Momentum, Mean-Reversion and Social Media: Evidence from StockTwits and Twitter*

Shreyash Agrawal[†], Pablo D. Azar[‡], Andrew W. Lo[§], Taranjit Singh[¶]

This Draft: 12 July 2018

We analyze the relation between stock market liquidity and real-time measures of sentiment obtained from the social-media platforms StockTwits and Twitter. Linear regression analysis shows that extreme sentiment corresponds to higher demand and lower supply of liquidity, with negative sentiment having a much larger effect on demand and supply than positive sentiment. An intraday event study shows that booms and panics end when bullish and bearish sentiment reach extreme levels, respectively. After extreme sentiment, prices become more mean-reverting and spreads narrow. To quantify the magnitudes of these effects, we conduct a historical simulation of a market-neutral mean-reversion strategy that uses social-media information to determine its portfolio allocations. Our results suggest that the demand and supply of liquidity are influenced by investor sentiment, and that market makers who can keep their transaction costs to a minimum are able to profit by using extreme bullish and bearish emotions in social media as a real-time barometer for the end of momentum and a return to mean reversion.

Keywords: Sentiment; Market Liquidity; Social Media; Twitter; StockTwits; Mean Reversion

JEL classification: G11, G12

---

*We thank James Crane-Baker at Psychsignal and Pierce Crosby at Stocktwits for providing the social-media sentiment data used in this study, and Shomesh Chaudhuri, Katy Kaminski, and Amir Khandani for helpful discussion and comments. We thank Jayna Cummings for editorial assistance. Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged.

†MIT Laboratory for Financial Engineering, shreyash@mit.edu (e-mail).

‡MIT Department of Economics and Laboratory for Financial Engineering, pabloazar@gmail.com (e-mail).

§MIT Sloan School of Management, Laboratory for Financial Engineering, and Computer Science and Artificial Intelligence Laboratory, alo-admin@mit.edu (e-mail). Corresponding author.

¶MIT Laboratory for Financial Engineering, taranjitsingh96@gmail.com (e-mail).

Contents

1 Introduction 1

2 Related Work 3

3 Data and Variables 4

4 Regression Analysis 6

5 Event Study Analysis 8

6 Trading-Strategy Analysis 9

7 Conclusion 14

A Regression Results 17

1 Introduction

Over the last decade, financial markets have witnessed the rise of new systemic risk factors such as the Quant Meltdown of August 2007 and the Flash Crashes of May 2010 and August 2015. These events were driven in part by the sudden disappearance of liquidity providers, leading to large price movements. As prices shift, market makers who rely on mean reversion will pull out of the market, triggering a self-reinforcing feedback loop where prices move even more sharply due to declining liquidity, causing more investors to liquidate their holdings in a panic because of increasing price volatility.

This same period has also been characterized by the rise of alternative sources of data, especially those that can be used to measure quantitatively news content and social-media sentiment. With the use of machine learning tools, it is now possible to process large amounts of text information about an asset and assign real-time investor-sentiment scores based on natural-language processing of these texts.

In this article, we study whether social media and news data can provide us with insights on market panics and manias that are not already captured by existing data. While these new data sources have the advantage of being available in real time and represent the beliefs of a wide variety of investors, inferring information from them requires separating useful signals about investor sentiment from everyday noise. In particular, we focus on the following questions:

- Given that social-media users represent a small fraction of market participants (and many non-market participants), do these sources contain relevant information about liquidity?
- To what extent can social media give us insights that cannot be inferred from more fundamental sources, such as traditional news feeds?
- Do positive and negative sentiment have asymmetric effects on markets?
- Can news and social media be used to predict future levels of liquidity?
- Can social-media information be used to improve trading strategies?

We answer these questions using three different approaches:

1. We regress several measures of volume and liquidity—log number of trades, log number of quotes, log number of trades outside the quoted bid-ask spread, log turnover, and average spreads—on news and social-media sentiment indicators.

2. We perform a series of intraday event studies on abnormal social-media sentiment.

3. We perform historical simulations of an intraday mean-reversion strategy that uses social-media messages to assign portfolio weights.

Our regression results show that social-media sentiment—as measured by Twitter and StockTwits messages—does exhibit correlation with liquidity measures that cannot be explained by other sources of data such as news sentiment. Furthermore, negative social-media sentiment has a much larger effect on liquidity measures than positive sentiment.[1] A one percent increase in bearish sentiment has twice as much impact on our trading and liquidity measures than a one percent increase in bullish sentiment. By using *pre-trading* measures of sentiment, we find that news and social-media data can be used to predict liquidity measures before the market opens.

Our event study shows that highly abnormal social-media sentiment—defined as a sentiment score that is at least three standard deviations above or below average for a given stock—is preceded by very strong momentum and followed by mean-reverting returns. Social-media sentiment can be used to detect the peak of an intraday boom or the trough of an intraday panic. In addition, our results show that these abnormal-sentiment events are followed in the next half-hour by a decrease in spreads. This decrease in spreads is slightly larger for positive events than for negative events.

Our results suggest that high volumes of social-media messages are followed by increases in liquidity and a return to mean reversion. We use this intuition to backtest an intraday mean-reversion trading strategy that trades in 30-minute intervals. Every 30 minutes, our strategy buys equities that had negative returns over the previous window, and shortsells equities that had positive returns. Our strategy puts more positive or negative weight on stocks that had a high amount of StockTwits and Twitter message volume on the previous 30-minute window, essentially betting on mean reversion after a high volume of social-media events. This strategy outperforms a benchmark mean-reversion strategy that does not use social-media data.

[1] As described below, messages are tagged with two independent scores, corresponding to bullish and bearish sentiment.

2 Related Work

While our work is perhaps the first to study the effect of social-media sentiment on intraday liquidity, it is part of a growing literature that studies the effect of social media and news on asset prices. Dredze et al. (2016) provide a survey of that literature.

The earliest papers in this literature considered the effect of social media on aggregate indices such as the S&P 500 and the VIX volatility index. Bollen, Mao, and Zeng (2011) show that Twitter sentiment can be used to predict the returns of the Dow Jones Industrial Average. Mao, Counts, and Bollen (2011) follow up and show that Twitter and Google Insight Search (GIS) sentiment can predict daily stock market index returns, including the Dow Jones Industrial Average, the S&P 500, and the Russell 2000 index. They find that Twitter sentiment leads GIS sentiment, and that bullish sentiment can predict positive returns in stock indices.[2] Zhang, Fuehres, and Gloor (2011) find that the mass use of highly emotional language (e.g., "hope","fear", and"anxious") in financial tweets predicts a low return for stock indices the next day, and an increase in the VIX. More recently, Karagozoglu and Fabozzi (2017) use social-media sentiment to develop a strategy that trades VIX futures, and show that this strategy outperforms a benchmark, even after taking into account transaction costs.

Several studies have considered the effect of social media on individual equities. Da, Engelberg, and Gao (2011) use Google search volume data and find that an increase in search volume predicts a short term increase in prices, followed by a reversal within the year. Ruiz et al. (2012) build daily social networks for each stock in their universe based on retweets and other characteristics (e.g., Alice and Bob are considered "connected" in the YHOO social network on a given day if Alice tweets about YHOO and Bob retweets her message that day), and show that the features of these social networks, such as the size of the largest connected component, correlate to trade volume and returns for the stocks. They build a daily trading strategy based on their results and show that it outperforms several benchmarks. Sul, Dennis, and Yuan (2014) find the surprising result that tweets that are not retweeted have the most impact on future returns. They use an event study to show that a strategy using tweet-based information from non-retweeted tweets outperforms a market benchmark. Curtis, Richardson, and Schmardebeck (2016) show that a high volume of social-media messages implies higher return sensitivity to earnings surprises, and a low volume of

[2]However, there is some debate as to the statistical validity of the results in these two papers; see Lachanski and Pav (2017) for further details.

social-media messages implies a larger magnitude in post-earnings announcement drift. Chen et al. (2014) use posts from the popular finance website `www.SeekingAlpha.com` to predict stock returns and earnings surprises. Sun, Lachanski, and Fabozzi (2016) use a sparse latent space model to jointly model social-media volume and asset-price movements which they use to construct an investment strategy.

More recent studies have focused on the interaction between social media and other sources of company information such as press releases by companies and government institutions. Blankespoor, deHaan, and Zhu (2017) show that the Associated Press' use of robo-journalism increased trading volume and liquidity. That is, even though articles written by robots added no information to the market, their introduction led investors to increase trading, strongly suggesting that investors respond to published news articles. Lee, Hutton, and Shu (2015) show that firms that tweet frequently after a product recall will experience lower drops in their stock prices than firms that allow the conversation to be driven by other users. This is additional evidence that social media is a dissemination channel for information that is relevant to asset prices. Ge, Kurov, and Wolfe (2017) show that President Trump's tweets about individual companies have significant effects on these companies' intraday returns and increase intraday volume. Azar and Lo (2016) use Twitter sentiment in anticipation of Federal Reserve meetings to predict the S&P 500's performance in reaction to announcements from the Federal Open Market Committee. They show that the content of tweets can be used to predict future returns, and that a tweet-based asset-allocation strategy outperforms several benchmarks.

3 Data and Variables

We use data from RavenPack to measure news sentiment. RavenPack provides a database of news events associated with equities, with each news event scored for relevance, novelty, and sentiment. We use the Composite Sentiment Score (CSS) as a measure of news sentiment. RavenPack determines this score by parsing news text and assigning positive and negative sentiment scores to the words and phrases in the news piece. The model used to assign scores is estimated to predict intraday stock-price reactions using 100 large cap stocks. This score ranges between 0 and 100, with 0 representing completely negative sentiment, 50 representing neutral sentiment, and 100 representing completely positive sentiment. We filter news events and retain only those with maximum relevance and novelty scores, events which have a completely neutral score of 50, and events reflecting developments that have already taken place.[3] To make meaningful comparisons with our other sentiment data, we use a logarithmic transformation of this measure, $\overline{\text{CSS}} \equiv \log(1 + \frac{\text{CSS}}{100})$.

We also use minute intraday data from PsychSignal to measure social-media sentiment. We use the combined StockTwits and Twitter-with-retweets feed, treating any message across both Twitter and StockTwits as relevant to our analysis. PsychSignal provides separate minute-level scores for bullish and bearish sentiment for equities, as well as a count of the number of messages. We normalize the bullish and bearish scores to be between 0 and 1 and then take logarithms of these variables.

For our regression analysis, our dependent variables (number of trades, returns, etc.) are measured at the daily level. Thus, we aggregate news and social-media sentiment to a daily variable which averages the intraday sentiment scores between 9:30am and 4:00pm on any given trading day. For every ticker and trading day, we use the logarithm of the total number of messages between 9:30am and 4:00pm as a measure of the total-social media message volume. We also aggregate the variables during the "pre-trade" interval between 4:00am and 9:30am to perform predictive regressions, where the independent variables are measured during normal market trading hours.

Summary statistics for these variables are given in Exhibit 1, including a Dickey-Fuller test statistic showing that all variables are stationary.[4] We use a universe of 4,544 equities for the years 2011 through 2014.

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**DF Statistic**	**N**
News Sentiment	0.497	0.057	0.040	1	−8.317	68,576
Bullish Social Sentiment	−2.268	0.929	−8.900	−0.111	−6.341	68,576
Bearish Social Sentiment	−2.998	1.195	−10.556	0	−7.756	68,576
Log Number of Messages	2.082	1.549	0	10.343	−4.241	68,576

Exhibit 1: Summary Statistics of Independent Variables

Exhibit 2 provides a summary of our intraday indicators,[5] including daily returns, turnover, number of trades, number of quotes, and value-weighted quoted spread (measured

[3] We remove the categories "order-imbalances","technical-analysis", and "insider-trading", following the RavenPack guide.

[4] The Dickey-Fuller test we use has a lag of 4 trading days.

[5] Our source of intraday indicators is the WRDS Intraday Indicators database.

in dollars). We also obtain from Wharton Research Data Services (WRDS) the number of "mini flash-crashes", which is measured as the number of trades outside the quoted spread. A trade outside the quoted spread is a proxy for either buyers who are desperate to buy or sellers who are panicking to sell.

Variable	Mean	Std. Dev.	Min.	Max.	DF Statistic	N
Return	0	0.040	−1.420	1.438	−13.940	189,827
Log Turnover	2.155	1.225	−8.661	8.493	−7.048	189,991
Log Number of Trades	8.359	2.016	0	14.218	−6.997	189,994
Log Number of Quotes	11.169	1.822	0	15.934	−5.617	189,967
Log Mini Flash Crash Count	6.163	2.167	0	13.404	−7.177	187,039
Dollar Quoted Spread	0.075	0.260	0	25.858	−12.960	189,963

Exhibit 2: Summary Statistics of Dependent Variables

4 Regression Analysis

Exhibits 3 and 4 contain the results of our regression analysis. For each one of our intraday indicators, $Y_{i,t}$, we estimate a linear regression with time and ticker fixed effects as follows:

$$Y_{i,t} = \beta_{\text{News}}\text{News}_{i,t} + \beta_{\text{Bull}}\text{Bull}_{i,t} + \beta_{\text{Bear}}\text{Bear}_{i,t} + \beta_{\text{Messages}}\text{Messages}_{i,t} + \gamma_i + \delta_t + \epsilon_{i,t} \quad (1)$$

where $\epsilon_{i,t}$ is an error term that is clustered by ticker. For expositional purposes, we include in the main body the results when $Y_{i,t}$ represents turnover and when $Y_{i,t}$ represents the number of mini-flash crashes. All other regression results are relegated to the Appendix.

Our main finding is that there is much more demand for liquidity, and much less supply, when social-media sentiment is negative than when it is positive. The coefficient β_{Bear} in the regressions for number of trades, number of quotes, and turnover is positive and about twice as large as the coefficient β_{Bull}, indicating that demand for liquidity is higher when investor sentiment is bearish. At the same time, the coefficient β_{Bear} in the regressions for spreads and the number of mini flash-crashes is also positive and twice as large as the coefficient β_{Bull}. One interpretation of these results is that, since spreads and mini flash-crashes increase when market makers withdraw from trades, high sentiment implies a low supply of liquidity.

The coefficient β_{News} in the liquidity indicators regression is negative, implying asymmetric reaction of liquidity measures to sentiment. Even though we do not have separate measures

Variable	(1) FE	(2) FE	(3) FE	(4) FE Predictive
News Sentiment	-1.384***	-1.274***	-1.018***	-0.006***
	(0.11)	(0.11)	(0.09)	(0.00)
Bullish Sentiment		1.158***	0.670***	0.114***
		(0.04)	(0.03)	(0.01)
Bearish Sentiment		1.951***	1.398***	0.138***
		(0.09)	(0.06)	(0.01)
Number of Messages			0.338***	0.004***
			(0.01)	(0.00)
Constant	2.778***	2.598***	1.840***	2.588***
	(0.04)	(0.04)	(0.04)	(0.04)
Observations	68,575	68,575	68,575	72,604
Number of tickers	4,544	4,544	4,544	4,370
Within R^2	0.00426	0.0393	0.238	0.0508
Between R^2	0.00158	0.193	0.328	0.0977
R^2	0.00390	0.0919	0.204	0.0501

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Exhibit 3: Turnover

for positive and negative news sentiment, the fact that this coefficient is negative in regressions for turnover and number of trades shows that demand for liquidity will increase more for negative news than for positive news. This coefficient is also negative when measuring the number of mini flash-crashes, implying that the amount of liquidity provided will be lower for negative news than for positive news.

These results are purely descriptive. However, we may want to know whether our liquidity measures can be predicted before markets open. Our predictive-regression results—using news and social media sentiment measured in the pre-market hours to predict liquidity during market hours—show that sentiment does have some predictive power. However, it should be noted that the coefficients and R^2 of these regressions are smaller than those in our descriptive regressions. As in our predictive results, the coefficients for β_{Bull}, β_{Bear}, and β_{Messages} will generally be positive and significant (with the regression for spreads being the only exception). This indicates that social-media activity before markets open predicts higher demand for liquidity during the trading day. However, higher social media sentiment also predicts a higher number of mini flash-crashes, indicating a lower supply of liquidity.

Variable	(1) FE	(2) FE	(3) FE	(4) FE Predictive
News Sentiment	-1.342***	-1.208***	-0.888***	-0.005***
	(0.12)	(0.12)	(0.10)	(0.00)
Bullish Sentiment		1.320***	0.703***	0.107***
		(0.05)	(0.04)	(0.01)
Bearish Sentiment		2.230***	1.534***	0.088***
		(0.11)	(0.07)	(0.01)
Number of Messages			0.426***	0.006***
			(0.01)	(0.00)
Constant	7.102***	6.892***	5.933***	6.605***
	(0.05)	(0.05)	(0.05)	(0.05)
Observations	67,554	67,554	67,554	72,016
Number of ticker_num	4,511	4,511	4,511	4,338
Within R^2	0.00281	0.0355	0.259	0.0513
Between R^2	8.48e-05	0.165	0.287	0.115
R^2	1.94e-05	0.0846	0.326	0.0715

Robust standard errors in parentheses
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Exhibit 4: Mini Flash-Crash Count

5 Event Study Analysis

In addition to our regression analysis, we perform a series of intraday event studies on abnormal social-media sentiment. We use a universe of 500 high-capitalization stocks and match intraday trading data from WRDS to intraday message data from Twitter and StockTwits. We then classify messages into positive and negative events. An abnormal positive (negative) event is one that has a social-media sentiment score at least three standard deviations above (below) the average social-media sentiment for that stock. Our event window is the interval 10 minutes (600 seconds) before and after the event.

For each stock i in our event study and time $t \in (-600, 600)$, we compute the second-by-second return, $R_{i,t}$. Our null hypothesis is that $R_{i,t}$ is normally distributed with mean 0 and variance σ^2. Because of this choice of null, the abnormal return, AR_{it}, is just R_{it}. This implies that the average abnormal return $\mathrm{AR}_t = \frac{1}{N}\sum_{i=1}^{N} R_{i,t}$ at every second t is normally distributed with mean 0 and variance $\frac{1}{N}\sigma^2$, and that the cumulative average abnormal return, $\mathrm{CAR}(\tau_1, \tau_2) \equiv \sum_{t=\tau_1}^{\tau_2} \mathrm{AR}_t$, over a time interval $[\tau_1, \tau_2]$ is normally distributed with mean 0 and variance $\sigma^2(\tau_1, \tau_2) = \frac{\tau_2 - \tau_1 + 1}{N}\sigma^2$. To test our hypothesis, we use the estimator $\widehat{\sigma^2}(\tau_1, \tau_2) =$

$\frac{1}{N^2}\sum_{t=\tau_1}^{\tau_2}\sum_{i=1}^{N} R_{it}^2$, and compute the t-statistic, $t(\tau_1,\tau_2) = \frac{\mathrm{CAR}(\tau_1,\tau_2)}{\widehat{\sigma^2}(\tau_1,\tau_2)}$.

Our main finding from these event studies is that highly abnormal social-media sentiment is preceded by very high momentum and followed by mean-reverting returns. The results of this event study are presented in Exhibit 5a for positive sentiment, and in Exhibit 5b for negative sentiment. The blue lines show cumulative average returns, while the green lines show cumulative winsorized average returns, where outliers have been winsorized at the 95% level. The shaded areas represent 95% confidence intervals, computed via 1,000 bootstrap samples. From Exhibits 5a,b, we can see that social-media sentiment can be used to detect the peak of an intraday boom or the trough of an intraday panic.

In Exhibits 5c and d, we show the results of spread event studies using abnormal social-media sentiment. Again, the blue lines show cumulative average spreads, while the green lines show cumulative winsorized average spreads, where outliers have been winsorized at the 95% level. In contrast with our regression results, this event study shows that abnormal-sentiment events are followed in the next half-hour by a decrease in spreads, although this decrease is not statistically significant. Both our returns and spread event studies show that extreme sentiment is followed by an increase in market liquidity (mean reversion and lower spreads). This increase in liquidity may be due to increased retail investor attention and participation, or it might be due to institutional market makers reacting to extreme sentiment.

Exhibit 6 contains test statistics for these event studies, testing the null hypotheses that the average return and the average spread change around the event window are normally distributed and have mean zero. In our tests, we consider three periods: the pre-event window $[-600, 0]$, the post-event window $[0, 600]$, and the full event window $[-600, 600]$. The null hypothesis is not rejected for our spread event studies, but it is rejected for our returns event studies for the pre-event and full-event windows, confirming that mean reversion returns to the markets after abnormally high volumes of tweets.

6 Trading-Strategy Analysis

We use the insights from our regressions and event studies to construct an intraday trading strategy that uses StockTwits and Twitter messages. Since a high volume of social-media messages implies future mean reversion, lower spreads, and high liquidity, we use a mean-reversion strategy that increases the portfolio weights on stocks that have had high social-media sentiment.

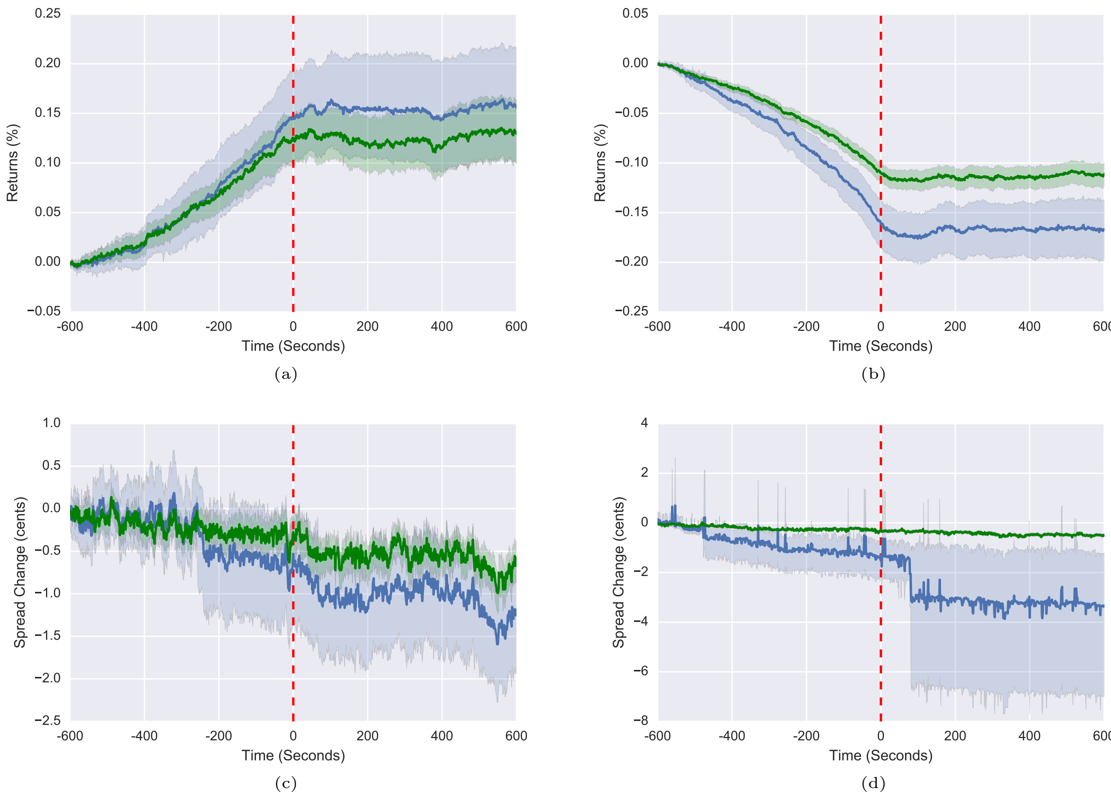

Exhibit 5: Return event studies for messages with sentiment score three standard deviations (a) above and (b) below the mean, and spread event studies for messages with sentiment score three standard deviations (c) above and (d) below the mean. The blue line represents average returns or spreads and the green line represents winsorized average returns or spreads. Confidence bands at the 95% level generated using 1,000 bootstrap samples.

Sentiment Direction	N	Window	CAR(%)	$\widehat{\sigma}$(%)	t-statistic
Hypothesis Tests for Returns Event Study					
Bullish	703	Pre-Event	0.146	0.031	4.736
Bullish	703	Post-Event	0.010	0.029	0.349
Bullish	703	Full Event	0.156	0.042	3.692
Bearish	7,097	Pre-Event	−0.161	0.018	−8.972
Bearish	7,097	Post-Event	−0.005	0.021	−0.256
Bearish	7,097	Full Event	−0.167	0.027	−6.071
Hypothesis Tests for Winsorized Returns Event Study					
Bullish	703	Pre-Event	0.124	0.023	5.452
Bullish	703	Post-Event	0.005	0.023	0.237
Bullish	703	Full Event	0.130	0.032	4.030
Bearish	7,097	Pre-Event	−0.110	0.008	−13.005
Bearish	7,097	Post-Event	−0.001	0.009	−0.107
Bearish	7,097	Full Event	−0.111	0.013	−8.839

Sentiment Direction	N	Window	CAR(cents)	$\widehat{\sigma}$(cents)	t-statistic
Hypothesis Tests for Spreads Event Study					
Bullish	1,040	Pre-Event	−0.531	1.523	−0.348
Bullish	1,040	Post-Event	−0.709	1.448	−0.489
Bullish	1,040	Full Event	−1.239	2.101	−0.590
Bearish	10,531	Pre-Event	−1.313	3.478	−0.378
Bearish	10,531	Post-Event	−2.021	4.493	−0.450
Bearish	10,531	Full Event	−3.334	5.681	−0.587
Hypothesis Tests for Winsorized Spreads Event Study					
Bullish	1,040	Pre-Event	−0.275	0.845	−0.326
Bullish	1,040	Post-Event	−0.389	0.879	−0.443
Bullish	1,040	Full Event	−0.664	1.218	−0.545
Bearish	10,531	Pre-Event	−0.269	0.296	−0.908
Bearish	10,531	Post-Event	−0.230	0.313	−0.735
Bearish	10,531	Full Event	−0.499	0.431	−1.158

Exhibit 6: Hypotheses Tests for Event Studies on Returns and Spreads. The null hypothesis that the average return is zero is rejected for the pre-event and full event windows, but not the post-event window. The null hypothesis that the change is zero around the event window is not rejected for any window.

We test our strategies using Quantopian (see `www.quantopian.com`). For any given month, our universe of stocks consists of the 500 companies with the largest average volume over the previous 200 days, with the constraint that no sector should contain more than 30% of our universe.[6] We consider the following two strategies:

1. A benchmark strategy, which trades every 30 minutes starting at 10:00am until the market closes. Every 30 minutes, we rebalance the portfolio as follows:

   (a) Close all open positions.

   (b) Let TOP be a set containing the 50 stocks in the top decile of our universe when ranked by returns in the previous 30-minute window. Let BOT be a set containing the 50 stocks in the bottom decile.

   (c) Set the weights $w_i = \begin{cases} -\frac{1}{50} \text{ if } i \in \text{TOP} \\ \frac{1}{50} \text{ if } i \in \text{BOT} \\ 0 \text{ otherwise.} \end{cases}$

   This strategy is market neutral and has a leverage ratio of $\sum_i |w_i| = 2$ in accordance with Regulation T.[7] We note that broker-dealers are not bound by Regulation T, and often take on more than 2:1 leverage.

2. A social-media-augmented strategy, which also trades every 30 minutes starting at 10:00am until the market closes. Every 30 minutes, we rebalance the portfolio as follows:

   (a) Close all open positions.

   (b) Let TWT be a set containing stocks which had more than 5 StockTwits and Twitter messages in the last 30 minutes, and where the number of messages in the last 30 minutes is above an exponentially weighted moving average. Let TOP be a set containing the 50 stocks in the top decile of our universe when ranked by returns in the previous 30-minute window. Let BOT be a set containing the 50 stocks in the bottom decile. The portfolio assigns weights so that the following conditions hold:

   - $w_i < 0$ for all $i \in \text{TOP}$, and $\sum_{i \in \text{TOP}} w_i = -1$,

[6] More concretely, we use the rules for the construction of the Quantopian 500, defined at `https://www.quantopian.com/help#quantopian_pipeline_filters_Q500US`.

[7] For a portfolio consisting of X dollars of long positions and X dollars of short positions, Regulation T requires X of initial capital, or 2:1 leverage.

- $w_i > 0$ for all $i \in \text{BOT}$, and $\sum_{i \in \text{BOT}} w_i = 1$,
- if $i \in \text{TWT}$ and $j \notin \text{TWT}$, then $|\frac{w_i}{w_j}| = 2$.

This strategy is market neutral, has a leverage ratio of 2:1, and assigns stocks that had a high number of tweets double the weight of stocks that did not have a high number of tweets. Essentially, the strategy "doubles down" on predicting that stocks with a high amount of social media messages will mean-revert in the next period.

We test these strategies from January 1, 2011 to December 31, 2014. Exhibit 7 reports the performance of the strategies during the entire period and individually for each year of the test period. All portfolios are initialized with $10 million in capital. To account for market impact, we take the following steps:

- When entering a position, our algorithm uses a limit order set at the current price. Buy orders must be filled at or below the current price. Sell orders must be filled at or above the current price.
- Orders are not filled until the second after they are placed.
- For any given stock and second, our algorithm cannot buy or sell more than the realized volume for that stock at that second. Orders that are only partially filled remain open and can be completed in subsequent periods.

We also evaluate a stricter algorithm that can only trade 10% of the volume of any given stock at any given second.

The strategy that doubles down on stocks with high message volume consistently outperforms the benchmark mean-reversion strategy. The annualized returns are 20.61% for the benchmark strategy and 24.10% for the social media strategy. When we restrict the algorithm to only being able to trade 10% of the volume at any given second, these annualized returns decrease to 12.64% and 14.77%, respectively (not shown in Exhibit 7).

While both of these strategies are profitable, we emphasize that, because of the frequency of trading, these profits are only achievable by market makers with minimal trading costs. Since a large number of our limit orders are being filled, sophisticated traders that follow this strategy could be a net liquidity provider, collecting fees that offset a portion of their trading costs.

Strategy	Period	Annualized Return	Alpha	Beta	Sharpe	Sortino	Max Drawdown	Volatility
Benchmark	2011–2014	20.61%	0.20	-0.03	1.30	1.89	-12.27%	0.15
	2011	35.46%	0.32	-0.00	1.70	2.53	-12.27%	0.18
	2012	5.75%	0.09	-0.16	0.46	0.67	-9.28%	0.14
	2013	35.74%	0.31	0.01	3.04	5.04	-4.48%	0.10
	2014	8.81%	0.09	-0.00	0.58	0.79	-11.22%	0.17
Social Media Strategy	2011–2014	24.10%	0.23	-0.04	1.47	2.16	-12.43%	0.16
	2011	36.33%	0.33	-0.01	1.71	2.55	-12.43%	0.19
	2012	6.84%	0.10	-0.18	0.53	0.78	-9.06%	0.14
	2013	43.76%	0.36	0.02	3.50	6.00	-4.29%	0.10
	2014	13.25%	0.14	-0.01	0.83	1.14	-10.74%	0.17

Exhibit 7: Performance Metrics for our Benchmark and Social Media Strategies. Positions are entered using limit orders at the current price. For any given stock and second, our algorithm cannot buy or sell more than the realized volume for that stock at that second. Orders that are only partially filled remain open and can be completed in subsequent periods.

7 Conclusion

In this article, we show that social-media activity can have a significant impact on liquidity at the intraday level, with negative sentiment having a much larger effect on liquidity measures than positive sentiment. This is consistent with the idea that panic and fear have much larger immediate effects on markets than booms and manias. In addition, event studies show that, at the intraday level, peak social-media sentiment corresponds to the end of momentum and a return to mean reversion. We used these results to construct a trading strategy that "doubles down" on predicting that stocks with high social-media sentiment will mean-revert, and show that this strategy outperforms a traditional intraday mean-reversion strategy.

The limitations of our study suggest several directions for future work. We did not address issues of causality, or the feedback effect between social media and markets. It is likely that sentiment on StockTwits and Twitter is driven by price movements, and plausible that this sentiment affects trades themselves, whether through users of these sites or through institutional investors who use the signals these sites provide. A key question is whether this feedback loop can amplify systemic risk, and what its role is in the triggering of flash crashes.

Another important open question is whether the social-network structure matters in how investors make decisions. Our analysis did not take into account the identity of users posting messages. It is likely that users who are more influential in the social network will have a higher impact on driving community sentiment, and through this sentiment, a higher impact on asset prices.

References

[1] Pablo D Azar and Andrew W Lo. 2016. "The wisdom of twitter crowds: Predicting stock market reactions to FOMC meetings via twitter feeds." *Journal of Portfolio Management*, 42(5):123.

[2] Elizabeth Blankespoor, Ed deHaan, and Christina Zhu. 2017. "Capital market effects of media synthesis and dissemination: evidence from robo-journalism." *Review of Accounting Studies*, 23(1):1–36.

[3] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. "Twitter mood predicts the stock market." *Journal of computational science*, 2(1):1–8.

[4] Hailiang Chen, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang. 2014. "Wisdom of crowds: The value of stock opinions transmitted through social media." *The Review of Financial Studies*, 27(5):1367–1403.

[5] Asher Curtis, Vernon Richardson, and Roy Schmardebeck. 2016. "Investor attention and the pricing of earnings news." In *Handbook of Sentiment Analysis in Finance*, edited by Gautam Mitra and Xiang Yu. Albury Books.

[6] Zhi Da, Joseph Engelberg, and Pengjie Gao. 2011. "In search of attention." *The Journal of Finance*, 66(5):1461–1499.

[7] Mark Dredze, Prabhanjan Kambadur, Gary Kazantsev, Gideon Mann, and Miles Osborne. 2016. "How twitter is changing the nature of financial news discovery." In *Proceedings of the Second International Workshop on Data Science for Macro-Modeling.* ACM.

[8] Qi Ge, Alexander Kurov, and Marketa Halova Wolfe. 2017. "Stock market reactions to presidential social media usage: Evidence from company-specific tweets." *Working Paper.* Available at SSRN: https://ssrn.com/abstract=2986350orhttp://dx.doi.org/10.2139/ssrn.2986350.

[9] Ahmet K Karagozoglu and Frank J Fabozzi. 2017. "Volatility wisdom of social media crowds." *The Journal of Portfolio Management*, 43(2):136–151.

[10] Michael Lachanski and Steven Pav. "Shy of the character limit: 'Twitter mood predicts the stock market' revisited." 2017. *Econ Journal Watch*, 14(3):302–345.

[11] Lian Fen Lee, Amy P Hutton, and Susan Shu. 2015. "The role of social media in the capital market: evidence from consumer product recalls." *Journal of Accounting Research*, 53(2):367–404.

[12] Huina Mao, Scott Counts, and Johan Bollen. 2011. "Predicting financial markets: Comparing survey, news, twitter and search engine data." *Working Paper.* Available at `https://arxiv.org/abs/1112.1051`.

[13] Eduardo J Ruiz, Vagelis Hristidis, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes. 2012. "Correlating financial time series with micro-blogging activity." In *Proceedings of the fifth ACM international conference on Web search and data mining*, 513–522. ACM.

[14] Hong Keel Sul, Alan R Dennis, and Lingyao Ivy Yuan. 2014. "Trading on twitter: The financial information content of emotion in social media." In *47th Hawaii International Conference on System Sciences (HICSS)*, 806–815. IEEE.

[15] Andrew Sun, Michael Lachanski, and Frank J Fabozzi. 2016. "Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction" *International Review of Financial Analysis*, 48:272–281.

[16] Xue Zhang, Hauke Fuehres, and Peter A. Gloor. 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.

A Regression Results

This Appendix presents the regression results when using Returns, Number of Trades, Number of Quotes, Number of Mini Flash-Crashes, and Dollar-Quoted Effective Spreads as our dependent variables.

Variable	(1) FE	(2) FE	(3) FE	(4) FE Predictive
News Sentiment	0.123***	0.111***	0.112***	0.000***
	(0.01)	(0.01)	(0.01)	(0.00)
Bullish Sentiment		0.032***	0.030***	-0.000
		(0.00)	(0.00)	(0.00)
Bearish Sentiment		-0.053***	-0.056***	-0.001***
		(0.00)	(0.00)	(0.00)
Number of Messages			0.002***	-0.000**
			(0.00)	(0.00)
Constant	-0.049***	-0.045***	-0.048***	-0.014***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	68,499	68,499	68,499	72,584
Number of tickers	4,543	4,543	4,543	4,366
Within R^2	0.0113	0.0217	0.0231	0.00268
Between R^2	0.00437	0.0107	0.00892	0.00922
R^2	0.0103	0.0202	0.0198	0.00425

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Exhibit A.1: Returns

Variable	(1) FE	(2) FE	(3) FE	(4) FE Predictive
News Sentiment	-1.066***	-0.967***	-0.712***	-0.004***
	(0.10)	(0.10)	(0.08)	(0.00)
Bullish Sentiment		1.030***	0.544***	0.093***
		(0.04)	(0.03)	(0.01)
Bearish Sentiment		1.750***	1.199***	0.105***
		(0.09)	(0.06)	(0.01)
Number of Messages			0.337***	0.004***
			(0.01)	(0.00)
Constant	9.141***	8.980***	8.225***	8.759***
	(0.04)	(0.04)	(0.04)	(0.04)
Observations	68,576	68,576	68,576	72,606
Number of tickers	4,544	4,544	4,544	4,370
Within R^2	0.00303	0.0366	0.272	0.0539
Between R^2	0.000633	0.155	0.272	0.116
R^2	2.47e-05	0.0769	0.292	0.0648

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Exhibit A.2: Number of Trades

Variable	(1) FE	(2) FE	(3) FE	(4) FE Predictive
News Sentiment	-0.373***	-0.312***	-0.200***	0.000
	(0.07)	(0.07)	(0.07)	(0.00)
Bullish Sentiment		0.486***	0.271***	0.039***
		(0.03)	(0.03)	(0.00)
Bearish Sentiment		0.923***	0.680***	0.052***
		(0.07)	(0.06)	(0.01)
Number of Messages			0.149***	0.002***
			(0.01)	(0.00)
Constant	11.644***	11.559***	11.226***	11.284***
	(0.03)	(0.03)	(0.03)	(0.03)
Observations	68,575	68,575	68,575	72,581
Number of tickers	4,544	4,544	4,544	4,370
Within R^2	0.000536	0.0131	0.0810	0.0181
Between R^2	0.00195	0.107	0.161	0.105
R^2	0.000814	0.0484	0.181	0.0449

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Exhibit A.3: Number of Quotes

Variable	(1) FE	(2) FE	(3) FE	(4) FE Predictive
News Sentiment	0.030***	0.034***	0.038***	0.000
	(0.01)	(0.01)	(0.01)	(0.00)
Bullish Sentiment		-0.006	-0.014**	-0.000
		(0.01)	(0.01)	(0.00)
Bearish Sentiment		0.018	0.007	-0.003***
		(0.02)	(0.02)	(0.00)
Number of Messages			0.006***	0.000***
			(0.00)	(0.00)
Constant	0.071***	0.070***	0.056***	0.069***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	68,571	68,571	68,571	72,605
Number of tickers	4,544	4,544	4,544	4,370
Within R^2	3.21e-05	5.18e-05	0.00153	0.00136
Between R^2	0.000551	0.000961	0.00301	6.36e-05
R^2	0.000163	0.000185	0.000283	0.00136

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Exhibit A.4: Dollar-Quoted Effective Spreads