/	/	/	/	$\$^{+}_{(30)}/$	/
fear	/	$\$^{-}_{(5)}/\$^{-}_{(1)}$	/	$\$^{-}_{(30)}/$	$\$^{-}_{(5)}/\$^{-}_{(5)}$	/
joy	/	/	/	/	$\$^{+}_{(1)}/$	/
sadness	$\$^{+}_{(5)}/\$^{+}_{(5)}$	$\$^{+}_{(5)}/$	$\$^{-}_{(5)}/\$^{-}_{(5)}$	/	$\$^{-}_{(1)}, \$^{+}_{(30)}/$	$\$^{-}_{(30)}/$
surprise	$\$^{-}_{(5)}, \$^{-}_{(30)}/$	/	/	/	/	$\$^{+}_{(1)}/$
trust	/	/	$\$^{+}_{(1,5,30)}/\$^{+}_{(1)}$	/	/	/
negative	/	/	/	$\$^{+}_{(5)}/\$^{+}_{(5)}$	$\$^{-}_{(30)}/$	/
positive	/	/	/	/	$\$^{-}_{(1)}/$	$\$^{+}_{(30)}/$
SMOG	/	$\$^{+}_{(5)}/$	$\$^{-}_{(30)}/$	$\$^{+}_{(30)}/$	$\$^{+}_{(1,5,30)}/$	/
Reaction to lagged firm specific sentiment						
anger	$\$^{-}_{(30)}/\$^{-}_{(30)}$	$\$^{-}_{(30)}/$	/	/	$/\$^{+}_{(30)}$	$/\$^{+}_{(5)}$
anticipation	$\$^{-}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	$\$^{-}_{(30)}/$	$/\$^{+}_{(1)}$	/	$\$^{-}_{(30)}/$
disgust	$\$^{-}_{(30)}/$	/	$\$^{-}_{(30)}/\$^{+}_{(1)}$	$/\$^{+}_{(30)}$	/	$\$^{-}_{(30)}/$
fear	$\$^{+}_{(30)}/$	/	$\$^{-}_{(30)}/\$^{+}_{(30)}$	$/\$^{+}_{(1)}$	/	$\$^{-}_{(30)}/\$^{+}_{(1)}$
joy	$\$^{-}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	/	$/\$^{+}_{(1)}$	/	$/\$^{+}_{(1)}$
sadness	$\$^{-}_{(30)}/$	/	$\$^{+}_{(30)}/$	$/\$^{+}_{(30)}$	/	$/\$^{+}_{(5)}$
surprise	$\$^{+}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	/	/	/	/
trust	$\$^{+}_{(30)}/\$^{+}_{(1)}, \$^{-}_{(30)}$	$/\$^{-}_{(1)}$	/	$/\$^{+}_{(1)}$	/	$\$^{+}_{(30)}/$
negative	$\$^{-}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	$/\$^{+}_{(30)}$	$/\$^{+}_{(1,5,30)}$	/	$/\$^{+}_{(1,5)}$
positive	$\$^{+}_{(30)}/\$^{-}_{(30)}$	$/\$^{-}_{(1)}$	$\$^{+}_{(30)}/$	$/\$^{+}_{(30)}$	/	/
SMOG	$\$^{-}_{(30)}/$	$/\$^{-}_{(1)}$	/	$/\$^{+}_{(1,5,30)}$	/	$\$^{-}_{(30)}/$
Reaction to lagged 'market' sentiment						
anger	/	$/\$^{-}_{(5)}$	/	/	/	/
anticipation	/	/	$/\$^{-}_{(30)}$	/	/	/
disgust	/	$/\$^{-}_{(30)}$	/	/	/	/
fear	/	$/\$^{-}_{(1)}$	/	/	$/\$^{-}_{(5)}$	/
joy	/	/	/	/	/	/
sadness	$/\$^{+}_{(5)}$	/	$/\$^{-}_{(5)}$	/	/	/
surprise	/	/	/	/	/	/
trust	/	/	$/\$^{+}_{(1)}$	/	/	/
negative	/	/	/	$/\$^{+}_{(5)}$	/	/
positive	/	/	/	/	/	/

(*continued on next page*)

Table 2 (*continued*)

SMOG	/$$^{-}_{(30)}$	/	/	/	/	/
Coefficients on market β(variable RM)						
1-min returns	0.750***	0.749***	1.452***	0.770***	0.777***	0.527***
5-min returns	0.778***	0.800***	1.903***	0.878***	0.660***	0.437***
30-min returns	0.690***	0.891***	2.396***	0.877***	0.943***	0.473***
*Diagnostic statistics***						
Observations						
1-min returns	7002	7002	7002	7002	7001	7002
5-min returns	1386	1386	1386	1386	1386	1386
30-min returns	216	216	216	216	216	216
Adjusted R^2						
1-min returns	0.14	0.05	0.03	0.06	0.09	0.07
5-min returns	0.14	0.08	0.07	0.09	0.07	0.06
30-min returns	0.25	0.13	0.12	0.25	0.28	0.17

Note: * 'full' refers to stepwise estimation of Eqn. (2), 'individual' refers to the partial sentiment regressions as in Eqn. (3).
** Diagnostic statistics reported for 'full' regressions only.

across the different sentiment types and stocks.

Two further features of the results deserve special attention, namely that the results differ when using different return intervals, and that there is a high degree of sign inconsistency in the effects. Taking the latter first, the observed significance (ignoring the sign for now) is strong evidence of co-movement, which given the significance is obtained across all stocks, would be difficult to attribute to patterns of 'incidentally significant noise'. The sign inconsistency, when using different returns intervals and across stocks, is a bit more of a mystery. We do not have a clear explanation for this pattern, though nor did we have any strong priors on the signs *a-priori*. While it seems intuitive for instance that negative sentiment conveys bad news which might depress stock prices, the same negative sentiment could potentially signal a period of temporary mis-pricing that could stimulate speculative trading, especially if analysts/ investors hold the belief that the sentiment is somehow erroneous.

5. Conclusion

This paper provides an alternative lens on the relationship between social-media based sentiment factors and intraday stock returns. The primary contribution stems from our treatment of the sources of sentiment, and our initial attempt to disentangle what we refer to as firm specific and market wide sentiment. Using a sample of US companies' stock returns measured at 1-, 5- and 30 min frequency, we run regressions based on an empirical CAPM and find clear significant effects attributable to firm specific and market wide sources of sentiment.

Although the results are slightly mixed, subject to the diversity among the stocks under investigation, they unambiguously demonstrate the pricing power of sentiment extracted from social-media. A number of high priority extensions from this study might include: expanding the set of stocks considered; exploring non-linearities and interactions between classes of sentiment; or establishing the causal structure and patterns of association between stocks and sentiment within a spillover modeling framework. The latter of these is in many ways the most intuitive extension of this letter, since it builds upon the premise that prices reflect and absorb price risk emanating from a range of direct and indirect sources.

Acknowledgements

This project is supported by the National Natural Science Foundation of China (NSFC grant No.71573214) and the Fundamental Research Funds for the Central Universities (grant No.JBK1802028). The corresponding author would also like to acknowledge Friends of the Earth (Hong Kong) Limited for continued support as an Environmental Economics Research Fellow.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.frl.2019.03.030.

References

Bartov, E., Faurel, L., Mohanram, P.S., 2017. Can twitter help predict firm-level earnings and stock returns? Account. Rev. 93 (3), 25–57.
Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. J. Comput. Sci. 2 (1), 1–8.
Borovkova, S., Mahakena, D., 2015. News, volatility and jumps: the case of natural gas futures. Quant. Finance 15 (7), 1217–1242.
Bukovina, J., 2016. Social-media big data and capital markets—An overview. J. Behav. Exp. Finance 11, 18–26.
Chen, H., De, P., Hu, Y.J., Hwang, B.-H., 2014. Wisdom of crowds: the value of stock opinions transmitted through social-media. Rev. Financ. Stud. 27 (5), 1367–1403.
Cohen-Charash, Y., Scherbaum, C.A., Kammeyer-Mueller, J.D., Staw, B.M., 2013. Mood and the market: can press reports of investors' mood predict stock prices? PLoS ONE 8 (8). E72031
Corea, F., 2016. Can twitter proxy the investors' sentiment? the case for the technology sector. Big Data Res. 4, 70–74.
de Jong, P., Elfayoumy, S., Schnusenberg, O., 2017. From returns to tweets and back: an investigation of the stocks in the dow jones industrial average. J. Behav. Finance 18 (1), 54–64.

Ji, Q., Zhang, D., 2018. China's crude oil futures: introduction and some stylized facts. Finance Res. Lett. https://doi.org/10.1016/j.frl.2018.06.005
Mazboudi, M., Khalil, S., 2017. The attenuation effect of social-media: evidence from acquisitions by large firms. J. Financ. Stab. 28, 115–124.
Renault, T., 2017. Intraday online investor sentiment and return patterns in the US stock market. J. Bank. Finance 84, 25–40.
Siganos, A., Evangelos, V.-N., Patrick, V., 2017. Divergence of sentiment and stock market trading. J. Bank. Finance 78, 130–141.
Sprenger, T.O., Sandner, P.G., Tumasjan, A., Welpe, I.M., 2014. News or noise? Using twitter to identify and understand companyspecific news flow. J. Busi. Finance Account. 41 (7–8), 791–830.
You, W., Guo, Y., Peng, C., 2017. Twitter'S daily happiness sentiment and the predictability of stock returns. Finance Res. Lett. 23, 58–64.
Zhang, Y., An, Y., Feng, X., Jin, X., 2017. Celebrities and ordinaries in social networks: who knows more information? Finance Res. Lett. 20, 153–161.