Information Diffusion on Social Media: Does It Affect Trading, Return, and Liquidity? *

Nitesh Chawla, Zhi Da, Jian Xu, and Mao Ye

November 2016

Abstract

We track information diffusion in real time by monitoring how the news is tweeted and retweeted on Twitter. We find that news diffusion is highly correlated with intraday trading, especially for retail trading. News diffusion leads to a lower bid-ask spread and price pressure on the news day that is completely reverted the next day. The result is robust when we employ an instrumental variables approach. Our results show that information diffusion via Twitter does not incorporate new information into the stock price. Rather, Twitter spreads stale news, albeit at a much higher speed than traditional media.

Key Words: Information Diffusion, Price Pressure, Social Media, Liquidity.

---

* Chawla: University of Notre Dame, Department of Computer Science & Engineering, South Bend, IN, 46556. Email: nchawla@nd.edu. Da: University of Notre Dame, Mendoza College of Business, South Bend, IN, 46556. Email: zda@nd.edu. Xu: University of Notre Dame, Department of Computer Science and Engineering, South Bend, IN, 46556. Email: jxu5@nd.edu. Ye: University of Illinois at Urbana-Champaign, College of Business, Champaign, IL, 61820. Email: maoye@illinois.edu. We thank Daniel Andrei, Robert Battalio, Tim Burch, Shane Corwin, David Hirshleifer, Thierry Foucault, Pengjie Gao, Tim Loughran, Maureen O'Hara, Rik Sen, Paul Tetlock, Charles Trzcinka, Bonnie Van Ness, and seminar participants at the University of Miami, the University of Notre Dame, the NFA 2014 meeting, the MFA 2015 meeting, the Research in Behavioral Finance Conference at Erasmus, the Central Bank Workshop on Market Microstructure in Rome for helpful comments. We thank TD Ameritrade for data through their academic collaboration program. We thank Zi Chao and Mehrnoush Shahhosseini for excellent research support. Ye acknowledges the support of National Science Foundation grant 1352936 (jointed with the Office of Financial Research at U.S. Department of the Treasury). Chawla and Xu acknowledge the support of NSF Grant IIS-1447795. This work also uses the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number OCI-1053575. We thank Robert Sinkovits of the San Diego Supercomputer Center for assistance with supercomputing, which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program.

We were witnessing another explosion of technological innovations that facilitate interpersonal communication, consisting of e-mail and chat rooms, and after 2000 these led to social media… These new and effective media for interactive (if not face-to-face) communication may have the effect of expanding yet again the interpersonal contagion of ideas. They may have allowed enthusiasm for the market to spread much more widely than it would otherwise have. Certainly we are still learning how to regulate the use of these new media in the public interest.

Robert Shiller

Irrational Exuberance, Third Edition (2015)

In the past, investors learned the news through reading newspapers, watching television, or by word-of-mouth. The advent of social media fundamentally changed how information is produced and disseminated in financial markets (Dugast and Foucault 2016). In this paper, we focus on the role of social media on information dissemination.[1] Take Twitter as an example; people receive short electronic messages known as "tweets" through the Twitter accounts they follow, and they can re-transmit the message by retweeting. Does information diffusion through social media affect investors' trading behavior? If so, what is the impact of such information diffusion on return and liquidity? We provide the first analysis of the impact of information diffusion in social media on the stock market by tracking the history of tweets and their retweets.

Under the semi-strong form of market efficiency, information diffusion through Twitter

[1] Dugast and Foucault (2016) consider the role of social media on information production. They show that a decline in the cost of noisy information contained in social media can reduce the demand for precise information and, for this reason, the informativeness of the asset price in the long run.

should be inconsequential. The information in tweets is considered to be stale or becomes stale almost instantly after the initial tweet. Rational traders should not react to tweets and/or retweets, and the transmission of news via Twitter should not affect trading volume, return, and/or liquidity.

The information contained in tweets and retweets is public by default, and public information can affect trading behavior, return, and/or liquidity through two channels: the new information channel and the stale news channel. Under the new information channel, public news is not incorporated into prices until investors pay attention (Peng 2005). Hirshleifer and Teoh (2005), Hirshleifer, Lim, and Teoh (2009), DellaVigna and Pollet (2009), and Peng and Xiong (2006) find that investors' limited attention causes *underreactions* to news, that is, the slow adjustment of price to fundamental value. The information channel implies a permanent adjustment of prices, although the particular path of adjustment depends on whether traders also learn from the price (Han and Yang 2013; Manela 2014), and the existence of other types of traders may cause overshooting and partial reversal (Hong and Stein 1999). Hong, Hong, and Ungureanu (2012) predict that return increases with the speed of information diffusion. Their model also implies that liquidity decreases with the speed of the diffusion, because faster diffusion increases the number of informed traders relative to uninformed traders, thereby the level of information asymmetry.

Under the stale news channel, tweets and retweets affect trading, return, and liquidity if investors *overreact* to stale news. Ho and Michaely (1988), Huberman and Regev (2001), Gilbert (2010), and Tetlock (2011) find that the stock market reacts to previously published news, suggesting that relevant information is neglected at the time of the previous news. Under the stale news channel, the irrational exuberance generated through social media can temporally move the price away from fundamentals. As stale news only generates noise trading, liquidity increases with the speed of diffusion.

We find that news diffusion through Twitter is associated with price pressure and then reversal, along with an increase in liquidity. We begin our analysis by constructing a unique dynamic measure of information diffusion. The construction of the measure starts from tracking the increase in the number of retweets as well as the number of followers reached by the retweets over time. The empirical cumulative distribution function (CDF) of the information diffusion is the fraction of retweets or the followers reached by retweets in a given interval (e.g., 10 minutes) relative to a terminal time (e.g., one hour or three hours). For example, if the tweet reaches 200,000 users in the first 10 minutes and 1 million users in an hour, the information diffusion speed in the first 10 minutes is 0.2. This CDF function is essential in theoretical models on information diffusion (Hong and Stein 1999; Hong, Hong and Ungureanu 2012; Manela 2014; Andrei and Cujean 2016), and we provide the first direct proxy for the CDF function over time.

We then define trading intensity as the CDF function of trading volume. We find a strong correlation between information diffusion speed and trading intensity, even after controlling for information diffusion through traditional media outlets, time of day, recent turnover and volatility, past and contemporaneous returns, and various fixed effects. A 1% increase in information diffusion speed is associated with a 0.14% increase in trading intensity. The number increases to 0.21% for retail trading intensity, and to 0.36% for 20% of retail investors with the largest stock holdings.

We also find a strong positive contemporaneous relationship between information diffusion and stock returns. The more Twitter users a tweet reaches, the higher the stock returns on that day (from 10 minutes after the tweet to market close). However, we find that the higher returns are completely reverted the next day. The price overshooting and reversal are driven mostly by smaller stocks. Qualitatively, our results are similar to price pressure led by retail attention in Barber and Odean (2008), Tetlock (2011), and Da, Engelberg, and Gao (2011), but the whole cycle of price pressure and reversal transpires at a much higher speed. A researcher using low-frequency data would find that social media has no impact on stock returns. Information transmitted via social media reduces the cost of paying attention, which expedites but does not eliminate the retail-induced price pressure and its reversal. A further support for the price pressure interpretation is the greater decrease in the bid-ask spread for stocks whose tweets reach more users. The spread decrease is consistent with lower adverse selection risk as retail trading picks up.

Certainly, it is possible that higher return leads to faster retweets, or both returns and information diffusion are driven by the same unobserved factors. Fortunately, the field of computer science discovers a number of predictors of diffusion speed that do not directly predict volume, return, and liquidity (Suh et al. 2010; Petrovic, Osborne, and Lavrenko 2011; Jenders, Kasneci, and Naumann 2013; Cheng et al. 2014). For example, the network structure affects information diffusion speed. A tweet retweeted by users with more followers in the first 10 minutes will diffuse more rapidly afterwards. We also find that a tweet that comes from an active Twitter account generating many new tweets per day diffuses slowly, consistent with the investor distraction hypothesis proposed by Hirshleifer, Lim, and Teoh (2009). We find that Tweets with pictures or hashtags diffuse more rapidly, and that a tweet with a URL link diffuses more slowly. One interpretation is that Tweets with pictures or hashtags attracts attention, but it takes time to read the linked article in URL. Since pictures, hashtags and URLs are independent of future returns and not directly related to trading or valuation, we use them to instrument our diffusion measure.

We multiply the predicted information diffusion rate using these instruments by the number of Twitter users a tweet reaches within 10 minutes. This generates the same price overshooting and reversal pattern. Surprisingly, the total number of Twitter users a tweet can reach within 10 minutes alone does not have predictive power on the return and liquidity. Therefore, it is the information diffusion speed that is predictive of the return and liquidity, which again highlights the importance of having a dynamic, not static measure of information diffusion.

We then check the robustness of our results using an out-of-sample exercise. We use only data from the first six months of our sample period (2013/11-2014/04) to run the predictive regression and then apply the regression coefficients to the next six months (2014/05-2014/10). The predicted information diffusion rate is therefore free of forward-looking bias and can be computed in real-time. We then detect positive price pressure and the subsequent reversal of the retweeting effects for the second half of our sample (2014/05-2014/10).

Taken together, we show that the role of Twitter is to spread stale news. Trading according to the information in Twitter harms an investor's wealth, unless the investor can react earlier in the diffusion process and quickly reverse the position within the day.

Our paper contributes to the burgeoning literature on social media. The unique feature of social media is the dynamic and interactive way information is distributed from user to user. This dynamic feature differentiates social media from other information dispersal methods (Shiller 2015). Yet most researchers treat Twitter as a traditional information distribution outlet. Sprenger and Welpe (2010), Bollen, Mao and Zeng (2011), Brown (2012), Mittal and Goel (2012), Rao and Srivastava (2012), Nann, Krauss and Schoder (2013), and Oliveira et al. (2013) apply text analysis to generate new sentiment measures. On the other hand, Blankespoor, Miller and White (2014), Bhagwat and Burch (2015), and Chen, Hwang, and Liu (2016) consider social media as alternative way to release information, which reduces information asymmetry, improves stock liquidity, and attracts more investors. Under these two approach, only the initial information release matters. A notable exception is Giannini, Irvine and Shu (2013), who use Twitter to track the change in investor disagreement, but they do not track the spread of information across agents. To the best of our knowledge, we are the first to analyze Twitter's dynamic features of information diffusion using a financial perspective. This dynamic feature, in turn, allows us to contribute to two other lines of literature: information diffusion and social interaction.

The key variable in the theoretical information diffusion literature is the portion of investors who know the information relative to the portion who do not. The change in the portion over time serves as the main driver for volume, return, and liquidity (Hong and Stein 1999; Hong, Hong, and Ungureanu 2013; Han and Yang 2013; Manela 2014; Andrei and Cujean 2016). Yet the existing proxies for information diffusion are all static. By tracking tweets and retweets over time, we contribute to the literature by proposing the first dynamic proxy for information diffusion.

Word-of-mouth communication has been shown to affect market outcome. In the theoretical literature, the information exchange can happen through a network of friends (Ozsoylev and Walden 2011; Bildik et al. 2013), or through information percolation, a random meeting of agents without a network (Duffie and Manso 2007; Duffie, Malamud, and Manso 2009). A fundamental challenge for empirical work is to document word-of-mouth communication, in which one agent receives a piece of information from another agent, and then transmits it to a third agent. There are indirect proxies of word-of-mouth communication such as physical proximity (Hong, Kubik, and Stein 2005; Ivkovich and Weisbenner 2007; Brown et al. 2008), sociability (Ivkovich and Weisbenner 2007), ownership (Shive 2010), common schooling (Cohen, Frazzini, and Malloy 2008), coworkers (Hvide and Östberg 2015), and correlated stock trades (Ozsoylev et al., 2014). One challenge of word-of-mouth communication literature is to differentiate it with homophily, in which investors act alike because they share similar backgrounds, not because they share information. To the best of our knowledge, documentation of word-of-mouth communication only exists as anecdotes through criminal investigations on insider trading (Shiller 2015; Ahern 2016) or Ponzi schemes (Rantala 2015). The primary drawback of this approach is generality, because the information diffusion in illegal activity can be substantially different from normal information diffusion (Ahern 2016). Tweets and retweets provide the first direct proxy of word-of-mouth communication under normal conditions. Besides finding direct evidence that word-of-mouth communication affects stock return and liquidity, we also provide a direct proxy of information diffusion. This proxy may be valuable in direct empirical tests on theories of information percolation and network effects.

Studies document that limited investor attention affects trading behavior, return, and liquidity. Limited investor attention can lead to underreaction to actual news (Hirshleifer and Teoh

2005; Peng 2005; DellaVigna and Pollet 2009; Hirshleifer, Lim, and Teoh 2009), but it can also lead to overreaction to stale news (Gilbert 2010; Tetlock 2011). Our empirical results show that Twitter mainly serves the second function. We also find that the attention generated by tweets and retweets has a large impact on retail traders. In this sense, our paper contributes to the literature on retail attention (Barber and Odean 2008; Da, Engelberg, and Gao 2011).

We do not argue that social media makes stock prices less efficient, because price pressure and reversal have been documented for the era before Twitter. Instead, we maintain that information dispersal through Twitter speeds up the price pressure and reversal. The cycle of price pressure and reversal usually lasts for several days or even weeks (Barber and Odean 2008; Da, Engelberg, and Gao 2011), but in our paper price pressure accumulates within a day and completely reverses the next day. Our speeding up finding provides supportive evidence for Shiller (2015), who finds that the invention of the telephone sped up word-of-mouth communication in last century. Apparently, Twitter also speeds up the word-of-mouth communication because it is fast and interactive.

The rest of the paper is organized as follows. In Section 1, we describe the data. In Section 2, we propose the measure of information diffusion and examine its relation with trading, return, and liquidity. In Section 3, we provide a robustness check of the results using an instrumental variable approach. We conclude in Section 4.

1. Data description

Using Twitter's Streaming APIs,[2] we track the history of tweets and retweets of 78 major media outlets (e.g., @WSJ and @CNBC), 56 active Twitter accounts of S&P 1,500 CEOs and CFOs (e.g., Elon Musk of Tesla Motors), (Chen, Hwang, and Liu, 2013), and 143 Twitter accounts of S&P 500 companies (e.g., @TysonFoods). We focus on news from trustworthy outlets to avoid potential noise or even rumors from social media.[3] We captured all original tweets posted by any of these 277 accounts and their retweets from November 1, 2013 through October 31, 2014.

Table 1 reports the summary statistics for these 277 accounts. The 78 Twitter accounts from media outlets tend to have more followers, with a mean of 888,545 followers and a median of 100,446 followers. For example, @nytimes has more than 11 million followers and @WSJ has more than 4 million followers. The 143 official Twitter accounts of S&P 500 companies also have many followers, with a mean of 601,931 followers and a median of 125,521 followers. Both @Google and @Starbucks have more than 5 million followers apiece. Firm CEOs and CFOs have fewer followers; the mean is 54,576 followers and the median is only 621 followers. @ericschmidt, @RalphLauren, and @MichaelDell attract the most followers (779K, 672K, and 628K, respectively).

Insert Table 1 about Here

Table 1 also reports that the average number of years since inception is 5.7 for media outlet accounts, 5.3 for company accounts, and 4.3 for CEO/CFO accounts. Twitter accounts by media outlets are the most active with almost 7,488 tweets per year per account, followed by S&P 500

[2] https://dev.twitter.com/docs/api/1.1/post/statuses/filter.
[3] Dugast and Foucault (2016) theoretically examine the impact of earlier but noisy signal from social media versus later but more precise information.

company accounts (3,334 tweets per year per account). The Twitter accounts of CEOs and CFOs are the least active, with only 264 tweets per year per account.

To identify potentially influential tweets, we apply the following filters to the tweets:

1) Having been retweeted more than 50 times.

2) Having been posted during extended trading hours (4:00 to 20:00 ET).

3) Mentioning at least one company that is in the Russell 3000 index.

4) If multiple events happened to the same company, we pick the one with the most retweets. If multiple tweets about the same event are captured, we pick the one that was sent the earliest.

We carry out steps (1) and (2) automatically using a computer script, and we perform steps (3) and (4) manually (e.g., distinguishing the mentions of the tech company "Apple" from the fruit "Apple"; identifying different tweets that discuss the same topic, etc.). The selection process leaves us with 1,261 tweets. These tweets originate from 115 Twitter accounts and cover 178 distinct stocks. Table 2 contains a few sample tweets, which cover a wide range of news (mergers and acquisitions, earnings announcements, product launches, independent research, etc.).

Insert Table 2 about Here

Of the 115 distinct Twitter accounts, @WSJ generates the most tweets in our sample (270), followed by @Forbes (129), and @CNBC (83). Of the 178 distinct stocks, Apple (AAPL) appears most frequently (92 times) in tweets, followed by Facebook (FB, 88 times), Google (GOOG, 82 times), Twitter (TWTR, 81 times), Microsoft (MSFT, 67 times), and Tesla (TSLA, 64 times). Table 3 presents summary statistics on the stocks in our sample. The average stock size is at the

90th percentile of the CRSP universe. The average institutional ownership is also large, at 57.7%, corresponding to the 80th percentile of the CRSP universe. The volatility of the average stock in our sample is similar to that of an average stock in the CRSP universe but has higher turnover.

Insert Table 3 about Here

In Panel A of Figure 1, we plot the number of retweets over time during the first hour after an original tweet, with each time interval representing 10 minutes. The median tweet in our sample will be retweeted 68 times by the end of the first hour. The small number of 68 retweets, however, reaches 3 million people, because many accounts that retweet the news also have a large number of followers (Panel B of Figure 1).

Insert Figure 1 about Here

We use the NYSE Daily Trade and Quote (DTAQ) database to construct the complete National Best Bid and Offer (NBBO). The DTAQ provides two files that contain official NBBO quotes. If a single exchange has both the best bid and the best offer, then the official NBBO will be recorded in the DTAQ Quotes File. Otherwise, the NBBO quotes will be recorded in the DTAQ NBBO file. Following the procedure proposed by Holden and Jacobsen (2014), we combined the NBBO quotes from both files to construct the complete official NBBO.[4] We then compute bid-ask spreads and intraday returns using bid-ask midpoints.

As market-wide intraday retail trading volume data are not directly available, we use the trading volume from the Trade Report Facility (TRF) as a proxy for the market-wide retail trading

[4] We exclude quotes with abnormal quote conditions (A, B, H, O, R, and W). We delete any quote with a bid that is greater than or equal to the ask. We also delete cases in which the quoted spread is greater than $5.00.

volume. The results are supplemented by a proprietary dataset on retail trading from TD Ameritrade (TDA). The market-wide proxy is constructed based on the empirical finding of Battalio, Corwin, and Jennings (2015) that non-direct limit and market orders are seldom routed to public exchanges but are often internalized by broker-dealers. Therefore, we use TRF volume (exchange symbol D in the TAQ dataset) as our proxy for market-wide retail trading. This measure has two limitations. First, TRF volume also contains volume from dark pools (Kwan, Masulis, and McInish 2015). Second, Battalio, Corwin, and Jennings (2016) find that some retail brokers route orders to public exchanges, including TDA. Therefore, we supplement our market-wide proxy of retail trading with a proprietary dataset from TDA. This dataset includes 331 million de-identified transactions made by 2.8 million clients from June 1, 2010 to June 10, 2014.

We use Ravenpack data to control for news coverage on other media outlets. Following Hafez (2009), we only keep news events with a novelty score of 100 and relevance score above 75. For each tweet in our sample, we then count the amount of news on the same stock on the day of the tweet up to the time of the tweet and also trace how this news count changes after the tweet. This allows us to measure both the amount of news coverage in other media outlets and how this coverage changes over time.

2. Regression Analysis

In this section, we first define our measure of the information diffusion speed. Next, we examine two questions: (1) How does the information diffusion speed affect trading intensity, especially among retail investors? (2) How does the information diffusion speed affect asset prices and stock liquidity?

2.1 Definitions of Information Diffusion Speed and Trading Intensity

The driver of the asset-pricing dynamics in information diffusion models is the proportion of agents who know the information earlier than others, which is characterized by the CDF function (Hong, Hong, and Ungureanu 2012). Yet the empirical literature does not include a dynamic proxy for the CDF function. In this paper, we provide the first direct proxy for the CDF function to fill this void.

For each tweet in our sample, we calculate the total number of its retweets within the first hour. Next, we divide the hour into six 10-minute intervals and calculate the number of retweets in the 10-, 20-, 30-, 40-, 50-, and 60-minute intervals relative to the total of retweets within the first hour. By construction, this number is 1 after 60 minutes. A fast diffusion of information implies a quick convergence to 1 over time. For simplicity, we classify a diffusion process as fast if more than 60% of total first-hour retweets occur in the first 10 minutes; we classify a diffusion process as slow if less than 40% of total first-hour retweets occur in the first 10-minute interval. The result is robust under other specifications. Panel A of Figure 2 presents the average information diffusion speed of the fast and slow diffusion in our sample.

Insert Figure 2 about Here

Similarly, trading intensity is also a normalized measure. We divide the cumulative volume in the first 10, 20, 30, 40, 50, and 60 minutes by the total volume in the first hour. Panel B of Figure 2 presents the average trading intensity of the fast and slow information diffusion in our sample.

2.2 Trading Intensity Increases with Information Diffusion Speed

We next examine how the information diffusion speed is related to trading intensity. We discuss the results for aggregated volume and then retail volume.

2.2.1 Information Diffusion Speed and Total Volume

Figure 2 shows that 25.0% of the first-hour trades take place in the first 10 minutes after a tweet for the fast diffusion case. In contrast, only 13.4% of the first-hour trading takes place in the first 10 minutes for the slow diffusion case.

To overcome any potential omitted variable bias, we control for other factors that can drive the correlation between diffusion rate and trading intensity. For example, both retweets and trading intensity may have intraday seasonality, and we control for this seasonality using time dummies. Also, extreme returns immediately following a tweet could trigger both retweets and trading. In addition, the information diffusion rate may be correlated with news coverage for the same stock in other media outlets.

We control these variables in Panel A of Table 4. The dependent variable, the percentage of first-hour total trading that occurs in the first 10 minutes following a tweet, measures trading intensity. The main independent variable, diffusion, measures the percentage of first-hour retweets that occurs in the first 10 minutes. Other control variables include a dummy variable equal to 1 if the tweet takes place before 9:30 ET (*pre-market*); a dummy variable equal to 1 if the tweet takes place between 12:30 and 16:00 ET (*afternoon*); a dummy variable equal to 1 if the tweet takes place after 16:00 ET (*post-market*); log market capitalization (*size*); turnover (*turn*); daily return volatility in the past 30 days (*volatility*); book-to-market ratio (*bm*); absolute stock returns over the market in the past hour (*abs past 1h ret*); absolute stock returns over the market in the first 10 minutes after the tweet (*abs 10m ret*); log number of media coverage on the same day but prior to the tweet (*media*); and percentage of first-hour media coverage for the stock that occurs in the first 10 minutes after the tweet (*media_diffusion*). We include stock and Twitter account fixed effects in our regression. The standard error is clustered by ticker. The sample covers 1,261 tweets during one year from 2013/11 to 2014/10.

Insert Table 4 about Here

In Panel A of Table 4, the results confirm the strong unconditional correlation between diffusion rate and trading intensity observed in Figure 2. A 1% increase in the diffusion rate leads to a 0.3% increase in trading intensity, with a *t*-value of 5.13. Once we control for time of day in column (2), the effect attenuates to 0.17% but is still highly significant. The number reduces to 0.14 but remains significant (*t*-value = 2.25) after we control for size, turnover, volatility, and book-to-market in Column (3).

Surprisingly, the results in column (4) show that neither the amount of news coverage in other media outlets nor its diffusion rate drives the trading intensity. After controlling for the information diffusion in Twitter, the coefficient of media coverage is negative, although not statistically significant. Plus, the inclusion of news coverage in other media and media diffusion has no impact on the coefficient before Twitter, which remains at 0.14, with a *t*-value of 2.27.

These results are consistent with the Shiller's (2015) conjecture that conventional media—print media, television, and radio—have limited ability to generate active behavior. We provide evidence to support Shiller's (2015) claim that interpersonal and interactive communications have the most powerful impact on behavior. To the best of our knowledge, there is no horse race between these two ways of spreading news in the literature other than a survey conducted by Pound and Shiller (1989). Our paper provides direct empirical evidence that the attention and action of investors is more stimulated by interactive communications.

2.2.2 Diffusion Speed and Retail Volume

We find an even stronger link between diffusion speed and trading intensity for retail investors. While retail investor trading volume is not directly available, we compute trading intensity using only TRF volume and use it as the dependent variable in the regressions.

In Panel B of Table 4, we find that a 1% increase in information diffusion speed is associated with a 0.21% increase in retail trading intensity, after controlling for other factors (Column 4). Therefore, the link between information diffusion speed and retail trading intensity is 50% stronger for retail investors than for all investors.

We then take advantage of a unique brokerage account dataset from TDA that includes 331 million transactions made by 2.8 million clients from June 1, 2010 to June 10, 2014. The data have been provided by TDA through an academic data collaborative agreement. Individual clients are not identified in the data, although demographic characteristics such as age and gender are included for each anonymous ID. We are also able to track the history of trading from an individual through this unique ID. While trades in the TDA data represent a subset of all trades, it is a relatively clean subset of retail trading.

We merge our tweet sample with TDA brokerage-account-level transaction data for the overlapping period from 2013/11 to 2014/06.[5] We focus only on stock trades from "individual" accounts in TDA.[6] Since investors at TDA rarely trade during after-hour sessions, we focus on tweets during market trading hours (9:30 to 16:00 ET). We examine only accounts that trade the corresponding stock at least once during the first hours after a tweet. Altogether, our selection criteria result in a merged dataset that contains 331 tweets and trades from 35,443 individual TDA accounts.

Insert Table 5 about Here

Panel A of Table 5 provides summary statistics for our merged sample. The average individual TDA account holder is ~49 years old in 2014. The first age quartile is 38 and the third quartile is 58. Their median stock holdings with TDA are worth $20,000 with 25% holding stocks worth more than $74,000. When they trade during the first hour after a tweet, they are more likely to buy. Both the mean and median of the net trade variable are positive (with *t*-value > 5.00). This finding provides direct support for Barber and Odean (2008), who argue that retail attention leads to positive price pressure on average since retail investors are less likely to short. On average, 20%

[5] No effort was made to cross-reference TDA accounts with Twitter accounts. The data set enables us to compare the behavior of TDA traders and Twitter users, but cannot indicate whether an individual trader did or did not have access to Twitter.

[6] TDA also records trading data for options, bonds, warrants, mutual funds, and other securities, although these transactions represent less than 30% of all trades in our sample. Other account types include "Joint Tenants WROS," "IRA," "Rollover IRA," "Trust," "Roth IRA," etc. Trades from "individual" accounts represent almost half of all trades in our sample.

of all first-hour trades take place during the first 10 minutes following the tweet.

The bottom half of Panel A of Table 5 reports the summary statistics for the 331 tweets in our merged sample. On average, we observe trades from almost 200 individual accounts following a tweet. Seventy-two percent of the trades come from male account holders. The average account holder is ~51 years old, holding about $91,000 in stocks, and is more likely to buy stocks (rather than sell them). Twenty percent of all first-hour trades take place during the first 10 minutes following a tweet.

We repeat the regressions in Table 4 for our merged sample and report the results in Panel B of Table 5. We measure trading intensity using the following regressions: (I) all trades from TAQ; (II) all TDA trades; (III) all TDA trades of female investors; (IV) all TDA trades of young investors (age <35); and (V) all TDA trades of "rich" investors (whose stock holding is greater than $100,000). We include the same set of control variables as those used in the regressions for Table 4. Since we focus on tweets during the normal trading hour, the pre-market and post-market dummies drop out.

Column I of the Panel B of Table 5 confirms the strong correlation between diffusion speed and trading intensity measured using all trades in our merged sample of 331 tweets. A 1% increase in the information diffusion rate leads to a significant 0.16% increase in trading intensity. The coefficient of 0.16 is higher than the corresponding coefficient of 0.14 using the larger 1,261 tweet sample (see Panel A of Table 4), possibly because we focus on trading hours.

The results of regression II suggest a much stronger link between information diffusion speed and retail trading intensity measured using all TDA trades. The coefficient for the information diffusion variable increases from 0.16 to 0.23. The results of regression III suggest an even stronger link between information diffusion speed and retail trading intensity among female investors, who account for less than 30% of all TDA investors in our sample.

The strongest link between information diffusion speed and retail trading intensity we find in our sample is among the 20% of TDA investors with the largest stock holdings. For them, a 1% increase in the information diffusion rate leads to a 0.36% increase in trading intensity. This is not surprising as traders with higher investments in stocks should be more attentive to financial news and thus react more quickly to that news.

Surprisingly, the weakest link between diffusion speed and retail trading intensity we find is among the 20% of TDA investors who are younger than age 35, who should be more frequent users of Twitter. For them, a 1% increase in the information diffusion rate leads to only a 0.17% increase in trading intensity, for two possible reasons. First, younger investors typically have fewer financial resources and therefore fewer investment assets and lower investment value, which means they may be constrained by commissions or other fixed transaction costs. Attention is one such cost. Investors with less valuable investments may have a weaker incentive to follow a particular firm. Second, compared with average TDA investors in our sample, who are close to retirement age, younger investors have to focus more on work during trading hours and thus may not trade immediately after a tweet.

Overall, the TDA data we examine provide direct support for the concept that diffusion speed measured using retweets is more strongly related to retail trading.

2.3 Diffusion Speed, Stock Returns, and Liquidity

Insofar as diffusion speed measured using retweets relates to trading intensity, we next examine how it affects prices and dollar bid-ask spreads. We measure contemporaneous stock returns (in excess of the market) from 10 minutes after an initial tweet until the end of the day (*CAR%*[*10m, close d0*]). We skip the first 10 minutes after the tweet to avoid mechanical correlation in the next section. We also measure stock returns (in excess of the market) on the next trading day (*CAR%*[*close d0, close d1*]). We then examine the change in stock liquidity as the average dollar bid-ask spread during the three hours after a tweet minus the average dollar bid-ask spread during the hour before the tweet.

We measure information diffusion speed using the total number of Twitter users the tweet can reach after three hours (*diffusion_3hr*).[7] If an influential Twitter user with 5,000 followers retweets, the number of Twitter users the tweet can reach will increase by 5,000. Focusing on a three-hour horizon after a tweet makes the measure comparable across tweets. We find similar results when we measure the level of retail attention until the end of the day, as most of the retweets take place during the first three hours after the initial tweet.

We then regress contemporaneous stock returns, stock returns the next day, and the dollar spread change on information diffusion speed in panel regressions. To avoid the mechanical effect that more breaking news leads to both higher returns and diffusion speed, we control for both stock returns over the market in the first 10 minutes after a tweet (*10m ret*) and absolute value of 10-

[7] Here the diffusion speed is in absolute amount but not the CDF function, because we are interested in the absolute level of return.

minute return (*abs 10m ret*). As a result, the return is measured relative to the return over the first ten minutes after a tweet. We also control for stock returns over the market in the past hour (*past 1h ret*) and absolute value of past hour returns (*abs past 1h re*t). The variable *media* measures the log number of media coverage on the same day and up to three hours after the tweet. Other control variables include *pre-market*, *afternoon*, *post-market*, *size*, *turn*, *volatility*, and *bm*.

Insert Table 6 about Here

Panel A of Table 6 reports the regression results from the full sample. We observe a positive and significant association between information diffusion speed and contemporaneous returns. A one-standard-deviation increase in our retail attention measures (*diffusion_3hr*) leads to a 23 bps increase in contemporaneous-day returns.[8] Interestingly, the higher returns that result due to a tweet revert completely the next day. Such temporary price overshooting and subsequent reversal is consistent with the stale news channel documented by Tetlock (2011). This finding is also consistent with the retail attention mechanism documented by Da, Engelberg, and Gao (2011).

The economic magnitude of the price pressure is similar to that in Da, Engelberg, and Gao (2011). Yet the cycle of price pressure and reversal ends at a much higher speed in Twitter. In Da, Engelberg, and Gao (2011), the price pressure and reversal cycle lasts for two weeks, but this cycle lasts for less than a day in our sample. Researchers using low-frequency data may not be able to detect the price pressure and reversal cycle. Taken together, our results are consistent with the hypothesis that tweets spread stale news among investors, which generate price pressure followed

[8] We multiply the coefficient on *diffusion_3hr* (21 bps) by its standard deviation (1.09).

by reversal. The Twitter platform, however, speeds up the decimation of stale news.

We also find that a one-standard-deviation increase in the information diffusion speed decreases the bid-ask spread by 5.45 bps points (1.09*0.05), which provides further support for the stale news channel. This result is in contrast to Hong, Hong and Ungureanu (2012), who find that fast information diffusion increases the number of informed investors relative to the number of uninformed investors, which reduces liquidity. Under the stale news channel, noise trading is generated through information diffusion, which can increase liquidity.

The return and liquidity results are much more pronounced among stocks with a market capitalization that is below the median.[9] Panel B of Table 6 shows that a one-standard-deviation increase in diffusion measures (*diffusion_3hr*) leads to a much higher 46 bps increase in contemporaneous-day returns.[10] Again, the price pressure completely reverted the day following a tweet. Panel B of Table 6 also shows an even greater decrease in the bid-ask spread after a tweet concerning smaller stocks.

3. Instrumental Variables Approach

In the previous section, we establish a correlation between information diffusion speed and trading, return, and liquidity. Yet, correlation does not necessarily imply causality. Fortunately, researchers in computer science have developed advanced machine-learning techniques that can be employed to predict information cascades on large social networks. A number of these

[9] Note that the average market capitalization of smaller stocks in our sample is still at the 82th percentile of the CRSP universe, which represents large stocks by traditional measures.
[10] We multiply the coefficient on *diffusion_3hr* (39 bps) by its standard deviation among smaller stocks (1.17).

predictors do not have a direct relation with trading, return, and liquidity. We use these instrumental variables to first generate a predictive value for the information diffusion speed, and then examine whether it can be used to forecast trading, return, and liquidity. We use ordinary least squares (OLS) to predict the diffusion speed so that our method is similar to the two-stage least squares (2SLS) regression.[11] More sophisticated machine-learning techniques such as support vector machine, neural networks, and decision tree-based algorithms provide stronger statistical power, and are available upon request.[12]

The dependent variable of interest is the future information diffusion rate on Twitter. Specifically, the growth rate is calculated as log(*diffusion_3hr*) – log(*diffusion_10m*), where *diffusion_10m* and *diffusion_3hr* are the number of users a tweet potentially reaches after 10 minutes and after three hours, respectively.

There are two types of variables that can be employed to predict information diffusion speed. Suh et al. (2010), Petrovic, Osborne and Lavrenko (2011), and Jenders, Kasneci and Naumann (2013), among others, rely mostly on the content and source of a tweet, while Cheng et al. (2014) suggest that how information diffuses in the first few minutes after a tweet (also known as temporal features) and the characteristics of people who have retweeted are also crucial factors in predicting information cascades.

For the content of a tweet, we include in the regression a dummy variable equal to 1 if the tweet contains a picture (*HasPicture*), a dummy variable equal to 1 if the tweet contains hashtags

[11] We thank Paul Tetlock for his suggestion to simplify our method so that our analysis is more approachable to a finance audience.

(*HasHashtags*), and a dummy variable equal to 1 if the tweet contains URL links (*HasURLs*). Tweets with pictures or hashtags should disseminate faster because they typically grab users' attention. On the other hand, a tweet with a URL link should diffuse more slowly as it takes time to read the linked material. Certainly, we cannot fully rule out the possibility that a tweet with a picture or hashtag is more important, whereas news with URL is less important. Yet we suggest that these three variables should not have a direct effect on trading, return, and liquidity other than through the information diffusion rate.

For the temporal features and characteristics of people, we first find the Twitter account for the last five retweets before the 10-minute cut-off time. The variable log(*# of followers of recent retweeters*) is defined as the log of the total number of followers for these five Twitter accounts. Intuitively, a retweet from an account with more followers tends to disseminate faster. As the network of followers is established well before news is tweeted, it is unlikely to be affected directly by trading, return, and liquidity. Motivated by the investor distraction hypothesis put forth by Hirshleifer, Lim and Teoh (2009), we also include the average daily number of tweets sent by a Twitter account (*Total # of tweets*).

The regressions for the results in Table 7 include the inverse of the average time lapse between the five most recent retweets before the 10-minute cut-off time (*Speed of recent retweets*), the hour of the tweet (*Hour*), a dummy variable equal to 1 if the tweet is sent from the West Coast (*IsWest*), and a dummy variable equal to 1 if the tweet is sent by a CEO (*IsCEO*). The results show that tweets with pictures or hashtags increase the information diffusion speed, whereas inclusion of a URL decreases information diffusion speed. If an initial tweet is retweeted quickly by users with more followers, the retweet will disseminate more quickly. Tweets sent from a Twitter account generating more tweets per day will disseminate slower, consistent with the "driven-to-distraction" hypothesis of Hirshleifer, Lim, and Teoh (2009). In addition, if recent retweets are posted in rapid-fire fashion, the initial tweet will disseminate faster. The predictive power of these temporal features is consistent with the findings in Cheng et al. (2014).

Insert Table 7 about Here

Our predictive variables are all observable during the first ten minutes after an initial tweet, and thus are independent of future returns measured after 10 minutes. In addition, most of these variables are not directly related to the value and liquidity of a stock. We therefore use them to instrument our information diffusion speed.

Specifically, we first compute the predicted attention diffusion rate from the regression. We then multiply the diffusion rate by the total number of Twitter users a tweet can reach after 10 minutes and use this product in our analysis. Intuitively, this product measures the expected number of users the tweet can reach using the information set available 10 minutes after the initial tweet. We then link the predictive retail attention to contemporaneous and future stock returns using the same panel regressions that we use for Table 6.

Insert Table 8 about Here

When we include the predicted diffusion rate in the regression, the results in Panel A of Table 8 exhibit the same price overshooting and reversal pattern as in Table 6. Interestingly, we do not detect this pattern when using only the total number of Twitter users the tweet can reach after 10 minutes in the regression for the results in Panel B of Table 8. In other words, the predicted information diffusion speed 10 minutes after a tweet is crucial for measuring the actual information diffusion speed.

Finally, we conduct a predictive out-of-sample exercise. We use data from the first six months of our sample period (2013/11-2014/04) to run the predictive regression and then apply the regression coefficients to the next six months (2014/05-2014/10) in computing the diffusion rate. The predicted retail attention measure is therefore free of forward-looking bias and can be computed in real time. We then link the predictive retail attention to contemporaneous and future stock returns using only the second half of our sample period.

Insert Table 9 about Here

The results in Table 9 show that the predicted diffusion speed forecasts the price pressure and subsequent reversal out-of-sample, thus providing even stronger evidence for the stale news hypothesis.

4. Conclusion

In this paper, we track the diffusion of news by monitoring how such news is tweeted and retweeted on Twitter. We find the diffusion speed to be highly correlated with intraday retail trading patterns. The resulting retail attention leads to lower bid-ask spreads and positive price pressure on the news day, but these effects are completely reverted the next day. The amount of retail attention that news generates on Twitter can be predicted using characteristics of users, accounts, and tweets. The fact that predicted retail attention generates similar results helps to alleviate concerns about reverse causality and endogeneity. Taken together, we show that the role of Twitter is to spread stale news. Twitter generates price pressure and reversal, albeit at a much faster speed than the cycle generated by traditional media. This finding sheds some light on the question raised by Shiller (2015) on the impact of social media on financial markets.

More broadly, we are among the first to construct a dynamic and direct measure of information diffusion and word-of-mouth communication. This measure can be applied in a number of ways to test the implications of information diffusion or social network theory. For example, we can test the differential impacts of learning from trading and learning from information diffusion. The advent of social media provides a unique opportunity for researchers to construct new measures that is hard to obtain using traditional media. Using our measure to test the implications of economic theory could prove very fruitful.

References

Ahern, Kenneth R., 2016, Information networks: Evidence from illegal inside trading tips, *Journal of Financial Economics*, forthcoming.

Andrei, Daniel, and Julien Cujean, 2016, Information percolation, momentum, and reversal, Working paper, University of California, Los Angeles and University of Maryland.

Bailey, Michael, Ruiqing Cao, Theresa Kuchler, and Johannes Stroebel, 2016, Social networks and housing markets, Working paper, Facebook Economist, Harvard University and New York University.

Barber, Brad M., and Odean Terrance, 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.

Battalio, Robert H., Shane A. Corwin, and Robert H. Jennings, 2016, Can brokers have it all? On the relation between make take fees & limit order execution quality, *Journal of Finance*, forthcoming.

Bhagwat, Vineet, and Timothy R. Burch, 2015, Pump it up? Tweeting to manage investor attention to earnings news, Working Paper, University of Oregon and University of Miami.

Blankespoor, Elizabeth, Brian P. Miller and Hal D. White, 2014, Initial evidence on the market impact of the XBRL mandate, *Review of Accounting Studies* 19, 1468-1503.Bollen, Johan, Huina Mao, and Xiaojun Zeng, 2011, Twitter mood predicts the stock market, *Journal of Computational Science* 2, 1-8.

Brown, Jeffrey R., Zoran Ivkovich, Paul A. Smith, and Scott Weisbenner, 2008, Neighbors matter: Causal community effects and stock market participation, *Journal of Finance* 63, 1509-1531.

Brown, Kenneth S., 2012, Cohomology of groups, Springer Science & Business Media.

Chen, Hailiang, Byoung-Hyoun Hwang, and Baixiao Liu, 2016, The economic consequences of having "Social Media Savvy" executives, Working paper, City University of Hong Kong, Cornell University and Florida State University.

Cheng, Justin, Lada A. Adamic, P. Alex Dow, Jon M. Kleinberg and Jure Leskovec, 2014, Can cascades be predicted?, *Proceedings of the 23rd international conference on World wide web*, 925-963. ACM..

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, 2008, The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116, 951-979.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *Journal of Finance* 66, 1461-1499.

DellaVigna, Stefano, and Joshua M. Pollet, 2009, Investor inattention and friday earnings announcements, *Journal of Finance 64,* 709-749.

Dugast, Jérôme and Foucault, Thierry, 2016, Data Abundance and Asset Price Informativeness, Working paper, Luxembourg School of Finance and HEC Paris

Duffe, Darrel, and Gustavo Manso, 2007, Information percolation in large markets, *American Economic Review* 97, 203-209.

Duffe, Darrel, Semyon Malamud, and Gustavo Manso, 2009, Information percolation with equilibrium search dynamics, *Econometrica* 77, 1513-1574.

Giannini, Robert Charles, Paul J. Irvine, and Tao Shu, 2015, The convergence and divergence of investors' opinions around earnings news: Evidence from a social network, Working paper, Blue Crest Capital Management, Neeley School of Business and University of Georgia.

Gilbert, Thomas, Shimon Kogan, Lars Lochstoer, and Ataman Ozyildirim, 2012, Investor inattention and the market impact of summary statistics, *Management Science* 58, 336-350.

Hafez, P.A., 2009, Construction of market sentiment indices using news sentiment. *RavenPack International S.L.*

Han, Bing, and Liyan Yang, 2013, Social networks, information acquisition, and asset prices, *Management Science* 59, 1444-1457.

Hong, Dong, Harrison Hong, and Andrei Ungureanu, 2013, An epidemiological approach to opinion and price-volume dynamics, Working paper, Singapore Management University and Princeton University .

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143-2184.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2005, Thy neighbor's portfolio: Word‐of‐mouth effects in the holdings and trades of money managers, *The Journal of Finance* 60, 2801-2824.

Hong, Harrison, Jeffrey D. Kubik, and Jeremy C. Stein, 2008, The only game in town: Stock-price consequences of local bias, *Journal of Financial Economics* 90, 20-37.

Huberman, Gur and Tomer Regev, 2001, Contagious speculation and a cure for cancer: A nonevent that made stock prices soar, *Journal of Finance* 56, 387–396.

Hvide, Hans K., and Per Östberg, Social interaction at work, 2015, *Journal of Financial Economics* 117, 628-652.

Ivkovich, Zoran., and Scott Weisbenner, 2007. Information diffusion effects in individual investors’ common stock purchases: Covet thy neighbors’ investment choices, *Review of Financial Studies* 20, 1327-1357.

Jenders, Maximilian, Gjergji Kasneci, and Felix Naumann, 2013, Analyzing and predicting viral tweets, *Proceedings of the 22nd international conference on World Wide Web companion*, 657-664. ACM.

Kwan, Amy, Ronald Masulis, and Thomas H. McInish, 2015, Trading rules, competition for order flow and market fragmentation, *Journal of Financial Economics* 115:330-348.

Manela, Asaf, 2014, The Value of diffusing information, *Journal of Financial Economics* 111, 181-199.

Mittal, Anshul, and Arpit Goel, 2011, Stock prediction using twitter sentiment analysis, Standford University, CS229 (http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf).

Nann, Stefan, Jonas Krauss, and Detlef Schoder, 2013, Predictive analytics on public data-the case of stock markets, Proceeding of 21st European Conference on Information Systems, 116.

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality, *Journal of Financial Economics* 100, 459-474.

Ozsoylev, Han N., and Johan Walden, 2011, Asset pricing in large information networks, *Journal* *of Economic Theory* 146, 2252-2280.

Ozsoylev, Han N., Johan Walden, M. Deniz Yavuz, and Recep Bildik, 2014, Investor networks in the stock market, *Review of Financial Studies* 27, 1323-1366.

Peng, Hanchuan, Fuhui Long, and Chris Ding, 2005, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Transactions on pattern analysis and machine intelligence* 27, 1226-1238.

Peng, Lin, and Wei, Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563-602.

Petrovic, Sasa, Miles Osborne, and Victor Lavrenko, 2011, RT to win! Predicting message propagation in Twitter, Proceeding of the Fifth International AAAI Conference on Weblogs and Social Media.

Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity*, 457-510.

Shiller, Robert J., 2015, Irrational exuberance. Princeton university press.

Shiller, Robert J., and John Pound, 1989, Survey evidence on diffusion of interest and information among investors, *Journal of Economic Behavior & Organization* 12, 47-66.

Shive, Sophie, 2010, An epidemic model of investor behavior, *Journal of Financial and Quantitative Analysis* 45, 169-198.

Suh, Bongwon, Lichan Hong, Peter Pirolli, and Ed H. Chi, 2010, Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network, *Social Computing (socialcom)*, 177-184. 2010 IEEE second international conference on Privacy, Security, Risk and Trust.

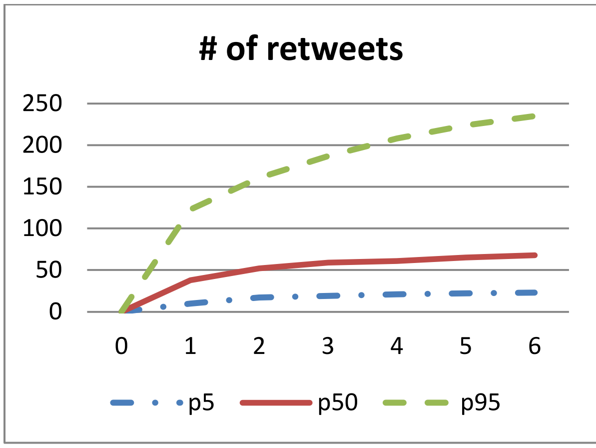
Tetlock, Paul C, 2011, All the news that's fit to reprint: Do investors react to stale information?, *Review of Financial Studies* 24, 1481-1512.

Tumasjan, Andranik, Timm O. Sprenger, Philipp G. Sandner, and Isabell M. Welpe, 2010, Predicting elections with Twitter: What 140 characters reveal about political sentiment, 178-185. Proceeding of the Fourth International AAAI Conference on Weblogs and Social Media.

Figure 1: Retweets During the First Hour

In Panel A, we plot the total number of retweets during the first hour after the original tweet, for the median case, for the 5th percentile, and for the 95th percentile. In Panel B, we also account for the number of followers of each Twitter account that posts the original tweet or the retweet. As a result, the number measures the number of potential users the tweet can reach in the first hour. Each time interval represents 10 minutes.

Panel A: Number of retweets



Panel B: Number of accounts reached

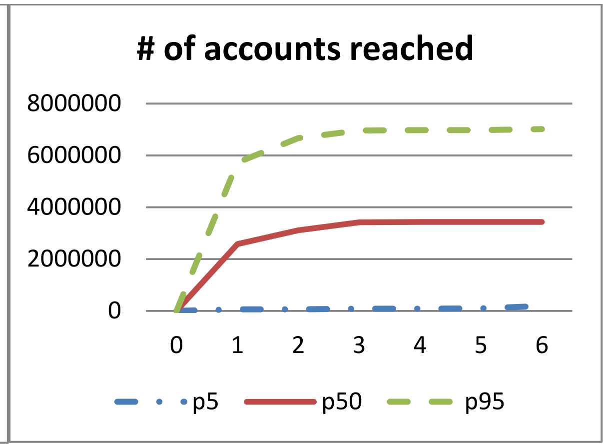


Figure 2: Fast and Slow Diffusion: Retweet Data

In Panels A and B, we plot the cumulative numbers of retweets and trading volumes for each of the six 10-minute intervals during the first hour following a tweet. Both variables are normalized by their totals during the first hour, so the plot resembles a cumulative distribution function (CDF). Rapid diffusion occurs when more than 60% of total first-hour retweets occur in the first 10 minutes; slow diffusion occurs when less than 40% of total first-hour retweets occur in the first 10 minutes.

Panel A: Diffusion Panel B: Trading volume

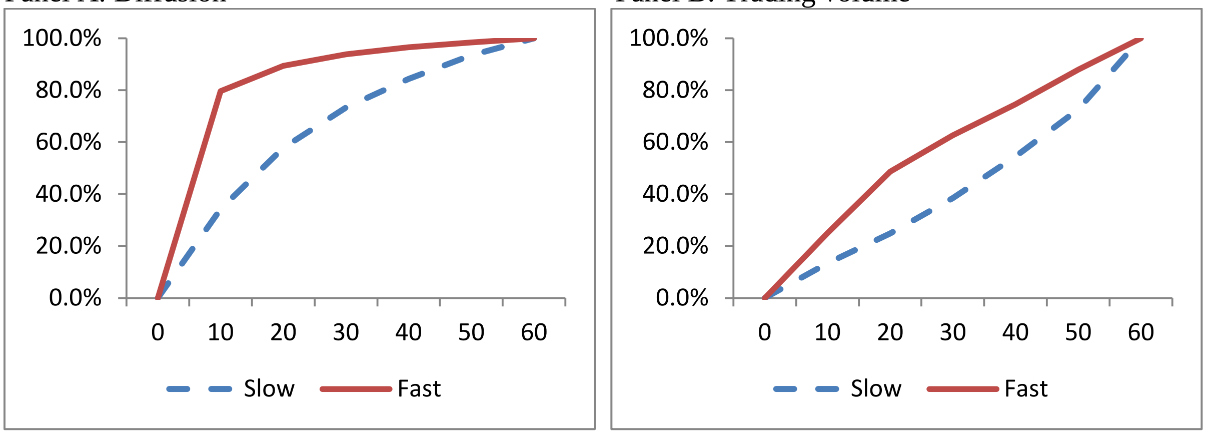


Table 1: Summary Statistics on Twitter Accounts in the Sample

This table reports summary statistics on the 277 Twitter accounts we monitored from 2013/11 to 2014/10. They include 78 major media outlets (e.g., @WSJ and @CNBC), 56 active accounts of S&P 1,500 CEOs and CFOs (e.g., Elon Musk), and 143 official Twitter accounts of S&P 500 companies (e.g., @TysonFoods).

	N	Mean	Std dev	Q1	Median	Q3
Total Number of Followers						
CEO/CFO	56	54,576	173,192	167	621	7034
Media	78	888,545	1,789,724	17,802	100,446	923,497
SP500	143	601,931	1,222,074	42,249	125,521	467,134
Number of Years Since Inception						
CEO/CFO	56	4.3	1.8	3.1	4.5	5.4
Media	78	5.7	1.7	5.1	5.6	6.6
SP500	143	5.3	1.7	4.9	5.4	6.0
Number of Tweets Per Year						
CEO/CFO	56	264	703	6	45	186
Media	78	7,488	6,312	2,157	6,076	11,821
SP500	143	3,334	6,987	788	1,413	2,985

Table 2: Examples of Tweets in Our Sample

This table contains examples of tweets in our sample. We report the date, source, and content associated with each tweet. The relevant tickers are also identified.

Date	Source	Tweet	Ticker
11/12/2013	*@WSJ*	AirWatch expresses interest in buying service division of Blackberry: http://t.co/R9vTFvfHkD	BBRY
11/14/2013	*@FordTrucks*	@Ford?F-150 EcoBoost hits 400,000 sales, saving 45 million gallons of gas annually: http://t.co/xYRgWVGoph?#BuiltFordTough	F
11/22/2013	*@paradimeshift*	Western Union and tradition bank wire transfers are dead! 11/23/13 $147 Million transferred for 37 CENTS! #bitcoin	WU
12/9/2013	*@ABC*	Just in: American Airlines/US Airways merger complete says company - @ABCaviation	AAL
12/19/2013	*@DavidJBarger*	Very cool @JetBlue's SJU Team welcomed N903JB, our first A321, "Bigger, Brighter, Bluer" to the airline! http://t.co/IU7JFJt9Y4	JBLU
1/9/2014	*@EMCcorp*	Congratulations to David Goulden - new CEO of #EMC. Joe Tucci will continue as Chairman & CEO of EMC Corporation http://t.co/no4P9BYOwT	EMC
1/29/2014	*@BreakingNews*	Facebook earnings: Q4 EPS $0.31 ex-items v. $0.27 estimate; revenues $2.59 billion v. $2.33 billion estimate - @CNBC http://t.co/sNqDbtfyzv	FB
2/5/2014	*@ReutersBiz*	Twitter reports revenue of $243 million, up 116 percent year-over-year	TWTR
2/19/2014	*@businessinsider*	TESLA EXPECTS 55% VEHICLE DELIVERY GROWTH IN 2014 http://t.co/aXQZAqHd0z	TSLA
3/4/2014	*@CNET*	2015 Lamborghini Huracan debuts with Nvidia-powered digital dashboard http://t.co/j7bvnt9JuH http://t.co/XlfBKsU85Q	NVDA

Table 3: Summary Statistics of Firms in our Sample

This table reports summary statistics on the stocks in our final sample. Market capitalization is measured in millions of dollars. Turnover and daily return volatility are measured over the past 30 days. Institutional ownership (IO) is measured using the 13f filing at the most recent quarter. The last row reports the average percentiles of the entire CRSP universe. Our sample covers 178 distinct stocks from 2013/11 to 2014/10.

	Mkt Cap (M$)	Turnover	Volatility	IO
Mean	136,668	4.20	0.022	0.577
Median	85,186	2.05	0.016	0.602
Std dev	144,755	4.80	0.019	0.175
CRSP percentile	89.9	62.5	50.5	80.0

Table 4: Diffusion Speed and Trading Intensity

This regression links retweets to trading during the first hour after a tweet. The dependent variable is the percentage of first-hour trading that occurs in the first 10 minutes. Panel A reports the regression results for trading volume in TAQ. Panel B reports the regression results for trading volume from TRFs (exchange symbol D from the TAQ dataset). The main independent variable, *diffusion*, measures the percentage of first-hour retweets that occur in the first 10 minutes. Other control variables include *pre-market* (a dummy variable equal to 1 if the tweet takes place before 9:30 ET); *afternoon* (a dummy variable equal to 1 if the tweet takes place between 12:30 and 16:00 ET); *post-market* (a dummy variable equal to 1 if the tweet takes place after 16:00 ET); *size* (log market capitalization); *turn* (turnover); *volatility* (daily returns volatility in the past 30 days); *bm* (book-to-market ratio); *abs past 1h ret* (absolute stock returns over the market in the past hour); *abs 10m ret* (absolute stock returns over the market in the first 10 minutes after the tweet); *media* (log number of media coverage on the same day but prior to the tweet); *media diffusion* (percentage of first-hour media coverage occurs in the first 10 minutes). We include stock and Twitter account fixed effects. Standard errors are clustered by ticker. The sample covers 1,261 tweets from 2013/11 to 2014/10. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Total Trading Volume

	(1)	(2)	(3)	(4)
intercept	0.02	0.11***	0.01	-0.04
	(0.65)	(3.27)	(0.12)	(-0.42)
diffusion	0.30***	0.17***	0.14**	0.14**
	(5.13)	(3.07)	(2.25)	(2.27)
pre-Market		-0.12***	-0.12***	-0.12***
		(-9.88)	(-7.70)	(-7.93)
afternoon		-0.02	-0.01	-0.01
		(-1.58)	(-1.02)	(-0.95)
post-Market		0.07***	0.07***	0.07***
		(3.67)	(2.86)	(2.96)
size			0.00	0.01*
			(1.00)	(1.71)
turn			0.00	0.00
			(-0.37)	(0.47)
volatility			1.43***	1.51***
			(10.09)	(10.49)
bm			0.00	0.00
			(-0.53)	(-0.43)
abs past 1 hr ret			-0.46***	-0.39**
			(-2.41)	(-2.07)
abs 10m ret			2.71***	2.89***
			(3.10)	(3.10)
media				-0.01
				(-1.35)
media diffusion				-0.03
				(-1.13)
stock FE	Y	Y	Y	Y
account FE	Y	Y	Y	Y
R^2	0.031	0.123	0.134	0.138

Panel B. TRF Trading Volume

	(1)	(2)	(3)	(4)
intercept	0.12***	0.24***	0.33**	0.35**
	(2.64)	(5.95)	(2.37)	(2.38)
diffusion	0.44***	0.25***	0.21**	0.21**
	(5.84)	(3.69)	(2.91)	(2.89)
pre-Market		-0.24***	-0.22***	-0.22***
		(-11.81)	(-8.23)	(-8.12)
afternoon		-0.02	-0.02	-0.02
		(-1.51)	(-1.17)	(-1.15)
post-Market		0.14***	0.14***	0.14***
		(4.89)	(3.81)	(3.67)
size			0.00	-0.01
			(-0.61)	(-0.73)
turn			0.00	0.00
			(1.29)	(1.30)
volatility			-1.87	-1.99
			(-1.19)	(-1.23)
bm			0.00	0.00
			(1.49)	(1.48)
abs past 1 hr ret			-0.81***	-0.79**
			(-3.35)	(-3.29)
abs 10m ret			2.21***	2.22***
			(1.97)	(1.93)
media				0.01
				(0.63)
media diffusion				-0.03
				(-0.72)
stock FE	Y	Y	Y	Y
account FE	Y	Y	Y	Y
R^2	0.036	0.215	0.248	0.249

Table 5: Analysis of TD Ameritrade Brokerage Account Data

We merge our tweet sample with the TD Ameritrade (TDA) brokerage-account-level transaction data during the overlapping period from 2013/11 to 2014/06. We focus on tweets during market hours and retail accounts that trade the corresponding stock at least once during the first three hours after a tweet. The merged sample contains 331 distinct tweets and 35,443 distinct TDA accounts. Panel A reports descriptive statistics across both accounts and tweets. To compute net trades, one buy (sell) is counted as 1 (-1). *Trade intensity* is again measured as the percentage of first-hour trading that occurs in the first 10 minutes. Panel B repeats the regressions of Table 4 for our merged sample of tweets. We measure *trade intensity* using (I) all trades from TAQ; (II) all TDA trades; (III) all TDA trades of female investors; (IV) all TDA trades of young investors (age <35); and (V) all TDA trades of "rich" investors (with stock holdings >$100K). Other control variables include *afternoon* (a dummy variable equal to 1 if the tweet takes place between 12:30 and 16:00 ET); *post-market* (a dummy variable equal to 1 if the tweet takes place after 16:00 ET); *size* (log market capitalization); *turn* (turnover); *volatility* (daily returns volatility in the past 30 days); *bm* (book-to-market ratio); *abs past 1h ret* (absolute stock returns over the market in the past hour); *abs 10m ret* (absolute stock returns over the market in the first 10 minutes after the tweet); *media* (log number of media coverage on the same day but prior to the tweet); *media_diffusion* (percentage of first-hour media coverage occurs in the first 10 minutes). We include stock and Twitter account fixed effects. Standard errors are clustered by ticker. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Descriptive Statistics of the Merged Sample

Across 35,443 accounts					
Avg char	Mean	Std dev	Q1	Median	Q3
Age	48.7	14.3	38.0	48.0	58.0
Stock holding ($)	78,063	167,449	3,481	20,146	74,288
Net trade	0.153	1.023	-1.000	0.167	1.000
Trade intensity	20.0%	36.6%	0.0%	0.0%	25.0%
Across 331 tweets					
Avg char	Mean	Std dev	Q1	Median	Q3
# of accounts	194.4	288.9	19.0	69.0	242.0
% of Male	72.1%	12.2%	67.9%	71.7%	76.2%
Age	51.0	5.0	48.8	50.4	52.9
Stock holding ($)	91,004	51,164	68,843	84,461	102,741
Net trade	0.060	0.430	-0.202	0.045	0.333
Trade intensity	20.4%	17.1%	10.2%	18.5%	26.1%

Panel B. The Link between Diffusion Speed and Trading Intensity

	All Trades	All TDA Trades	TDA, Female	TDA, Young	TDA, Rich
	(I)	(II)	(III)	(IV)	(V)
intercept	-0.06	-0.03	-0.07	-0.09	0.30
	(-0.44)	(-0.12)	(-0.24)	(-0.19)	(0.94)
diffusion	0.16***	0.23**	0.27**	0.17	0.36***
	(2.78)	(1.99)	(1.99)	(0.97)	(2.69)
afternoon	-0.01	0.01	0.04	0.00	0.06**
	(-0.50)	(0.57)	(1.40)	(-0.04)	(2.01)
size	0.01	0.00	0.00	0.01	-0.01
	(1.16)	(0.31)	(0.32)	(0.23)	(-0.84)
turn	0.00	0.00	0.00	0.00	0.00
	(0.84)	(0.91)	(-0.25)	(-0.77)	(1.48)
volatility	-0.19	-1.99	1.40	8.45	-8.36*
	(-0.10)	(-0.55)	(0.25)	(1.00)	(-1.69)
bm	0.00*	0.00**	0.00	0.00	0.00*
	(1.93)	(2.46)	(0.52)	(-0.40)	(1.74)
abs past 1h ret	-0.17	-0.12	0.55	-1.43	2.08
	(-0.15)	(-0.09)	(0.30)	(-0.91)	(0.86)
abs 10m ret	3.35	7.71*	6.84*	6.78	2.88
	(1.38)	(1.70)	(1.71)	(1.36)	(0.90)
media	0.00	0.01	-0.02	0.0	0.01
	(0.39)	(0.42)	(-0.85)	(0.33)	(0.36)
media diffusion	-0.05***	-0.03	-0.04	-0.12***	0.02
	(-2.86)	(-0.85)	(-1.08)	(-2.95)	(0.39)
stock FE	Y	Y	Y	Y	Y
account FE	Y	Y	Y	Y	Y
R^2	0.081	0.079	0.064	0.054	0.092

Table 6: Retail Attention, Stock Returns, and Change in Dollar Spread

The dependent variables are *stock returns* (in excess of the market and by percentage) 10 minutes after a tweet until the end of the day (CAR%[10m, close d0]); *stock returns* (in excess of the market and by percentage) on the next trading day (CAR%[close d0, close d1]); and the change in the average dollar spread from the one hour before the tweet to the one hour after. The main independent variable is *diffusion_3hr*, which measures the log number of users the tweet can potentially reach three hours after the tweet. Other control variables include *pre-market* (a dummy variable equal to 1 if the tweet takes place before 9:30 ET); *afternoon* (a dummy variable equal to 1 if the tweet takes place between 12:30 and 16:00 ET); *post-market* (a dummy variable equal to 1 if the tweet takes place after 16:00 ET); *size* (log market capitalization); *turn* (turnover); *volatility* (daily returns volatility in the past 30 days); *bm* (book-to-market ratio); *past 1h ret* (stock returns over the market in the past hour); *abs past 1h ret* (absolute value of past 1h ret); *10m ret* (stock returns over the market in the first 10 minutes after the tweet); *abs 10m ret* (absolute value of 10m ret); *media* (log number of media coverage on the same day and up to three hours after the tweet); *Isbreaking* (a dummy variable equal to 1 if the tweet contains “breaking”). We include stock and Twitter account fixed effects. Standard errors are clustered by ticker. Panel A reports results for all 1,261 tweets from 2013/11 to 2014/10. Panel B reports the results for tweets on firms with market capitalization below the median of all stocks in our sample.

Panel A: All Stocks

	CAR% [10m, close d0]	CAR% [close d0, close d1]	Spread_chg
intercept	-4.28***	-5.16***	0.04
	(-2,68)	(2.85)	(0.07)
diffusion_3hr	0.21***	-0.21***	-0.05*
	(3.16)	(-2.74)	(-1.84)
Pre-market	0.13	-0.23	-0.78***
	(0.56)	(-1.00)	(-9.52)
afternoon	-0.03	-0.25	0.34***
	(-0.16)	(-0.93)	(7.32)
post-market	0.02	-0.12	0.23***
	(0.11)	(0.51)	(4.62)
size	0.07	-0.02	0.02
	(1.02)	(-0.31)	(1.00)
turn	0.00	0.00*	0.00*
	(0.15)	(1.85)	(-1.84)
volatility	-17.75	-137.13**	13.70**
	(-0.30)	(-2.29)	(2.21)
bm	0.01	0.06**	-0.01
	(0.29)	(2.11)	(-1.62)
past 1h ret	14.09	-6.53	-1.61
	(-0.45)	(-1.07)	(-1.27)
abs past 1h ret	16.01	1.68	-1.47
	(1.52)	(0.21)	(-0.88)
10 ret	-6.11	-38.37*	-4.81
	(-0.29)	(-1.67)	(-1.92)
abs 10m ret	64.81***	20.61	2.25
	(3.15)	(0.65)	(0.64)
media	-0.02	-0.02	-0.01
	(-0.21)	(-0.16)	(-0.48)
isbreaking	-0.20	-0.88***	-0.06
	(-0.95)	(-2.87)	(-0.67)
stock FE	Y	Y	Y
account FE	Y	Y	Y
R^2	0.117	0.081	0.341

Panel B: Small Stocks

	CAR% [10m, close d0]	CAR% [close d0, close d1]	Spread_chg
intercept	-7.59***	7.93***	0.70
	(-2.52)	(2.07)	(0.86)
diffusion_3hr	0.39***	-0.38***	-0.06*
	(3.24)	(-2.99)	(-2.03)
Pre-market	0.33	-0.63	-0.89***
	(0.79)	(-1.48)	(-9.01)
afternoon	0.01	-0.28	0.36***
	(0.04)	(-0.57)	(6.80)
post-market	0.13	-0.51	0.21***
	(0.43)	(-1.09)	(3.92)
size	0.10	-0.01	0.01
	(0.67)	(-0.08)	(0.20)
turn	0.00	0.00*	0.00*
	(0.08)	(1.72)	(-1.08)
volatility	-13.09	-150.53**	9.19**
	(-0.20)	(-2.19)	(1.26)
bm	0.01	0.07**	-0.01
	(0.42)	(2.28)	(-1.54)
past 1h ret	6.35	-7.88	-1.76
	(-0.67)	(-1.25)	(-1.45)
abs past 1h ret	17.44	5.00	0.82
	(1.56)	(0.58)	(-0.50)
10 ret	-5.94	-44.17*	-4.91
	(-0.25)	(-1.80)	(-1.92)
abs 10m ret	67.05***	22.47	2.91
	(3.01)	(0.66)	(0.81)
media	0.01	0.05	-0.04
	(0.08)	-0.28)	(-1.15)
isbreaking	-0.24	-1.46***	-0.09
	(-0.56)	(-2.51)	(-0.77)
stock FE	Y	Y	Y
account FE	Y	Y	Y
R^2	0.148	0.113	0.370

Table 7: Diffusion Prediction

In the OLS regression, we use Twitter characteristics observable 10 minutes after a tweet to predict the growth rate in diffusion from 10 minutes to three hours after the tweet. The growth rate is defined as log(*diffusion_3hr*) – log(*diffusion_10m*), where *diffusion_10m* and *diffusion_3hr* are the number of Twitter users the tweet potentially reaches after 10 minutes and three hours, respectively. The Twitter characteristics include *Total number of tweets* (the average daily number of tweets sent by that Twitter account); *log(# of followers of recent retweeters)*—the total number of followers in log, of the five most recent Twitter accounts that retweeted the tweet; *Speed of recent retweets* (inverse of the average time lapse between the five most recent retweets); *Hour* (calendar hour of the tweet); *IsWest* (a dummy variable equal to 1 if the tweet is sent from the West Coast); *IsCEO* (a dummy variable equal to 1 if the tweet is sent by the CEO of the company); *HasPicture* (a dummy variable equal to 1 if the tweet contains a picture); *HasURLs* (a dummy variable equal to 1 if the tweet contains URL links); *HasHashtags* (a dummy variable equal to 1 if the tweet contains Hashtags). The sample covers all 1,261 tweets from 2013/11 to 2014/10. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Variable	CAR% [10m, close d0]
Intercept	-4.15***
	(-10.87)
Total # of tweets	-0.01***
	(-3.71)
Log (# of followers of recent re-tweeters)	0.02**
	(2.17)
Speed of recent retweets	4.31***
	(5.16)
Hour	-0.05***
	(-4.95)
Iswest	0.75***
	(5.80)
IsCEO	1.24**
	2.35
HasPicture	0.63***
	(6.28
HasURLs	-0.37**
	(-2.52)

HasHashtags	0.41**
	2.46
R^2	0.190

Table 8: Predicted Retail Attention and Stock Return Predictions

The dependent variables are *stock returns* (in excess of the market and by percentage) 10 minutes after a tweet until the end of the day (CAR%[10m, close d0]) and *stock returns* (in excess of the market and by percentage) on the next trading day (CAR%[close d0, close d1. In Panel A, the main independent variable is *predicted diff*, which measures the log number of users the tweet is predicted to reach three hours after the tweet. It is computed by summing the log number of Twitter users the tweet reaches after 10 minutes (*diffusion_10m*) and the predicted log growth rate from the regression shown in Table 6. In Panel B, the main independent variable is simply *diffusion_10m*. Other control variables include *pre-market* (a dummy variable equal to 1 if the tweet takes place before 9:30 ET); *afternoon* (a dummy variable equal to 1 if the tweet takes place between 12:30 and 16:00 ET); *post-market* (a dummy variable equal to 1 if the tweet takes place after 16:00 ET); *size* (log market capitalization); *turn* (turnover); *volatility* (daily returns volatility in the past 30 days); *bm* (book-to-market ratio); *past 1h ret* (stock returns over the market in the past hour); *abs past 1h ret* (absolute value of past 1h ret); *10m ret* (stock returns over the market in the first 10 minutes after the tweet); *abs 10m ret* (absolute value of 10m ret); *media* (log number of media coverage on the same day and up to three hours after the tweet); *Isbreaking* (a dummy variable equal to 1 if the tweet contains "breaking"). We include stock and Twitter account fixed effects. Standard errors are clustered by ticker. The sample covers all 1,261 tweets from 2013/11 to 2014/10. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Using Predicted Diffusion

	CAR% [10m, close d0]	CAR% [close d0, close d1]
Intercept	-4.20***	4.62
	(-2.61)	(2.56)
Predicted diff	0.20***	-0.17**
	(3.14)	(2.22)
Pre-market	0.211	-0.23
	(0.94)	(-0.95)
afternoon	-0.02	-0.27
	(-0.11)	(-1.04)
post-market	0.02	-0.06
	(0.11)	(-0.26)
size	0.08	-0.03
	(1.07)	(-0.31)
turn	0.00	0.00*
	(0.20)	(1.88)
volatility	-15.69	-139.40**
	(-0.26)	(-2.31)
bm	0.00	0.06**
	(0.07)	(2.10)
past 1h ret	-3.95	-38.31*
	(-0.19)	(-1.65)
abs past 1h ret	-2.46	-6.01
	(-0.28)	(-0.98)
10 ret	16.74	1.63
	(1.59)	(0.02)
abs 10m ret	62.56***	20.26
	(3.04)	(0.63)
media	-0.04	-0.06
	(-0.44)	(-0.49)
isbreaking	-0.23	-0.91***
	(-1.09)	(-2.95)
stock FE	Y	Y
account FE	Y	Y
R^2	0.126	0.081

Panel B. Using Predicted Diffusion after 10 minutes

	CAR% [10m, close d0]	CAR% [close d0, close d1]
Intercept	-3.64**	2.48
	(-2.18)	(1.25)
Predicted diff	0.11	-0.04
	(1.62)	(-0.44)
Pre-market	0.04	-0.17
	(0.15)	(-0.71)
afternoon	-0.12	-017
	(-0.54)	(-0.62)
post-market	0.07	-0.05
	(0.42)	(-0.18)
size	0.12	-0.01
	(1.50)	(-0.07)
turn	0.00	0.00*
	(0.30)	(1.90)
volatility	-15.12	-129.72**
	(-0.27)	(-2.20)
bm	0.01	0.04
	(0.44)	(1.50)
past 1h ret	-9.49	-40.37*
	(-0.38)	(-1.81)
abs past 1h ret	-10.00	-8.10
	(-1.08)	(-1.31)
10 ret	9.36	4.08
	(0.80)	(0.56)
abs 10m ret	74.37***	23.46
	(2.89)	(0.81)
media	-0.07	-0.16
	(-0.74)	(-1.45)
isbreaking	-0.22	-0.95***
	(-0.80)	(-2.88)
stock FE	Y	Y
account FE	Y	Y
R^2	0.082	0.069

Table 9: Predicted Retail Attention and Stock Returns: Out-of-Sample Predictions

We break our one-year sample period (2013/11-2014/10) into an in-sample period (2013/11-2014/04) and an out-of-sample period (2014/05-2014/10). We estimate the predictive regression from Table 6 during the in-sample period only. We then take the estimated coefficients and apply them to the out-of-sample period to compute predicted diff. In other words, *predicted diff* is observable 10 minutes after a tweet. We then link predicted diff to future returns in the out-of-sample period. The dependent variables are *stock returns* (in excess of the market and by percentage) 10 minutes after the tweet until the end of the day (CAR%[10m, close d0]) and *stock returns* (in excess of the market and by percentage) on the next trading day (CAR%[close d0, close d1. Other control variables include *pre-market* (a dummy variable equal to 1 if the tweet takes place before 9:30 ET); *afternoon* (a dummy variable equal to 1 if the tweet takes place between 12:30 and 16:00 ET); *post-market* (a dummy variable equal to 1 if the tweet takes place after 16:00 ET); *size* (log market capitalization); *turn* (turnover); *volatility* (daily returns volatility in the past 30 days); *bm* (book-to-market ratio); *past 1h ret* (stock returns over the market in the past hour); *abs past 1h ret* (absolute value of past 1h ret); *10m ret* (stock returns over the market in the first 10 minutes after the tweet); *abs 10m ret* (absolute value of 10m ret); *media* (log number of media coverage on the same day and up to three hours after the tweet); *Isbreaking* (a dummy variable equal to 1 if the tweet contains "breaking"). We include stock and Twitter account fixed effects. Standard errors are clustered by ticker. The regression uses tweets during the out-of-sample period from 2014/05 to 2014/10. T-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	CAR% [10m, close d0]	CAR% [close d0, close d1]
Intercept	-0.93	4.17*
	(-0.48)	(1.77)
Predicted diff	0.15*	-0.22**
	(1.88)	(-2.21)
pre-market	0.56*	-0.44**
	(1.94)	(-1.35)
afternoon	0.14	-0.69**
	(0.67)	(-2.17)
post-market	0.15	-0.40
	(0.87)	(-1.22)
size	-0.06	0.07
	(0.75)	(0.63)
turn	0.00	0.00**
	(0.41)	(2.07)
volatility	-41.89	-159.92**

	(-0.54)	(-2.24)
bm	0.01	0.05
	(0.26)	(0.90)
past 1h ret	29.07	-121.99***
	(0.67)	(-2.86)
abs past 1h ret	-4.08	-1.49
	(-0.41)	(-0.22)
10 ret	14.52	1.53
	(1.27)	(0.16)
abs 10m ret	76.43	99.81*
	(1.50)	(1.85)
media	0.04	0.01
	(0.39)	(0.04)
Isbreaking	-0.17	-1.18***
	(-0.70)	(-2.85)
stock FE	Y	Y
account FE	Y	Y
R^2	0.134	0.187