Twitter activity, investor attention, and the diffusion of information

David Rakowski [a], Sara E. Shirley [b], Jeffrey R. Stark [b, *]

[a]*University of Texas at Arlington, College of Business, Arlington, TX 76019, USA*

[b]*Middle Tennessee State University, Jones College of Business, Murfreesboro, TN 37132, USA*

ABSTRACT

Measures of information traditionally arise from curated, professional sources such as newspapers, analyst coverage, earnings announcements, and business news wires. We utilize Twitter activity to examine the impact of attention generated by individuals. Causal evidence shows that Twitter activity has a direct impact on trading volume in financial markets. Furthermore, increases in Twitter activity are associated with positive abnormal returns and, when occurring in conjunction with traditional information supply events, increase the diffusion of information to investors. Our results identify conditions under which attention generated by individuals drives price discovery and trading activity.

JEL classification: G10, G12, G14, G23

Keywords: Twitter, Investor attention, Social media, Behavioral finance, Asset pricing, Spread of information

Funding: This work did not receive any grant from funding agencies.

[*] Corresponding author. Tel.: 615-494-8715, jeff.stark@mtsu.edu, MTSU Box 27 – BAS N324, 1301 East Main St., Murfreesboro, TN 37132.

Introduction

> "These websites [Facebook and Twitter] can provide perhaps the most powerful mechanisms available to a private citizen to make his or her voice heard."
>
> – US Supreme Court associate justice Anthony M. Kennedy
>
> *Packingham v. North Carolina* 582 US 8 (2017)

To have an impact on asset prices, information must be viewed by investors. While traditional, professional sources of attention have been studied at great length (Barber and Odean, 2008; Fang and Peress, 2009; Ben-Rephael et al., 2017; Ben-Rephael et al., 2018; Frijns and Huynh, 2018), technological advances are transforming the environment for sharing and viewing information. With the advent of social media, sources of information are shifting away from curated, top-down providers to a more democratized setting in which an individual can share almost anything using little more than an Internet connection.[1] The emergence of crowd-sourced social media and news platforms has transformed technologies once intended for social communication into potential channels for price discovery in financial markets. We examine the conditions under which activity on the social media platform Twitter impacts investor trading in modern US stock markets.

One of the primary challenges of working with any measure of social media attention is to disentangle its impact from that of traditional sources of investor attention. We attempt to overcome this hurdle through a series of tests that provide evidence of the direct impact Twitter

[1] Cogent Research LLC in 2013 polled four thousand investors with more than $100,000 in investable assets and found that 34% used social media for investment decisions, 36% said that social media research had shaped the questions they asked their financial advisers, and 70% changed some aspect of their investment process because of news they read on social media. (See http://www.businesswire.com/news/home/20130222005037/en/Cogent-Research-Trending---Social-Media-Fuels.)

has on the financial markets. Using data from Twitter, we identify instances in which individuals increase or decrease attention paid to a particular security and estimate the impact these changes have on a stock's trading activity.

We first utilize a quasi-natural experiment in which we capture unexpected Twitter outages. We proxy for Twitter outages through daily Google search volume for the phrase "Twitter down." Our results show that Twitter outages are unrelated to past stock market characteristics, confirming outages as an exogenous shock to the attention generated through Twitter activity. We then show that outages of Twitter significantly reduce trading but have no effect on other sources of information such as news coverage or earnings announcements. Our results provide the first causal evidence that Twitter activity has a significant impact on daily trading volume, thus showing that Twitter plays a role in financial markets.

Next, we utilize the price impact from investors' Twitter-motivated trading to identify the channels through which Twitter impacts financial markets: directly on investor trading and stock prices and indirectly though the diffusion of the information contained in traditional information sources. We find that a one standard deviation increase in Twitter activity leads to a 9.6 basis point (bps) characteristic-adjusted daily excess return, a result that increases to 24.3 bps for small, less visible securities.[2] By exploring this relation further, we show that Twitter activity on its own results in a short-term return spike followed by a partial reversal, whereas increased Twitter activity occurring in conjunction with a traditional information event, such as a news release or earnings announcement, results in a large and lasting additional daily excess return between 13.9 and 37.7 bps.

[2] As explained in Section 2.2, we measure excess returns as the difference between the security's returns and the average daily return of a size and book-to-market matched portfolio based on 5 x 5 sorts.

Next, we make novel use of our data by simultaneously identifying changes in the supply of information (Twitter activity) and the consumption of the information (retweet activity). This dual channel allows us to determine how information supply and consumption interact. We show that the consumption of information increases the magnitude of the price impact, particularly among small stocks (Da et al., 2011; Ben-Rephael et al., 2017). The effect of information consumption is approximately twice the magnitude of information supply, and the impact of a simultaneous increase in both consumption and supply is nearly three times larger than the magnitude of the daily effect relative to a change in information supply alone. Finally, we examine the differences between Twitter activity and traditional measures of investor attention and information, providing evidence that Twitter differs from attention measures previously studied.

To determine whether it is retail or institutional investors who respond to Twitter activity, we evaluate how two hypotheses explain our observed effects. The investor recognition hypothesis (Merton, 1987) posits that the Twitter effect will be largest in stocks that are least visible to investors. As Twitter attention increases, stock visibility and buying pressure grow, driving prices upward.[3] The impediments-to-trade hypothesis (Fang and Peress, 2009) says that the Twitter effect will be largest in illiquid and difficult-to-arbitrage stocks because large institutional investors are not present to take the other side of uninformed trades, leading to temporary deviations from fundamental values. Our findings are consistent with both hypotheses, indicating that Twitter activity is primarily a measure of retail investor attention. We then show that our Twitter-based

[3] Consistent with the Barber and Odean (2008) attention hypothesis, when attention is generated about a stock, potential new investors become aware of that stock for the first time. Because retail investors are net buyers, and observations about a stock could make investors aware of it for the first time, buying pressure increases. An investor can sell a stock only after owning it. This underlying mechanism results in attention growth leading to an increase in returns regardless of the sentiment or content of the new attention.

excess returns cannot be explained by traditional risk factors, and that our results are consistent across time-periods.

1. Literature and contribution

Research into social media and asset pricing gained popularity in 2004 with the examination of Yahoo! Finance message board activity by Antweiler and Frank (2004). In this paper the authors explored message board sentiment during the year 2000 and found no meaningful relationship with stock returns. However, the lack of a significant relation with stock returns may be the result of the internet becoming mainstream only just before the time period examined, and the relative infancy of the social media aspect of the internet. This led to an exploration of the impact of spam e-mails on stock returns and turnover (Hanke and Hauser, 2008). Then in 2014 Chen et al. (2014), provided evidence that reports and commentary on the popular crowd-sourced analyst site, Seeking Alpha, help to explain future stock returns and earnings surprises.

Our paper differs from these studies in multiple ways. First, due to the popularity of the internet and social media, and widespread internet availability, examining Twitter activity over the period 2011 – 2015 provides a more conducive environment for this type of analysis. Second, sites such as Seeking Alpha encourage well thought-out analysis and stock valuation that must then be approved by an editor prior to becoming publicly available. In contrast to the immediate and social nature of Twitter, sites like Seeking Alpha do not capture investor attention in the short-term. Although early studies of social media message board interaction and asset prices exist, the environment and conditions of their analyses are distinctly different from that of our current study.

More recently, studies have started to take advantage of the immense growth in social media use by exploring various aspects of Twitter activity, exploring the use of social media for the dissemination, guidance, or manipulation of firm financial information (Guindy, 2017; Renault, 2017). Papers such as Liew and Wang (2016) show that for a sample of 325 stock IPOs, Twitter sentiment helps explain the contemporaneous first day trading returns, Bartov et al. (2017) examine how Twitter sentiment around earnings announcements predict returns, and Renault (2018) shows that StockTwits sentiment can explain broad index returns at 30-minute horizons. In the paper most closely related to ours, Sprenger et al. (2014) find evidence that Twitter activity is associated with trading volume but not returns. Whereas the Sprenger et al. paper provides evidence of an association between tweeting and trading, our paper differs, in that we provide the first causal evidence in support of their trading volume result. The most likely cause for the difference between the return results in Sprenger et al. (2014) and our paper is the samples examined. Our sample covers over 21 million tweets from over 2.2 million tweet-stock days, encompassing all Russell 3000 stocks over a 5-year period, whereas the Sprenger et al. (2014) study focuses on 250,000 tweets on S&P100 stocks over a 6-month period in 2010. Results from Sprenger et al., therefore, apply only to large-cap stocks, while our sample provides a more generalizable analysis of stocks from firms of different size.

Our paper contributes to the literature in several ways. In the presence of an information event such as news coverage or an earnings announcement, we show that increases in Twitter activity improve the diffusion of information. Fang and Peress (2009) and Peress (2014) detail how mass media coverage significantly affects stock returns. Engelberg et al. (2017) find that news coverage is generally associated with a positive contemporaneous stock return, and Ben-Rephael et al. (2018) show how earnings announcements and news coverage are positively associated with same-day returns. In our analysis, we combine the foci of these papers to demonstrate how social media activity interacts with other instances of increased information supply. We show that Twitter activity assists in spreading the information contained within traditional information events to a wider range of investors, leading to larger and more lasting price responses.

While the literature on Twitter's impact on asset prices continues to grow, the approach and methodology utilized in our paper differs from prior studies. First, to the best of our knowledge, this study is the largest and most comprehensive study to date of the impact of Twitter activity on individual stock prices. Second, our paper focuses on the intensity of Twitter activity rather than the direction of Twitter sentiment, the specific content of the tweet, the type of user creating the tweet, or whether a tweet is spam. In doing so, we focus on a pure measure of investor attention, rather than trying to extract some form of underlying information from the tweet. We thus allow for a broad analysis of how changes in social media attention impact financial markets through trading activity. Our approach has more in common with past research on the factors that measure investor attention, such as the studies by Barber and Odean (2008), Fang and Peress (2009), Peress (2014), Da et al. (2011, 2015), and Vozlyublennaia (2014) than many of the papers looking at Twitter sentiment. Finally, while prior studies document strong correlations between Twitter and market activity, they generally do not differentiate social media attention from other sources of information. Through a quasi-natural exogenous shock to Twitter availability, we provide causal evidence of the impact social media attention has on the trading volume of financial markets.

2. Data, sample, and hypotheses

In this section, we describe Twitter, cashtags, and our measure of the change in daily Twitter activity. We describe our data sources and sample creation process, we present descriptive statistics and develop our hypotheses.

2.1. Twitter

Twitter is a social media platform in which users post short messages (140 characters or fewer), known as tweets.[4] Since its founding in 2006, Twitter has grown rapidly; increasing from 70 million users in 2011 to more than 300 million active users by the end of 2015 (see Appendix Figure A.I for annual growth). As of 2014, approximately 350,000 tweets were posted every minute.[5] With such a large and active userbase, Twitter is an ideal mechanism to disseminate information about any topic, including financial securities. Furthermore, the design, tagging system, and searchability of Twitter allows tweets to be used as a measure of investor attention that is directly tied to a particular stock or topic.

We obtain Twitter data on individual securities at daily frequencies over the period 2011 through 2015.[6] To accurately assign tweets to the correct day, we utilize Coordinated Universal Time (UTC), which allows us to account for different time zones, as well as daylight saving time.[7] Twitter utilizes a hashtag system known as "cashtags" to help users more easily identify tweets referencing financial securities and avoid having to sift through millions of tweets for relevant information. With cashtags, users place a dollar sign ($) before a ticker, such as $AAPL, relaying

[4] On November 7, 2017 the character limit was increased to 280 for all languages except Chinese, Japanese, and Korean (https://www.theverge.com/2017/11/7/16616076/twitter-280-characters-global-rollout).

[5] See about.twitter.com for the number of active users, and see https://www.dsayce.com/social-media/tweets-day/ for number of tweets sent per minute.

[6] We thank PsychSignal for providing us with its Twitter and retweets data. The tweets (and retweets) are aggregated from the entire Twitter universe, not tied directly to the users creating them. In this sense, our data are truly a measure of crowd-sourced attention, as the attention creation is not attributed to any one Twitter user.

[7] We explore alternative timing setups in robustness tests with similar outcomes. We detail why we chose calendar timing as our primary measure in Section 3.4.

that the tweet is about Apple stock.[8] Figure 1 is an example of an anonymized tweet containing the $AAPL cashtag.

[Insert Figure 1 near here]

Because specific companies can be found through cashtags, consumers of information on Twitter can more easily sift through an enormous amount of data and focus on those tweets pertaining to financial securities, thus helping to reduce the information overload problem (Bawden and Robinson, 2008). This unique feature identifies a tweet, for example, as being about Mednax, Inc., with its ticker "MD," and not the state of Maryland, a medical doctor, or something else entirely. Our Twitter data contain the company referenced in the tweet, the date of the tweet, and the aggregate number of tweets for each security on each day during our sample period.

To construct our measure of increased or decreased attention resulting from tweets, we calculate the daily change in tweets in a manner similar to how Da et al. (2015) calculate abnormal Google Search Volume (GSV). We define our daily change in tweets as

$$\Delta Tweets_{i,t} = Ln\big(Number\ of\ Tweets_{i,t} + 1\big) - Ln\big(Number\ of\ Tweets_{i,t-1} + 1\big), \quad (1)$$

where $Ln(Number\ of\ Tweets_{i,t} + 1)$ is the natural log of one plus the number of tweets about security i on day t. We deseasonalize each series by regressing the daily change in tweets on day-of-the-week and month-of-the-year dummies. The resulting residual represents the deseasonalized daily

[8] Twitter users utilize cashtags throughout our sample period. However, Twitter officially announced the ability to click on cashtags on July 30, 2012. In unreported robustness tests, we confirm that our results do not change based on this cutoff date.

change in tweets, or the daily change in the supply of information. We construct a measure of the daily change in retweets with the same method.

Our subsequent results are robust to replacing the change in tweets with the number of tweets, when utilizing a dummy variable for Twitter attention shocks in the same manner as abnormal institutional attention (AIA) in Ben-Rephael et al. (2017), without deseasonalizing our daily change in tweets, by standardizing the change in tweets by the time series standard deviation of tweets (Da et al., 2015), and when calculating the daily change relative to day *t* - 2. [9]

2.2. Data, sample creation, and descriptive statistics

From our sample of all securities mentioned in tweets with cashtags, we require valid observations of several additional control variables. We limit our sample to only those securities that belong to the Russell 3000 index at some point over our sample period. We do not require a security's inclusion in our sample to match with the period it was listed in the Russell 3000 index, only that the security appear in the index at any time between 2011 and 2015. We further require that a security contain tweets on at least 10% of its observations. [10]

We obtain daily security-level data from the Center for Research in Security Prices (CRSP) database, including market capitalization, trading volume, and returns. From stock returns, we calculate characteristic-adjusted excess returns by ranking securities into quintile portfolios based on market capitalization and book-to-market value, resulting in 25 characteristic-adjusted

[9] In our main measure of change in tweets, we sum the number of tweets across days the market is closed and aggregate them into one daily count. The fact that Monday is on average an aggregation of tweets across Saturday, Sunday, and Monday is addressed through the day-of-the-week dummy. We also calculate the change in tweets across days when the market is open (for example, a standard week is the change in tweets between Friday and Monday) and across every day, regardless of whether the market is open (for example, the change between Sunday and Monday for our Monday variable). These alternative measures result in stronger findings.

[10] We repeat our tests with cutoffs ranging from 25% to no threshold, and find results consistent with those reported.

portfolios (similar to Daniel et al., 1997).[11] We reconstitute these portfolios daily on the basis of their prior month (day t – 22) characteristics and calculate excess returns as the daily stock return minus the matched portfolio's daily return.

The Institutional Bankers' Estimate System (I/B/E/S) provides the number of analysts covering each security and the earnings announcement dates. Compustat offers information on advertising expenses, sales, and dividends. Our measure of news coverage comes from Bloomberg Professional and is the daily count of the number of news stories about a stock. For a security to remain in our sample, it must contain all relevant variables of interest, have at least 24 months of observations, and have a price greater than $5. The exception is advertising expenses, for which we fill missing observations with a value of zero.

We measure institutional and retail investor demand for information as in Ben-Rephael et al. (2017). Institutional investor attention comes from Bloomberg queries about a stock and takes a value of one if the daily institutional interest is in the top 6% of attention measures over the prior 30 trading days and zero otherwise. Retail investor attention comes from Google Trends data for daily search volume about a stock and takes on a value of one if the daily search volume is in the top 6% of attention measures over the prior 30 trading days and zero otherwise.[12] We define all variables used throughout our analyses in Appendix Table A.1. The final sample contains 1,976 unique securities and 2,215,535 security-day observations from more than 21 million tweets between 2011 and 2015.

[11] To avoid extreme outliers, we winsorize our daily returns at the top and bottom 0.01%. To avoid the impact of nonsynchronous trading, we require that each stock within our sample trade on all days that the stock is within our sample.

[12] More detailed discussions of these variables can be found in Ben-Rephael et al. (2017). We thank the authors for providing us with the daily Google search volume data used in their paper.

We calculate descriptive statistics by taking the time series average values for each security in our sample and averaging these values across securities. Table 1 shows that firms in our sample, on average, receive 9.50 daily tweets, tally $7.89 billion in market capitalization, are covered by 7.14 analysts, and have a share price of $42.54. We divide our sample into securities with above and below median daily tweet counts to explore sample differences further. Securities with above median tweets average 17.13 daily tweets; below median tweets, just 1.88. Above median securities are also larger ($14.35 billion versus $1.43 billion) and have greater analyst coverage (9.96 analysts versus 4.31 analysts). These statistics indicate that larger, more visible securities garner greater attention, when measured by Twitter activity. The univariate relations hold under a multivariate framework presented in Appendix Table A.2.

[Insert Table 1 near here]

2.3. Motivations and hypotheses

Our hypotheses focus on how Twitter attention impacts financial markets, if Twitter activity is consistent with retail or institutional investors, and how Twitter attention reduces informational frictions by expanding the reach of traditional information events.

2.3.1. Measures of investor attention

Direct measures of investor attention are challenging to find (Da et al., 2011). Some studies use stock characteristics to indirectly infer levels of attention from investors (Odean, 1999; Barber and Odean, 2008), while others measure attention from mentions in the mass media (Fang and Peress, 2009), or Internet search activity (Da et al., 2011). The attention that investors pay to specific stocks is a major factor driving investor trading and short-term price pressure in financial markets (Baker and Wurgler, 2006; Barber and Odean, 2008).

Our first two hypotheses about investor attention are derived from the work of Barber and Odean (2008). The authors argue that, due to cognitive limitations, investors often purchase securities based on attention-driven psychological factors. We first hypothesize that an increase in Twitter attention increases trading activity, and second, that the resulting imbalance in trading activity is reflected in increased short-term returns.

While the supply of information has been shown to affect stock prices (Engelberg et al., 2017), we attempt to move further and identify if the reception, or "consumption", of information has an association with trading activity that is discernable from that of information supply. We utilize retweets as a proxy for the consumption of new information from Twitter activity. For a retweet to occur, a Twitter user must view the original tweet and then decide that it is worth sharing with others. Many users view the information within a tweet without choosing to share that information via a retweet. In this regard, retweets underestimate the true level of information consumption and therefore bias against finding strong results when using retweets. For our third hypothesis, we predict that when a stock experiences an increase in both the supply of information (an increase in Twitter activity) and the consumption of new information (an increase in retweeting activity), the market impact will be larger in magnitude, compared to the supply of information alone.

2.3.2. Stock characteristics and the impact of Twitter attention

Ben-Rephael et al. (2017), Ben-Rephael et al. (2018), Da et al. (2011), and Brennan et al. (1993) estimate the varying impact of retail investor and institutional investor attention on stocks.

Our fourth hypothesis focuses on which type of investor Twitter activity most closely proxies for. We utilize a theory on investor recognition (Merton, 1987) and one on impediments of trade (Fang and Peress, 2009) to better understand which investors drive our Twitter effect. We use these two established theories to disentangle the effect of retail and institutional investors.

2.3.3. Increasing the diffusion of informative events

Fang and Peress (2009) show that news coverage can alleviate informational asymmetries, particularly among smaller stocks; Peress (2014) shows that the media helps disseminate information to a larger number of investors; and Ben-Rephael et al. (2017) find a magnifying impact on the spread of information from increased institutional investor attention. Therefore, our final hypothesis is that in the presence of an event adding information to the market, an increase in Twitter attention will magnify awareness of the new information and result in a larger and more lasting price reaction. We utilize information supply events from Ben-Rephael et al. (2017), i.e., news coverage and earnings announcements, to explore this association.

3. Asset trading, pricing, and investor attention

We explore how changes in the supply of information provided through Twitter impact trading and return characteristics. We then look at the impact of information consumption. This section concludes with tests estimating the direct impact Twitter has on the financial markets.

3.1. The causal impact of Twitter attention on trading volume

One challenge when working with social media data such as Twitter is the difficulty of disentangling the impact of Twitter from that of other information sources. Following the approach of Peress (2014) with newspaper strikes, we provide causal evidence of Twitter's impact on financial markets through a quasi-natural experiment in which we observe exogenous Twitter outages. We proxy for market-wide (rather than stock-specific) Twitter outages through daily Google search volume for the phrase "Twitter down." By capturing when individuals search for "Twitter down," we identify days when individual's experience difficulty accessing Twitter and when we expect a reduction in Twitter activity. Whereas the newspaper strikes examined by Peress (2014) provide clear outages, complete outages of Twitter are extremely rare, oftentimes only lasting for a fraction of the day. Although we do not know the underlying cause or exact severity of the outages, we establish a proxy for severity through a greater number of Google searches, with greater search volume representing more severe outages. While directly testing the exclusion restriction is not possible here, the ability to conduct a Google search does preclude the possibility that users experience an internet outage, providing support that a particular outage is unique to Twitter.

Before we explore the causal association between Twitter and investor actions, we first show that Twitter outages are unrelated to past market characteristics. Though logically unlikely, if Twitter outages follow periods of large volume or high excess returns, the exogenous nature of our outages would be unclear. To test whether stock market characteristics occur before Twitter outages, we follow the lead-lag methodology detailed in Bertrand and Mullainathan (2003). We determine how lag values of volume traded and stock returns explain future Twitter outages through the multivariate regression:

$$Tweets_{i,t} = \alpha + \beta_z Volume_{i,t-n} + \beta_w Excess\ Returns_{i,t-n} + \sum_{j=11}^{m} \beta_j Controls_{i,t} + FEs + \varepsilon_{i,t}\ , (2)$$

where we measure volume traded (*Volume*) and excess returns (*Excess Returns*) for stock *i* on day *t* while varying values of *n* from 5 to 1. We include day of the week and stock level fixed effects (*FEs*) and standard errors are clustered across each day to control for within day correlations. Control variables for each stock *i* include a measure of stock size (*Size*), daily news coverage (*News Coverage*), scaled advertising expenditures (*Advertising/Sales*), an indicator for dividend-paying firms (*Dividend Paying*), analyst coverage (*Ln(Analysts)*), an indicator for earnings announcement days (*EDay*), volatility of stock returns (*Risk of Returns*), retail (*DADSVI*) and institutional (*DAIA*) investor demand, and the absolute value of excess returns on days $t-1$ through $t-5$ to account for investor preference for prior performance (Glaser and Weber, 2009) and autocorrelation in daily returns (*AER*), as defined in Appendix Table A.1. This model allows us to determine if trading volume or excess returns occurring before day $t = 0$ result in an increase in search activity for "Twitter Down."

Results presented in Table 2 provide clear evidence that prior trading and return activity have no economically meaningful impact on Twitter outages. No measures of prior trading volume are significantly different than zero. The results for excess return do show significance, although the coefficients are so small in magnitude that the economic interpretation is that a one standard deviation increase in the prior day excess returns decreases Twitter down search volume by 0.13%. We attribute the observed statistical significance to the over 2.2 million observations in our dataset.

[Insert Table 2 near here]

Having established the exogenous nature of our Twitter outage proxy, we now standardize *Twitter Down* to have a mean of zero and a standard deviation of one, with larger values representing greater search activity. We estimate the following model:

$$Event_{i,t} = \alpha + \beta_1 TwitterDown_t + \sum_{j=2}^{m} \beta_j Controls_{i,t} + FEs + \varepsilon_{i,t}\ , (3)$$

where *Event* measures *Ln(Dollar Trades)* in columns 1 and 2, *Volume* in columns 3 and 4, and *Excess Returns* in columns 5 and 6. We include day of the week and stock level fixed effects (*FEs*) and standard errors are clustered across each day. We include controls as defined and calculated in Equation (3) and described in Appendix Table A.1.

We present causal evidence of the impact of Twitter attention on the financial markets in Table 3. Whereas Twitter activity is potentially endogenous to stock and market level factors, Twitter outages provide an exogenous shock only impacting Twitter. By utilizing Google search volume for our Twitter outages proxy, we further ensure that the outage is Twitter related rather than an internet outage, as a Google search requires power and internet access. Columns 1 through 4 provide causal evidence that Twitter activity impacts trading behavior in a similar same way to how newspaper strikes impact trading volume (Peress, 2014), although through a different channel. We observe a coefficient of -0.015 for log dollar volume traded and of -0.022 for log volume. These coefficients translate into a 1.0% reduction in dollar volume traded and a 1.5% reduction in total volume for a one standard deviation increase in *Twitter Down* search volume. Columns 5 and 6 present the causal impact on stock returns and show a negative and significant coefficient. However, the coefficient on returns is statistically, but not economically meaningful, at just 0.4 bps. The results presented in Table 3 mirror those in Peress (2014), support our first hypothesis, and provide direct evidence of the causal impact of Twitter attention on the trading activities of investors, though stop short of providing economically meaningful causal evidence of the impact on stock returns.

[Insert Table 3 near here]

3.2. Asset pricing and Twitter

Barber and Odean (2008), Ben-Rephael et al. (2017), Da et al. (2011), and Fang and Peress (2009) study the impact of investor attention on asset prices through measures of stock characteristics, institutional investor attention, retail investor attention, and mass media coverage. We now build on this line of study and explore the ability of changes in Twitter activity to explain investor attention, expecting an increase in tweets about a company to result in an increase in buying pressure and excess returns, estimated through the following regression:

$$Excess\ Returns_{i,t+n} = \alpha + \beta_1 \Delta Tweets_{i,t} + \sum_{j=2}^{m} \beta_j Controls_{i,t} + FEs + \varepsilon_{i,t}\ , (4)$$

where $Excess\ Returns_{i,t+n}$ is the daily value of a characteristic-adjusted excess return measured in bps for stock *i,* on day *t,* with varying values of *n*. We calculate this measure of return over days *t*, $t+1$, $t+2$, $t+3$, and the cumulative return over days $t+4$ through $t+40$. Our variable of interest, $\Delta Tweets_{i,t}$, measures the daily change in tweets as in Equation (1). We include variables defined in Equation 2 as controls and unless otherwise stated, we model control variables as defined in Appendix Table A.1. We cluster standard errors each day and include day of week and stock level fixed effects (*FEs*).

Columns 1, 3, 5, and 7 of Table 4, Panel A, show the isolated impact of a change in Twitter attention. Columns 2, 4, 6, 8, and 9 include control variables. When a stock experiences a one standard deviation increase in the supply of information, daily excess returns increase by 9.6 bps. This increase continues through the next day, followed by a small longer-term reversal.[13]

Da et al. (2011) show that the impact of GSV is concentrated among small stocks. Therefore, we include an interaction term between *ΔTweets* and *Size*, in Panel B. After accounting for the effect of stock size on this relation, we find an increase of 24.3 bps for a one standard deviation increase in daily Twitter activity. The effect decreases by about 60% for every standard deviation increase in stock size. Like the results for GSV on stock returns, the Twitter effect is economically largest for smaller, less visible securities. We find a slightly reduced, but economically meaningful, impact across all stock sizes.

Moving forward one day, the remaining effect is fully concentrated among small stocks. A one standard deviation increase in Twitter activity results in a day 1 return of 9.8 bps, and the offsetting size effect amounts to a 9.4 basis point reduction. This size effect is similar to what we observe on day 3. Small stocks experience a limited reversal amounting to just 4.5 bps. Our results provide support for our second hypothesis.

[Insert Table 4 near here]

We next consider that supplying information about stocks is not a guarantee that information will have a meaningful impact on asset prices. Investors must read and acknowledge a tweet before the tweet can meaningfully be said to generate attention. While it is not possible to

[13] We directly examine the precise timing of tweets relative to trading hours in Subsection 3.4.

observe the consumption of information with many traditional measures of investor attention, one unique feature of Twitter that addresses this limitation is the retweet. A retweet occurs when a user shares an original tweet instead of creating new content on her own. For a retweet to occur, an individual must view the original tweet and determine that the content warrants further recognition. Because many Twitter users read a tweet without retweeting it, this measure provides a lower bound for the consumption of stock information supplied on Twitter. In Panel C of Table 4, we include variables measuring the daily change in retweets (*ΔRetweets*) and interaction terms between changes in retweets and stock size and between changes in retweets and changes in tweets.

Consistent with our third hypothesis and the finding that the consumption of information is strongly associated with stock returns (Ben-Rephael et al., 2017), increased retweeting has a significant and economically meaningful association with returns. The inclusion of both changes in tweets and retweets shows a large impact from an increase in Twitter activity (18.0 bps) as well as retweeting (37.1 bps). When a stock experiences an increase in both simultaneously, we observe an additional increase of 51.0 bps (totaling a same day increase of 106.1 bps). Similar to our earlier results, the effect is driven by small, less visible stocks. A smaller increase appears during the following day across all three variables, though the day 3 reversal is confined to Twitter activity. When a stock experiences an influx in the supply of information that is simultaneously consumed, the price impact is larger and more long-term compared with when information is only supplied.

3.3 Twitter coverage and unique information

We consider three final tests to show that attention measured through Twitter activity provides information beyond traditional measures of investor attention. First, we evaluate Twitter by comparing it with other widely used measures of investor attention: daily trading volume, excess returns, absolute value of excess returns, daily news coverage, retail investor demand for information, and institutional investor demand for information. We measure these associations by calculating the correlations across securities and present the results in Appendix Table A.3.

All measures of daily attention are positively correlated with each other, though at relatively low levels. Tweets have statistically significant correlations with traditional measures of attention that range from a low of 1.7% with retail investor demand for information to a high of 12.4% with institutional demand for information and 11.9% with news coverage. These correlations indicate that changes in Twitter activity are positively related to other measures of attention but still contain unique information.[14]

Next, we look at how well alternative measures of investor attention explain our measure of *ΔTweets*. Appendix Table A.4 shows that an increase in Twitter activity is positively related to alternative measures of the supply of information (*News Coverage* and *EDay*), to performance, and to demand for information measures (*DADSVI* and *DAIA*). Of the models presented, we observe R^2 values no larger than 3.7%, indicating that alternative measures of investor attention explain little of the variation in our measure of daily changes in Twitter attention.

We further distinguish between measures of attention in Appendix Table A.5. To isolate the Twitter effect, we orthogonalize *ΔTweets* with respect to other information events by regressing *ΔTweets* on *News Coverage* and *EDay* and retain the residual as our new variable of interest. We thus remove the confounding impact of other information events, similar to how Da et al. (2014) remove the confounding season and day-of-the-week effects from Google search data.

[14] To further differentiate between news coverage and Twitter activity, we calculate the percent of days containing each attention measure. We observe news coverage on 37% of days in our sample and Twitter activity on nearly 70% of days.

Appendix Table A.5 confirms two results: first, the original findings from Table 4 hold, and second, Twitter measures investor attention above and beyond the effects of news and earnings announcements. We also observe a positive significant coefficient on *News Coverage* and on *EDay* after the orthogonalization of *ΔTweets*. Results from these three additional tests provide greater evidence that Twitter activity provides a unique source of attention that differs from traditional news-creating events. Although utilizing our orthogonalized *ΔTweets* variable more clearly disentangles the effects attributed to attention changes on Twitter versus other information-rich events, it also further distances our measure of Twitter activity from what everyday investors observe directly on the Twitter platform. For that reason, we elect to use our orthogonalized measure of Twitter activity as robustness rather than as our primary variable of interest.

3.4. Robustness tests

Guindy (2017) shows how the sentiment of Twitter activity from official corporate Twitter accounts has an impact on stock returns only after the April 2, 2013 adoption of a Securities and Exchange Commission (SEC) regulation allowing firms to use social media as a news outlet. Utilizing this date as a natural inflection point, we confirm our results hold over both periods. Ben-Rephael et al. (2018) use a news day dummy variable as their measure of news coverage. We replace our news variable with an indicator variable and find similar results.

We report results primarily for the change in log Twitter activity because a focus on the levels of Twitter activity is likely to lead to inaccurate inferences when a small number of high-activity firms dominate the estimation of model parameters. However, we do examine the robustness of our results for a variety of alternative measures of Twitter activity, including the level of Twitter activity and lagged Twitter activity. Appendix Table A.6 reports the results of a replication of Table 4 including the natural log of one plus the number of daily tweets and five lags of change in Twitter activity included as additional control variables. We show a significant positive coefficient estimate on the term for the level of tweets, but, importantly, our findings for changes in Twitter activity remain significant and qualitatively unchanged.

We test the association between increased Twitter attention and stock returns across alternative samples. We examine the relation across industries (Appendix Table A.7) and find that the same pattern of returns holds for stocks in all major industry categories. Appendix Table A.8 considers a sample of Russell 1000 and Russell 2000 stocks separately to further explore the impact of size and shows results generally consistent with our observed size effect. Stocks at the small end of the Russell 1000 index experience the greatest impact. This indicates that stocks of a reasonable size with lower levels of institutional ownership experience the largest impact from Twitter attention. We explore this in more detail in Section 4.

Next, we examine how the timing of tweets throughout the day affects this relation. Our main tests follow the variable timing structure defined in Ben-Rephael et al. (2017) which allows for measurement of our Twitter variable to coincide with the timing of variables such as Google Search Volume (*DADSVI*) and Institutional demand for information (*DAIA*). Although tweets do occur after the market closes, the majority of cashtagged tweets are posted during the trading hours of 9:30 a.m. Eastern Standard Time (EST) and 4:00 p.m. EST (Appendix Figure A.2). For robustness, we reassign Twitter activity on the basis of trading-day hours instead of calendar time. All tweets occurring after 4:00 p.m. EST are assigned to the next trading day. Under this new timing scheme, our results remain unchanged (Appendix Table A.9). In Appendix Table A.10, we further address how after-market Twitter activity affects our results by dropping all tweets posted when the markets are closed. The same relation exists as in Table 5. These results show that after-hours activity does not drive our observed Twitter effect.

4. The source of the Twitter effect

Recent research on investor attention focuses on retail investor attention measured through Google searches (Da et al., 2011) or institutional investor attention measured through Bloomberg searches (Ben-Rephael et al., 2017). Although our results thus far lend themselves to an explanation based on attention from retail investors, the crowd-sourcing aspects of Twitter allow for the possibility of a wider range of investors acting on Twitter posts. As such, it is not yet clear which type of investor is responsible for the Twitter effect. We use stock characteristics associated with the Twitter effect to identify whether the actions of retail or institutional investors provide a better explanation for our observed Twitter effect.

We first make use of the investor recognition hypothesis from Merton's (1987) model of informationally incomplete markets. In this model, stocks attract differing levels of investor attention, and those with less attention experience lower demand in equilibrium, resulting in higher expected returns. We build on this line of research by exploring if stocks with lower investor recognition are the same stocks that experience the greatest impact from increased Twitter attention.

Following the current literature on attention, we use stock size, stock age, analyst coverage, and individual ownership as proxies for investor recognition, in which lower visibility is consistent with a stronger retail investor presence (Baker and Wurgler, 2006; Barber and Odean, 2008; Fang and Peress, 2009). Results presented in Table 5, Panel A, come from double-sorted quintile portfolios across Twitter activity and each of our measures of investor attention. We create portfolios each day based on the variable calculations presented in Appendix Table A.1 and measure the contemporaneous performance. We retain the quintile of stocks showing the greatest change in Twitter activity and report daily excess returns across all five quintiles of each investor recognition variable, as well as the difference between the high and low rank portfolios. Across all proxies for stock visibility, except for analyst coverage, we find support for the investor recognition hypothesis, suggesting that retail investors drive the Twitter effect.

[Insert Table 5 near here]

The impediments-to-trade hypothesis, tested by Fang and Peress (2009), posits that limits to arbitrage and illiquidity concerns prevent rational market participants from actively taking the other side of trades for certain stocks. With no large institutional investors, stocks become more susceptible to the impact of retail investor actions. Baker et al. (2011) examine risk and beta relations with institutional portfolio ownership, and Baker and Wurgler (2006) show that less visible, and harder-to-arbitrage, securities are more likely to be affected by retail investor attention. We use these proxies for institutional investor involvement to explore the type of investor responsible for the Twitter effect.

Utilizing stock price levels and bid-ask spreads as measures of illiquidity, and stock volatility and betas as measures of limits to arbitrage, we repeat the analysis from Table 5, Panel A. We measure volatility and beta as defined in Appendix Table A.1. Baker et al. (2011) show that holding stocks with betas deviating from the market (beta of 1) makes tracking benchmarks more difficult for institutional investors. We thus construct our beta-sorted portfolios based on the absolute value of each stock's beta minus one, or how much a stock's beta differs from the market beta.

We report the results of our impediments-to-trade test in Table 5, Panel B. Stocks with low levels of institutional appeal, with low prices, high bid-ask spreads, high volatility, and betas deviating the most from a value of one all experience the largest impact from Twitter activity on excess returns. Combined, our investor recognition and impediments-to-trade results explore our third hypothesis and provide evidence consistent with the Twitter effect being driven by attention from retail investors.

5. Common risk factors and the Twitter effect over time

Within this section, we explore the economic magnitude of Twitter-associated returns through the construction of a long-short zero-cost portfolio. Results thus far demonstrate a cross-sectional relation, wherein stocks that experience an increase in Twitter activity also experience outsized contemporaneous returns. During the following trading day, small stocks continue to experience positive excess returns. Therefore, we sort stocks into deciles each day ($t = 0$) based on changes in Twitter activity and market capitalization. To allow for portfolio creation after observing Twitter activity, we take long positions in the set of stocks that rank in the top change-in-Twitter-activity decile and the bottom market capitalization decile, and we take short positions in the set of stocks that rank in the bottom change-in-Twitter-activity decile and the bottom market capitalization decile and hold the portfolio for trading day ($t + 1$). This yields a time series of characteristic-adjusted excess daily returns for our portfolio. We regress the time series of returns on four risk factors from Fama and French (1993) and Carhart (1997) and include lagged factors to account for autocorrelation and nonsynchronous trading (Busse, 1999). If known risk factors fully explain the returns from our long-short portfolio, we expect to find an insignificant intercept.

We report the portfolio regression results in Table 6. Across a one-factor, three-factor, and four-factor model, we observe a significant alpha coefficient from daily excess returns ranging from 11.1 bps to 10.4 bps, equivalent to an annualized excess return of approximately 27.75% (based on 250 trading days). Within Panels B and C, we explore the long and short components individually. Our Twitter effect is primarily driven by the long position generating an alpha of 9.4 bps per day. The short positions generate positive, but nonsignificant, alphas ranging from just 1.0 bps to 1.4 bps per day. As robustness, we replicate this analysis using raw returns and observe a similar relation. Our results confirm that the excess returns derived from changes in Twitter attention are not subsumed by traditional risk factors.

[Insert Table 6 near here]

[Insert Figure 2 near here]

From our daily portfolios, we plot in Figure 2 the cumulative excess returns generated from the long only side, short only side, and the long-short portfolio to illustrate how the impact of Twitter attention persists through time. Profits from this strategy are consistently positive over the period 2011 through 2015, with no one subperiod driving our results. Because this strategy is based on small securities that are often unavailable or prohibitively expensive to short, it is important that the short position not drive our observed relation. Our result is robust to single-sorted portfolios by only Twitter activity. The resulting portfolio generates excess returns of 43% over the five-year period. We also consider returns calculated from the midpoint of the bid-ask spread. This helps to remove concerns about illiquidity, especially for the small stocks that may drive these results. Our findings utilizing midpoint returns produce similar results.

6. The effect of Twitter activity on the diffusion and impact of information

In a 2009 paper, Fang and Peress (p. 2023) investigate how the "breadth of information dissemination affects stock returns" and Peress (2014) shows how newspaper coverage assists in the diffusion of information. We use the previously documented positive impact of news and earnings to test our final hypothesis; how our measure of attention via Twitter activity assists in the diffusion of information contained in news coverage and earnings announcements. [15] Our focus in this section is not on explaining the association between news, earnings announcements, and stock returns, but on determining how changes in Twitter attention affect this relation by magnifying the impact on stock returns. We expect that increased Twitter attention helps to quickly spread information to a wider audience, resulting in larger price impacts that persist over time, relative to the price impacts that occur when news releases or earnings announcements are not accompanied by investor interest. We test this within the multivariate regression framework of Equation (4).

We report our results on the role of news coverage and earnings announcements occurring in conjunction with increases in Twitter attention in Table 7. We reestimate the model from Equation (4) including additional interaction terms between *News Coverage*, *EDay*, *Tweets*, and *Size*. The additional interaction terms allow us to examine the resulting impact that changes in

[15] News stories appear on 37% of stock-day observations. However, stocks within the smallest quintile of market capitalization have news releases on only 3.3% of stock-day observations. The infrequent news coverage is similar to the coverage shown in Barber and Odean (2008).

Twitter activity have on alternative information events. Our results provide strong support for our final hypothesis. An increase in Twitter activity facilitates a greater spread of information than would otherwise occur in the presence of an informational event on its own. An increase in *News Coverage* combined with an increase in Twitter activity results in an additional 37.7 bps in excess returns, while an increase in Twitter activity occurring on an earnings date increases excess returns by 13.9 bps. The increase in excess returns a stock experiences when an increase in Twitter activity accompanies an information event is consistent with a reduction in informational asymmetries by improving the diffusion of information to investors, particularly for small firms that are often most susceptible to a lack of attention.

[*Insert Table 7 near here*]

We consider, as robustness, the impact of news and earnings announcements separately in Appendix Table 11. The results from each individual impact confirm the findings of Table 7; an increase in Twitter activity increases the diffusion of information contained in *News Coverage* and earnings dates (*EDay*). As a final robustness test, we consider the impact that orthogonalized *ΔTweets,* as created in Appendix Table A.5, have on the interaction between Twitter activity and information events. After removing the effects of news and earnings from changes in Twitter activity, we observe results consistent with those presented in Table 7. Results from Appendix Table A.12 show a strong magnification effect from an increase in Twitter activity combined with news coverage or an earnings announcement day. As a result of removing the impact of news and earnings from Twitter activity, we now observe a strong positive association between stock returns and each of these three events. These results support our finding that Twitter removes informational barriers faced by less visible stocks and increases the diffusion of information to investors.

7. Conclusion

Attention stemming from the social media platform Twitter has wide-ranging implications for financial markets. We show that Twitter activity is a unique and meaningful source of investor attention that has a statistically and economically significant impact on asset prices. This impact is best explained by the interpretation that Twitter generates investor attention and facilitates the diffusion of information amongst investors in ways that are not captured by traditional measures of information or attention.

Through a quasi-natural experiment capturing random Twitter outages, we provide the first causal evidence between Twitter attention and economically meaningful changes in daily trading activity. We observe a strong association between Twitter attention and stock returns, which increases in magnitude when Twitter activity corresponds to an information event such as an increase in news coverage or an earnings announcement. Twitter activity not only increases the magnitude of the contemporaneous day's returns, but also this price change persists over time. Because the longer-term effect is greatest for less visible stocks, this finding is consistent with Twitter removing informational asymmetries experienced by stocks that investors are less aware of. Although we do not claim a causal relation with returns, we do provide strong anecdotal evidence of Twitter's direct impact on stock prices.

We explore whether institutional or retail investors drive the association between Twitter and trading activity and show that Twitter activity proxies for retail investor attention. The impact of an increase in Twitter attention is largest for small, less visible stocks with greater individual ownership, as well as for illiquid and difficult-to-arbitrage stocks. These findings are consistent with individual or retail trading responding to activity on Twitter. Furthermore, we observe this relation across our full sample period, with no single year driving our results.

User-generated content from Twitter affects stock returns in ways that are distinct from other previously considered measures of investor attention. As highlighted by the quotation from US Supreme Court associate justice Anthony M. Kennedy, social media and online self-sharing information sources provide individuals with an important medium to voice their opinions and spread information about any topic, including financial markets. Because of the large economic impact, coupled with the rising interest in social media platforms such as Twitter, both academics and practitioners need to better understand how individual investors can influence financial markets using nothing more than an app on their mobile device.

Acknowledgements

We are thankful to Vineet Bhagwat, Jesse Blocher, Keith Gamble, Svetlana Gavrilova, Mark Huson, Benjamin Jansen, Timothy Marlo, Jeffrey Pontiff, Adam Rennhoff, Sanjiv Sabherwal, Salil Sarkar, Hai Tran, Weike Xu, and seminar participants at the Academy of Behavioral Finance and Economics annual meeting, the Midwest Finance Association annual meeting, Middle Tennessee State University, and the University of Texas at Arlington for their helpful comments. We also thank Ryan Israelsen for sharing the daily Google search volume data used in Ben-Rephael, Da, and Israelsen (2017) and James Crane-Baker from PsychSignal for assistance with the Twitter data. We are responsible for all remaining errors. This paper was previously titled "Is all that Twitters gold? Social media attention, stock returns, and the spread of information."

References

Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise? The information content of Internet stock message boards. Journal of Finance 59, 1259–1294.

Baker, M., Bradley, B., Wurgler, J., 2011. Benchmarks as limits to arbitrage: understanding the low-volatility anomaly. Financial Analysts Journal 67, 40–54.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross section of stock returns. Journal of Finance 61, 1645–1680.

Barber, B. M., Odean, T., 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21, 785–818.

Bartov, E., Faurel, L., Mohanram, P. S., 2017. Can Twitter help predict firm-level earnings and stock returns? The Accounting Review 93, 25–57.

Bawden, D., Robinson, L., 2008. The dark side of information: overload, anxiety, and other paradoxes and pathologies. Journal of Information Science 35, 180–191.

Ben-Rephael, A., Carlin, B. I., Da, Z., Israelsen, R. D., 2018. Demand for information and asset pricing. Unpublished working paper no. w23274. National Bureau of Economic Research, Cambridge, MA.

Ben-Rephael, A., Da, Z., Israelsen, R. D., 2017. It depends on where you search: institutional investor attention and underreaction to news. Review of Financial Studies 30, 3009–3047.

Bertrand, M., Mullainathan, S., 2003. Enjoying the quiet life? Corporate governance and managerial preferences. Journal of Political Economy 111, 1043–1075.

Busse, J. A., 1999. Volatility timing in mutual funds: evidence from daily returns. Review of Financial Studies 12, 1009–1041.

Carhart, M. M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

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

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. Journal of Finance 66, 1461–1499.

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all fears: investor sentiment and asset prices. Review of Financial Studies 28, 1–32.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance 52, 1035–1058.

Engelberg, J., Parsons, C., 2011. The causal impact of media in financial markets. Journal of Finance 66, 67-97.

Engelberg, J., McLean, R. D., Pontiff, J., 2017. Analysts and anomalies. Unpublished research paper no. 2939174. Georgetown University, McDonough School of Business, Washington, DC.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fang, L., Peress, J., 2009. Media coverage and the cross section of stock returns. Journal of Finance 64, 2023–2052.

Frijns, B., Huynh, T. D., 2018. Herding in analysts’ recommendations: The role of media. Journal of Banking and Finance 91, 1-18.

Guindy, M. A., 2017. Is corporate tweeting informative or is it just hype? Evidence from the SEC social media regulation. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2824668.

Hanke, M., Hauser F., 2008. On the effects of stock spam e-mails. Journal of Financial Markets 11, 57–83.

Liew, J., Wang, G., 2016. Twitter sentiment and IPO performance: a cross-sectional examination. Journal of Portfolio Management (Summer), 129–135.

Merton, R. C., 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483–510.

Odean, T., 1999. Do investors trade too much? American Economic Review 89, 1279–1298.

Peress, J., 2014. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. Journal of Finance 69, 2007–2043.

Petersen, M. A., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Review of Financial Studies 22, 435–480.

Renault, T., 2017. Market manipulation and suspicious stock recommendations on social media. Unpublished working paper. IÉSEG School of Management and Pathéone-Sorbonne University, Paris, France. http://www.thomas-renault.com/wp/market-manipulation-suspicious.pdf.

Renault, T., 2017. Intraday online investor sentiment and return patterns in the U.S. stock market. Journal of Banking and Finance 84, 25-40.

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. Journal of Finance 52 (1), 35–55.

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

Vozlyublennaia, N., 2014. Investor attention, index performance, and return predictability. Journal of Banking and Finance 41, 17-35.

Table 1
Descriptive statistics

This table provides sample descriptive statistics for all stocks, stocks with above median tweets, and stocks with below median tweets. We calculate one time-series value for each stock within our sample and present the cross-sectional mean, median, and standard deviation. We define all variables as detailed in Appendix Table A.1.

	All stocks (N = 1,976)			Above median tweet stocks (N = 988)			Below median tweet stocks (N = 988)		
Variable	Mean	Median	Standard deviation	Mean	Median	Standard deviation	Mean	Median	Standard deviation
Number of Tweets	9.502	3.089	45.015	17.128	6.769	62.752	1.875	1.820	0.750
Market Cap (billions)	7.891	1.744	25.703	14.350	4.238	35.145	1.433	0.825	1.807
Shares Traded (millions)	1.408	0.451	3.385	2.517	1.192	4.505	0.299	0.182	0.422
Price (US dollars)	42.539	32.740	47.400	49.721	39.829	47.188	35.356	26.127	46.538
Number of Analysts	7.136	5.733	5.380	9.961	8.967	5.860	4.311	3.701	2.757
Excess Return (basis points)	-0.027	0.048	5.710	0.698	0.559	6.234	-0.752	-0.451	5.031
Twitter Down	13.601	11.228	14.706						

Table 2
Lead and lag relation

This table reports regression coefficients with the dependent variable measured as the daily change in Tweets on day 0. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Column 1 contains stock excess returns and stock volume calculated over days [*t-5*, *t-1*]. Control variables in Column 2 include *Size, News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, and *DAIA*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Variable	[1]	[2]
Excess Returns (t - 5)	0.000	-0.001
	(0.214)	(-1.321)
Excess Returns (t - 4)	0.000	-0.001**
	(1.143)	(-2.403)
Excess Returns (t - 3)	0.000	-0.001**
	(1.386)	(-2.310)
Excess Returns (t - 2)	0.000	-0.001**
	(0.998)	(-2.217)
Excess Returns (t - 1)	0.000	-0.001**
	(0.215)	(-2.486)
Volume (t - 5)	-0.002	-0.009
	(-0.173)	(-0.754)
Volume (t - 4)	-0.008	-0.020
	(-0.449)	(-1.166)
Volume (t - 3)	0.001	-0.011
	(0.076)	(-0.558)
Volume (t - 2)	0.004	-0.007
	(0.214)	(-0.340)
Volume (t - 1)	0.009	0.002
	(0.711)	(0.096)
Stock Controls	No	Yes
Day of Week FE	Yes	Yes
Stock FE	Yes	Yes
Cluster	Daily	Daily
Observations	2215535	2215535

Table 3
Twitter outages

This table reports regression coefficients with the dependent variables measured as *Ln(Dollar Trade)* in columns 1 and 2, as *Volume* in columns 3 and 4, and as *Excess Returns* in columns 5 and 6. All variables are defined in Appendix Table A.1. As controls, we include *Size, News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We standardize all independent variables to have a mean of zero and a standard deviation of one. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	Ln(Dollar Trade)		Volume		Excess Returns	
Variable	[1]	[2]	[3]	[4]	[5]	[6]
Twitter Down	-0.044***	-0.015***	-0.036***	-0.022***	0.041***	-0.368***
	(-5.893)	(-2.771)	(-5.783)	(-4.005)	(2.774)	(-4.485)
Size		1.756***		0.519***		-22.079***
		(124.802)		(35.872)		(-10.520)
News Coverage		0.092***		0.090***		3.822***
		(45.784)		(43.253)		(12.096)
Volume						1.665***
						(5.080)
Advertising / Sales		-0.003*		-0.005***		-1.071**
		(-1.870)		(-3.796)		(-2.041)
Dividend Paying		0.004*		-0.015***		1.739***
		(1.849)		(-6.499)		(3.333)
Ln(Analysts)		0.046***		0.051***		-0.413
		(10.103)		(10.536)		(-1.004)
EDay		0.073***		0.073***		0.957**
		(71.736)		(71.199)		(2.192)
Risk of Returns		0.095***		0.106***		-0.350
		(27.407)		(29.236)		(-0.719)
DADSVI		0.003***		0.004***		0.461***

		(2.672)		(2.968)		(3.308)
DAIA		0.088***		0.089***		1.676***
		(53.035)		(51.427)		(4.808)
5 Lags AER	No	Yes	No	Yes	No	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2215535	2215535	2215535	2215535

Table 4
Change in tweets and stock returns

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Panel A examines how changes in Twitter activity affect stock returns. Columns 1, 3, 5, and 7 include *ΔTweets*, and Columns 2, 4, 6, 8 and 9 include stock control variables: *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. Panel B accounts for size through an interaction term between *ΔTweets* and *Size*. Panel B follows the control variable setup from Panel A. Panel C examines the impact of Retweeting. Panel C follows the control variable setup from Panel A and includes additional variables *ΔRetweets*, *ΔRetweets* interacted with *Size*, *ΔRetweets* interacted with *ΔTweets*, and finally *ΔRetweets* interacted with *ΔTweets* and *Size*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Twitter activity and stock returns

	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	10.105***	9.573***	0.719***	0.473***	0.063	0.041	-0.305**	-0.299**	0.231
	(43.817)	(47.453)	(4.471)	(3.129)	(0.481)	(0.314)	(-2.329)	(-2.229)	(0.249)
Size		-21.716***		-22.011***		-22.051***		-22.137***	-793.219***
		(-10.379)		(-10.581)		(-10.600)		(-10.682)	(-48.239)
News Coverage		1.972***		1.573***		0.216		-0.131	-9.463***
		(6.286)		(5.341)		(0.955)		(-0.611)	(-7.250)
Volume		0.871***		1.555***		0.445***		0.287*	0.380
		(2.715)		(8.932)		(2.687)		(1.878)	(0.289)
Advertising / Sales		-1.082**		-1.028*		-0.948*		-0.943*	-29.100***
		(-2.074)		(-1.950)		(-1.794)		(-1.785)	(-10.348)
Dividend Paying		3.563***		3.531***		3.332***		3.326***	106.959***
		(3.351)		(3.315)		(3.129)		(3.128)	(15.516)
Ln(Analysts)		-0.417		-0.285		-0.315		-0.294	-16.952***
		(-1.007)		(-0.704)		(-0.800)		(-0.743)	(-5.636)
EDay		7.625**		-6.207***		-2.187*		1.638	19.096***
		(2.156)		(-3.529)		(-1.724)		(1.460)	(2.919)

Risk of Returns		-0.488		-0.562		-0.409		-0.467	-26.306***
		(-1.010)		(-1.169)		(-0.865)		(-0.986)	(-12.968)
DADSVI		1.737**		0.245		0.976		1.101	38.322***
		(2.407)		(0.350)		(1.523)		(1.643)	(9.710)
DAIA		0.846***		-0.542***		-0.144		-0.113	-4.874***
		(2.725)		(-2.687)		(-0.783)		(-0.663)	(-4.754)
5 Lags AER	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Panel B: Effect of size

	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	22.330***	24.275***	9.686***	9.839***	1.715	1.754	-4.577**	-4.542**	-7.310
	(8.390)	(9.023)	(4.327)	(4.356)	(0.932)	(0.952)	(-2.578)	(-2.547)	(-0.694)
*ΔTweets * Size*	-12.255***	-14.747***	-8.990***	-9.394***	-1.659	-1.718	4.279**	4.255**	7.560
	(-4.686)	(-5.589)	(-4.111)	(-4.246)	(-0.920)	(-0.951)	(2.456)	(2.427)	(0.727)
Stock Controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes
5 Lags AER	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Panel C: Retweeting and stock returns

	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	8.663***	18.009***	0.356**	8.878***	0.049	0.940	-0.304**	-4.026**	-5.475
	(44.264)	(7.214)	(2.380)	(4.007)	(0.369)	(0.490)	(-2.214)	(-2.147)	(-0.491)
*ΔTweets * Size*		-9.563***		-8.597***		-0.906		3.751**	5.692
		(-3.907)		(-3.958)		(-0.480)		(2.029)	(0.516)
ΔRetweets	2.547***	37.110***	0.354**	4.824**	-0.041	2.363	0.035	-1.822	-5.426
	(13.156)	(14.321)	(2.248)	(2.152)	(-0.292)	(1.155)	(0.251)	(-0.932)	(-0.482)
*ΔRetweets * Size*		-34.424***		-4.359**		-2.395		1.815	4.981
		(-13.885)		(-2.012)		(-1.209)		(0.958)	(0.454)
*ΔTweets * ΔRetweets*	4.646***	51.011***	0.465*	10.463***	0.045	0.570	-0.080	-0.462	5.268

	(11.498)	(9.511)	(1.835)	(2.982)	(0.264)	(0.240)	(-0.505)	(-0.196)	(0.455)
*ΔTweets * ΔRetweets * Size*		-46.539***		-10.024***		-0.526		0.377	-2.354
		(-8.980)		(-2.952)		(-0.230)		(0.166)	(-0.209)
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Table 5
Investor recognition, impediments to trade, and the Twitter effect

This table presents double-sorted quintile portfolios. We first sort securities into quintiles by their daily change in Twitter activity. We retain the top quintile and further sort by size, age, analyst coverage, and individual ownership (Panel A) and by price, bid-ask spread, volatility, and absolute value of beta – 1 (Panel B). All sort variables are defined in Appendix Table A.1. From each double-sorted portfolio, we present the day 0 excess returns in basis points and the high minus low difference. *, **, and *** represent statistical significance from zero at the 10%, 5%, and 1% level, respectively.

Panel A: Investor recognition				
	t = 0			
Portfolio	*Size*	*Age*	*Number of Analyst*	*Individual Ownership*
1 = low	21.287	22.551	15.791	14.876
2	17.618	17.883	16.368	17.441
3	15.002	15.342	15.325	16.637
4	14.489	13.773	15.942	16.426
5 = high	14.130	12.907	19.413	17.081
5 minus 1	-7.157***	-9.644***	3.622***	2.206*
Panel B: Impediments of trade				
	t = 0			
Portfolio	*Price*	*Bid-Ask Spread*	*Volatility*	*Absolute Value (Beta - 1)*
1 = low	24.758	12.058	5.940	11.940
2	17.238	13.593	8.218	13.761
3	14.032	14.551	11.689	14.645
4	13.292	16.508	18.099	16.490
5 = high	13.211	24.477	38.569	26.653
5 minus 1	-11.547***	12.419***	32.629***	14.713***

Table 6
Change in Tweets and investor alpha

This table explores the profits earned from a daily long-short trading strategy. We form portfolios, at the end of each day (t = 0), of the top and bottom decile of stocks based on their change in tweets, conditional upon the stock being in the smallest decile of market capitalization. We go long the top decile and short the bottom decile on the following day (t + 1). We regress this time series of returns on the market excess returns (*Mkt-Rf*), small minus big (*SMB*), high minus low (*HML*), and momentum factors (*UMD*). To account for autocorrelation and nonsynchronous trading, we include the prior day's factors as well. Column 1 uses a one-factor model [capital asset pricing model (CAPM)], Column 2 uses a three-factor model (FF-three), and Column 4 uses a four-factor model (FF-four). We report t-statistics below coefficient estimates. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively. FF = Fama and French.

Variable	CAPM (1)	FF-three (2)	FF-four (3)
Panel A: Long top ΔTweet stocks, short bottom ΔTweet stocks			
Intercept	0.111***	0.104***	0.104***
	(4.405)	(4.157)	(4.131)
Mkt-Rf	-0.022	0.012	0.012
	(-0.864)	(0.460)	(0.459)
SMB		-0.173***	-0.173***
		(-3.247)	(-3.227)
HML		-0.182***	-0.183***
		(-2.958)	(-2.589)
UMD			0.000
			(0.001)
One-day lagged factors	Yes	Yes	Yes
Number of observations	1,252	1,252	1,252
Panel B: Long top ΔTweet stocks			
Intercept	0.097***	0.094***	0.094***
	(4.828)	(4.687)	(4.666)
One-day lagged factors	Yes	Yes	Yes
Panel C: Short bottom ΔTweet stocks			
Intercept	0.014	0.010	0.010
	(1.000)	(0.726)	(0.711)
One-day lagged factors	Yes	Yes	Yes

Table 7
Change in tweets, information events, and stock prices

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Columns 1, 3, 5, and 7 of Panel A show the effect of *News Coverage*, earnings days (*EDay*), and *ΔTweets* on performance as well as their interactions with each other and *Size*. Columns 2, 4, 6, 8, and 9 include stock controls: *Size*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	12.141***	17.833***	5.390**	6.381**	-2.021	-1.187	-5.152**	-4.159*	-3.546
	(4.456)	(6.457)	(2.115)	(2.493)	(-0.862)	(-0.507)	(-2.236)	(-1.822)	(-0.261)
*ΔTweets * Size*	-6.060**	-6.942**	-5.224**	-6.217**	2.078	0.948	5.044**	3.845*	0.235
	(-2.214)	(-2.517)	(-2.037)	(-2.418)	(0.889)	(0.407)	(2.194)	(1.691)	(0.017)
News Coverage	22.698***	6.482*	39.057***	24.718***	22.958***	9.407***	24.472***	10.831***	28.431
	(5.613)	(1.659)	(8.707)	(5.777)	(6.554)	(2.932)	(7.350)	(3.608)	(1.630)
*News Coverage * Size*	-20.974***	-7.418**	-35.890***	-22.303***	-21.635***	-8.796***	-23.286***	-10.378***	-38.652**
	(-5.608)	(-2.046)	(-8.626)	(-5.616)	(-6.603)	(-2.934)	(-7.475)	(-3.705)	(-2.365)
*News Coverage * ΔTweets*	38.607***	37.669***	5.577	7.167	0.700	2.623	-6.141**	-4.173*	25.124*
	(7.536)	(7.385)	(1.253)	(1.610)	(0.279)	(1.050)	(-2.437)	(-1.667)	(1.867)
*News Coverage * ΔTweets * Size*	-32.832***	-32.245***	-4.916	-6.462	-0.922	-2.702	5.786**	3.977	-19.964
	(-6.659)	(-6.562)	(-1.138)	(-1.494)	(-0.376)	(-1.106)	(2.353)	(1.628)	(-1.510)
EDay	-1.261	-11.168***	-6.103***	-6.575***	-4.171***	-3.419**	1.927*	2.661**	14.415**
	(-0.362)	(-3.219)	(-3.451)	(-3.621)	(-3.205)	(-2.562)	(1.658)	(2.236)	(2.019)
*EDay * Size*	-12.824***	-14.480***	0.738	0.620	1.605	1.348	2.374*	2.092*	9.985*
	(-3.839)	(-4.331)	(0.484)	(0.408)	(1.301)	(1.096)	(1.951)	(1.725)	(1.691)
*EDay * ΔTweets*	12.551***	13.890***	-1.928	-2.095*	2.715***	2.399***	-0.880	-1.224*	0.314
	(4.599)	(5.095)	(-1.586)	(-1.726)	(3.322)	(2.918)	(-1.231)	(-1.711)	(0.075)
*EDay * ΔTweets * Size*	-9.992***	-9.924***	1.788	1.513	0.969	0.709	-0.642	-0.912	5.295
	(-3.792)	(-3.768)	(1.545)	(1.309)	(1.114)	(0.815)	(-0.783)	(-1.115)	(1.321)

Stock Controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes
5 Lags AER	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Figure 1. Cashtag. This figure shows an anonymized example of a cashtag from Twitter.

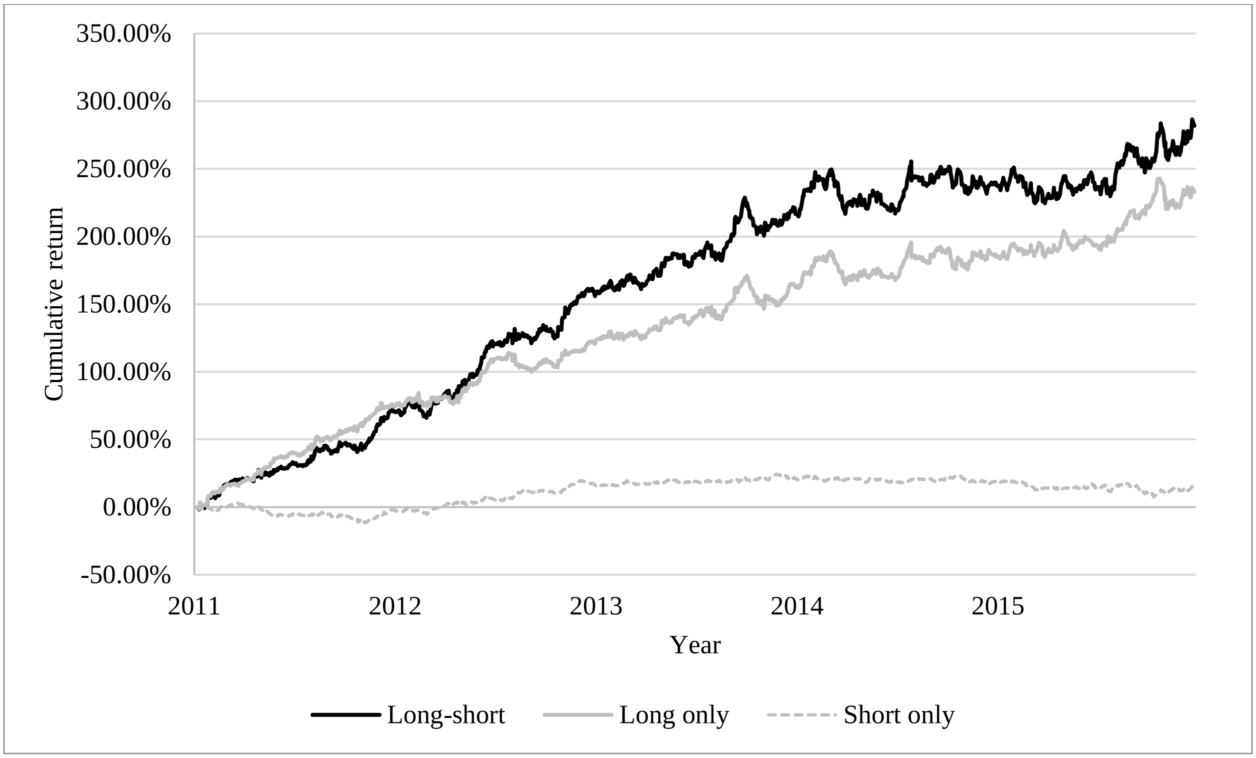

Figure 2. Cross-sectional long-short trading strategy. This figure shows the cumulative excess returns earned from a long-short trading strategy. The strategy is to go long the top decile of change in tweet securities on day 1 and short the bottom decile of change in tweet securities on day 1, both conditional on the stock being within the bottom decile of market capitalizations.

APPENDIX

Twitter activity, investor attention, and the diffusion of information

This Appendix presents variable descriptions as well as results from additional tests mentioned, but not presented, in the paper. The focus is on differentiating Twitter from traditional news as a measure of attention, the impact of Twitter activity on asset prices, results across various subsamples, and the importance of defining trading days.

Table A.1
Variable definitions

This table displays the names, sources, and brief definitions for all variables that appear in our paper. CRSP = Center for Research in Security Prices; I/B/E/S = Institutional Brokers' Estimate System.

Variable	Source	Definition
ΔTweets	PsychSignal	Deseasonalized daily change in the natural log of one plus the number of tweets about a stock, measured over Eastern Standard Time calendar days.
Absolute Value (*Beta* - 1)	CRSP	Absolute value of market excess return factor coefficient from a rolling 90-day one-factor model, including a one-day lagged market excess return factor to account for nonsynchronous trading, minus one.
Advertising/Sales	Compustat	Ratio of reported advertising expense to sales.
AER	CRSP	Absolute value of daily excess returns.
Age	CRSP	Age of a stock from inception, listed in fractional years.
Beta	CRSP	The coefficient on the market excess return factor from a rolling 90-day one-factor model, including a one-day lagged market excess return factor to account for nonsynchronous trading (Busse, 1999).
Bid-Ask Spread	CRSP	Difference between the bid and ask prices.
DADSVI	Google	Indicator variable equal to one if a stock's daily search volume index value is in the top 6% of search values for that stock from the prior 30 trading days, measured over Eastern Standard Time calendar days.
DAIA	Bloomberg	Indicator variable equal to one if a stock's daily institutional investor attention (Bloomberg queries) measure is in the top 6% of attention measures for that stock from the prior 30 trading days, measured over Eastern Standard Time calendar days. See Ben-Rephael, Da, and Israelsen (2017) for a detailed description.
Dividend Paying	CRSP	Indicator variable that takes on the value of one if a stock pays a dividend and zero otherwise
EDay	I/B/E/S	Indicator variable that takes on the value of one if a daily observation takes places on an earnings announcement day and zero otherwise. Announcements after market close are assigned to the following trading day.
Excess Return	CRSP	Daily excess return in basis points from a size and book-to-market (5 x 5) matched portfolio of stocks.
Individual Ownership	Factset	One minus total institutional ownership of a stock as reported in 13F filings.
Ln(Analysts)	I/B/E/S	Natural log of number of analysts covering a stock.
Ln(Dollar Trade)	CRSP	The natural log of the dollar value of all trades for a stock during the day calculated as the number of shares traded x stock price.
Ln(Tweet Count)	PsychSignal	Natural log of the total number of daily cashtagged tweets about a stock plus one, measured over Eastern Standard Time calendar days.
Market Cap	CRSP	Total market capitalization of a stock, measured in billions of US dollars.
News Coverage	Bloomberg	Number of news articles published about a stock on an Eastern Standard Time calendar day.
Number of Analysts	I/B/E/S	Number of analysts covering a stock.
Number of Tweets	PsychSignal	Total number of daily cashtagged tweets about a stock, measured over Eastern Standard Time calendar days.

O-ΔTweets	PsychSignal	Orthogonalized and deseasonalized daily change in the natural log of one plus the number of tweets about a stock, measured over Eastern Standard Time calendar days. To orthogonalize, we regress *ΔTweets* on *News Coverage* and *EDay* and keep the residual as our change in tweets variable.
Price	CRSP	Daily closing price of a stock listed in US dollars.
ΔRetweets	PsychSignal	Deseasonalized daily change in the natural log of one plus the number of retweets about a stock, measured over Eastern Standard Time calendar days.
Risk of Returns	CRSP	Standard deviation of daily stock returns measured over days $t - 6$ to $t - 27$.
Shares Traded	CRSP	Number of shares traded daily for a stock, measured in millions.
Size	CRSP	Natural log of market capitalization of a stock, measured as the average market capitalization over days $t - 6$ to $t - 27$.
Tweets	PsychSignal	Natural log of 1 + the daily number of tweets about a company.
Twitter Down	Google	Natural log of 1 + the daily measure of daily Google search volume of the phrase “Twitter down.”
Volatility	CRSP	Root mean squared error from a rolling 90-day one-factor model, including a one-day lagged market excess return factor to account for nonsynchronous trading (Busse, 1999).
Volume	CRSP	Natural log of the number of shares traded daily for a stock, measured in millions.

Table A.2
Stock characteristics and the number of tweets

This table reports regression coefficients with the dependent variable measured as the natural log of one plus the number of daily tweets for a security. We examine the supply of information (Column 1), stock characteristics (Column 2), return characteristics (Column 3), demand for information (Column 4), and the full model (Column 5). All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All variables are defined in Appendix Table A.1. Information supply variables are *News Coverage* and *EDay*. Stock characteristics are *Size*, *Advertising/Sales*, *Dividend Paying*, and *Ln(Analysts)*. Market characteristics are *Volume* and *AER*. Information demand variables are *DADSVI* and *DAIA*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Variable	[1]	[2]	[3]	[4]	[5]
News Coverage	0.183***				0.134***
	(48.125)				(51.666)
EDay	0.132***				0.072***
	(39.819)				(38.530)
Size		0.217***			0.220***
		(19.922)			(21.911)
Advertising / Sales		0.004**			0.003
		(2.014)			(1.600)
Dividend Paying		-0.000			0.004**
		(-0.164)			(2.438)
Ln(Analysts)		0.033***			0.023***
		(4.461)			(3.145)
Volume			0.189***		0.140***
			(34.666)		(35.233)
AER			0.111***		0.077***
			(32.640)		(40.624)
DADSVI				0.228***	0.033***
				(19.766)	(4.648)
DAIA				0.174***	0.102***
				(49.380)	(63.214)
Day of Week FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2215535	2215535	2215535

Table A.3
Correlations

This table presents the contemporaneous correlations between measures of attention supply and demand at daily frequencies. All correlations are statistically significant at better than the 1% level. All variables are defined in Appendix Table A.1.

Variable	*ΔTweets*	*Volume*	*ER*	*AER*	*News Coverage*	*DADSVI*
Volume	0.066					
ER	0.057	-0.014				
AER	0.113	0.449	0.054			
News Coverage	0.119	0.223	0.010	0.144		
DADSVI	0.017	0.011	0.002	0.013	0.019	
DAIA	0.124	0.271	0.017	0.197	0.382	0.008

Table A.4
Changes in Twitter and investor attention measures

This table reports regression coefficients with *ΔTweets* as the dependent variable. We examine the supply of information (Column 1) and add daily stock and return characteristics (Column 2), longer frequency characteristics (Column 3), and demand for information (Column 4). All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All variables are defined in Appendix Table A.1. Information supply variables are *News Coverage* and *EDay*. Stock and return characteristics are *Size*, *AER*, *Volume*, *Ln(Analysts)*, and *Advertising/Sales*. Information demand variables are *DADSVI* and *DAIA*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Variable	[1]	[2]	[3]	[4]
News Coverage	0.141***	0.130***	0.130***	0.109***
	(50.212)	(47.373)	(47.373)	(45.580)
EDay	0.214***	0.051***	0.051***	-0.007
	(17.106)	(3.909)	(3.909)	(-0.573)
Size		0.028*	0.028	0.025
		(1.664)	(1.640)	(1.484)
AER		0.073***	0.073***	0.064***
		(46.265)	(46.206)	(43.297)
Volume		0.003*	0.003*	-0.006***
		(1.748)	(1.743)	(-3.314)
Ln(Analysts)			0.000	-0.002
			(0.022)	(-0.332)
Advertising / Sales			-0.000	0.000
			(-0.004)	(0.039)
DADSVI				0.051***
				(5.616)
DAIA				0.070***
				(42.755)
Day of Week FE	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2215535	2215535
R^2	0.023	0.030	0.030	0.037

Table A.5
Stock returns and orthogonalized Twitter activity

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1, with the exception of *O-ΔTweets*. *O-ΔTweets* is orthogonalized with respect to *News Coverage* and *EDay* prior to use in this regression. By construction, this measure is independent of any effect from our *News Coverage* and *EDay* variables. Panel A presents changes in Twitter activity affecting stock returns. Columns 1, 3, 5, and 7 include *O-ΔTweets*, and Columns 2, 4, 6, 8, and 9 include stock control variables: *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. Panel B accounts for size through an interaction term between *O-ΔTweets* and *Size*. Panel B follows the control variable setup from Panel A. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Changes in Twitter activity and stock returns									
	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Orthogonalized ΔTweets	9.766***	9.629***	0.525***	0.423***	0.041	0.030	-0.222*	-0.212	-0.233
	(47.389)	(48.547)	(3.699)	(2.967)	(0.326)	(0.240)	(-1.716)	(-1.617)	(-0.254)
News Coverage		0.949**		0.151		-0.192		-0.432**	-5.354***
		(2.151)		(0.496)		(-1.035)		(-2.474)	(-5.558)
EDay		13.457***		-3.997**		-1.551		2.027*	14.365**
		(3.700)		(-2.218)		(-1.209)		(1.780)	(2.185)
Stock Controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes
5 Lags AER	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495
Panel B: Effect of size									
	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]

Orthogonalized ΔTweets	15.413***	16.544***	5.316***	5.695***	0.255	0.413	-5.335***	-5.278***	-11.156
	(6.726)	(7.193)	(2.698)	(2.884)	(0.140)	(0.228)	(-3.024)	(-2.989)	(-1.065)
*ΔTweets * Size*	-5.659**	-6.931***	-4.801**	-5.285***	-0.214	-0.384	5.124***	5.078***	10.945
	(-2.518)	(-3.080)	(-2.490)	(-2.735)	(-0.120)	(-0.216)	(2.958)	(2.926)	(1.059)
News Coverage		0.944**		0.147		-0.192		-0.428**	-5.347***
		(2.139)		(0.483)		(-1.036)		(-2.453)	(-5.551)
EDay		13.425***		-4.022**		-1.552		2.050*	14.417**
		(3.692)		(-2.232)		(-1.211)		(1.800)	(2.193)
Stock Controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes
5 Lags AER	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster 2	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Table A.6
Stock returns, number of tweets, lagged tweets, and Twitter activity

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. All models include the natural log of one plus the number of daily Tweets about a stock (*Ln(Tweet Count)*) as well as five lags of the daily change in Twitter activity. Columns 1, 3, 5, and 7 include *ΔTweets*, the interaction term between *ΔTweets* and *Size*. Columns 2, 4, 6, 8, and 9 include stock control variables: *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	[t + 0]		[t + 1]		[t + 2]		[t + 3]		[t + 4, t + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	23.194***	23.479***	8.909***	10.087***	0.775	1.890	-5.303***	-4.382**	16.693
	(8.710)	(8.897)	(3.961)	(4.403)	(0.418)	(1.016)	(-2.983)	(-2.439)	(1.580)
*ΔTweets * Size*	-13.672***	-13.842***	-9.123***	-9.305***	-1.610	-1.687	4.157**	4.220**	-3.077
	(-5.234)	(-5.379)	(-4.169)	(-4.182)	(-0.892)	(-0.926)	(2.387)	(2.396)	(-0.299)
Ln(Tweet Count)	9.685***	11.697***	2.202***	0.090	1.466***	-0.814**	1.606***	-0.516	-31.161***
	(20.136)	(27.140)	(5.648)	(0.277)	(3.856)	(-2.495)	(4.316)	(-1.630)	(-13.047)
Stock Controls	No	Yes	No	Yes	No	Yes	No	Yes	Yes
5 Lags AER	No	Yes	No	Yes	No	Yes	No	Yes	Yes
5 Lags of *ΔTweets*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Table A.7
Industries and stock returns

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Our independent variables of interest are *ΔTweets*, the daily change in tweets across 12 industries (*Nondurables*, *Durables*, *Manufacturing*, *Energy*, *Chemicals*, *Business Equipment*, *Telecommunications*, *Utilities*, *Shops*, *Health Care*, *Finance*, and *Other*). We include, as stock control variables, *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We also include an interaction term between *ΔTweets* and *Size*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	[*t* + 0]	[*t* + 1]	[*t* + 2]	[*t* + 3]	[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]
Non Durables	27.922***	13.189***	-0.343	-4.358**	0.063
	(9.806)	(5.378)	(-0.173)	(-2.255)	(0.006)
Durables	26.221***	10.411***	-0.016	-5.202**	-0.095
	(9.102)	(4.190)	(-0.007)	(-2.557)	(-0.008)
Manufacturing	26.573***	11.200***	0.682	-4.770**	-1.451
	(9.692)	(4.724)	(0.348)	(-2.500)	(-0.133)
Energy	34.266***	14.581***	-2.623	-2.976	-1.993
	(10.043)	(5.134)	(-1.148)	(-1.253)	(-0.149)
Chemicals	27.866***	11.184***	0.223	-4.846**	-0.180
	(9.597)	(4.488)	(0.110)	(-2.402)	(-0.015)
Business Equipment	29.961***	11.605***	-0.286	-4.529**	-1.224
	(10.994)	(4.831)	(-0.150)	(-2.417)	(-0.112)
Telecom	30.430***	11.125***	0.214	-4.289**	-1.572
	(10.200)	(4.372)	(0.104)	(-2.109)	(-0.132)
Utilities	21.373***	11.334***	-0.023	-4.334**	-0.075
	(7.548)	(4.596)	(-0.012)	(-2.175)	(-0.007)
Shops	26.181***	11.159***	0.053	-4.188**	-0.090
	(9.659)	(4.703)	(0.027)	(-2.212)	(-0.008)
Healthcare	29.834***	12.179***	-0.374	-3.896**	-1.723
	(10.425)	(4.971)	(-0.187)	(-2.049)	(-0.156)
Finance	22.618***	10.239***	0.860	-3.935**	-0.990
	(8.598)	(4.488)	(0.465)	(-2.198)	(-0.095)
Other	28.609***	11.159***	0.725	-4.618**	-0.328
	(10.603)	(4.756)	(0.384)	(-2.515)	(-0.031)
Stock Controls	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes

Cluster	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2213559	2211583	2209607	2136495

Table A.8
Change in tweets and stock returns for Russell 1000 and Russell 2000 securities

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Panel A considers a sample of Russell 1000 index stocks, and Panel B considers a sample of Russell 2000 index stocks. Our independent variable of interest is *ΔTweets*. We include, as stock control variables, *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We also include an interaction term between *ΔTweets* and *Size*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Russell 1000						
	[*t* + 0]		[*t* + 1]	[*t* + 2]	[*t* + 3]	[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]
ΔTweets	10.060***	44.288***	2.939	3.272	-5.626*	-7.750
	(42.744)	(10.104)	(0.781)	(1.010)	(-1.746)	(-0.430)
*ΔTweets * Size*		-34.298***	-2.615	-3.241	5.501*	7.345
		(-7.922)	(-0.704)	(-1.013)	(1.730)	(0.412)
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily
Observations	1107602	1107602	1106648	1105694	1104740	1069442
Panel B: Russell 2000						
	[*t* + 0]		[*t* + 1]	[*t* + 2]	[*t* + 3]	[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]
ΔTweets	10.645***	17.083***	18.553***	2.427	-6.636**	0.780
	(43.136)	(3.761)	(4.679)	(0.742)	(-2.081)	(0.042)
*ΔTweets * Size*		-6.451	-17.873***	-2.384	6.325**	-1.341
		(-1.433)	(-4.553)	(-0.733)	(1.991)	(-0.073)
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily
Observations	1566214	1566214	1564797	1563380	1561963	1509534

Table A.9

Change in tweets and stock returns by trading day hours

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Our independent variable of interest in *ΔTweets*, in which we assign tweets to days based on trading day hours, not by a 24-hour calendar time frame. Tweets occurring after market close on trading days are assigned to the following day. We include, as stock control variables, *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We also include an interaction term between *ΔTweets* and *Size*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	[*t* + 0]		[*t* + 1]	[*t* + 2]	[*t* + 3]	[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]
ΔTweets	7.877***	25.937***	8.259***	-0.832	-6.046***	0.793
	(40.500)	(8.914)	(4.211)	(-0.460)	(-3.330)	(0.074)
*ΔTweets * Size*		-18.119***	-8.116***	0.762	5.990***	-1.125
		(-6.354)	(-4.220)	(0.429)	(3.344)	(-0.108)
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215534	2215534	2213558	2211582	2209606	2136494

Table A.10
Change in tweets and stock returns with no after-hour tweets

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Our independent variable of interest is *ΔTweets*, in which we drop any tweets that do not occur during trading hours. We include, as stock control variables, *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We also include an interaction term between *ΔTweets* and *Size*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	[*t* + 0]		[*t* + 1]	[*t* + 2]	[*t* + 3]	[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]
ΔTweets	12.498***	57.972***	7.003***	3.052	-7.039***	1.723
	(51.157)	(18.636)	(3.404)	(1.586)	(-3.649)	(0.158)
*ΔTweets * Size*		-45.594***	-6.921***	-2.949	6.836***	-1.982
		(-15.139)	(-3.443)	(-1.564)	(3.615)	(-0.185)
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2214815	2214815	2212839	2210863	2208887	2135775

Table A.11

Change in tweets, news, earnings dates, and stock prices

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. Columns 1, 3, 5, and 7 of Panel A show the effect of *News Coverage* and *ΔTweets* on performance as well as their interaction. Columns 2, 4, 6, 8, and 9 include *News Coverage * Size* and *News Coverage * ΔTweets * Size* as additional interaction terms. Columns 1, 3, 5, and 7 of Panel B show the effect of earnings days (*EDay*) and *ΔTweets* on performance as well as their interaction. Columns 2, 4, 6, 8, and 9 include *EDay * Size* and *EDay * ΔTweets * Size* as additional interaction terms. Stock controls include *Size*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: The impact of news and size

	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	51.937***	16.656***	12.737***	6.329**	2.182	-0.429	-5.544***	-4.223*	-1.092
	(15.365)	(6.085)	(4.500)	(2.526)	(1.055)	(-0.185)	(-2.831)	(-1.859)	(-0.081)
*ΔTweets * Size*	-41.339***	-5.801**	-12.599***	-6.155**	-2.475	0.151	5.264***	3.926*	-2.302
	(-12.028)	(-2.121)	(-4.380)	(-2.451)	(-1.193)	(0.065)	(2.674)	(1.737)	(-0.171)
News Coverage	-21.795***	-20.869***	-22.041***	-21.211***	-22.073***	-21.788***	-22.135***	-21.794***	-792.379***
	(-10.387)	(-9.962)	(-10.598)	(-10.249)	(-10.604)	(-10.497)	(-10.677)	(-10.547)	(-48.186)
*News Coverage * Size*	-1.069***	18.084***	1.482***	23.972***	0.259	7.872**	-0.015	10.049***	19.438
	(-3.350)	(4.264)	(5.069)	(5.935)	(1.042)	(2.568)	(-0.063)	(3.495)	(1.149)
*News Coverage * ΔTweets*		-18.631***		-21.583***		-7.309**		-9.623***	-29.952*
		(-4.686)		(-5.776)		(-2.554)		(-3.594)	(-1.893)
*News Coverage * ΔTweets * Size*	6.783***	53.323***	0.790**	5.212	0.182	2.310	-0.288	-4.308*	18.140
	(14.444)	(10.276)	(2.270)	(1.309)	(0.918)	(0.958)	(-1.480)	(-1.819)	(1.430)
Stock Controls		-47.412***		-4.601		-2.199		4.041*	-12.705
5 Lags AER		(-9.454)		(-1.191)		(-0.932)		(1.749)	(-1.019)
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily

Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Panel B: The impact of earnings dates and size

	[*t* + 0]		[*t* + 1]		[*t* + 2]		[*t* + 3]		[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
ΔTweets	27.195***	18.542***	10.087***	10.024***	1.391	1.564	-4.462**	-4.681**	-5.706
	(10.122)	(7.576)	(4.454)	(4.363)	(0.753)	(0.836)	(-2.496)	(-2.570)	(-0.526)
*ΔTweets * Size*	-13.407***	-4.750**	-9.493***	-9.430***	-1.644	-1.817	4.057**	4.276**	5.218
	(-5.096)	(-1.987)	(-4.291)	(-4.203)	(-0.910)	(-0.993)	(2.314)	(2.394)	(0.488)
EDay	-13.721***	-12.022***	-5.857***	-6.118***	-3.125**	-3.167**	2.714**	2.969**	14.405**
	(-3.954)	(-3.443)	(-3.218)	(-3.378)	(-2.317)	(-2.382)	(2.318)	(2.511)	(2.024)
*EDay * Size*		-17.742***		-2.219		0.296		1.169	3.867
		(-5.286)		(-1.554)		(0.252)		(0.999)	(0.676)
*EDay * ΔTweets*	21.830***	22.772***	-0.949	-0.679	2.480***	2.468***	-1.395**	-1.575**	8.563**
	(8.258)	(8.551)	(-0.888)	(-0.654)	(3.102)	(3.183)	(-2.057)	(-2.414)	(2.250)
*EDay * ΔTweets * Size*		-11.529***		0.627		0.251		-0.840	4.035
		(-4.449)		(0.594)		(0.296)		(-1.070)	(1.063)
Stock Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily	Daily
Observations	2215535	2215535	2213559	2213559	2211583	2211583	2209607	2209607	2136495

Table A.12
Orthogonalized change in tweets, news, earnings dates, and stock prices

This table reports regression coefficients with the dependent variable measured as the daily excess returns from a size and book-to-market matched sample measured over day 0, 1, 2, and 3, as well as days 4 through 40. All non-indicator independent variables are standardized to have a mean of zero and a standard deviation of one. All other variables are defined in Appendix Table A.1. *O-ΔTweets* is orthogonalized with respect to *News Coverage* and *EDay* prior to use in this regression. By construction, this measure is independent of any effect from our *News Coverage* and *EDay* variables. Additional variables of interest are *O-ΔTweets * Size*, *News Coverage * Size*, *News Coverage * O-ΔTweets * Size*, *EDay * Size*, *EDay * O-ΔTweets*, and *EDay * O-ΔTweets * Size* as interaction terms. We include, as stock control variables, *Size*, *News Coverage*, *Volume*, *Advertising/Sales*, *Dividend Paying*, *Ln(Analysts)*, *EDay*, *Risk of Returns*, *DADSVI*, *DAIA*, and five lags of *AER*. We also include an interaction term between *ΔTweets* and *Size*. We include day of the week and stock level fixed effects in all models. We report, below coefficient estimates, *t*-statistics clustered across each day of our sample. *, **, and *** represent statistical significance at the 10%, 5%, and 1% level, respectively.

	[*t* + 1]	[*t* + 2]	[*t* + 3]	[*t* + 4]	[*t* + 4, *t* + 40]
Variable	[1]	[2]	[3]	[4]	[5]
ΔTweets	13.216***	6.689***	0.129	-5.109***	0.791
	(5.952)	(3.217)	(0.067)	(-2.720)	(0.074)
*ΔTweets * Size*	-5.546**	-6.556***	-0.129	4.920***	-1.945
	(-2.513)	(-3.186)	(-0.068)	(2.664)	(-0.184)
News Coverage	38.497***	16.604***	5.017*	0.684	7.280
	(6.782)	(3.758)	(1.925)	(0.289)	(0.565)
*News Coverage * Size*	-36.614***	-16.529***	-5.258**	-1.092	-9.451
	(-6.648)	(-3.799)	(-2.062)	(-0.469)	(-0.745)
*News Coverage * ΔTweets*	47.306***	7.087*	-0.196	-3.009	2.395
	(8.281)	(1.851)	(-0.100)	(-1.437)	(0.233)
*News Coverage * ΔTweets * Size*	-43.139***	-6.247*	0.164	2.956	-1.541
	(-7.978)	(-1.699)	(0.087)	(1.446)	(-0.157)
EDay	9.496**	-5.930***	-2.413*	1.594	-2.376
	(2.508)	(-3.122)	(-1.825)	(1.352)	(-0.385)
*EDay * Size*	-19.247***	-1.005	1.464	1.022	6.884
	(-5.967)	(-0.705)	(1.342)	(0.989)	(1.321)
*EDay * ΔTweets*	61.630***	-3.251**	1.707	-0.118	8.304
	(14.852)	(-2.069)	(1.550)	(-0.109)	(1.493)
*EDay * ΔTweets * Size*	-17.086***	-2.898*	-1.909	-2.834**	3.934
	(-4.337)	(-1.846)	(-1.533)	(-2.427)	(0.724)
Stock Controls	Yes	Yes	Yes	Yes	Yes
5 Lags AER	Yes	Yes	Yes	Yes	Yes
Day of Week FE	Yes	Yes	Yes	Yes	Yes

Stock FE	Yes	Yes	Yes	Yes	Yes
Cluster	Daily	Daily	Daily	Daily	Daily
Observations	2215593	2213617	2211641	2209665	2136553

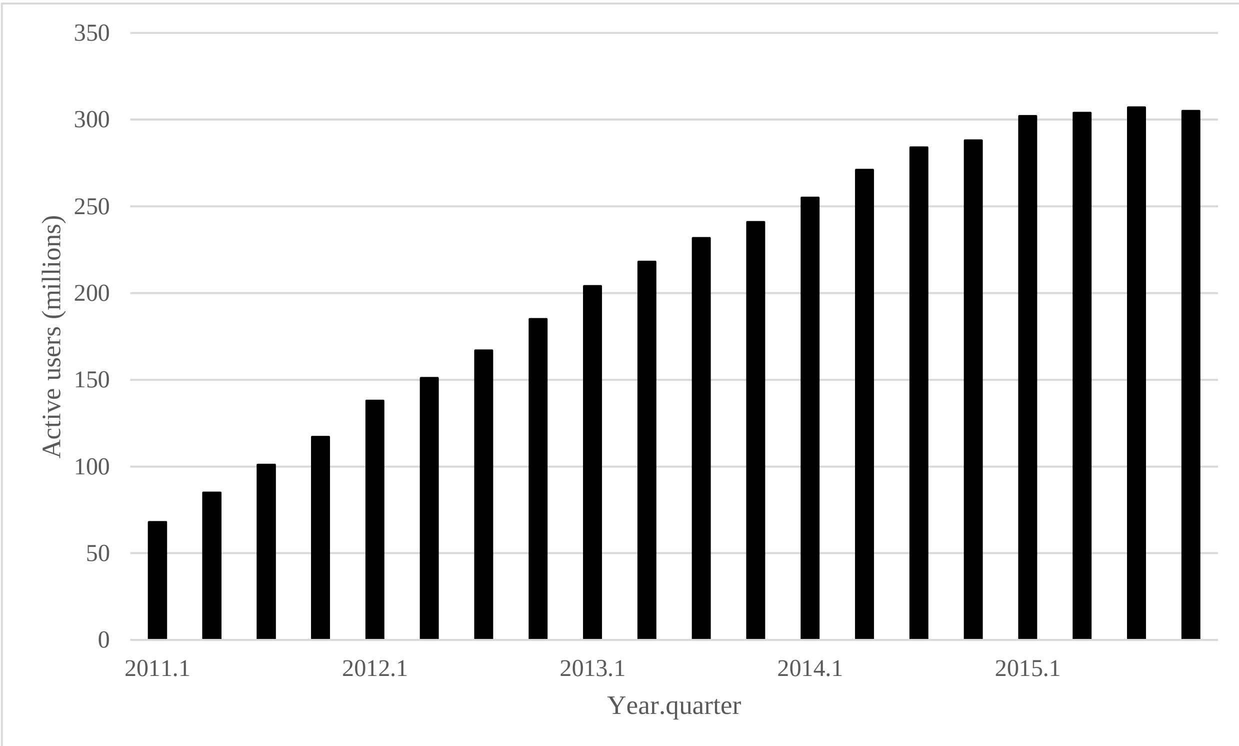

Figure A.1. Growth of Twitter. This figure shows the growth of active Twitter users in millions during our sample period, 2011 through 2015.

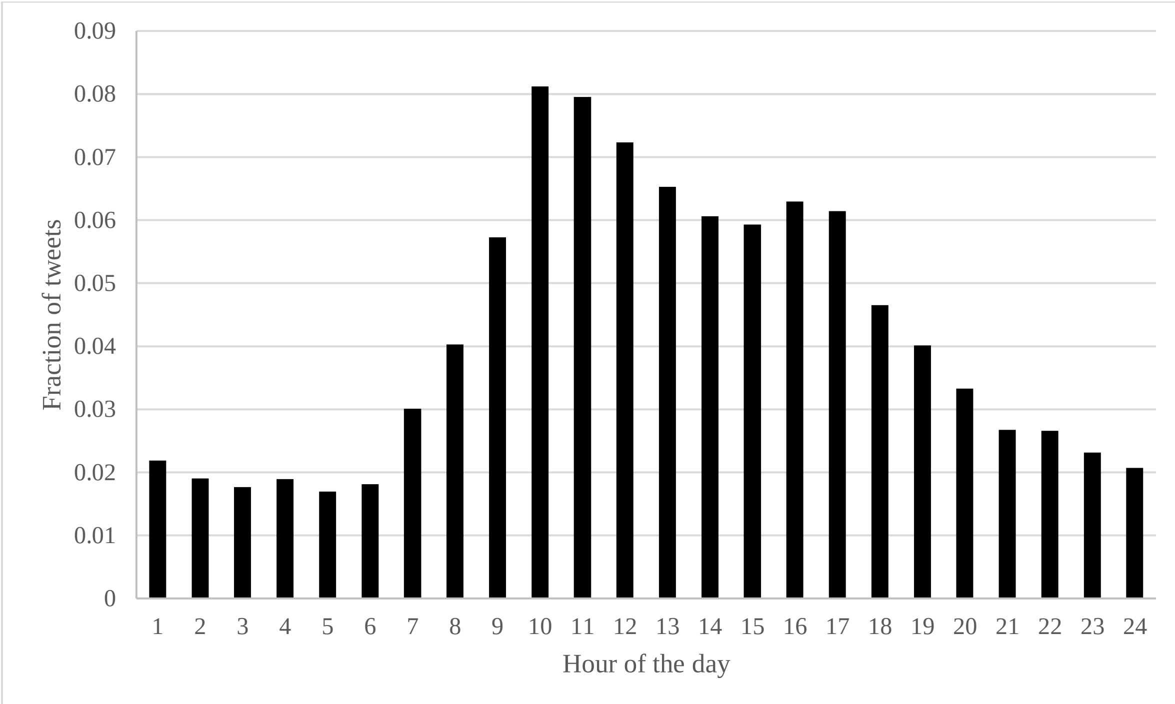

Figure A.2. Fraction of tweets during the day. This figure shows the distribution of tweets that are posted during each hour of the day, based on 24 hours, Eastern Standard Time.

Appendix References

Ben-Rephael, A., Da, Z., Israelsen, R. D., 2017. It depends on where you search: institutional investor attention and underreaction to news. Review of Financial Studies 30, 3009–3047.

Busse, J. A., 1999. Volatility timing in mutual funds: evidence from daily returns. Review of Financial Studies 12, 1009–1041.