Social Media and Financial News Manipulation

Shimon Kogan*
The Wharton School, University of Pennsylvania
Interdisciplinary Center Herzliya

Tobias J. Moskowitz
Yale School of Management
NBER
AQR Capital Management

Marina Niessner
The Wharton School, University of Pennsylvania

September 2021

*We thank Tony Cookson, Kent Daniel, Anastassia Fedyk, Diego Garcia, Matt Gentzkow, Gary Gorton, Bryan Kelly, Elizabeth Kempf, Bonnie Moskowitz, James Pennebacker, Jesse Shapiro, Kelly Shue, Eric So, Denis Sosyura, Sam Hartzmark, as well as conference and seminar participants at UCLA (Anderson), Rice University (Jones), University of Miami Business School, ASU Sonoran Winter Finance Conference, 3rd Annual News & Finance Conference, University of Colorado at Boulder, Northwestern University (Kellogg), FSU SunTrust Beach Conference, MIT Sloan, Yale SOM, Catolica-Lisbon, University of Kentucky Finance Conference, FEB, 3rd Rome Junior Finance Conference, the 2018 WFA meetings, the U.S. Securities and Exchange Commission Division of Economic and Risk Analysis, the Red Rock Conference, Behavioral NBER, Jackson Hole Finance Conference, AQR Capital Management, the NBER Asset Pricing meetings, 2020 AFA meetings in San Diego, and the Global Virtual Seminar Series on Fintech for their helpful comments and suggestions. We also thank Elli Hoffmann and Keren Ben Zvi for providing and helping organize the data. AQR Capital Management is a global investment management firm, which may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR. Contact emails: skogan@upenn.edu, tobias.moskowitz@yale.edu, and marina.niessner@gmail.com. A previous version of this paper was titled, "Fake News in Financial Markets."

Abstract

We dissect an undercover SEC investigation into the manipulation of financial news on social media to study the indirect effects of market manipulation. While fraudulent news had a direct impact on retail trading and prices, revelation of the fraud caused market participants to discount all news, including legitimate news, from these platforms. The results highlight the indirect consequences of fraud and its spillover effects that reduce the social network's impact on information dissemination, especially for small firms. The effect appears to dissipate over time, becoming insignificant a year later. The results highlight the importance of social capital for financial activity.

We study a unique setting involving fraudulent news in financial markets that culminates in a Securities and Exchange Commission (SEC) investigation of shared financial news networks. We use this setting to document the indirect effects of market manipulation, tracing the responses of consumers and producers of information. The SEC investigation provides a shock to market participants' *awareness* of fraud, providing an opportunity to examine the indirect spillover effects of misinformation and manipulation on the impact of legitimate news and markets. We interpret the results through the lens of models of trust and the importance of social capital for financial activity.

The vast literature on stock market manipulation focuses on the direct effects of manipulation.[1] We seek to study their broader implications, such as deterioration of trust in markets, merging the literature on stock market manipulation with the literature on social capital. The idea that agent's trust matters for economic and financial activity is developed formally in Knack and Keefer (1997), Guiso et al. (2004), and Guiso et al. (2008), and its role in households' financial decisions is surveyed in Gomes et al. (2021). Guiso et al. (2010), and Sapienza and Zingales (2012) empirically relate trust to financial activity. However, little evidence exists of how trust evolves in markets and the consequences of changes in trust. Our setting provides a shock to trust in information from social media financial news networks and an opportunity to measure its broader implications.

We examine social media platforms for financial news, which are crowd-sourced providers of information that allow for the democratization of news. These networks grew in popularity and impact over the past decade (Chen et al. (2014)),[2] which, coupled with lax regulatory monitoring, created incentives for manipulation. In 2014, an industry whistle blower, Rick Pearson, who was an author on one of the most prominent crowd-sourced websites, Seeking Alpha, was approached by a public relations firm to write paid-for false content to promote a stock. Instead of rejecting the offer he went undercover and turned over to the SEC 171 articles written by 20 authors about 47 companies who knowingly wrote false information about these firms. The SEC launched an investigation that became public in March 2014 and ultimately led to legal action against these firms and the articles' authors. We use the SEC investigation as our empirical laboratory to assess the indirect consequences of fraud coming from a

[1] Aggarwal and Wu (2006) examine SEC litigation cases, Frieder and Zittrain (2007), Hanke and Hauser (2008), Hu et al. (2009), Nelson et al. (2013), and Hu and McInish (2013) examine the efficacy of email spams touting stocks, Delort et al. (2009) and Sabherwal et al. (2011) examine the use of online message boards to manipulate prices, and Ullah et al. (2014) examine third-party false information releases. Benabou and Laroque (1992) and Gale and Allen (1992) model how market manipulation through the release of information and trading can be sustained in equilibrium. There is also a large literature in finance and accounting on the strategic release (or omission) of information with, presumably, the intent of temporarily affecting prices (e.g., Niessner (2018)). For example, Ahern and Sosyura (2014) show that bidders in stock mergers originate more news stories during merger negotiations, but before the public announcement.

[2] According to a survey from the Pew Research Center (Gottfried and Shearer (2016)), 62% of American adults get news from a social media site. Allcott and Gentzkow (2017) argue that social media platforms enable content to be disseminated with no significant third party filtering or monitoring, allowing false information to be spread quickly through a vast social network. Vosoughi et al. (2018) find that fake news diffuses faster, deeper, and more broadly than actual news, in part because the fake news is often more extreme and exaggerated in order to increase diffusion.

shock to trust.[3]

We first confirm that these social media networks have impact on markets. Articles written on these platforms are associated with higher subsequent retail trading activity and price volatility. We focus on retail trading volume, measured by Boehmer et al. (2020), because retail investors are the primary participants on these social networks. Later, we also examine institutional investor trading volume as a placebo test. Obtaining the 171 fraudulent articles identified by Rick Pearson and the SEC, we show that these false promotional articles increased abnormal retail trading volume by more than 55% over the three days following the article's publication relative to legitimate articles written by the same author. These results are concentrated among small firms. In addition, the fraudulent media campaign was effective at manipulating prices, causing an average 8% rise in affected firm share prices immediately following the articles, that eventually reverses in the long run and ultimately becomes cumulatively negative at −2.5%. These results are consistent with investors being deceived by the fraud that eventually is corrected in markets. These findings provide evidence for the premise behind a theory of trust: investors being manipulated by false news that impacts markets requires a level of trust in these platforms in the first place.

We use the public revelation of the SEC investigation, and awareness of the social media manipulation, as a shock to trust in this news source. We use the event of the announced SEC investigation and the media coverage of the scandal in February and March 2014 as a shock to the public's awareness of fraud on these platforms, and measure fraud's broader, indirect effects on markets. Specifically, we collect 203,545 articles from Seeking Alpha from 2005 to 2015 and a competitor platform, Motley Fool (147,916 articles from 2009 to 2014), that cover 7,700 publicly traded firms. We test an implication from theory that fraud imposes externalities on other, legitimate news (Aymanns et al. (2017) and Allcott and Gentzkow (2017)) and relate this test more generally to trust in markets (Guiso et al. (2004)). We find that abnormal retail trading volume and price volatility drop significantly for *any* news article written on these platforms after the announcement, particularly for small firms with high retail interest. Retail abnormal trading volume drops by 4.1% for an article published in the 6 months after the SEC announcement versus the 6 months before the announcement. For the smallest firms, the drop is 23.5%. Price volatility for small firms also drops by 1.3%, with no significant price response for larger firms, consistent with externalities of manipulation on trust in information networks. The effect of the SEC investigation is both immediate and, surprisingly, transitory. We document a reduction in retail investors' response to all news as quickly as one week into the event, but the effect appears to go down

[3] In fact, the 1934 Act that formed the SEC was originally motivated by a desire to restore the public's faith in capital markets following the 1929 stock market crash, with one of its mandates to monitor and investigate price manipulation.

over time. A year into the event window, the change in response appears to be small and statistically insignificant.

Of course, identifying the impact of trust on markets, versus other possible channels impacting markets, is empirically challenging. There are likely many factors driving news and trading activity simultaneously that may also coincide with the SEC event. For example, trends in news, changes in the market's response to news, or unobservables affecting trading that may happen to coincide with the SEC event.

To address this identification challenge we conduct a number of tests. First, we note that the SEC investigation was an unanticipated event, having been spawned by Rick Pearson's self-initiated undercover investigation. This fact makes the event more likely to be exogenous to the outcome variables. Second, we also employ a number of controls for other sources of news beyond the social platforms (corporate filings, press releases, and Wall Street Journal and New York Times articles) to help capture the market's response to news more generally. Third, we show that there are no pre-trends in news or in the market's response to news prior to the announcement. Fourth, we run several "placebo" or falsification tests designed to pick up general effects associated with news and the market's response to news, but that are otherwise unrelated to these platforms or the SEC's investigation. For example, we look at institutional trading activity in response to articles published pre- versus post-SEC event. While we find a significant decline in *retail* trading volume for articles published after the SEC event, we find no significant change to institutional trading volume to articles published before versus after the event. Since retail and institutional investors should both be affected by general trends in news or responses to news, or other omitted variables driving news and trading, but only retail investors should be affected by shocks to the social network platforms, this finding strongly suggests it is the effect of the scandal, and not other factors, driving the precipitous drop in retail trading activity.

As another placebo test, we look at news media unlikely to be affected by the SEC investigation. For instance, newspaper articles from the Wall Street Journal (WSJ) and New York Times (NYT) are unlikely to be affected by the fraud on social media platforms – WSJ or NYT journalists were not being paid to write false promotional articles about firms. However, general trends in news and the market's response to news should be evident for these media as well, and so we should see a decline in retail trading to WSJ and NYT articles, too if general market conditions are driving these effects. We do not find a commensurate decline in trading volume in response to news from NYT or WSJ articles before versus after the SEC announcement. This result is consistent with the SEC event only affecting responses to these social media platforms and not being a symptom of wider trends in news. Conducting similar falsification exercises for corporate filings and press releases – sources of news that are also unlikely to be polluted by fraud or be interpreted differently after the SEC investigation – we find no significant change in market response either. These results suggest that other news trends or unobservables are not likely confounding the event study. The result is consistent with investors' increased distrust of news from the social media platforms, but not other forms of media.

Another way to address the identification challenge is to find an instrument that moves investor attention to news and/or news coverage, but is otherwise unrelated to market activity. Peress and Schmidt (2020) show that big-news events crowd out coverage and attention to other news and distracts retail traders from trading. We use mass shootings as a shock to investors' attention. We find that in our full sample, mass shootings coverage lowers the retail trading response to the social network's articles, consistent with a crowding out effect for investor attention. However, since mass shootings should be unrelated to the SEC scandal, the crowding out effect should be no different before versus after the SEC event. We find that to be the case – the crowding out effect from mass shootings on the social media networks is the same 6 months before versus 6 months after the event. While this particular instrument is useful in shifting investor attention that is distinct from any economic event or information, it does not instrument for any changes associated with the SEC event itself (e.g., potential changes in news or response to news that coincided with the event). However, this result highlights that the market's muted response to financial news on days of sensational non-economic news (i.e., mass shootings) is no different after the SEC event. In other words, an exogenous shock to investor attention has the same impact on articles published on these networks after the scandal, suggesting that nothing has changed with respect to the market's response to news generally. Thus, the decline in retail trading activity to articles after the SEC announcement is unlikely coming from a shift in investor response to news or attention to news generally.

Finally, we provide more direct corroborating evidence for the trust channel driving these results. We show that the reduction in response to news is tempered for authors with better reputations and larger followings. In addition, articles that contain more numbers, such as articles about earnings, receive a more muted trading response reduction from investors after the shock. These results are consistent with a trust mechanism. Conversely, measures of disagreement, such as the number and length of comments on an article, and mentions of the word "disagree" in the comments section have no *differential* impact on trading volume after the SEC announcement, despite having an unconditional impact on trading. We also find that when comments contain the words "fake" or "fraud," the impact on volume declines even further, but *only* after the SEC announcement. Before the event, claims of "fraud" or "fake" in the comments have no impact. Finally, author disclosure of a position they have in the stock they are writing about has a negative effect on volume after the scandal, but no effect prior to the scandal. These results are consistent with increased mistrust in news after the scandal and not changes in disagreement or other drivers of trading activity.

We interpret our results through the lens of several theories. We reject a rational expectations equilibrium response to the SEC announcement in favor of theories of trust (Guiso et al. (2004)), where the SEC announcement altered investors' trust in news from these platforms. The evidence is consistent with broader themes about social trust and institutions affecting economic activity. The spillover effects of market manipulation result in a new market equilibrium, where investors discount all news, including legitimate news, from these platforms, after the ensuing scandal was made public. To further corroborate that story, we apply natural language processing (NLP) to the comments section of the articles. We find a significant increase in use of the words "fake" and "fraud" in the comments section after the SEC investigation, consistent with participants being more concerned or sensitive to fraud. We also find a much more negative tone in the comments after the SEC announcement, consistent with readers having less trust in the articles.

We also examine the response from producers of news on these platforms, which could also contribute to our findings. Applying the linguistic algorithm to the articles themselves, we find no differences in tone, clout, or other linguistic cues after the SEC shock, with one notable exception: authenticity, a measure designed to detect deception in language. Articles score higher on authenticity after the SEC scandal, suggesting that producers of news on these platforms are optimally responding to the new equilibrium and trying to increase trust. However, this response by producers of news should *increase*, not decrease, trading activity in response to their articles. Hence, the equilibrium response of news producers cannot explain our results and go in the wrong direction, thus underestimating the response from consumers we document. We also find that articles written after the scandal focus more on earnings and other "hard" information that is more easily verifiable and hence more trustworthy, which should also increase, not decrease, retail trading activity. Hence, the changing characteristics of articles after the scandal is not driving the change in trading activity and is consistent with market participants optimally responding to a shock to trust.

Our results are consistent with theories of trust being an important driver of financial activity. The shock to trust from the SEC investigation had a significant impact on investor activity that resulted in a new equilibrium response to news on these platforms, including legitimate news. These findings highlight the spillover effects and indirect consequences of market manipulation. Our results also relate to the implications of fake news and media bias. Misleading information can impact social, political, and economic relationships. While analysis of these issues has primarily been theoretical,[4] our findings

[4] Allcott and Gentzkow (2017) model fake news as an extension of Gentzkow and Shapiro (2005) and Gentzkow et al.

provide novel empirical evidence. Our findings may be consistent with news being tailored to readers' priors and news-producers sacrificing longer-term reputational capital in lieu of short-term gains (Allcott and Gentzkow (2017)). The decline in trading activity to all news, including legitimate news, following the public's awareness of the SEC investigation is also consistent with Aymanns et al. (2017) and Allcott and Gentzkow (2017), where fake news increases distrust of media in general.[5]

Finally, our setting suggests reasons to be both cautious and optimistic in generalizing the findings. One of the benefits of financial markets is the ability to quantify outcomes. On the other hand, arbitrage forces and market efficiency may mute its effects. If market manipulation can impact U.S. equity markets, where competition for information is fierce, markets are liquid, and arbitrage activity exists, then it could have even greater influence in settings where information costs are higher and the ability to correct misinformation is more limited (e.g., online consumer or political markets).

The rest of the paper is organized as follows. Section 1 briefly motivates our analysis through the lens of several theories. Section 2 describes the shared financial networks and discusses the scandal of fraudulent paid-for articles that led to the SEC's investigation. Section 3 documents the direct impact on markets from articles on these platforms, including the fraudulent articles from the scandal. Section 4 uses the SEC event and the public's awareness of manipulation as a shock to investor trust, and examines the indirect impact of manipulation on the market. Section 5 concludes.

1. A Brief Theoretical Motivation

We briefly discuss several theories that guide the empirical analysis and help interpret the results. Our sample of fraudulent articles from the SEC are authored by paid agents acting on behalf of a principal to manipulate the stock price.

Financial markets provide a useful setting to examine the impact of manipulation because they provide high frequency outcomes, such as trading volume and market prices. Trading volume provides a measure of whether investors pay attention to, and act upon, the news in the articles, which may or may not result in price impact. For example, news can cause trading without impacting prices if markets are informationally efficient (Fama (1970)), or prices can move without trading (Milgrom and

(2015) on media bias, where fake news occurs in equilibrium when agents cannot costlessly verify the truth and the news matches the agent's priors. Aymanns et al. (2017) provide an equilibrium model of an adversary using fake news to target agents with a biased private signal, where knowledge of the adversary causes agents to discount all news. Kshetri and Voas (2017) discuss the pervasiveness of fake news and its dissemination across news consumers. False content may impose private and public costs by making it more difficult for readers to infer the truth, reduce positive social externalities from shared-information platforms, increase skepticism and distrust of legitimate news, and potentially cause resource misallocation. Readers may also derive utility from fake news (as entertainment or if slanted toward their biases, as in Mullainathan and Shleifer (2005)).

[5] See also "Trust in Social Media Falls – Raising Concerns for Marketers," by Suzanne Vranica, Wall Street Journal, June 19, 2018, which discusses research by Edeleman, the world's largest public relations firm, that found trust in social media has fallen world-wide and particularly in the U.S. over the last year.

Stokey (1982)). We also use the SEC event and its public announcement as a shock to trust in news on the social media platforms and assess the implications of that trust.

A representative investor with fully rational expectations will place no weight on the manipulated signal and hence such signals will have no impact on markets. Similarly, any shock to the awareness of fraudulent news from the SEC investigation should also have no impact under rational expectations.

Under a model of heterogeneous beliefs, the manipulated news may influence some agents more than others. To take an extreme case, suppose there are two types of investors: naive investors who are persuaded by the news and rational arbitrageurs who are not manipulated by the news, but may be limited in the size of their positions due to risk aversion or limited capital. In this case, the fraudulent news will increase trading and may or may not result in increased price impact and volatility, depending on the relative dominance of naive versus rational traders. Any price movement should reverse if driven by naive traders being fooled by fraudulent news.

A shock to the awareness of fraudulent news on these platforms may alter the beliefs of naive investors who now may discount or distrust the signals they receive from these platforms. Moreover, the awareness of fraud may cause investors to discount all news from these platforms, including legitimate news, causing a spillover effect from the announcement on all news from this source.

In our empirical tests, we examine the direct effects of fraudulent articles on the trading volume and price movements of firms who are the subject of the articles, and whether that price movement is temporary or permanent. We also examine the indirect effects of awareness of the news scandal from the SEC announcement on other news and on other firms to document the indirect spillover effects of market manipulation.

2. Knowledge-Sharing Platforms and the Social Media Scandal

We describe the knowledge-sharing financial news platforms and the SEC-investigated scandal of manipulated articles.

2.1 Knowledge-sharing platforms

Our sample comes from a website called Seeking Alpha, supplemented with a competitor platform, Motley Fool. Seeking Alpha is an online news service provider for financial markets, whose content is provided by independent contributors. The company has distribution partnerships for its content with MSN Money, CNBC, Yahoo Finance, MarketWatch, NASDAQ, and TheStreet. The Motley Fool is a multimedia financial-services company that provides financial advice for investors through a shared-knowledge platform. They are the two largest financial crowd-sourced sites.

As the popularity of these platforms grew (Seeking Alpha grew from two million unique monthly visitors in 2011 to over nine million in 2014, generating 40 million visits per month), concerns of their susceptibility to fraud surfaced due to these sites being virtually unregulated, frequented by retail investors, and because authors can use pseudonyms (though the platforms claim they know the true identity of each author, which was eventually subpoenaed by the SEC). Authors are allowed to talk up or down a stock that they are long or short, provided they disclose any positions they have in the stock in a disclaimer accompanying the article. Failure to disclose has legal ramifications. According to Section 17b of the securities code, it is illegal to fail to disclose any direct or indirect compensation that the author received from the company, its representative, a broker-dealer, or an underwriter. Appendix A details how authors on these sites contribute and are compensated.[6]

2.2 Rick Pearson and the SEC Investigation

In 2014, an industry insider, Rick Pearson, who was a regular contributor to Seeking Alpha, was approached by an investment-relations firm, DreamTeam, to promote certain stocks looking for "good news" by writing articles with false information for a fee without disclosing the payment. Specifically, Mr. Pearson was asked to write paid promotional articles on Galena Biopharma and CytRx Corporation. Instead, Mr. Pearson went undercover to investigate how rampant this practice was and uncovered more than one hundred fraudulent, paid-for articles by other authors who did not disclose their compensation.[7] He turned the evidence over to the SEC, who investigated each of these cases starting in February 2014. The first lawsuit (against Galena Biopharma) was brought on October 31, 2014, in a claim against the authors, the promotion firms who paid them, and the companies and their executives who hired the promotion firms.[8]

We use the announcement of the initial SEC investigation as a shock to the public's awareness of fraud, and show that the timing of this shock was unlikely confounded by other effects, recent trends, or unobservables likely to affect our outcome variables. Since the investigation stemmed from Rick Pearson's undercover work, the timing of the announcement should be unrelated to any outcome variable we analyze, except from the event itself. Below, we briefly describe some of the details of the first case, against Galena Biopharma, which provides a microstudy of the direct impact these fraudulent articles had on the stock's trading activity and prices, as well as the motivation behind them.

[6] In June 2012, Seeking Alpha announced it would no longer permit publication of articles for which outside compensation had been paid.

[7] See https://seekingalpha.com/article/2086173-behind-the-scenes-with-dream-team-cytrx-and-galena and also https://www.barrons.com/articles/seeking-alpha-needs-to-take-stock-of-its-policies-1395420277.

[8] Subsequent lawsuits were also filed on April 10, 2017 and September 26, 2018. See filing documents at: http://securities.stanford.edu/filings-documents/1051/GBI00_01/20141031_r01c_14CV00367.pdf; https://www.sec.gov/litigation/complaints/2017/comp23802-lidingo.pdf.

2.3 *Galena Biopharma*

On October 31, 2014 the SEC filed a lawsuit in the United States District Court on behalf of all persons who bought Galena's common stock between August 6, 2013 and May 14, 2014.[9] Figure 1 depicts the stock price of Galena from April 2013 to May 2014, as well as the events that led to the lawsuit. According to the lawsuit, Galena worked with investment relations companies Lidingo and DreamTeam to publish a series of promotional articles on Seeking Alpha that Galena paid for, where the payments were not disclosed by the authors and in some cases authors falsely claimed *not* to have received any payment. Appendix B contains an example of one of the fake articles written about Galena. Figure 1 shows that Galena's share price rose from about \$2 to over \$7 between the summer of 2013 and January of 2014. The publications of the promotional articles are highlighted on the graph by green boxes and often coincide with a bump in stock price on that day and a steady increase in price for several days after. The motivation behind the campaign seems to be a pump-and-dump scheme, as Galena insiders took advantage of the price rise through corporate actions and their own personal trading. On September 18, 2013 Galena sold 17,500,000 units of stock in a seasoned equity offering for net proceeds of \$32.6 million. On November 22, 2013, Galena held a board meeting and granted stock options to executives and directors with a strike price of \$3.88. In January 2014, after the stock price reached its highest level since 2010, seven Galena insiders sold most of their stock in less than a month, for more than \$16 million. These events are highlighted in Figure 1.

In February and early March 2014, several investigative journalists published exposé articles documenting the fraud, including in Barron's and Fortune. On March 17, 2014 Galena revealed in a 10-K filing that it was the target of an SEC investigation over the promotion. The SEC brought charges against Galena and its former CEO Mark Ahn "regarding the commissioning of internet publications by outside fake firms." Mr. Ahn was fired in August 2014 over the controversy, and in December 2016, the SEC, Galena, and Mr. Ahn reached a settlement. Appendix B reports the 8-K form documenting the settlement. By that point Galena's stock price had dropped to \$2 a share.[10] We exploit the timing of the SEC announcement and subsequent media attention as a shock to investors' awareness of fraudulent news on these platforms, and examine the impact on markets before versus after the event.

[9](Case 3:14-cv-00558-SI): http://securities.stanford.edu/filings-documents/1051/GBI00_01/20141031_r01c_14CV00367.pdf.

[10]Interestingly, while Galena is a relatively small firm, it was not an obscure one. For example, in July 2013, before the promotion started, it had a market cap of approximately \$350 million, and it was followed by analysts at Cantor Fitzgerald, JMP Securities, Oppenheimer & Co., and others. Furthermore, according to the SEC lawsuit, more than a hundred market makers facilitated trading in the company's stock.

2.4 Fraudulent Articles

Mr. Pearson kindly provided the articles to us that he determined to be fraud: 111 articles by 12 authors covering 46 publicly traded companies. We also obtained a second set of known fraudulent articles that the SEC identified during their investigation.[11] Seeking Alpha kindly shared 147 of those articles with us, as they had been removed from the platform. Among those articles, we match 60 to firms publicly traded on U.S. exchanges to obtain price and volume information from the Center for Research in Security Prices (CRSP). The rest of the articles pertain to firms traded over the counter. Combining all of the data sources, our final sample consists of 171 fraudulent articles written by 20 different authors about 47 publicly traded firms.[12]

It is useful to define what we mean by *fraudulent* articles. In the sample from Rick Pearson and the SEC, the articles were paid for by a promotional firm to deceive the market and manipulate the stock price. Consequently, these articles contained information of some kind that authors knew to be incorrect at the time. How false or wrong that information turned out to be is difficult to assess. Our sample of fraudulent articles is about *intent* to deceive, where the articles contain information that the authors knew at the time to be false.

To provide some insight into the content of these promotional articles, we highlight a recent example from our sample that was part of the SEC's lawsuit filed in September 2018. One of the fraudulent, paid-for articles in this case was a publication that appeared on Seeking Alpha on September 26, 2013 about the company Biozone. The article stated,

> *Biozone has developed a new method of drug delivery, QuSomes that provides improved efficacy, reduced side effects, and lower costs. This technology will allow Biozone to reformulate and sell certain FDA approved drugs at a reduced cost, which should help Biozone capture a large percentage of these drug markets.*

From the SEC lawsuit filed in September 2018 in the District Court of New York City:

> *Keller misleadingly stated that Company A had a formulation ready for testing to be brought to the billion-dollar injectable drug market. Yet, as Keller knew, as of summer 2012, all R&D efforts had been shut down without the successful formulation of an injectable drug and Company A had ceased all efforts to develop this technology in mid-2012.*

"Keller" refers to Brian Keller, the co-founder and Chief Scientific Officer of Biozone, who had paid for the promotional article. Many of the paid-for articles involve similar issues.

Our focus, however, is on investor *trust* in these platforms, and not necessarily the actual content of the articles. The scandal of fraudulent articles and the SEC investigation provide a shock to trust

[11] The full list can be found here:
https://ftalphaville-cdn.ft.com/wp-content/uploads/2017/04/10231526/Stock-promoters.pdf.

[12] While we gain 60 additional articles from the SEC, we only gain one additional firm. Most of the additional articles pertain to firms already covered by Rick Pearson, and hence simply give us more articles about the same firms.

in this medium, allowing us to assess the indirect impact of market manipulation. The content of the articles themselves embody direct manipulation, which requires knowing the content of the articles, how that deviates from the truth, and the market's expectations. Many studies have focused on the direct effects of manipulation (see footnote 1), but not the indirect effects through trust.

3. Direct Impact of Social Media Articles

Before proceeding to the indirect effects of manipulation and trust on markets, we first establish that articles on these platforms matter and that investors pay attention to them, and then examine the direct impact of the fraudulent articles on markets.

3.1 Do articles on the platforms have impact?

We first address whether articles posted on these social platforms impact markets at all, or whether these social platforms are just a side show. Recent research (Chen et al. (2014), Cookson and Niessner (2020)) finds that social media financial platforms predict trading volume, stock price, and earnings surprises. We add to those findings with our sample, which further supports that these information platforms have import.

We manually download all articles: 203,545 articles on Seeking Alpha over the period 2005 to 2015 and 147,916 articles on Motley Fool from 2009 to 2014, to obtain the content of the articles, authorship, and in the case of Seeking Alpha, comments from other users. This broader sample serves two purposes. First, it allows us to gauge how markets respond to articles on these platforms in general. Second, it is used to assess the broader response of consumers and producers of information on these platforms to the public awareness of fraudulent news from the scandal in the next section.[13]

We examine *abnormal* trading volume around the publication of articles on these platforms to capture whether investors "react" to the articles. It is likely the reverse is true as well, that articles react to trading activity. To try to establish some causal interpretation, we examine *abnormal* or unexpected changes in *future* trading volume from when the article is written. A reverse causality story would imply that authors are writing articles in anticipation of *future unexpected* trading activity, which we would call "news." We also control for lagged abnormal volume from the previous trading day to capture reaction to all events up to day $t-1$. Establishing a direct causal link between articles and trading activity is challenging due to omitted variables likely driving both. However, establishing causality here is not the goal. Rather, our aim is to examine the equilibrium response of market participants, consumers, and

[13] In a previous version of the paper, we also used these articles for a third purpose to apply a linguistic algorithm to detect false content, that we validate and calibrate using Rick Pearson's known fraud and legitimate articles as a training sample. The algorithm had some success, but with a lot of noise. We relegate those results to an internet appendix, as they are tangential to our focus here.

producers of news to the shock from the SEC event.

We further breakdown trading activity into retail (versus institutional) trading. We conjecture that retail traders, in particular, are participating and responding to these platforms and hence expect a stronger response from retail trading activity. We identify retail trades using TAQ data and the algorithm proposed in Boehmer et al. (2020). Their method uses the fact that most retail trades do not take place on registered exchanges. Instead, they are often filled internally by the broker or are sent to a wholesaler (e.g., Citadel). Those trades must be reported to a FINRA Trade Reporting Facility (TRF). These TRF executions are reported in TAQ with exchange code "D." Many orders that are executed off-exchange are given a small price improvement relative to the National Best Bid of Offer. Following Boehmer et al. (2020), we classify trades with TAQ exchange code of "D" and prices with just above or below a round penny as retail trades.

Table 1 examines the relation between future three-day abnormal retail trading volume and article publication on these sites. We define abnormal retail trading volume for stock i on day t as $RetVol(i,t)/\frac{1}{T}\sum_{k=1}^{140} RetVol(i,t-k)$, which is the retail trading volume for stock i on day t relative to the average daily retail trading volume in stock i over the last 6 months.[14] We sum abnormal retail volume over days $t=0$, $t+1$, and $t+2$, where $t=0$ is the date the article appears on the website. We then regress the natural logarithm of abnormal retail volume on an indicator variable for whether an article on these sites appeared about firm i on date $t=0$. We include year-month fixed effects in the regression. We examine only firms that had at least one article published on Seeking Alpha or Motley Fool over the sample period (about 7,700 unique firms and 9.8 million firm-day observations).

As the first column of Table 1 shows, an article published on these platforms is associated with a 36.5% increase in abnormal retail trading volume over the three days following publication. This result implies that investors are either trading in direct response to the articles or, more generally, in response to whatever news is coming out that day that these articles are also discussing. While the increase seems large, this first regression controls for no other variables, except time fixed effects. As we show below, the bulk of the effect is also concentrated in very small and illiquid firms, where percentage trading volume changes can be large.

Articles are often written following press releases or corporate filings with the SEC, so in the second column we control for whether there is an SEC filing (10-K, 10-Q, or 8-K), a company-issued press release, or a media article (WSJ and NYT) in the three days leading up to the article. We also include lagged abnormal retail trading volume as a regressor to help capture other events affecting trading activity. With these controls, the effect on abnormal retail trading volume over the next three days is

[14] Results are identical defining abnormal volume relative to the last 30, 60, or 180 days.

19.5%. To make sure that these results are not all coming from the day the news is released, Table C1 in Appendix C reports the effect on retail trading volume separately for the same day and for one and two days after the article's publication. Of the 19.5% rise in abnormal retail trading volume, 8.8% occurs on the day the article is published, 6.2% the following day, and 4.5% two days later. The abnormal retail trading volume following an article increases for about two weeks before returning back to its previous 6-month daily average.

The third column of Table 1 reports results for small firms separately, where we interact a small firm dummy (defined as firms smaller than the bottom 10^{th} percentile of NYSE firms based on beginning of month market cap) with the article dummy. The effect on abnormal retail trading volume rises strongly for small firms, with the effect being nearly three times larger for small firms. This result is consistent with small firms having less volume and liquidity, less active large investors, and a more opaque information environment.

3.2 How impactful are the fraudulent articles?

The last column of Table 1 adds a dummy variable for the 171 fraudulent articles identified by Rick Pearson and the SEC. The estimated coefficient on these fraudulent articles from the scandal is 0.554 with a t-stat of 2.22, indicating that the fraudulent articles increased retail abnormal trading activity by 55.4%, which is 2.5 times larger than the effect from a typical article published on these platforms. This result indicates that the fraudulent promotional articles were successful in having impact on investor trading. This effect seems large, but is reasonable considering that the SEC selects cases ex-post that had the largest impact. Moreover, the effect is in line with excerpts from recent SEC lawsuits.[15] The stronger impact on trading activity may also be driven by fraudulent articles being more sensational and diffusing more quickly across readers (Vosoughi et al. (2018)). The larger impact from the promotional articles may also indicate that they are different than other articles along other dimensions that might also affect trading volume, which we investigate later.

3.3 Return reaction event study

If the fraudulent articles were successful in pumping up the stock price, then we should see an impact on share prices as a direct result of these promotional articles. To test this conjecture, we focus on the sample of firms from Rick Pearson and the SEC and compare the fraudulent articles to a set of legitimate articles written by the same authors in order to difference out unobservable heterogeneity in author style

[15] From Case 1:18-cv-08175 filed on September 7, 2018 in the U.S. District Court, Southern District of New York: *"The market reacted strongly to the Company A promotion: the trading volume of Company A stock rose from approximately 1,100 shares on September 25, 2013 to over 4.5 million shares on September 27, 2013 and to more than 6 million shares on October 2, 2013."* And, in the same case about another firm, *"The article did not disclose that the author had been paid by Company B – at Honig's direction – to write the article. After the article was published on February 3, 2016, there was a 7000% increase from the previous day's trading volume, and an intraday price increase of over 60%."*

and reputation to better identify the impact of the promotional articles controlling for other effects that may also affect investor attention. We obtain 334 non-fake or legitimate articles published on the same platform, covering 171 companies, from the same set of authors that were investigated by the SEC.

Figure 2 plots the difference between cumulative abnormal returns (CARs) for days with fraudulent articles, relative to days with legitimate articles, from 20 days before the article's publication to 250 days after, for small and non-small firms separately. CARs are defined as the return on the stock minus a portfolio of stocks with similar size, BE/ME, and momentum characteristics following the procedure of Daniel, et al. (1997).

Abnormal returns for small firms increase after a fraudulent article is published (relative to legitimate articles), reaching as much as 13% cumulatively after about 60 days, before giving up all the gains and ending with a cumulative −5% return after a year. This finding supports another implication that the initial price impact from fraudulent news should eventually reverse. This pattern matches that of Galena Biopharma, the first SEC-prosecuted case, in Figure 1. The permanent price impact of −5% for small firms indicates either that once the market figured out there was fraud, investors viewed this as a bad sign about the firm, or that the true price should have dropped by 5% initially, but the promotional campaign temporarily propped up the price. For non-small firms, the price starts dropping immediately after the fraudulent article comes out and continues to decrease throughout the year. This result is consistent with the market figuring out the news is false immediately for larger firms, where the cost of information is lower. The results are consistent with the fraudulent articles uncovered by Rick Pearson and the SEC being part of pump-and-dump schemes designed to prop up share prices, that eventually reverses.

3.4 Robustness

For robustness, Table C2 in Appendix C reruns the regressions in Table 1 replacing abnormal retail volume with abnormal total volume as the dependent variable in Panel A and with the idiosyncratic price volatility of the stock in Panel B. We measure idiosyncratic volatility as the square of the difference between the return of the stock and a matched-portfolio of stocks with similar size, book-to-market equity (value), and past 12-month returns (momentum) following the procedure of Daniel et al. (1997). The dependent variable is the sum of daily idiosyncratic volatility on the day the article is published plus the next two days. We examine price volatility as opposed to signed returns because it is difficult to sign the direction of the content of the articles.[16] If the market already incorporated the news, then

[16] Textual analysis used to derive sentiment (Antweiler and Frank (2005), Tetlock (2007), Das and Chen (2007), Jegadeesh and Wu (2013), Heston and Sinha (2017), Boudoukh et al. (2018)) is notoriously challenging and noisy. In addition, price movements are deviations from expectations, so a "positive" article that is less positive than expected would predict a negative return. Not knowing expectations makes signing the price movement even more difficult.

the expected absolute return change should be zero.

We find that daily price volatility rises following articles published on these platforms, even after controlling for recent SEC filings, firm press releases, media articles, and lagged return volatility in the days leading up to the article. The effect is also strongest for the smallest firms. The magnitude of these effects is large but not unreasonable. Across all articles, the effect of a published article on idiosyncratic stock volatility is about 7.6% over the three days, which is roughly an additional 40% of the stock's normal price movement on days when there is no news about the firm on these platforms. For the smallest firms, the effect is an order of magnitude larger, which is consistent with extreme price movement for the smallest stocks (Frazzini et al. (2018)).

The last column of Panel B of Table C2 reports results for the impact on idiosyncratic volatility from the fraudulent articles. If arbitrageurs dominate in markets, then there will be little price response from the false articles despite the impact on trading volume. Conversely, if naive investors dominate, then prices will move with volume. As the table shows, the fraudulent articles have an additional significant impact on price movements relative to legitimate articles. The magnitude is large, at almost 3% over the three days following the promotional articles, which is not surprising considering this scandal grabbed the attention of the SEC.

The results indicate that articles written on these platforms coincide with increased retail trading volume and larger price movement for the stocks mentioned in the articles several days after the article's publication. This evidence is consistent with either the articles causing increased investor attention and/or coinciding with other news and being a symptom of that attention, which influences trading and prices. The impact on volume and volatility are even greater for fraudulent articles.

As a further direct link between articles published on these platforms and trading activity, we also obtain a proprietary supplemental dataset from Seeking Alpha on the readership of articles. The data only covers calendar year 2017, but contains daily number of "clicks" (i.e., number of times a given article is opened in a browser) and the number of times the article is "read," which is the instances in which a reader scrolled to the end of the article. In total, the dataset covers 25,596 articles about 3,118 publicly traded firms.

Table C3 in Appendix C presents results from regressing abnormal retail trading volume following the release of the article on the readership circulation of the article over the first three days after the article is published. The table shows that future abnormal retail trading volume is positively related to the number of clicks and number of times the article is read by investors, suggesting articles that influence circulation and readership are also associated with more retail trading activity in the stock. While establishing causality is difficult, and both readership and trading activity may be driven by the importance of the news covered in the article, this evidence more directly links the articles to future trading activity.

The results highlight that articles on these platforms are associated with greater trading activity and more price volatility, consistent with the literature that these platforms matter. Moreover, the fraudulent articles identified by Rick Pearson and the SEC have even greater impact. These two sets of findings indicate that the SEC announcement of the scandal could have a significant impact on markets if it changes the way participants consume news from these platforms (e.g., trust) and/or how news is subsequently produced on these platforms. We investigate whether these relations change after the SEC announcement and document the equilibrium response to the shock of trust in this news source.

4. Indirect Impact of Fraud

Our main and novel test is on the indirect impact of fraudulent news. The public revelation of the SEC's investigation and subsequent media attention around it provides a shock to investor awareness of fraud and hence trust in the news source. We use this event to test the spillover effects of fraud and how changes in trust can impact financial activity. The indirect effects of fraud are novel to the manipulation literature, which focuses primarily on the direct effects of misinformation. This analysis also distinguishes the role of trust from other theories, which do not imply any spillover effects (Allcott and Gentzkow (2017), Kshetri and Voas (2017), and Aymanns et al. (2017)).

4.1 Spillover effects from a shock to investor awareness of fraud

We use the period from February to March 2014 as the event that provides a shock to investors' awareness of fraudulent news on knowledge-sharing platforms. We examine the abnormal retail trading activity associated with published articles over the six months prior versus the six months after the event (i.e., August 2013 to January 2014 versus April 2014 to September 2014).

Panel A of Table 2 examines the impact of all published articles on future abnormal retail trading volume before versus after the scandal. The first column reports results from a regression of future abnormal retail volume on an article indicator, plus controls for SEC filings, press releases, print media, and lagged volume. Confirming the results from Table 1 over this one-year snapshot of our sample (6 months before to 6 months after the SEC announcement), published articles on these platforms are associated with significantly higher future abnormal retail trading activity. The second column adds the 6-month post event indicator, and its interaction with the article dummy. The negative interaction term with the post-event dummy shows that the effect of articles on future retail trading volume decreases significantly after the scandal by about 3.8% relative to before the scandal. This result is consistent with investors becoming aware of fraudulent content and muting their response to news in general on these platforms.

The third column adds the small firm dummy and interacts the small firm dummy with the article indicator and the post-event dummy. First, future retail trading volume associated with a published article declines by 4.1% after the scandal's announcement (t-stat of -2.30). Second, the retail volume response to articles is still much stronger for small firms (coefficient of 0.643 with a t-stat of 10.00). But, third, the triple interaction of article $\times$ small $\times$ post-event indicates that retail trading volume declines even more so for articles about small firms *after* the scandal (coefficient of -0.235 with a t-stat of -2.94). This last result indicates that retail trading volume for small stocks is the most sensitive to news from these platforms and hence is affected most by the announcement of fraud on these platforms. Retail trading activity in the firms where it matters most – small firms – are the most responsive to the shock of the scandal.

4.2 Identification and addressing alternative explanations

While we interpret our findings as the shock of fraudulent news on these platforms subsequently affecting the impact of future articles on financial markets, we consider here alternative explanations. For instance, the market environment surrounding the SEC announcement may have changed, which could also have caused variation in the relationship between news and market activity. In addition, the news generation process itself could have changed. In short, there is an identification challenge in understanding the relationship between news and trading activity, since there are many potential omitted variables affecting both.

To address alternative explanations for our results, we run a number of additional tests. First, we remind the reader that the timing of the SEC event should be exogenous to the outcome variables we study, since the source of the scandal was Rick Pearson's undercover work, which was initiated by him being idiosyncratically approached by a PR firm. In addition, we control for other news sources (SEC filings, press releases, and print media) happening at the same time in order to account for variation in news in general. However, to more directly and convincingly address alternative hypotheses and potential omitted variables, we examine pre-trends in trading activity in response to news and test whether reaction to news is simply lower in the post-scandal period for other, unrelated reasons. We also conduct several placebo, or falsification, tests to rule out alternative explanations by looking at institutional trading activity (who are largely not attending to these social media cites) and examining the market's response to news sources unrelated to and unlikely affected by the scandal (such as the WSJ and NYT or press releases). We also use an instrumental variable approach for news coverage –

mass shootings – as an instrument for attention to these platforms and their articles that is otherwise unrelated to our outcome variables. Finally, we provide some other corroborating evidence that helps directly identify that change in trust in this news source drove the market's changing response to articles on these platforms.

4.2.1 Pre-trends

We first examine pre-trends in the market's reaction to news leading up to the event. We regress daily abnormal retail trading volume on dummy variables for an article being published, the post-scandal period, and their interaction:

$$Log(RetAbVol)_t = \alpha + \beta_1 Article \times PostEvent + \beta_2 Article + \beta_3 PostEvent + Controls + \epsilon.$$

The controls include other sources of news about the firms at the same time from the WSJ, NYT, press releases, and corporate filings, as well as lagged volume and firm fixed effects. Figure 3 plots the coefficient β_1 at the daily level (with 95% confidence error bars drawn on the graph), which represents the difference-in-difference reaction to published articles after versus before the scandal, from seven days before to four trading weeks after each article's publication. A week prior to an article's publication, there is no difference in retail trading volume response after versus before the scandal, indicating no pre-trend in response to news before the scandal. Hence, the decline in retail trading volume we find after the SEC announcement does not appear to be driven by trends in the market's response to news that were happening at the same time. After the scandal, however, there is a significant drop in retail volume response to articles on the publication date and for the next two trading weeks. These results suggest that investors' reaction to articles on these platforms decreases significantly after the scandal, lasting for two weeks before returning to normal trading levels, and appears to be a direct result of the announcement rather than any trends in market activity.

Another possibility is that articles became less timely after the scandal, and thus investors react less to them, as that information has already been incorporated into the market. Table C4 in Appendix C looks at the timeliness of articles before versus after the scandal by regressing the incidence of an article on the lagged abnormal return (with respect to size, value, and momentum benchmarks) of the firm discussed in the article, using lagged returns over the previous day, week, and month. We find no evidence that articles are responding to returns of firms any faster or slower before versus after the scandal. Hence, the timeliness of article publication (at least with respect to abnormal price movement) appears to be the same after the scandal and hence is unlikely driving the lower trading volume we find after the SEC announcement.

4.2.2 Falsification tests

We conduct falsification exercises to rule out many alternative explanations for our results. We first examine institutional trading activity, which should not be affected by the social networks, and hence the SEC announcement, since institutional traders do not participate on those networks. However, if the SEC announcement is confounded by omitted factors that impact trading and news in the economy generally, then we expect institutional trading to show just as large an effect. Second, we examine the retail trading response to other news sources unlikely to be impacted by the SEC announcement, as they were not part of the scandal, such as articles published in the NYT and WSJ, press releases, and corporate filings. We examine the retail response to these unrelated news sources before versus after the SEC announcement, where we should expect no differential effect coming from the announcement itself, but would find an effect if the announcement happened to coincide with other news or changes in the market's response to news generally.

Falsification test #1: Institutional trading.

The previous results on trading activity pertain to retail trading volume, since retail traders are primary consumers of these social network platforms. As a test to address omitted variables related to news and trading activity, we examine institutional trading volume. The idea is that institutional trading should also be affected by omitted variables that are driving news and general trading activity in the economy, but because institutional traders are not primary participants on these social networks, the effect of the scandal should not otherwise impact institutional trading.

To measure institutional investor trading, we do not simply subtract retail volume from total trading volume, recognizing that the Boehmer et al. (2020) method for identifying retail trades will not capture all retail trading. Instead, we follow Bushee et al. (2020) and use large trades (trades greater than or equal to $50,000) as a measure of institutional volume. While some institutional investors split their trades into small orders (Frazzini et al. (2018)), implying we might miss those, we are confident that large trades should reflect only institutional investor activity. The point here is to capture trades we are confident are institutional trades and not to necessarily capture all institutional trading.

Panel B of Table 2 examines the impact of all published articles on abnormal *institutional* trading volume before versus after the scandal. As the first column shows, future abnormal institutional trading volume also rises following the publication of an article on these platforms. Since institutions do not participate on these platforms, we conclude that this effect is likely driven by omitted variables driving trading and news generally. However, the second column indicates that the response of institutional trading was unaffected by the SEC's announcement of the scandal. The interaction between the Post-event dummy and an article is −0.013 with an insignificant t-stat of −0.59. Looking at the third column of Panel B, we also find no evidence that institutional trading dropped in response to the SEC announcement for the smallest firms either. Hence, unlike for retail trading volume, we find no evidence that institutional trading volume in response to articles on these platforms was influenced by the SEC announcement. These results are consistent with the scandal having an impact on those market participants who pay attention to the social networks and helps rule out concerns that omitted variables associated with trading and news, which should affect both retail *and* institutional traders, is driving these results.

Falsification test #2: News sources unrelated to the scandal.

To rule out other potentially confounding effects associated with news or investor reaction at the same time, we examine whether the scandal impacted the response to other, unrelated news sources that should not have been affected by the scandal and SEC announcement. We examine SEC corporate filings, press releases, and articles published in the WSJ and NYT. Panels A and B of Table 2 show that these other news sources all have strongly significant and positive impact on abnormal retail and institutional trading volume, respectively, even after controlling for lagged abnormal trading volume in the day prior to the article's publication. However, we do not expect the market's reaction to these news sources to change after the SEC announcement, unless the SEC announcement is confounded by omitted unobservables driving trading and news. To test whether reaction to these news sources changed after the scandal, we interact the Post-event dummy with SEC filings, firm press releases, and other news media (WSJ and NYT articles). The interaction terms serve as falsification tests or "placebo" tests of whether the market responds to other news differently after the scandal. These tests address alternative explanations for the decline in trading volume after the scandal. For example, if the post-event period happened to coincide with less information content, less news, less firm activity, or less trading to news, then the Post-event reaction to *all* news sources should be significant. If, however, investors responded to the scandal itself, which just pertained to the social knowledge-sharing platforms, then we should only see a decline in trading response to news on these platforms and not to other media sources or corporate actions.

Panel C of Table 2 reports the results of these falsification tests. As the first column shows, the interaction terms for retail trading volume are statistically and economically zero, with two of the three coefficients having the wrong sign to be consistent with the alternative story. We find no discernible difference in abnormal retail trading volume response to other news sources before versus after the scandal. This evidence is inconsistent with trends in news, trends in trading activity, or changes in response to news generally explaining our findings. The scandal only seems to affect the market's response to articles contained on the social networks.

The second column of Panel C of Table 2 repeats these tests for abnormal institutional volume, which also shows no effects, economically or statistically, after the scandal. Institutional trading in response to news sources unrelated to the social platforms exhibit no change after the SEC announcement of fraud on the social networks.

The evidence from these falsification tests indicates that the market's response to news in general – both from retail and institutional investors – exhibits no changes after the SEC announcement, rejecting the alternative hypothesis that omitted variables are driving the decline in response to social media articles after the announcement. Rather, our results appear to be most consistent with retail investors, who primarily participate on the social platforms, discounting all news on the social media platforms, including legitimate news, due to distrust after the revelation of fraud on those networks. The magnitude of the drop in abnormal retail volume is even larger and more significant (-5.6% drop per article with a t-stat of -2.97) after accounting for interactions of the other news sources with the Post-event dummy. In other words, controlling for potential changes after the SEC announcement in the market's response to other news sources unlikely affected by the scandal, our results on retail trading activity become stronger, suggesting that, if anything, omitted variables may be understating our findings. The results for institutional trading volume remain statistically no different from zero, consistent with institutional traders being largely unaffected by the social media news source.[17]

4.2.3 Instrumental variables

Another way to address the identification challenge is to find an instrument that moves investor attention to news and/or news coverage, but is otherwise unrelated to market activity. Peress and Schmidt (2020) show that big-news events crowd out coverage and attention to other news and distracts retail traders from trading. We use mass shootings as a shock to investors' attention. Since mass shootings should be unrelated to the SEC scandal, the crowding out effect should be no different before versus after the SEC event. While this instrument is useful in shifting investor attention that is distinct from any economic event or information, it does not instrument for any changes associated with the SEC event itself (e.g., potential changes in news or response to news that coincided with the event). However, this test serves to highlight whether the market's muted response to financial news on days

[17]Panel B of Table C5 reports results from regressions using idiosyncratic price volatility as the dependent variable. There is also significantly reduced impact on price volatility from articles after the scandal for small firms. In addition, we find an *increase* in price impact from press releases, SEC filings, and other news media after the scandal, which indicates that other trends in news or omitted variables in the market's response to news are, if anything, biasing our results in the opposite direction.

of sensational non-economic news (e.g., mass shootings) is any different before versus after the SEC event. In other words, we test whether an exogenous shock to investor attention has the same impact on articles published on these networks before versus after the scandal, providing a further test of whether retail trading activity shifted in any way in response to news or attention to news generally over time.

Table 3 examines whether investors' reaction to mass shootings in the US changes after the public announcement of the SEC investigation. Panel A examines the impact on abnormal retail trading volume from mass shootings news over the entire sample. We regress abnormal retail volume over the subsequent three days on a dummy variable *Shooting*, which equals 1 if there was a mass shooting in the US on day t (as listed on https://en.wikipedia.org/wiki/List_of_mass_shootings_in_the_United_States), and 0 otherwise. The regressions include the usual controls. We find that in our full sample, mass shootings coverage lowers the retail trading response to the social network's articles over the next three days. The first column of Panel A shows that there is a 1.9% drop in retail trading activity (t-stat of -10.99) over the three days following a mass shooting in the U.S. These results are consistent with a crowding out effect for investor attention Peress and Schmidt (2020). The second column shows the effect on retail trading from mass shootings is no different for small firms. The third column of Panel A examines only those days when articles are published on the social media platforms, which show an even larger decline on retail trading activity when mass shootings occur of 2.7%.

Panel B of Table 3 examines the period from 6-months before the SEC announcement to 6-months after and interacts the mass shootings dummy with the Post-event dummy to test for the differential response to articles on mass shooting days before versus after the SEC announced investigation. In this one year subsample, we first confirm that mass shootings have a negative impact on overall retail trading volume (columns 1 and 2) and then examine whether the response to mass shootings is any different before versus after the SEC event by looking at the interaction between Shooting $\times$ Post. We find that the crowding out effect from mass shootings on the social media networks is the same 6 months before versus 6 months after the event, as the interaction term in column 3 is statistically indistinguishable from zero. Even for the smallest firms (column 4), the interaction term is insignificant (and switches sign). This evidences shows that exogenous shocks to attention of financial news coming from mass shootings, impact retail trading volume no differently after the SEC announcement, which suggests that investor attention and response to news generally has not changed following the announcement. The only significant change appears to be the direct effect of the announcement itself on trust in these social network news platforms (the previous results in Panel A of Table 2).

4.3 Further corroborating evidence

Finally, in addition to attempting to rule out alternative explanations for our results, we also provide corroborating evidence for the trust channel driving these results. We show evidence for increased mistrust following the scandal by looking at cross-sectional characteristics that are consistent with the awareness of fraud causing the decline in retail activity.

4.3.1 Article and author characteristics

We test additional implications by examining cross-sectional variation across articles and authors. We provide additional evidence that the decline in volume response to these articles after the scandal is due to investors becoming aware of the fraud, consistent with a shock to trust.

Panel A of Table 4 reports results from regressing abnormal retail trading volume associated with an article on various article and author characteristics (controlling for other news from corporate filings, press releases, WSJ and NYT articles, lagged abnormal trading volume, and year-month and firm fixed effects, and in some specifications, author fixed effects). We first look at the number of followers of the author, a proxy for author popularity. Conditional on an article being published, authors with greater past followings have a significant positive impact on retail trading volume (an additional 1.3% with a t-stat of 8.96). We also show (second column) that authors who have written more articles in the past have a bigger impact on retail volume (increase of 0.5% with a t-stat of 3.29). These results suggest that articles written by authors with better reputations and more experience are associated with more subsequent retail trading volume.

Columns 3 and 4 analyze whether the retail trading volume response to an article is higher when the article is more quantitative in nature and/or references accounting data, where presumably the information is easier to verify from other sources. We look at the fraction of the article text comprised of numbers, and fraction of text containing the bigram "earn." These articles have a bigger impact on retail trading volume of 2.6% and 4.5% (t-stats of 9.07 and 12.65), respectively. The results are consistent with investors putting more trust in articles written by more followed and experienced authors, as well as articles with hard numbers. The alternative explanation is that such articles are associated with bigger news. We try to distinguish between these two hypotheses below by looking at how they relate to retail volume before versus after the scandal, which we argue was a shock to trust in these platforms.

We also examine the comments section to each article. The number of comments is strongly positively related to retail trading volume (increase of 5.4% for a one standard deviation increase in number of comments, with a t-stat of 10.84), as is the average length of comments (increase of 2.3% with a t-stat of

6.68). The number and length of comments may proxy for disagreement about or interest in the articles, or may be a symptom of more activity associated with articles containing more news. To more directly look at disagreement, we use natural language processing on the comments and create a dummy variable for whether disagreement words ("disagree," "differ," "counter," "oppose," or "argue") or "wrong" words ("wrong" or "not right") show up in the comments section. In both cases, associated abnormal retail trading volume accompanying these articles rises. Viewed as proxies for disagreement, these results are consistent with theories arguing that disagreement generates trading.

We also add up mentions of the words "fake" or "fraud" in the comments section and compute a dummy variable equal to one if readers use these words. We find an additional positive impact on retail trading volume when comments contain these words. Thus, comments of "fake" and "fraud" are associated with *more* trading activity. However, we will show that this result changes after the SEC announcement. Finally, we also include a dummy variable for authors who disclose they have a position in the stock they are writing about. The impact on volume is negative but insignificant.

Panel B of Table 4 examines whether the author and article characteristics have any differential impact after the SEC announcement of the scandal. We rerun the regressions from Panel A, examining the time period six months before to six months after the SEC scandal, and interact each characteristic with the Post-event dummy. The interaction terms suggest that mistrust in these platforms as a result of the public's awareness of the fraud, likely drives the reduction in retail trading activity we document after the event. For example, as the first two interactions show, the drop in trading volume is not as large if the author has more followers and has written more articles in the past. In other words, reputational cues of the author became more important after the scandal, and do not appear to be important before the scandal. The results are consistent with market participants maintaining more trust in articles written by authors with better reputations. The alternative hypothesis that certain authors endogenously match with certain types of news articles should be unaffected by the event, unless the event caused a shift in the matching of authors to news. We will look at the responses of consumers and producers of articles in the next subsection.

Further support for the investor trust channel is that we also find positive interactions with the Post-event dummy for articles that discuss earnings or offer hard numbers, suggesting these articles are not discounted as much after the scandal.

Looking at measures of disagreement, such as the number and length of comments and comments containing "disagree" or "wrong" words, we find no significant interactions with the Post-event dummy. Although these measures of disagreement are shown to have a strong positive overall association with retail trading volume (Panel A), their impact on trading is no different after the scandal. These results indicate that disagreement is no more influential on trading activity before versus after the scandal, which makes sense since the SEC announcement should have no effect on disagreement per se. In essence, the results for disagreement provide another placebo test that shows that investor disagreement and its impact on trading had no coincidental shift with the SEC announcement. This result suggests that other influences on trading besides trust had no change after the scandal. Only measures of reputation or trust seem to have more prominence after the scandal.

Finally, we also interact the Post-event dummy with the dummy for "fake" or "fraud" appearing in the comments section. While such comments have a positive impact on retail trading volume before the scandal, they are associated with even larger declines in retail trading volume after the scandal. Hence, concerns of fraudulent content echoing in the comments to an article significantly reduces the article's impact on retail trading activity, but only *after* the scandal came to light. Finally, disclosure of an author's position in the stock also has a significant negative impact on volume, but only after the scandal as well (-10.3% with a t-stat of -3.10). Overall, these results further support that the decline in retail trading activity for articles published after the SEC announcement is a response to the shock of distrust in news from these platforms because of the scandal.

4.4 Equilibrium response of authors and readers

Another explanation for a muted volume response after the scandal is that readers are responding to changes in author and article characteristics, rather than trading less due to erosion in trust. Here, we examine the equilibrium response of readers and authors to the SEC announcement. Table 5 reports results from regressions of article and comment characteristics – fraction of numbers used in the text, mentions of earnings, number and length of comments and use of disagreement, "wrong," and "fake" words in comments – on the Post-event dummy, looking only at articles published 6 months before to 6 months after the SEC announcement period. We also control for other news about the firms through SEC filings, press releases, NYT and WSJ articles, and lagged abnormal volume, and include firm fixed effects. We run the regression separately for small firms (smallest NYSE decile) and non-small firms, reported in Panels A and B, respectively.

For small firms, we find that after the scandal, articles are more likely to be written about earnings or accounting information, which makes sense if authors recognize that the scandal may have eroded trust from their readers. Note, however, that our previous results show that these changes should *increase* rather than decrease retail trading volume. Hence, authors' equilibrium response to focus on more hard accounting information cannot explain our findings and actually understates our retail trading result.

We also examine the comments to articles to gauge the equilibrium response of readers to articles after the scandal. Consistent with readers distrusting these platforms more after the scandal, incidences of the words "fake" and "fraud" in comments increased significantly after the SEC event. Conversely, measures of disagreement, such as the number and length of comments, use of "disagree" or "wrong" words, is no more prevalent after versus before the scandal. These results further augment the retail trading volume results from Table 2.

We can also see if author characteristics changed after the scandal, either due to self-selection or perhaps the platforms sought to improve their reputation. In Panels C and D of Table 4, we regress the number of followers an author has, as well as the number of articles the author has written in the past, on the Post-event dummy. We find that both measures are higher in the six months after the scandal relative to the six months prior to the scandal for small firms. Hence, producers of news on these platforms do seem to respond to a new equilibrium created by the SEC shock. We interpret that response as being consistent with a shock to trust and a response to try to improve that trust. Note, however, that this response by news producers should increase retail trading (evidenced in Table 4, Panel A), which therefore underestimates our finding that retail trading activity decreases after the scandal, going the opposite direction to be able to explain it.

4.5 Article Characteristics and Authenticity

We can further explore the equilibrium response of consumers and producers by analyzing the actual text of the articles and comments. We apply linguistic tools to analyze the content of the comments and the articles themselves. We use the Linguistic Inquiry Word Count model (LIWC2015) from Pennebaker et al. (2015), which focuses on individuals' writing or speech style to measure individuals' cognitive and emotional states across various domains. The LIWC model outputs the percentage of words that fall into one of more than 80 linguistic, psychological, and topical categories. We use the broad categories of writing style measured by LIWC: "clout" (a measure of confidence or expertise in expression), "analytical" (formal, logical, and hierarchical as opposed to informal, personal, and narrative), and "emotional tone" (positive or upbeat),[18] to see if articles and/or comments change after the scandal along these linguistic dimensions.

Table 6 reports the results from regressing the linguistic scores of the comments and articles on the Post-event dummy, including all the controls for other news and firm fixed effects. Panel A reports results for small firms and Panel B for non-small firms. The comments show a much more negative emotional tone after the scandal. The more negative tone could be consistent with readers being more skeptical of these platforms after the scandal.

[18] Tone here should not be confused with "sentiment" as used in the finance literature (Tetlock (2007)). The former relates to emotional expression, while the latter relates to whether the news for prices is expected to be positive or negative.

Turning to the articles themselves, the linguistic algorithm detects no difference in tone, clout, or analytic characteristics of the articles before versus after the scandal. Hence, the change in tone from the comments does not appear to be driven by a change in writing style of the articles after the scandal. The linguistic characteristics of the articles appear to remain the same after the event, with the exception of one category we discuss now.

One of the LIWC categories is "authenticity," designed to detect deception in expression. Pennebaker (2011) describes which linguistic traits are associated with authenticity. In particular, truth-tellers tend to use more self-referencing words and communicate through longer sentences compared to liars. When people lie, they tend to distance themselves from the story by using fewer self-referencing words ("I" or "me"). Furthermore, liars use fewer insight words such as *realize*, *understand*, and *think*, and include less specific information about time and space. Liars also tend to use more discrepancy verbs, like *could*, that assert that an event might have occurred, but possibly did not. Newman et al. (2003) use an experimental setting to develop an authenticity score based on expression style components using similar techniques, and the Central Intelligence Agency and Federal Bureau of Investigation use similar methods to assess authenticity.

As an example, consider the two statements by former U.S. congressman Anthony Weiner before and after his admission in the "sexting" scandal.

Before admission:

> ***We*** *know for sure I didn't send this photograph. [...]* ***We don't know*** *where the photograph came from.* ***We don't know*** *for sure what's on it. [...]* ***If*** *it turns out there's something larger going on here,* ***we****'ll take the requisite steps.*

After admission:

> ***I*** *would like to make it* ***clear*** *that* ***I*** *have made terrible* **mistakes***, that* ***I*** *have* ***hurt*** *the people* ***I*** *care about the most, and* ***I*** *am deeply sorry.* ***I*** *have not been honest with* ***myself****,* ***my*** *family,* ***my*** *constituents,* ***my*** *friends,* ***my*** *supporters and the media.*

The use of "we" versus "I" and "my", the discrepancy words "don't know" and "if", and the lack of insight words like "mistakes," "clear," and "hurt" are all more prevalent in his statements when he was lying.

The algorithm uses a combination of these linguistic traits to generate the authenticity measure. We apply this measure to the articles' text and examine whether the authenticity of articles changed before versus after the SEC scandal. As the last column of Table 6 shows, the authenticity of articles about small firms increases significantly after the scandal. This finding suggests that authors responded to the scandal by providing more authentic content or that the bad actors were removed from these platforms. The results are also consistent with small companies, who engage or were willing to engage in promotional articles before the scandal, ceasing or decreasing this activity after the scandal.[19]

While this equilibrium response by authors could also affect retail trading volume after the scandal, please note again, that this response from authors should *increase* trading volume, which cannot explain the total decrease in abnormal retail trading we find after the SEC announcement. Hence, this author response likely underestimates the decline in retail trading in response to articles after the scandal. However, it also indicates that one of the fallouts from the scandal was a shift to try and increase trust in these platforms.

Finally, another test to rule out that the equilibrium response of news producers is causing all of the volume decline post-event, is to shorten the event window to a period where it is less plausible that the platforms could alter their behavior. For example, if looking immediately after the SEC announcement, it is plausible consumers of news could respond immediately by discounting news from these platforms due to distrust, but less plausible that producers of news would be able to respond that quickly to alter their articles. Reducing the time-band around the scandal period reduces statistical power, but also makes it less likely that our results are driven by platform changes. In Figure 4 we examine the three-day retail volume response for articles before versus after the scandal over different horizons, ranging from articles published one week after the announcement up to 18 months after the scandal (with 95% confidence error bars drawn on the graph). As the figure shows, even defining the event to be very narrow – articles published a week after the scandal – the three-day abnormal retail trading volume response to an article is lower after the scandal. The decline is sharpest when looking one month after, but is still significant even if we define the event from a year before to a year after the scandal. While the decline in trading activity a year after the scandal is likely due to both demand and supply effects from news on these platforms, the results only a week or two after the scandal seem much more likely to come from the demand side, which we interpret as a decrease in trust from the fraud.[20]

5. Conclusion

We investigate a novel setting of fraud and market manipulation on socially shared financial news networks. Using the announced SEC investigation as a shock to investors' awareness of fraud on these platforms, we analyze the broader effects of stock market manipulation and trust on subsequent investor behavior. After the shock, we find significant spillover effects from investors discounting their response to all news from these platforms, including legitimate news, where trading activity from news on these

[19] In a previous version of the paper, we also tried to identify fake content among all of the articles in our sample, using the known fraudulent articles identified by Rick Pearson and the SEC to calibrate the model. Our ability to detect fake content was somewhat successful, the details of which are provided in an online Internet Appendix.

[20] Panel B of Figure 4 repeats the plot for abnormal institutional trading volume that exhibits no reliable changes after the SEC event for any window length.

platforms declines significantly. We attempt to rule out alternative explanations for these results, including omitted variables that may drive changes in news, trading activity, and investor response to news. The evidence points to the SEC event itself as having a significant impact on the market's subsequent response and not other explanations or trends that happen to coincide with the event. These findings provide new evidence on the *indirect* effects of market manipulation, consistent with models of trust and social capital.

Our findings may provide some of the first evidence of indirect spillover effects of fraud that match theory (Allcott and Gentzkow (2017), Aymanns et al. (2017), and Kshetri and Voas (2017)) and provide novel evidence for the role of social capital and trust (Guiso et al. (2004)). Our setting of financial markets may underestimate the broader impact of disinformation for settings where information costs are higher and the ability to take corrective action (i.e., through trading) is more limited.

References

Aggarwal, R. and G. Wu (2006). Stock market manipulations. *The Journal of Business 79*(4), 1915–1953.

Ahern, K. R. and D. Sosyura (2014). Who writes the news? corporate press releases during merger negotiations. *The Journal of Finance 69*(1), 241–291.

Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *The Journal of Economic Perspectives 31*(2), 211–235.

Antweiler, W. and M. Z. Frank (2005). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance 59*(3), 1259–1294.

Aymanns, C., J. Foerster, and C. Georg (2017). Fake news in social networks. *CoRR abs/1708.06233.*

Benabou, R. and G. Laroque (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility. *The Quarterly Journal of Economics 107*(3), 921–958.

Boehmer, E., C. M. Jones, X. Zhang, and X. Zhang (2020). Tracking retail investor activity. *Journal of Finance, Forthcoming*.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson (2018, 07). Information, Trading, and Volatility: Evidence from Firm-Specific News. *The Review of Financial Studies 32*(3), 992–1033.

Bushee, B., M. Cedergren, and J. Michels (2020). Does the media help or hurt retail investors during the ipo quiet period? *Journal of Accounting and Economics 69*(1), 101261.

Chen, H., P. De, Y. J. Hu, and B.-H. Hwang (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies 27*(5), 1367–1403.

Cookson, J. A. and M. Niessner (2020). Why don't we agree? evidence from a social network of investors. *The Journal of Finance 75*(1), 173–228.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance 52*(3), 1035–1058.

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

Delort, J.-Y., B. Arunasalam, M. Milosavljevic, and H. Leung (2009, 11). The impact of manipulation in internet stock message boards. *International Journal of Banking and Finance 8.*

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance 25*(2), 383–417.

Frazzini, A., R. Israel, and T. J. Moskowitz (2018). Trading costs.

Frieder, L. and J. Zittrain (2007, 03). Spam works: Evidence from stock touts and corresponding market activity. *Social Science Research Network Working Paper Series*.

Gale, D. and F. Allen (1992). Stock price manipulation. *Review of Financial Studies 5*, 503–529.

Gentzkow, M. and J. Shapiro (2005, October). Media bias and reputation. Working Paper 11664, National Bureau of Economic Research.

Gentzkow, M., J. Shapiro, and D. Stone (2015). Media bias in the marketplace: Theory. handbook of media economics, simon anderson, david stroemberg and joel waldfogel, eds.

Gomes, F., M. Haliassos, and T. Ramadorai (2021). Household finance. *Journal of Economic Literature Forthcoming.*

Gottfried, J. and E. Shearer (2016). *News Use Across Social Medial Platforms 2016.* Pew Research Center.

Guiso, L., P. Sapienza, and L. Zingales (2004, June). The role of social capital in financial development. *American Economic Review 94*(3), 526–556.

Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the stock market. *The Journal of Finance 63*(6), 2557–2600.

Guiso, L., P. Sapienza, and L. Zingales (2010, March). Civic capital as the missing link. Working Paper 15845, National Bureau of Economic Research.

Hanke, M. and F. Hauser (2008, 02). On the effects of stock spam e-mails. *Journal of Financial Markets 11*, 57–83.

Heston, S. L. and N. R. Sinha (2017). News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal 73*(3), 67–83.

Hu, B. and T. McInish (2013). Greed and fear in financial markets: The case of stock spam e-mails. *Journal of Behavioral Finance 14*(2), 83–93.

Hu, B., T. H. Mcinish, and L. Zeng (2009). The can-spam act of 2003 and stock spam emails.

Jegadeesh, N. and D. Wu (2013). Word power: A new approach for content analysis. *Journal of Financial Economics 110*(3), 712–729.

Knack, S. and P. Keefer (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly Journal of Economics 112*(4), 1251–1288.

Kshetri, N. and J. Voas (2017, November). The economics of fake news. *IT Professional 19*(6), 8–12.

Milgrom, P. and N. Stokey (1982). Information, trade and common knowledge. *Journal of Economic Theory 26*(1), 17–27.

Mullainathan, S. and A. Shleifer (2005, September). The market for news. *American Economic Review 95*(4), 1031–1053.

Nelson, K. K., R. A. Price, and B. R. Rountree (2013). Are individual investors influenced by the optimism and credibility of stock spam recommendations? *Journal of Business Finance & Accounting 40*(9-10), 1155–1183.

Newman, M. L., J. W. Pennebaker, D. S. Berry, and J. M. Richards (2003). Lying words: Predicting deception from linguistic styles. *Personality and social psychology bulletin 29*(5), 665–675.

Niessner, M. (2018). Strategic disclosure timing and insider trading. *Available at SSRN 2439040*.

Pennebaker, J., R. Booth, R. Boyd, and M. Francis (2015). Linguistic inquiry and word count: Liwc 2015 operator's manual. retrieved april 28, 2016.

Pennebaker, J. W. (2011). The secret life of pronouns: what our words say about us. Chapter 6. New York: Bloomsbury Press.

Peress, J. and D. Schmidt (2020). Glued to the tv: Distracted noise traders and stock market liquidity. *The Journal of Finance 75*(2), 1083–1133.

Sabherwal, S., S. K. Sarkar, and Y. Zhang (2011). Do internet stock message boards influence trading? evidence from heavily discussed stocks with no fundamental news. *Journal of Business Finance & Accounting 38*(910), 1209–1237.

Sapienza, P. and L. Zingales (2012). A trust crisis. *International Review of Finance 12*(2), 123–131.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance 62*(3), 1139–1168.

Ullah, S., N. Massoud, and B. Scholnick (2014, March). The Impact of Fraudulent False Information on Equity Values. *Journal of Business Ethics 120*(2), 219–235.

Vosoughi, S., D. Roy, and S. Aral (2018). The spread of true and false news online. *Science 359*(6380), 1146–1151.

Figure 1. **The Case that Launched the SEC Investigation: The Pump-and-Dump Scheme of Galena Biopharma Inc.**

This figure depicts the stock price of Galena Biopharma Inc. from April 2013 to May 2014, as well as occurrences of fake articles being published on Seeking Alpha, instances of SEO and stock options being granted to senior executives, as well as instances of insider trading and exposé articles about the promotional articles. This information was obtained from the SEC Lawsuit filed against Galena on 31 October, 2014 in the United States District Court (Case 3:14-cv-00558-SI). According to the lawsuit, the fake articles were published on August 6 and 22, 2013, September 26 and 30, 2013, November 12, 13, and 22, 2013, December 4, 10, 16, 2013, January 15, 2015, and February 5, 2014. While this was happening, Galena sold on September 18, 2013 in an SEO 17,500,000 units of stock for net proceeds to Galena of $32.6 million. On November 22, 2013, Galena held a board meeting and granted stock options to executives and directors with a strike price of $3.88. The CEO received 600,000 options, the CMO and COO 300,000 options, the CAO 150,000 options and each of the six directors received 200,000 options. Galena has historically awarded options either at the end of December or in early January. During the board meeting on January 16, 2014, where the board reviewed the preliminary 2013 earnings which had not been made public yet, the CEO declared that insiders could trade the company's stock immediately. Between January 17 and February 12, 2014 insiders sold over $16 million of their stock. On January 24 and 27, 2014 attention was drawn to the large insider trades. Then on February 1, 13, 14 and on March 13, 2014 articles started to appear on Seeking Alpha and TheStreet, documenting the promotional scheme. Finally on March 17, 2014, Galena disclosed in it's 10-K form the SEC investigation.

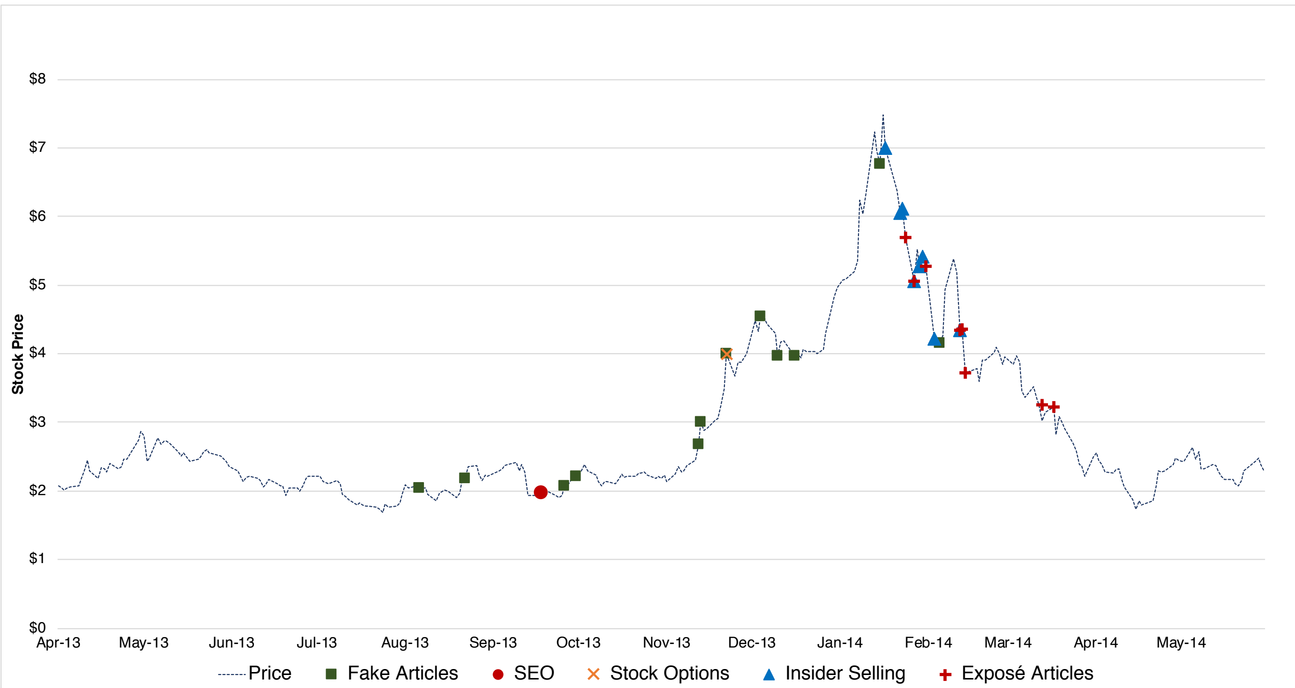


Figure 2. **Abnormal Returns for Fake Articles**

The figure depicts the difference in cumulative abnormal returns (residuals from a matched portfolio of stocks on size, BE/ME, and momentum) between days with fake articles and days with non-fake articles separately for small and non-small firms in our sample. The figure plots the returns for the for-sure fake articles from Rick Pearson and the SEC. Cumulative returns are measured starting with the day after the article's publication until 251 trading days after the article's publication. Before the article's publication, we measure cumulative returns starting with day -20 and ending on the day before publication ($t = -1$). Small firms are defined as firms in the bottom 10th percentile of NYSE firms.

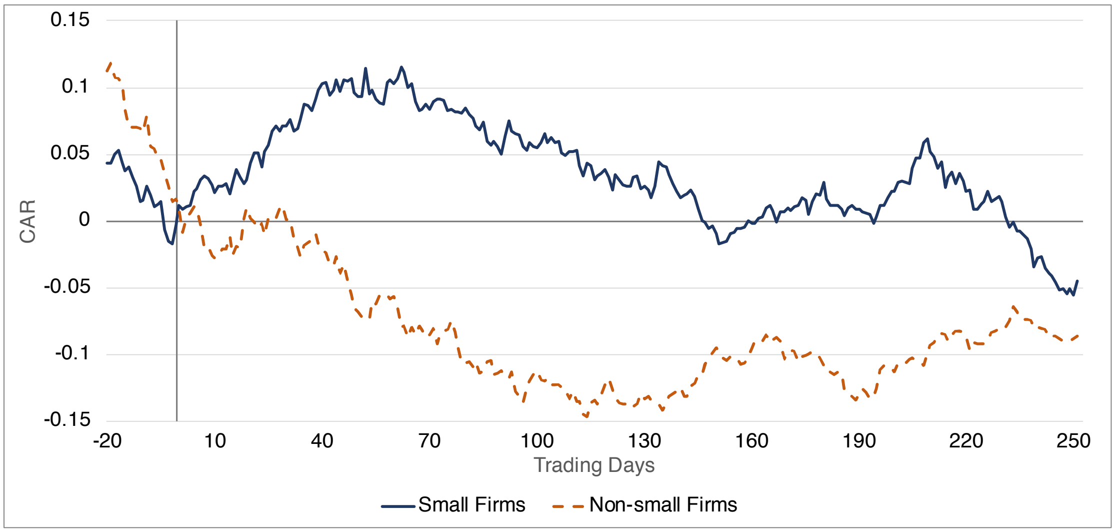


Figure 3. **Retail Trading Volume Response to Articles during the Six months *After* the 2014 SEC Investigation**

This figure plots the retail investors' reaction to articles published on the social media platforms during the six months *after* the SEC Lawsuit disclosure that publicly announced the presence of fake news on these platforms, and in the days around the article. We proxy for the market's reaction to the articles using daily abnormal retail trading volume. We plot the difference in reaction of abnormal retail volume around articles in the 6 months before versus 6 months after the disclosure period of the SEC announcement and press coverage (February-March 2014). We obtain retail trading volume from TAQ using Boehmer et al. (2020) method.
We estimate the following model for every day an article is published on Seeking Alpha and Motley Fool, and for days around the article publication:

$$Log(AbRetVol)_t = \alpha + \beta_1 Article \times \text{Post-Event} + \beta_2 Article + \beta_3 \text{Post-Event} + Controls + \epsilon$$

where $Log(AbRetVol)_t$ is defined as $RetVol(t)/AvgRetVol(t-146, t-20)$. *Post-Event* is a dummy variable equal to 1 if the article was published during the six month period after news of the SEC investigation broke, April 1 to September 30, 2014. The sample compares the period August 1, 2013 to January 31, 2014 to the period April 1 to 30 September, 2014. The figure plots the daily estimates of β_1 from seven days before the article's publication to four weeks after (trading days $t = -7, ..., 19$). Hence, $\beta_{1,t}$ represents the differential impact of the average retail trading volume response of an article after the SEC investigation relative to before the investigation on day t following the publication of an article on these social media platforms. The bars represent 90% confidence bands around the point estimates.
We also control for the presence of SEC filings, press releases, and print media coverage over the prior three days. SEC filing is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and press releases is a dummy variable if there was at least one press release issued by the firm over the past three trading days, and print media is a dummy if there was at least one WSJ or NYT article about the firm in the past three trading days.

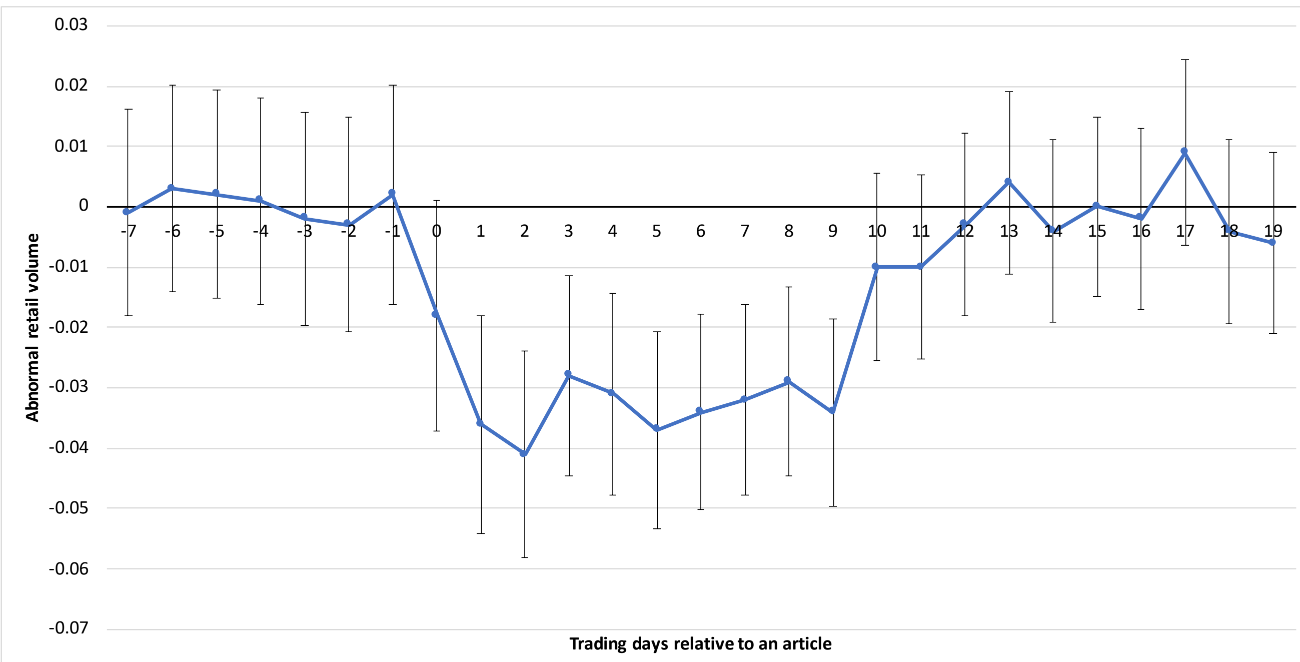


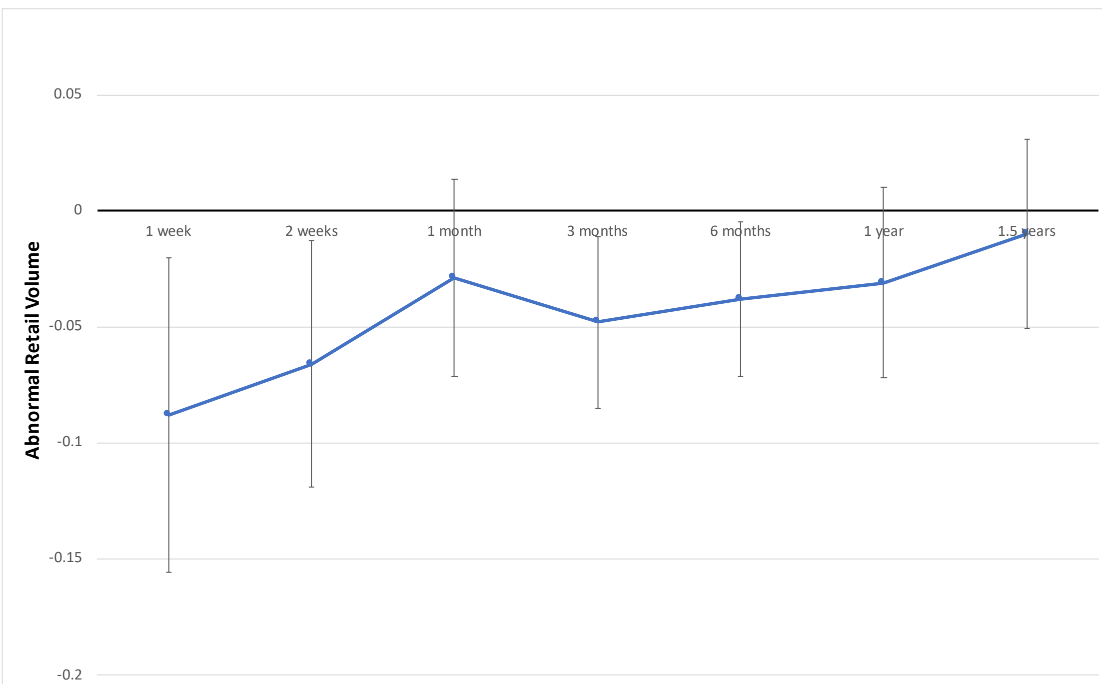
Figure 4. **Trading Volume Response to Articles over different Horizons *After* the 2014 SEC Investigation**

This figure plots the market's reaction to articles published on the social media platforms at various time bands *after* the SEC Lawsuit disclosure that publicly announced the presence of fake news on these platforms, and how that reaction changes over time after the scandal. We proxy for the market's reaction to the articles using abnormal daily trading volume. We plot the differences in reaction of abnormal volume to articles during 1 week, 2 weeks, 1 month, 3 months, 6 months, 1 year, and 1.5 years around the disclosure period of the SEC announcement and press coverage (February-March 2014). In Panel A we examine retail trading volume obtained from TAQ using Boehmer et al. (2020) method, and in Panel B we focus on institutional trading, proxied for by trades greater than \$50,000.
We estimate the following model for the day of, the day after, and two days after an article is published on Seeking Alpha and Motley Fool:

$$Log(AbVol)_t = \alpha + \beta_1 Article \times \text{Post-Event} + \beta_2 Article + \beta_3 \text{Post-Event} + Controls + \epsilon$$

where $Log(AbVol)_t$ is defined as $Vol(t)/AvgVol(t-146, t-20)$. *Post-Event* is a dummy variable equal to 1 if the article was published during a given time band (1 week, 2 weeks, 1 month, 3 months, 6 months, 1 year, and 1.5 years) after news of the SEC investigation broke, April 1 to September 30, 2014. The figure plots the estimates of β_1 summed over days $t = 0$, $t+1$, and $t+2$, which represent the cumulative effect of the scandal on the abnormal trading volume reaction over the given time period. We estimate the regression for 1 week, 2 weeks, 1 month, 3 months, 6 months, 1 year, and 1.5 years after the scandal (i.e., after April 1, 2014). The bars represent 90% confidence bands around the point estimates.
We also control for the presence of SEC filings, press releases, and print media coverage over the prior three days. SEC filing is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and press releases is a dummy variable if there was at least one press release issued by the firm over the past three trading days, and print media is a dummy if there was at least one WSJ or NYT article about the firm in the past three trading days.

Panel A: Retail Trading Volume



Panel B: Institutional Trading Volume

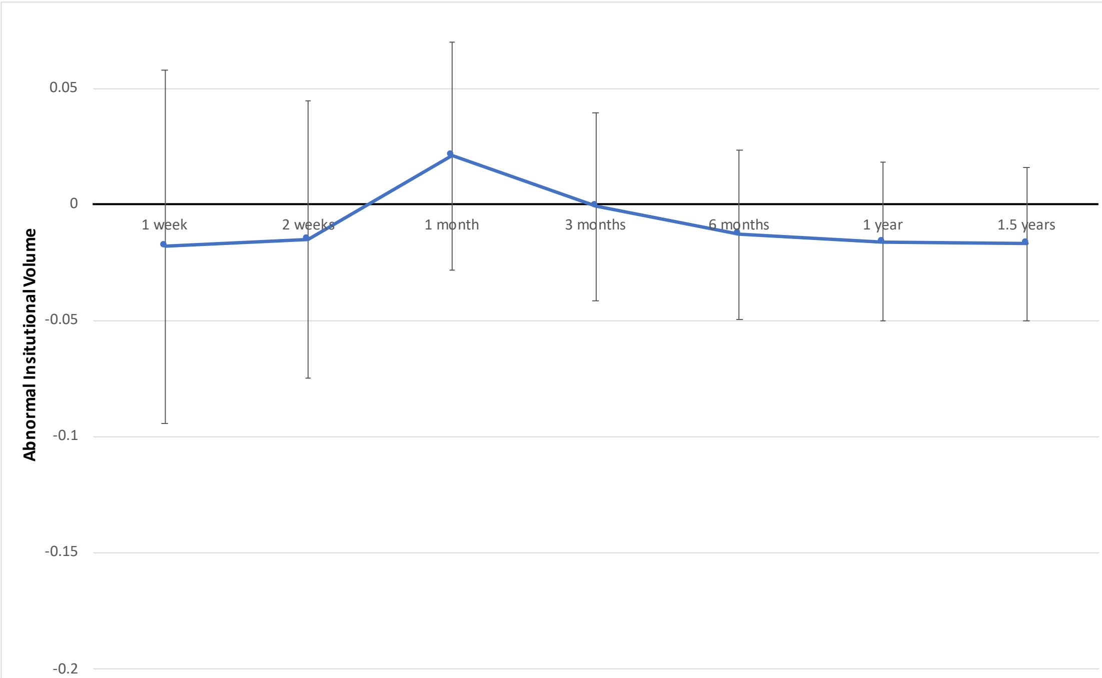


Table 1. **Relationship Between Articles and Subsequent Retail Trading Volume**

Reported are retail trading volume response to articles posted on Seeking Alpha and Motley Fool. We report results from regressions of the log of abnormal retail trading volume on the three days following the publication of an article on these platforms. We obtain retail trading volume from TAQ using Boehmer et al. (2020) method. Abnormal retail volume is defined as the log of $RetVol(t)/AvgRetVol(t-146, t-20)$, summed over days $t = 0$, $t+1$, and $t+2$. We examine all firms that have ever had an article written about them on Seeking Alpha or Motley Fool. The main independent variable is *Article*, a dummy variable equal to 1 if there was at least one article published about the firm on day $t = 0$. *Small firm* equals 1 if the firm is in the bottom 10th percentile of NYSE firms, and 0 otherwise. Firm size is measured in the prior trading month. In addition, we include as controls *SEC Filing*$_{t-3,t}$, which is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days and *Press release*$_{t-3,t}$, which is a dummy variable if there was at least one press release issued by the firm over the past three trading days. *Print Media*$_{t-3,t}$ is a dummy variable if there was at least one WSJ or NYT article about the firm in the past three trading days. We also control for lagged abnormal retail trading volume on day $t-1$. We include year-month fixed effects, and indicate statistical significance at the ten, five, and one percent levels with *, **, and ***, respectively. Standard errors are clustered at the firm level, with t-statistics in parentheses.

Dependent variable =	Log(Abnormal daily retail volume$_{t,t+2}$)			
Article	0.365***	0.195***	0.161***	0.194***
	(42.96)	(34.11)	(29.78)	(34.07)
Small firm			-0.028***	
			(-11.48)	
Article × Small firm			0.313***	
			(19.73)	
Fraudulent article				0.554**
				(2.22)
SEC filing$_{t-3,t}$		0.048***	0.047***	0.048***
		(26.60)	(26.13)	(26.60)
Press release$_{t-3,t}$		0.114***	0.111***	0.114***
		(51.70)	(50.37)	(51.72)
Print media$_{t-3,t}$		0.023***	0.020***	0.023***
		(4.81)	(4.39)	(4.82)
Abnormal retail volume$_{t-1}$		1.220***	1.219***	1.220***
		(174.78)	(174.76)	(174.66)
Observations	9,796,231	9,789,124	9,789,124	9,789,124
R-squared	0.023	0.284	0.284	0.284
Year-month F.E.	Y	Y	Y	Y

Table 2. **Shock to Awareness of Fake News: The 2014 SEC Lawsuit and Trading Volume**

The table examines whether the salience of the presence of fake news on the platforms, stemming from the public announcement of the SEC investigation and lawsuit, impacted investors' reaction to articles on these platforms. In Panel A we examine retail trading volume obtained from TAQ using Boehmer et al. (2020) method, and in Panel B we focus on institutional trading, proxied for by trades greater than $50,000, over the three days following the publication of an article on these platforms. We compare the volume response to articles in the six months before the SEC investigation and six months after the investigation, where we identify the February-March 2014 period as the period when the SEC investigation was announced and covered in the press. We include all firms that have ever had at least one article written about them on Seeking Alpha or Motley Fool during that time period. Abnormal volume are defined as in Table 1. The regressors include the dummy variable *Article*, which equals 1 if there was at least one article published about the firm on day t, and 0 otherwise, and we include the dummy variable *Post-event*, which equals 1 if the time period is April 1 to September 30, 2014 and is zero if the article was published from August 1, 2013 to January 31, 2014. We exclude all observations prior to August 2013 and after September 2014. We then interact the *Post-event* dummy with the *Article* dummy in the regressions to test for the differential response to articles before versus after the SEC announced investigation. *Small firm* equals 1 if the firm is in the bottom 10th percentile of NYSE firms, and 0 otherwise. Firm size is measured in the prior trading month.We also include controls for SEC filings, press releases, and abnormal volume over the previous three days before the article, plus the day of the article. *SEC filing*$_{t-3,t}$ is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and *Press release*$_{t-3,t}$ is a dummy variable if there was at least one press release issued by the firm over the past three trading days. *Print media*$_{t-3,t}$ is a dummy variable if there was at least one WSJ or NYT article about the firm in the past three trading days. In Panel C we control for any changes in the reaction of investors to news sources unrelated to the scandal on the social media platforms (SEC filings, Press releases, and articles in the WSJ and NYT) after the investigation announcement. We indicate statistical significance at the ten, five, and one percent levels with *, **, and ***, respectively (t-statistics in parentheses). Standard errors are clustered at the firm level.

Panel A: Effect on Abnormal Retail Volume

Dependent variable =	Log(Abnormal daily retail volume$_{t,t+2}$)		
Article	0.201***	0.239***	0.188***
	(19.12)	(15.03)	(13.50)
Post		-0.095***	-0.029***
		(-17.10)	(-5.56)
Article × Post		-0.038*	-0.041**
		(-1.88)	(-2.03)
Small firm			0.114***
			(12.79)
Article × Small firm			0.643***
			(10.00)
Post × Small firm			-0.174***
			(-13.91)
Article × Post × Small firm			-0.235***
			(-2.94)
SEC filing$_{t-3,t}$	0.052***	0.057***	0.058***
	(12.89)	(13.99)	(14.30)
Press release$_{t-3,t}$	0.124***	0.122***	0.128***
	(25.51)	(25.15)	(26.74)
Print media$_{t-3,t}$	-0.009	-0.011	0.009
	(-0.83)	(-1.05)	(0.81)
Abnormal retail volume$_{t-1}$	1.239***	1.232***	1.224***
	(88.63)	(88.81)	(89.47)
Observations	1,314,455	1,314,455	1,314,455
R-squared	0.275	0.277	0.280

Panel B: Effect on Abnormal Institutional Volume			
Dependent variable =	Log(Abnormal daily institutional volume$_{t,t+2}$)		
Article	0.369***	0.399***	0.242***
	(30.03)	(20.69)	(14.95)
Post		-0.153***	-0.129***
		(-23.61)	(-18.48)
Article × Post		-0.013	0.012
		(-0.59)	(0.60)
Small firm			-0.422***
			(-31.61)
Article × Small firm			0.837***
			(8.47)
Post × Small firm			-0.072***
			(-4.90)
Article × Post × Small firm			-0.005
			(-0.93)
SEC filing$_{t-3,t}$	0.062***	0.069***	0.050***
	(9.83)	(10.96)	(8.57)
Press release$_{t-3,t}$	0.228***	0.226***	0.174***
	(32.92)	(32.51)	(26.51)
Print media$_{t-3,t}$	0.227***	0.224***	0.127***
	(14.43)	(13.85)	(8.08)
Abnormal institutional volume$_{t-1}$	0.862***	0.853***	0.789***
	(82.42)	(81.79)	(69.41)
Observations	1,283,716	1,283,716	1,283,716
R-squared	0.163	0.166	0.194

Panel C: Effect from Unrelated News Sources

Dependent Variable =	Log(Ab Retail Vol)$_{t,t+2}$	Log(Ab Inst Vol)$_{t,t+2}$
Article	0.248***	0.408***
	(16.06)	(21.48)
Post	-0.097***	-0.147***
	(-17.12)	(-22.01)
Article × Post	-0.056***	-0.029
	(-2.97)	(-1.35)
SEC filing$_{t-3,t}$	0.062***	0.088***
	(9.34)	(9.71)
SEC filing$_{t-3,t}$ × Post	-0.010	-0.035
	(-1.20)	(-0.07)
Press release$_{t-3,t}$	0.111***	0.244***
	(15.72)	(25.88)
Press release$_{t-3,t}$ × Post	0.023	-0.05
	(0.48)	(-0.92)
Print media$_{t-3,t}$	-0.052***	0.159***
	(-4.18)	(8.91)
Print media$_{t-3,t}$ × Post	0.08	0.026
	(0.70)	(1.03)
Abnormal retail volume$_{t-1}$	1.232***	
	(88.83)	
Abnormal institutional volume$_{t-1}$		0.853***
		(81.79)
Observations	1,314,455	1,283,716
R-squared	0.277	0.167

Table 3. **Shocks to investor attention: mass shootings**

The table examines whether investors' reaction to mass shootings in the US changes after the public announcement of the SEC investigation and lawsuit. Panel A examines the reaction of abnormal retail trading volume in the overall sample and Panel B examines the reaction of abnormal retail trading volume in the six months before and after the SEC investigation announcement, where we identify the February-March 2014 period as the period when the SEC investigation was announced and covered in the press. The regressors include the dummy variable *Shooting*, which equals 1 if there was a mass shooting in the US on day t (as listed on https://en.wikipedia.org/wiki/List_of_mass_shootings_in_the_United_States), and 0 otherwise, and we include the dummy variable *Post*, which equals 1 if the time period is April 1 to September 30, 2014 and is zero if the article was published from August 1, 2013 to January 31, 2014. We exclude all observations prior to August 2013 and after September 2014. We then interact the *Post* dummy with the *Shooting* dummy in the regressions to test for the differential response to articles before versus after the SEC announced investigation. *Small firm* equals 1 if the firm is in the bottom 10th percentile of NYSE firms, and 0 otherwise. Firm size is measured in the prior trading month.We also include controls for SEC filings, press releases, and abnormal volume over the previous three days before the article, plus the day of the article. *SEC filing*$_{t-3,t}$ is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and *Press release*$_{t-3,t}$ is a dummy variable if there was at least one press release issued by the firm over the past three trading days. *Print media*$_{t-3,t}$ is a dummy variable if there was at least one WSJ or NYT article about the firm in the past three trading days. We indicate statistical significance at the ten, five, and one percent levels with *, **, and ***, respectively (t-statistics in parentheses). Standard errors are clustered at the firm level.

Panel A: Entire Sample			
Dependent variable =	Log(Abn. daily retail volume$_{t,t+2}$)		
	All Days	All Days	Article Days
Shooting	-0.019***	-0.020***	-0.027**
	(-10.99)	(-10.78)	(-2.06)
Small firm		-0.029***	
		(-12.34)	
Shooting × Small firm		0.004	
		(0.96)	
SEC filing$_{t-3,t}$	0.052***	0.051***	0.122***
	(28.95)	(28.36)	(15.59)
Press release$_{t-3,t}$	0.116***	0.113***	0.157***
	(54.33)	(52.15)	(14.23)
Print media$_{t-3,t}$	0.051***	0.043***	-0.128***
	(12.34)	(10.24)	(-12.83)
Abnormal retail volume$_{t-1}$	1.244***	1.244***	1.499***
	(178.23)	(177.90)	(75.85)
Observations	9,778,195	9,778,195	211,774
R-squared	0.276	0.276	0.366

Panel B: Six months before and after the SEC scandal				
Dependent variable =	Log(Abn. daily retail volume$_{t,t+2}$)			
	All Days	Article Days	Article Days	Article Days
Shooting	-0.009**	-0.033*	-0.011*	-0.018*
	(-2.07)	(-1.86)	(-1.87)	(-1.86)
Post			-0.113***	-0.059***
			(-6.04)	(-3.15)
Shooting × Post			-0.046	-0.062
			(-0.58)	(-0.90)
Small firm				0.675***
				(11.01)
Shooting × Small firm				-0.057
				(-0.16)
Small firm × Post				-0.369***
				(-4.81)
Shooting × Post × Small firm				0.300
				(0.74)
SEC filing$_{t-3,t}$	0.055***	0.120***	0.124***	0.134***
	(13.51)	(6.64)	(6.90)	(7.63)
Press release$_{t-3,t}$	0.129***	0.182***	0.182***	0.200***
	(26.60)	(8.64)	(8.61)	(9.59)
Print media$_{t-3,t}$	0.033***	-0.157***	-0.159***	-0.108***
	(3.09)	(-6.74)	(-6.79)	(-4.62)
Abnormal retail volume$_{t-1}$	1.241***	1.516***	1.505***	1.457***
	(88.78)	(39.83)	(39.21)	(39.00)
Observations	1,314,455	30,678	30,678	30,678
R-squared	0.274	0.358	0.360	0.374

Table 4. **Retail Volume Impact Across Article and Author Characteristics**

The table reports the relationship between author and article characteristics and the abnormal retail trading volume response to articles published on the social media platforms. Abnormal retail trading volume is defined in Table 1. Panel A reports results for the entire sample period and Panel B focuses on the time period six months before and six months after the SEC announced investigation (February/March 2014). The independent variables include the log of the number of followers the author of the article has, the log of one plus the number of articles that the author has written prior to the article, the log of one plus the number of comments that the article received, the standardized fraction or % of how often the author uses numbers in the article, and the standardized fraction or % of how often the word "earning" or the stem "earn" are used in the article. In addition, we include the number of comments, the average length of comments (word count) on the article, and dummy variables for whether disagreement words ("disagree", "differ", "counter", "oppose", or "argue"), "wrong" words ("wrong" or "not right"), or "fake" words ("fake" or "fraud") are used in the comments section, and Disclosed position is 1 if the author disclosed an existing position, and zero otherwise. Panel B interacts the *Post-Event* dummy, which equals 1 if the time period is April 1 to September 30, 2014, the six month period after the SEC announced investigation, with each of the independent variables. We also include controls for SEC filings, press releases, and print media coverage over the prior three days, and abnormal trading volume on day $t-1$. SEC filing is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and press releases is a dummy variable if there was at least one press release issued by the firm over the past three trading days, and print media is a dummy if there was at least one WSJ or NYT article about the firm in the past three trading days. We do not report the coefficient estimates on these controls for brevity and because their estimates are very similar to those in Table 2. We include fixed effects at the year-month, author, and firm level (where appropriate), and statistical significance is denoted at the ten, five, and one percent levels by *, **, and ***, respectively (t-statistics in parentheses). Standard errors are clustered at the firm level.

Panel A: Full Sample Period

Dependent variable =	Log(Abnormal daily retail volume$_{t,t+2}$)									
Log(Num followers)	0.013***									
	(8.96)									
Log(Num past articles)		0.005***								
		(3.29)								
% Numbers in text			0.026***							
			(9.07)							
% "Earn" mentions				0.045***						
				(12.65)						
log(Num comments)					0.054***					
					(10.84)					
Log(avg comment length)						0.023***				
						(6.68)				
Disagree comments							0.024***			
							(3.25)			
Wrong comments								0.047***		
								(5.16)		
Fake/Fraud comments									0.034**	
									(2.06)	
Discloses position										-0.006
										(-0.69)
Observations	174,377	320,228	317,443	317,443	172,002	123,005	172,002	172,002	172,002	171,525
R-squared	0.494	0.444	0.484	0.485	0.536	0.562	0.534	0.535	0.534	0.534
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year-Month F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Author F.E.			Y	Y	Y	Y	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Comparing 6-months before versus 6-months after the SEC investigation

Dependent variable =	Log(Abnormal daily retail volume$_{t,t+2}$)									
Log(Num followers)	-0.006									
	(-1.05)									
Log(Num followers) x Post-event	0.016**									
	(2.20)									
Log(Num past articles)		-0.007								
		(-1.32)								
Log(Num past articles) x Post-event		0.051***								
		(7.41)								
% Numbers in text			-0.002							
			(-0.26)							
% Numbers in text x Post-event			0.003							
			(0.17)							
%"Earn" mentions				0.002						
				(0.15)						
% "Earn" mentions x Post-event				0.050**						
				(2.25)						
log(Num comments)					0.037***					
					(2.60)					
log(Num comments) x Post-event					0.004					
					(0.13)					
Log(avg comment length)						0.034***				
						(2.73)				
Log(avg comment length) x Post-event						-0.009				
						(-0.33)				
Disagree comments							0.048			
							(1.61)			
Disagree comments x Post-event							0.005			
							(0.07)			
Wrong comments								0.078**		
								(2.22)		
Wrong comments x Post-event								-0.010		
								(-0.14)		
Fake/Fraud comments									0.206**	
									(2.46)	
Fake/Fraud comments x Post-event									-0.125	
									(-1.39)	
Discloses position										0.146***
										(5.92)
Discloses position x Post-event										-0.103***
										(-3.10)
Post-event	-0.193***	-0.281***	-0.048	-0.042	-0.092**	-0.018	-0.085***	-0.081***	-0.077*	-0.034
	(-2.65)	(-3.42)	(-0.69)	(-0.61)	(-2.00)	(-0.13)	(-3.70)	(-2.78)	(-1.83)	(-0.85)
Observations	28,392	55,258	55,257	55,257	28,392	25,595	28,392	28,392	28,392	28,229
R-squared	0.395	0.394	0.390	0.391	0.396	0.410	0.395	0.395	0.395	0.392
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 5. **Response of Authors and Readers to the 2014 SEC Investigation**

The table reports regression results of author and reader responses to the SEC announced investigation by looking at the characteristics of articles and comments on those articles in the six month period after the SEC investigation. Specifically, we regress a host of article and comment characteristics on the dummy variable *Post-event*, which equals 1 if an article was published from April 1 to September 30, 2014 and zero if the article was published from August 1, 2013 to January 31, 2014. These periods correspond to six months after and six months before the February-March 2014 SEC investigation period. The dependent variables are the log of one plus the number of comments that the article received, the standardized fraction or % of how often the author uses numbers in the article, the standardized fraction or % of how often the word "earning" or the stem "earn" are used in the article, the average length of comments (word count) on the article and dummy variables for whether the disagreement words ("disagree", "differ", "counter", "oppose", or "argue"), "wrong" words ("wrong" or "not right"), or "fake" words ("fake" or "fraud") are used in the comments section. Each of these dependent variables are regressed on the *Post-event* dummy plus controls for SEC filings, press releases, and print media coverage over the prior three days. SEC filing is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and press releases is a dummy variable if there was at least one press release issued by the firm over the past three trading days, and print media is a dummy if there was at least one WSJ or NYT article about the firm in the past three trading days. We also controls for lagged abnormal volume from the previous day, defined in Table 1. Panel A reports results for articles written about small firms only (smallest 10% of stocks based on NYSE breakpoints) and Panel B reports results for all other firms (non-small firms). Panels A and B include firm fixed effects. In Panels C and D, we examine whether author characteristics have changed after the SEC investigation. The dependent variables are the log number of followers that the authors have and the log number of articles that the author has written in the past. Statistical significance is denoted at the ten, five, and one percent levels by *, **, and ***, respectively (t-statistics in parentheses).

Dependent variable =	# Comments	% Numbers in text	% "Earn" in text	Length of comments	"Disagree" words in comments	"Wrong" words in comments	"Fake" words in comments
Panel A: Articles about small firms							
Post-event	0.083	-0.051	0.123***	-0.101	-0.003	0.020	0.006**
	(1.46)	(-1.13)	(4.71)	(-1.43)	(-0.11)	(0.84)	(2.10)
SEC filing$_{t-3,t}$	-0.097	-0.018	0.037	-0.047	-0.041	-0.026	-0.027
	(-1.48)	(-0.35)	(1.26)	(-0.57)	(-1.42)	(-0.95)	(-1.60)
Press release$_{t-3,t}$	-0.098	-0.015	0.136***	-0.147	-0.019	-0.014	0.002
	(-1.29)	(-0.28)	(4.26)	(-1.58)	(-0.58)	(-0.46)	(0.09)
Print media$_{t-3,t}$	-0.266	0.063	0.009	-0.387*	0.029	-0.057	-0.029
	(-1.47)	(0.44)	(0.11)	(-1.79)	(0.37)	(-0.77)	(-0.63)
Abnormal volume$_{t-1}$	0.097***	-0.007	0.006	0.076***	0.019*	0.031***	0.016***
	(4.29)	(-0.37)	(0.53)	(2.74)	(1.90)	(3.32)	(2.71)
Observations	2,257	2,845	2,845	2,257	2,257	2,257	2,257
R^2	0.606	0.400	0.445	0.592	0.422	0.432	0.517
Firm F.E.	Y	Y	Y	Y	Y	Y	Y
Panel B: Articles about non-small firms							
Post-event	-0.060***	-0.178***	-0.057***	-0.069***	-0.017***	-0.006	0.004
	(-4.86)	(-1.22)	(-6.26)	(-4.21)	(-3.00)	(-1.16)	(1.45)
SEC filing$_{t-3,t}$	-0.034**	0.040***	0.169***	-0.054***	-0.022***	-0.007	-0.006
	(-2.31)	(3.91)	(17.80)	(-2.74)	(-3.14)	(-1.12)	(-1.49)
Press release$_{t-3,t}$	-0.077***	0.017	0.142***	-0.097***	-0.018**	-0.021***	-0.004
	(-4.91)	(1.62)	(14.25)	(-4.63)	(-2.45)	(-2.97)	(-1.10)
Print media$_{t-3,t}$	0.022	-0.007	0.029**	0.023	0.015*	-0.000	0.003
	(1.24)	(-0.60)	(2.57)	(0.97)	(1.77)	(-0.06)	(0.57)
Abnormal volume$_{t-1}$	0.029***	0.003	0.067***	0.032**	0.006	0.009**	0.004
	(2.84)	(0.40)	(9.77)	(2.40)	(1.20)	(1.99)	(1.46)
Observations	25,635	52,273	52,273	25,635	25,635	25,635	25,635
R^2	0.596	0.148	0.185	0.515	0.312	0.288	0.253
Firm F.E.	Y	Y	Y	Y	Y	Y	Y

Dependent variable =	log(Num followers)	log(Num past articles)
Panel C: Articles about small firms		
Post-event	0.380***	0.908***
	(4.79)	(12.51)
Observations	2,866	3,427
R-squared	0.008	0.044
Panel D: Articles about non-small firms		
Post-event	-0.168***	0.335***
	(-7.58)	(17.80)
Observations	26,413	52,891
R-squared	0.002	0.006

Table 6. **Linguistic Analysis of Comments and Articles**

The table reports regression results of reader and author responses to the SEC announced investigation by looking at the linguistic characteristics of comments on the articles and the articles themselves using the linguistic algorithm LIWC. We compare linguistic characteristics six months before versus six months after the SEC investigation by regressing a host of article and comment characteristics on the dummy variable *Post-event*, which equals 1 if an article was published from April 1 to September 30, 2014 and zero if the article was published from August 1, 2013 to January 31, 2014. The dependent variables are LIWC measures of Analytic, Clout, and Tone of the comments on the articles (first three columns) and the measures of Analytic, Clout, and Tone on the articles themselves (next three columns) plus the LIWC measure of authenticity of the articles (last column). Analytic is how analytical the language sounds, Clout is how authoritative the language sounds, and Tone is the sentiment of the article, where higher scores mean more positive tone. Authenticity measures how the article is using authentic linguistic cues, where a positive number indicates more authenticity. Each of these dependent variables are regressed on the *Post-event* dummy plus controls for SEC filings, press releases, and print media coverage over the prior three days. SEC filing is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and press releases is a dummy variable if there was at least one press release issued by the firm over the past three trading days, and print media is a dummy if there was at least one WSJ or NYT article about the firm in the past three trading days. We also controls for lagged abnormal volume from the previous day, defined in Table 1. Panel A reports results for articles written about small firms only (smallest 10% of stocks based on NYSE breakpoints) and Panel B reports results for all other firms (non-small firms). All regressions include firm fixed effects and statistical significance is denoted at the ten, five, and one percent levels by *, **, and ***, respectively (t-statistics in parentheses).

Panel A: Articles about small firms

	Comments			Articles			
Dependent variable =	Analytic	Clout	Tone	Analytical	Clout	Tone	Authenticity
Post-event	0.053	-0.300	-4.345***	0.333	0.569	0.550	3.079***
	(0.07)	(-0.44)	(-3.70)	(0.25)	(1.43)	(0.64)	(5.50)
SEC filing$_{t-3,t}$	0.015	-0.716	1.490	-0.600**	0.265	1.676*	-0.181
	(0.02)	(-0.91)	(1.10)	(-2.02)	(0.52)	(1.74)	(-0.26)
Press release$_{t-3,t}$	0.680	0.391	1.746	-0.581*	1.414**	-0.093	2.520***
	(0.73)	(0.44)	(1.13)	(-1.79)	(2.53)	(-0.09)	(3.35)
Print media$_{t-3,t}$	-4.007*	1.472	-5.695	-0.551	-1.366	1.524	-3.823**
	(-1.85)	(0.71)	(-1.58)	(-0.65)	(-0.94)	(0.55)	(-2.31)
Abnormal volume$_{t-1}$	-0.506*	0.114	-0.478	0.089	0.182	-0.137	0.175
	(-1.80)	(0.42)	(-1.03)	(0.84)	(1.00)	(-0.40)	(0.79)
Observations	2,257	2,257	2,257	2,845	2,845	2,845	3,421
R^2	0.312	0.333	0.384	0.347	0.399	0.424	0.014
Firm F.E.	Y	Y	Y	Y	Y	Y	Y

Panel B: Articles about non-small firms

	Comments			Articles			
Dependent variable =	Analytic	Clout	Tone	Analytical	Clout	Tone	Authenticity
Post-event	0.202	-0.406**	-1.232***	-0.019	0.007	0.652	-1.683***
	(1.24)	(-2.45)	(-4.29)	(-0.34)	-0.23	(1.43)	(-4.02)
SEC filing$_{t-3,t}$	0.303	-0.105	0.119	0.236***	0.396***	0.686***	1.615***
	(1.55)	(-0.53)	(0.35)	(3.95)	(4.33)	(3.46)	(10.26)
Press release$_{t-3,t}$	-0.098	0.087	0.434	0.162***	0.587***	0.566***	1.095***
	(-0.47)	(0.41)	(1.19)	(2.59)	(6.14)	(2.73)	(7.08)
Print media$_{t-3,t}$	-0.245	0.340	0.096	-0.190***	-0.299***	-0.159	-1.546***
	(-1.06)	(1.45)	(0.24)	(-2.71)	(-2.79)	(-0.69)	(-10.03)
Abnormal volume$_{t-1}$	-0.163	-0.217	-0.460*	0.020	0.076	-0.372***	0.980***
	(-1.21)	(-1.59)	(-1.94)	(0.47)	(1.17)	(-2.62)	(9.06)
Observations	25,635	25,635	25,635	52,273	52,273	52,273	52,871
R^2	0.151	0.169	0.191	0.122	0.180	0.157	0.010
Firm F.E.	Y	Y	Y	Y	Y	Y	Y

APPENDIX

Appendix A: Contributors and compensation for authorship on shared-knowledge platforms

For authors on Seeking Alpha, base payment is $35 plus $10 per 1,000 page-views. For analysis of stocks that have a large number of followers, Seeking Alpha has three additional payment tiers, from $150 to $500 per article. Finally, two articles are selected each week for a $2,500 "outstanding performance" prize on the basis of how well the stock idea played out. The articles are published as Premium articles, Standard articles, and Instablogs. Standard articles are allowed to be published elsewhere, and are unpaid, but also undergo a selection process. Instablogs are published instantly and with no pay.

The Motley Fool offers a wide range of stock news and analysis at its free website, www.fool.com, as well as through a variety of paid investment advice services, which provide online stock analysis and research with interactive discussion boards. The discussion boards are used heavily to recruit future Motley Fool staffers, where frequent posters are first awarded free subscriptions and then can receive a small stipend. The Motley Fool Blog Network was a stock analysis and news site that provided a platform for non-Motley Fool staff writers to submit articles. They received compensation ranging from $50 to $100 for each article submitted and additional compensation for how many recommendations or "editors picks" they received. Eventually the company merged the Blog Network with its primary site in 2014.

Appendix B: Documents from Galena Biopharma, Inc.

Example of a for-sure fake article about Galena Biopharma, Inc.

Seeking Alpha α

Galena Biopharma Has A Promising Pipeline For Revenue Growth

Feb. 5, 2014 5:32 AM ET
by: John Mylant

Galena Biopharma (GALE) is presently trading at $4.22 a share after a favorable reactionary move in the market when it announced a recent acquisition that has a potential for generating good revenue in the years ahead. I will write more about that later, but here is a company that is highly favored by investors and analysts alike. Roth Capital increased its price target on the company to $11.00 because of the acquisition of Mills Pharmaceuticals which would lead to the development of GALE-401. Let's take a look at where this company is presently and the "pipeline potential" that makes this company a good long-term growth investment.

Fundamentally Speaking

The company is transforming from a "research and development" firm to a revenue-producing firm. The revenue in the 3Q of 2013 ($1,170k) was from its recent launch of "Abstral." Even though it is now producing revenue, the company is still a long-term growth investment because it will take a little bit more time for revenue to outgrow expenditures. In the company's 3Q of 2013 10Q report, research and development was still ($3,633k), so it's going to take a little bit more time for the company to be profitable.

Observing the company's balance sheet over the last four quarters from Yahoo Finance, we can see that it has been "cash strong" since its September offering. Presently, the company has $32 million in cash and equivalents. Its current "burn rate" is about $2 million per month according to Wall Street Cheat Sheet.

This means the company should have good working capital through 2015 and longer if revenues increase like the company plans.

Period Ending	Sep 30, 2013	Jun 30, 2013	Mar 31, 2013	Dec 31, 2012
Assets				
Current Assets				
Cash And Cash Equivalents	51,496	21,021	17,583	32,908
Short Term Investments	2,837	5,786	9,709	2,678
Net Receivables	1,543	-	-	-
Inventory	425	352	-	-
Other Current Assets	485	389	210	535
Total Current Assets	**56,786**	**27,548**	**27,502**	**36,121**
Long Term Investments	-	-	-	-
Property Plant and Equipment	545	467	27	29
Goodwill	5,898	5,898	5,898	5,898
Intangible Assets	15,032	15,083	15,086	-
Accumulated Amortization	-	-	-	-
Other Assets	12,993	13,006	12,938	12,938
Deferred Long Term Asset Charges	-	-	-	-
Total Assets	**91,254**	**62,002**	**61,451**	**54,986**
Liabilities				
Current Liabilities				
Accounts Payable	10,779	10,536	9,601	4,014
Short/Current Long Term Debt	1,221	307	6	6
Other Current Liabilities	24,514	14,909	16,772	11,899
Total Current Liabilities	**36,514**	**25,752**	**26,379**	**15,919**
Long Term Debt	8,613	9,437	51	51
Other Liabilities	6,454	6,529	6,656	6,207
Deferred Long Term Liability Charges	5,053	5,053	5,053	5,053
Minority Interest	-	-	-	-
Negative Goodwill	-	-	-	-
Total Liabilities	**56,634**	**46,771**	**38,139**	**27,230**

In this type of industry, the financial position is important, but so is the debt load. I believe one thing analysts like about the company is its long-term debt is negligible compared to its cash position. This is what gives it such a small debt-equity ratio. According to MSN Money, it also has a very healthy "current ratio" of 1.55.

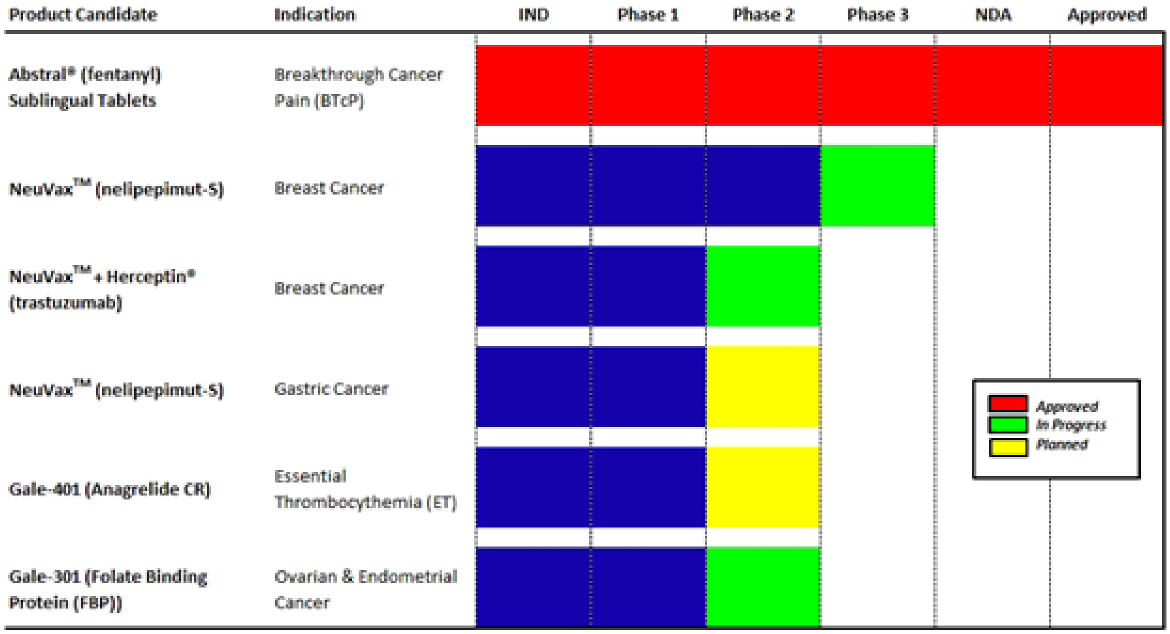
Presently, the company has 105.2 million outstanding shares of stock trading at $4.22, which gives it a market cap of $441.10 million.

It also has a good cash position as it starts to bring revenue to the company with Abstral. As I stated before though, this is a good long-term growth investment because it has two other products that look promising to bring to market. Let's take a look at all three of the company's pipeline products.

Product for Revenue

What does the company's pipeline look like?

Presently, the company has one product on market and two others in its research stage. The one already brought to market, I would conservatively say has a revenue potential of at least $40 million while the other two could conservatively top $120 million or more when they come to the market. (combined)



Abstral (fentanyl)

When it was announced in March 2013 that Galena bought Abstral, this was part of its growth plan to acquire drugs with good revenue potential for its pipeline. The drug treats "breakthrough cancer pain" which occurs in (40%-80%) of patients receiving treatment for cancer as well as pain management. When the drug was introduced, it was and remains the only fast-acting sublingual tablet for cancer treatment on the US market. The market value for this product in the United States is about $400 million. The 3Q of 2013 was the first quarter the company recorded revenue and it came from Abstral. It generated net revenue of $1.2 million for the first time.

This particular market is not overly crowded, and it would not be surprising for the company to capture 10% of the market which could see a potential revenue generation of $40 million a year.

NeuVax

As a second-line treatment, NeuVax focuses on the prevention of recurrence of breast cancer (and other tumors) around the body. It is not uncommon for some breast cancer cells to remain and possibly migrate to other parts of the body. To prevent them from growing and becoming tumors, NeuVax is treatment that seeks out cancer cells that are high in the HER2 protein, neutralizing and destroying the tumor cells. The HER2 protein is highly overexpressed by 85% in breast cancer cells. Studies have identified the NeuVax peptide sequence as being highly effective and clinical data has indicated the ability to maintain a long-term elevated level of NeuVax specific T cells, could potentially provide long-term prevention against the possibility of a tumor recurrence.

NeuVax is currently enrolling breast cancer patients for the NeuVax™ Phase 3 PRESENT (Prevention of Recurrence in Early-Stage Node-Positive Breast Cancer with Low to Intermediate HER2 Expression with NeuVax™ Treatment) study. The FDA granted NeuVax a Special Protocol Assessment (SPA) for its Phase 3 study.

How big is this HER2 Breast Cancer market?

HER2 accounts for close to 25% of the total breast cancer patients, but has 55% of the research breast cancer market and is expected to increase to 65% by 2021. The market itself is fairly busy with activity. Roche (OTCQX:RHHBY) and Novartis (NVS) dominate the first-line treatment market and Roche's drug, "Herceptin" earned more than $3 billion from back in 2011. The breast cancer market as a whole is close to $9 billion right now and expected to top out at $10.9 billion by 2018.

Decision Resources is a research and advisory firm for pharmaceutical and healthcare issues. They put out a report in 2013 that surveyed oncologists from the United States. 50% of them said they would prescribe a second line treatment for HER2 spastic breast cancer.

While the Phase 3 study is expected to observe and track survival rates three, five and 10 years out in vaccine controlled groups. Revenue generation from this product is a couple years out. Even though the market is crowded studies have proven that there is an interest in this type of treatment. Capturing but 1% of this market would generate $90 million in revenue.

Alliance to Open Market in India

Galena Biopharma and Dr. Reddy's Laboratories Ltd. have developed a strategic partnership for commercialization of NeuVax in India. When the drug is approved, it has the potential for doubling the patient population. By 2016, the pharmacological market for breast cancer is expected to reach INR $10,000 million, which translates into a USD $1.6 billion industry in a country that has a very high mortality rate that sees 50,000 women dying each year.

What the Mills Pharmaceuticals Acquisition means for Investors

Galena's long-term business strategy is to add therapies to their pipeline that will strengthen their hematology-oncology portfolio. The recent acquisition of Mills Pharmaceuticals is an example of that plan in action.

This is an acquisition with a long-term investment in mind for a market which potentially could reach $200 million in the United States alone. The market is for the treatment of Essential Thrombocythemia (ET). This is a rare disease that is characterized by a person's body manufacturing an overabundance of platelets in bone marrow.

Mills Pharmaceuticals owned the worldwide rights to GALE-401 which is a controlled-release formulation of a drug called anagrelide. The treatment has shown great promise reducing the side effects of anagrelide while maintaining efficacy for the patients. This is important because a significant amount of patients are unable to tolerate fully effective doses of anagrelide. They either stop treatment or the dose is reduced and becomes inefficient to achieve the target platelet levels.

Presently the drug is still in the trial phase, and Galena believes it will eventually be eligible for orphan status which enhances its regulatory process. A Phase 2 study is expected to be initiated in mid-2014 and the FDA indicated that only a single Phase 3 trial will be required for approval.

In a $200 million industry where many physicians are unhappy with the current treatment for ET, there is great potential here for Galena. Presently physicians are faced with the treatment which leaves patients with unmanageable side effects. If GALE-401 continues to prove effective in reducing the adverse effects on patients, physicians will notice quickly.

This is only in the Phase 2 study so it's a long-term vision. If the clinical trials continue to go well, physicians should embrace this therapy quickly. It is not out of the question to conservatively see the company capture 15% of this market which could translate into $30 million a year in revenue.

Outlook and Investment Risks

Galena Biopharma is just turning the road to profitability. It may take a little more time to get there, but with three strong drugs in its pipeline (Abstral already to market), the potential revenue base can conservatively be estimated at $160 million between the three.

The company appears to be managed well, has minimal debt compared to the industry as a whole and a strong asset to liability ratio which is important for small companies like this. This would make a good long-term growth investment for those who enjoy this industry. Its present drug, Abstral, could potentially bring the company into profitability by itself before the other two drugs are introduced to the market in the coming years ahead.

With all companies in this arena, potential growth is based upon FDA approval of the drugs going through trials. Abstral has a good market potential in itself, but there is no guarantee that the other two I described in this article will reach the market. This is the risk that investors face in this industry.

Author's note: The chart in this article came from the company's investor presentation in January.

8-K form documenting the settlement between the SEC, Galena, and Mr. Ahn

8-K 1 gale-201612228xk.htm 8-K

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
WASHINGTON, D.C. 20549

FORM 8-K

CURRENT REPORT
PURSUANT TO SECTION 13 OR 15(d) OF THE
SECURITIES EXCHANGE ACT OF 1934

Date of report (Date of earliest event reported): December 22, 2016

GALENA BIOPHARMA, INC.

(Exact name of registrant as specified in its charter)

Delaware	001-33958	20-8099512
(State or other jurisdiction of incorporation or organization)	(Commission File Number)	(I.R.S. Employer Identification No.)

2000 Crow Canyon Place, Suite 380, San Ramon, CA 94583
(Address of Principal Executive Offices) (Zip Code)

Registrant's telephone number, including area code: (855) 855-4253

Check the appropriate box below if the Form 8-K filing is intended to simultaneously satisfy the filing obligation of the registrant under any of the following provisions (see General Instruction A.2. below):

o Written communications pursuant to Rule 425 under the Securities Act (17 CFR 230.425)

o Soliciting material pursuant to Rule 14a-12 under the Exchange Act (17 CFR 240.14a-12)

o Pre-commencement communications pursuant to Rule 14d-2(b) under the Exchange Act (17 CFR 240.14d-2(b))

o Pre-commencement communications pursuant to Rule 13e-4(c) under the Exchange Act (17 CFR 240.13e-4(c))

Item 8.01 Other Events.

SEC Investigation

On December 22, 2016, Galena Biopharma, Inc. (Galena) and its former Chief Executive Officer (CEO) reached an agreement in principle to a proposed settlement that would resolve an investigation by the staff of the Securities and Exchange Commission (SEC) involving conduct in the period 2012-2014 regarding the commissioning of internet publications by outside promotional firms.

Under the terms of the proposed settlement framework, Galena and the former CEO would consent to the entry of an administrative order requiring that we and the former CEO cease and desist from any future violations of Sections 5(a), 5(b), 5(c), 17(a), and 17(b) of the Securities Act of 1933, as amended, and Section 10(b), 13(a), and 13(b)(2)(A) of the Securities Exchange Act of 1934, as amended, and various rules thereunder, without admitting or denying the findings in the order. Based upon the proposed settlement framework, the Company will make a $200,000 penalty payment. In addition to other remedies, the proposed settlement framework would require the former CEO to make a disgorgement and prejudgment interest payment as well as a penalty payment to the Commission. To address the issues raised by the SEC staff's investigation, in addition to previous governance enhancements we have implemented, we have voluntarily undertaken to implement a number of remedial actions relating to securities offerings and our interactions with investor relations and public relations firms. The proposed settlement is subject to approval by the Commission and would acknowledge our cooperation in the investigation and confirm our voluntary undertaking to continue that cooperation. If the Commission does not approve the settlement, we may need to enter into further discussions with the SEC staff to resolve the investigated matters on different terms and conditions. As a result, there can be no assurance as to the final terms of any resolution including its financial impact or any future adjustment to the financial statements.

A special committee of the board of directors has determined in response to an indemnification claim by the former CEO that we are required under Delaware law to indemnify our former CEO for the disgorgement and prejudgment interest payment of approximately $750,000 that he would be required to pay if and when the settlement is approved by the Commission. Any penalty payment that the former CEO will be required to make in connection with this matter ($600,000 under the proposed settlement framework) will be the responsibility of the former CEO.

SIGNATURES

Pursuant to the requirements of the Securities Exchange Act of 1934, the registrant has duly caused this report to be signed on its behalf by the undersigned hereunto duly authorized.

GALENA BIOPHARMA, INC.

Date: December 22, 2016

By: /s/ Mark W. Schwartz

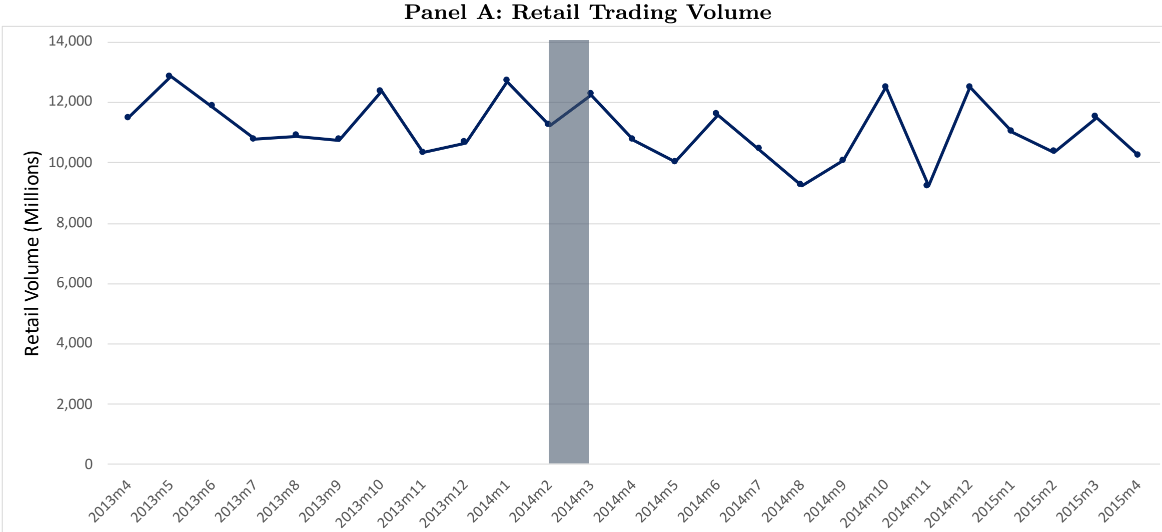
Mark W. Schwartz Ph.D.
President and Chief Executive Officer

Appendix C: Supplemental Tables and Figures for "Social Media and Financial News Manipulation"

Figure C1. **Time Trend in Trading Volume**

This figure plots the trading volume in the year before and the year after the SEC Lawsuit disclosure that publicly announced the presence of fake news on these platforms (February and March 2014). In Panel A we examine retail trading volume obtained from TAQ using Boehmer et al. (2020) method, and in Panel B we focus on institutional trading, proxied for by trades greater than $50,000, over the three days following the publication of an article on these platforms. In grey we highlight the months in which news about the presence of fake articles on these platforms was disclosed.

Panel A: Retail Trading Volume



Panel B: Institutional Trading Volume

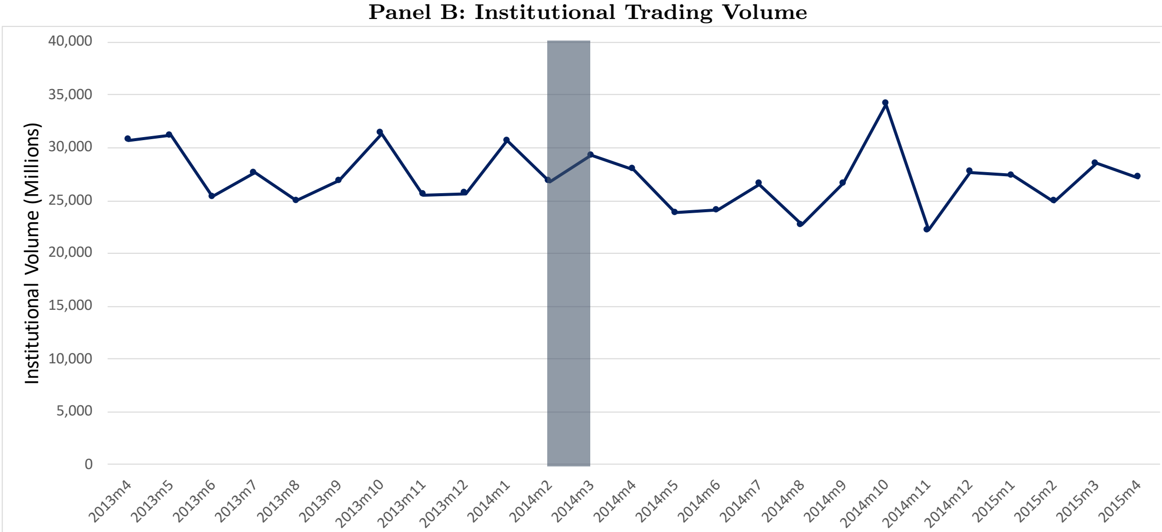


Table C1. **Effect of articles on retail trading volume - Daily Reaction**

Reported are daily retail trading volume responses to articles posted on Seeking Alpha and Motley Fool over the first three days after the article is published. The table reports results from regressions of the log of abnormal trading volume on the three days following the publication of an article on these platforms on the dummy variable *Article*, which equals 1 if there was at least one article published about the firm on day $t = 0$. In addition, we include as controls $SEC\ Filing_{t-3,t}$, which is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days. $Press\ release_{t-3,t}$, which is a dummy variable if there was at least one press release issued by the firm over the past three trading days. $Print\ media_{t-3,t}$ is a dummy variable if there was at least one WSJ or NYT article about the firm in the past three trading days. We obtain retail trading volume from TAQ using Boehmer et al. (2020) method. Abnormal retail volume is defined as the log of $RetVol(t)/AvgRetVol(t-146, t-20)$, summed over days $t = 0$, $t+1$, and $t+2$. We also control for lagged abnormal trading volume on day $t-1$. We include year-month fixed effects, and indicate statistical significance at the ten, five, and one percent levels with *, **, and ***, respectively (t-statistics in parentheses). Standard errors are clustered at the firm level.

Dependent variable =	Log(Abnormal retail trading volume)		
	Day 0 ($t = 0$)	Day 1 ($t+1$)	Day 2 ($t+2$)
Article	0.088***	0.062***	0.045***
	(32.26)	(32.89)	(31.30)
SEC filing$_{t-3,t}$	0.028***	0.017***	0.003***
	(47.49)	(25.75)	(4.25)
Press release$_{t-3,t}$	0.050***	0.043***	0.021***
	(63.94)	(50.37)	(28.19)
Print media$_{t-3,t}$	0.009***	0.006***	0.008***
	(4.80)	(3.40)	(5.47)
Abnormal retail volume$_{t-1}$	0.468***	0.394***	0.358***
	(193.75)	(169.08)	(156.71)
Observations	9,789,254	9,783,738	9,778,195
R^2	0.237	0.171	0.141
Year-month F.E.	Y	Y	Y

Table C2. **Relationship Between Articles and Subsequent Overall Trading Volume and Price Volatility**

Reported are retail trading volume response to articles posted on Seeking Alpha and Motley Fool. Panel A reports results from regressions of the log of abnormal trading volume on the three days following the publication of an article on these platforms (days $t = 0$, $t + 1$, and $t + 2$). Panel B reports results from the same regressions using three-day squared price movements (volatility) as the dependent variable. We examine all firms that have ever had an article written about them on Seeking Alpha or Motley Fool. Abnormal retail volume is defined as the log of $RetVol(t)/AvgRetVol(t-146, t-20)$, summed over days $t = 0$, $t + 1$, and $t + 2$. We examine all firms that have ever had an article written about them on Seeking Alpha or Motley Fool. The main independent variable is *Article*, a dummy variable equal to 1 if there was at least one article published about the firm on day $t = 0$. *Small firm* equals 1 if the firm is in the bottom 10th percentile of NYSE firms, and 0 otherwise. Firm size is measured in the prior trading month. In addition, we include as controls *SEC Filing*$_{t-3,t}$, which is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days and *Press release*$_{t-3,t}$, which is a dummy variable if there was at least one press release issued by the firm over the past three trading days. *Print media*$_{t-3,t}$ is a dummy variable if there was at least one WSJ or NYT article about the firm in the past three trading days. We also control for lagged abnormal retail trading volume on day $t-1$. We include year-month fixed effects, and indicate statistical significance at the ten, five, and one percent levels with *, **, and ***, respectively (t-statistics in parentheses). Standard errors are clustered at the firm level.

Panel A: Effect on abnormal trading volume from articles

Dependent variable =	Log(Abnormal daily trading volume$_{t,t+2}$)			
Article	0.856***	0.347***	0.203***	0.326***
	(63.57)	(37.47)	(27.36)	(105.28)
Small Firm			-0.471***	
			(-67.74)	
Article × Small Firm			0.599***	
			(26.60)	
Fraudulent article				0.502***
				(3.78)
Observations	13,951,256	13,890,419	13,890,419	13,890,419
R-squared	0.024	0.371	0.381	0.370
Year-month F.E.	Y	Y	Y	Y
Controls	Y	Y	Y	Y

Panel B: Effect on price volatility from articles

Dependent variable =	Return Volatility$_{t,t+2}$			
Article	0.076***	0.104***	0.123***	0.066***
	(2.90)	(4.21)	(10.72)	(3.40)
Small Firm			0.484***	
			(29.10)	
Article × Small Firm			0.855***	
			(2.98)	
Fraudulent article				2.968***
				(3.97)
Observations	10,617,750	10,617,750	10,617,750	10,617,750
R-squared	0.001	0.001	0.002	0.001
Year-month F.E.	Y	Y	Y	Y
Controls	Y	Y	Y	Y

Table C3. **Article Readership Analysis**

The table reports regression results on the relation between Seeking Alpha readership and abnormal daily trading volume. The readership variables include the log of the number of clicks and number of reads (measured as those that scroll through the entire article) on the day the article is published plus the following two days. All regressions include date and firm fixed effects and indicate statistical significance at the ten, five, and one percent levels, with *, **, and ***, respectively (t-statistics in parentheses).

Dependent variable =	Abnormal daily volume			
log(# Clicks)	0.053***		-0.137***	
	(10.68)		(-6.24)	
log(# Reads)		0.060***	0.191***	
		(12.43)	(8.89)	
Fraction of reads				0.460***
				(8.51)
Observations	14,567	14,567	14,567	14,567
R^2	0.89	0.89	0.89	0.89
Daily F.E.	Y	Y	Y	Y
Firm F.E.	Y	Y	Y	Y

Table C4. **Timeliness of stories**

The table reports regression results on the relation between an article being posted on Seeking Alpha or Motley Fool and the returns in the prior periods. We examine whether that effect differs before and after the SEC investigation became public. AbRet prior day is abnormal return (residuals from a matched portfolio of stocks on size, BE/ME, and momentum) the day $t-1$ before the article is published. AbRet prior week and month are cumulative abnormal returns over $t-7$ or $t-20$ to $t-1$, respectively. All regressions indicate statistical significance at the ten, five, and one percent levels, with *, **, and ***, respectively (t-statistics in parentheses).

Dependent variable=	Article
AbRet prior day	0.009
	(1.15)
AbRet prior day × Post-event	-0.004
	(-0.35)
AbRet prior week	0.005
	(1.00)
AbRet prior week × Post-event	0.001
	(0.08)
AbRet prior month	0.009***
	(4.37)
AbRet prior month × Post-event	-0.005
	(-1.47)
Post-event	0.016***
	(46.93)
Observations	981,489
R-squared	0.002

Table C5. **Shock to Awareness of Fake News: The 2014 SEC Lawsuit on Trading Volume and Return Volatility**

The table examines whether the salience of the presence of fake news on the platforms, stemming from the public announcement of the SEC investigation and lawsuit, impacted investors' reaction to articles on these platforms. Panel A examines abnormal trading volume and Panel B examines return volatility over the three days following the publication of an article on these platforms. We compare the market's response to articles (in terms of trading volume and price volatility) in the six months before the SEC investigation and six months after the investigation, where we identify the February-March 2014 period as the period when the SEC investigation was announced and covered in the press. We include all firms that have ever had at least one article written about them on Seeking Alpha or Motley Fool during that time period. The regressors include the dummy variable *Article*, which equals 1 if there was at least one article published about the firm on day t, and 0 otherwise, and we include the dummy variable *Post-event*, which equals 1 if the time period is April 1 to September 30, 2014 and is zero if the article was published from August 1, 2013 to January 31, 2014. We exclude all observations prior to August 2013 and after September 2014. We then interact the *Post-event* dummy with the *Article* dummy in the regressions to test for the differential response to articles before versus after the SEC announced investigation. *Small firm* equals 1 if the firm is in the bottom 10th percentile of NYSE firms, and 0 otherwise. Firm size is measured in the prior trading month.We also include controls for SEC filings, press releases, and abnormal volume over the previous three days before the article, plus the day of the article. *SEC filing*$_{t-3,t}$ is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past three trading days, and *Press release*$_{t-3,t}$ is a dummy variable if there was at least one press release issued by the firm over the past three trading days. *Print media*$_{t-3,t}$ is a dummy variable if there was at least one WSJ or NYT article about the firm in the past three trading days. In Panel C we control for any changes in the reaction of investors to traditional information (SEC filings, Press releases, and articles in the WSJ and NYT) after the investigation announcement. We report results separately by firm size and retail ownership. We indicate statistical significance at the ten, five, and one percent levels with *, **, and ***, respectively (t-statistics in parentheses). Standard errors are clustered at the firm level.

Panel A: Effect on Abnormal Volume

Dependent variable =	Log(Abnormal daily volume$_{t,t+2}$)		
Article	0.338***	0.389***	0.235***
	(23.50)	(18.24)	(12.61)
Post		-0.179***	-0.079***
		(-17.75)	(-8.80)
Article × Post		-0.041	-0.068**
		(-1.52)	(-2.57)
Small Firm			-0.217***
			(-13.25)
Article × Small Firm			1.000***
			(11.66)
Post × Small Firm			-0.265***
			(-11.56)
Article× Post×Small Firm			-0.277***
			(-2.62)
SEC filing$_{t-3,t}$	0.122***	0.131***	0.118***
	(16.74)	(17.86)	(16.52)
Press release$_{t-3,t}$	0.287***	0.285***	0.246***
	(34.80)	(34.53)	(30.57)
Print media$_{t-3,t}$	0.074***	0.070***	-0.006
	(4.58)	(4.25)	(-0.37)
Abnormal Retail volume$_{t-1}$	1.486***	1.479***	1.450***
	(136.69)	(136.94)	(131.01)
Observations	1,401,509	1,401,509	1,401,509
R-squared	0.366	0.367	0.375

Panel B: Effect on Return Volatility			
Dependent variable =	Return volatility$_{t,t+2}$		
Article	0.138***	0.195***	0.073***
	(4.02)	(4.32)	(4.44)
Post		-0.055***	-0.014
		(-4.90)	(-1.41)
Article × Post		-0.076	0.056
		(-1.33)	(1.29)
Small Firm			0.362***
			(13.81)
Article × Small Firm			1.809***
			(4.59)
Post × Small Firm			-0.113***
			(-3.87)
Article× Post×Small Firm			-1.282***
			(-3.03)
SEC filing$_{t-3,t}$	0.032***	0.035***	0.038***
	(2.59)	(2.83)	(3.13)
Press release$_{t-3,t}$	0.043***	0.042***	0.077***
	(3.02)	(2.94)	(5.52)
Print media$_{t-3,t}$	-0.172***	-0.174***	-0.074***
	(-11.72)	(-11.87)	(-5.71)
Return Volatility$_{t-1}$	4.664***	4.654***	4.379***
	(3.24)	(3.24)	(3.20)
Observations	1,005,279	1,005,279	1,005,279
R-squared	0.002	0.002	0.004

Internet Appendix for "Social Media and Financial News Manipulation"

1. Detecting fake content

We develop a probability function for detecting fake content using the LIWC authenticity score combined with our sample of for-sure fake articles from Rick Pearson and the SEC. A critical advantage of our study is that we use the for-sure fake articles from Rick Pearson and the SEC to validate the linguistic algorithm and calibrate the authenticity score into a probability of fake news.[21] Our unique sample of 171 for-sure fake articles and 334 non-fake articles written by the same authors provides a test of the generalizability of the linguistic algorithm. We compare the LIWC authenticity score, normalized between 0 and 100 (where a high score denotes a higher level of authenticity), for fake and non-fake articles, controlling for author fixed effects to capture heterogeneity in author style, content, and reputation, and any matching of authors to types of articles.

Panel A of Table IA1 reports the difference in the LIWC authenticity scores for the fake and non-fake articles. Relative to an average authenticity score of 33 for non-fake articles, fake articles have a much lower average score of 19 (statistically significant at the 1% level). A plot of the distribution of authenticity scores for fake and non-fake articles in Figure IA1, Panel A, highlights the differences. Panel B of Figure IA1 provides more specific examples for two authors: *John Mylant* and *Equity Options Guru*. The distribution of authenticity scores across fake and non-fake articles for each author are quite different. While some of the non-fake articles also have low authenticity scores, the majority of the fake articles have very low authenticity scores.

Panel A of Table IA1 also reports summary statistics on language characteristics associated with authenticity as described in Pennebaker (2011) for the for-sure fake and non-fake articles. We report the average use of *1st person singular* (examples: I, me, mine), *Insight* (examples: think, know), *Relativity* (examples: area, bend, exit), *Time* (examples: end, until, season), *Discrepancy* (examples: should, would), and the average number of words per sentence. According to Pennebaker (2011) and Pennebaker et al. (2015), when people lie they also tend to use fewer words per sentence. Fake articles have lower authenticity scores due to fewer self-referencing, lower insight, lower relativity, and higher discrepancy scores. These differences suggest that fake articles that are posted on the knowledge sharing platforms are written differently and hence are potentially detectable based on linguistic cues.

1.1 Probability an article is fake

The sample of for-sure fake and non-fake articles allows us to calibrate and quantify the authenticity scores into a probability of an article being fake. While the LIWC authenticity score is statistically different between fake and non-fake articles, it is not easy to interpret the cardinal nature of the score – what does a 14 point difference in authenticity score mean? To provide a more direct interpretation of the results and their economic meaning, we develop a mapping of the authenticity score into probability space. This exercise is only possible because we have a set of known fake articles. Using the smaller sample of for-sure fake and non-fake articles, we map the authenticity score into the frequency of fake articles and apply Bayes rule to convert authenticity scores into a conditional probability of an article being fake for the broader sample of all articles on the platforms.

Specifically, let S be the authenticity score and F (T) denote a fake (true) article. We compute $Prob(S|F)$ and $Prob(S|T)$, where, crucial to this exercise, we use the smaller validation sample, where we know which articles are F and T, in order to measure the probabilities. From Bayes rule,

$$Prob(F|S) = \frac{Prob(S|F)Prob(F)}{Prob(S|F)Prob(F) + Prob(S|T)Prob(T)}.$$

If we integrate $Prob(F|S)$ over the empirical distribution of scores S, we get $Prob(F)$. The issue, of course, is that $Prob(F)$ is also an input in the calculation. The solution is found by solving the fixed point problem in which the observed $Prob(F)$ in the sample is representative of $Prob(F)$ in the overall

[21] A byproduct of this analysis is that since the LIWC authenticity score was not developed in the context of financial media, it is useful to assess its ability to distinguish fake from non-fake articles in our setting.

population.[22]

Figure IA2 plots the mapping of LIWC authenticity scores (S) into the conditional probability of being fake ($Prob(F|S)$) for the entire sample of 203,545 Seeking Alpha articles published between 2005 and 2015. The relation between the LIWC authenticity score and the probability of being fake is highly nonlinear. Specifically, the sharp increase in probability in the very low authenticity range suggests that articles may be more efficiently and better classified into fake and non-fake using a probability cutoff. We use a cutoff of $Prob(F|S) > 0.20$ to classify articles as being fake and classify articles with $Prob(F|S) < 0.01$ as being non-fake, with the remaining articles ($0.01 \leq Prob(F|S) \leq 0.20$) being classified as ambiguous or "other."[23] This cutoff implies an authenticity score that is even lower (about half) than the average authenticity score for the known fake articles in the SEC sample. Hence, this cutoff is conservative and designed to reduce type II errors. This conservative cutoff comes at a cost, however, where many articles will fail to be classified as either fake or non-fake.

We first examine how accurate our method is at identifying fake news from our small sample of 171 for-sure fake and 334 non-fake articles written by the same authors from the SEC investigation. We generate an authenticity score for each article, and calculate its probability of being fake. We tabulate the type I and type II errors of our classification at the bottom of Figure **??**. Our algorithm classifies 18 articles as being fake, of which 17 are actually fake, indicating that the Type II error rate is very low. Our method, being very conservative, misses a lot of fake articles, however, with 137 fake articles not being classified. Hence, the LIWC algorithm appears useful at identifying the most extreme content, where the linguistic cues are clear and discerning. If the aim is to identify fake news accurately, then we are less concerned with identifying fake content accurately, and less concerned with capturing *all* fake news and more concerned with identifying false content.

The algorithm also identifies 165 articles as being non-fake, of which 148 are non-fake and 17 are actually fake, implying an error of only 10%, which is quite low considering our methodology is designed to minimize type II errors of fake news. However, again, the conservative methodology results in 185 non-fake articles not being classified. Articles falling into the ambiguous region (with $0.01 \leq Prob(F|S) \leq 0.20$) comprise 64% of the sample. But, for the 36% of articles we do classify, we are confident in their classification. The tradeoff in how confident our classification is versus how many articles we classify highlights the challenges facing social media platforms in flagging fake content.

We apply our calibrated probability model to the broader sample of all articles. Table IA1 Panel A, columns 3 to 8, show summary statistics for the *Fake*, *Non Fake*, and *Other* articles identified by the algorithm. The number of articles in each category, the mean of the authenticity measure that we use to construct the probabilities, and the components of that authenticity measure from the LIWC algorithm are reported. The difference in authenticity measures translates into large differences in the estimated probability of being fake from our calibrated probability function: the articles we identify as fake have an average 44.5% $Prob(F)$ based on their authenticity score, while the average probability for articles we identify as non-fake is less than 1%. The authenticity scores for the two groups of articles are 5.5 for fake articles versus 49.1 for non-fake articles.

As another validation exercise we analyze a particular set of articles written by a Motley Fool author, Seth Jayson, who has been working for Motley Fool full-time since 2004 as a journalist, and has written over 31,000 articles. Mr. Jayson's articles are a good test case because he works directly for Motley Fool and has done so for a long time. Hence, it is unlikely he has written fake articles on their platform and unlikely that promotional firms would approach him to do so. Using Mr. Jayson's articles as an extreme test of our classification methodology, we classify 18,361 of Mr. Jayson's articles as reliably (99% probable) non-fake and only 2 of his articles (0.006%) as probabilistically fake. These findings are consistent with our prior that Mr. Jayson did not write fake articles for Motley Fool and suggests our classification methodology works well. However, once again, the tradeoff of our conservative methodology is that 38.8% of Mr. Jayson's articles cannot be classified.

Panel C of Table IA1 reports summary statistics on the firms covered in these articles. The average

[22] While we could estimate $P(F|S)$ directly in our smaller sample and then integrate over the distribution of S, a concern might be that our smaller sample is highly selected by the SEC and, therefore, will not give an accurate picture of the frequency of fake news.

[23] Our results are not sensitive to different cutoffs in the 0.10 to 0.30 probability range for fake, where 0.20 was chosen based on the steep increase in probability in Figure IA2.

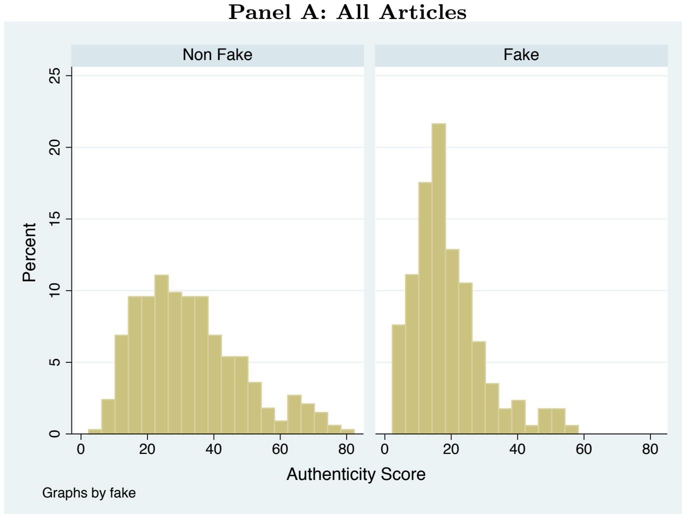
fraction of retail investors, the average number of analysts covering the firm, and the average firm size (in \$US millions) for each article group are reported. For-sure fake articles from Rick Pearson's sample tend to cover really small firms (average market capitalization of \$7.4 million) with a high fraction of retail investor ownership and very low analyst coverage.

We show that the probabilistically fake articles have a more muted impact on retail trading activity and stock price movement than the for-sure fake articles, with similar sign but smaller magnitude and precision, which makes sense since our algorithm is a noisy measure of fake content.

Figure IA1. **Authenticity Scores**

This figure depicts the distribution of authenticity scores from the LIWC algorithm for fake and non-fake articles. In Panel A, we plot authenticity scores for all the articles in our validation sample provided by Rick Pearson and the SEC, consisting of 171 for-sure fake and 334 non-fake articles. In Panel B, we plot authenticity scores for two authors in our validation sample with the most articles written.

Panel A: All Articles



Panel B: Author-Specific

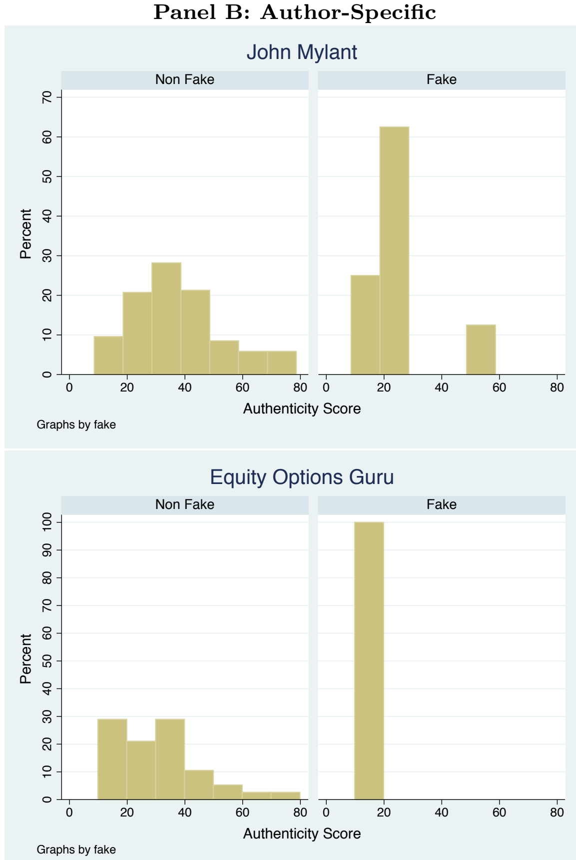
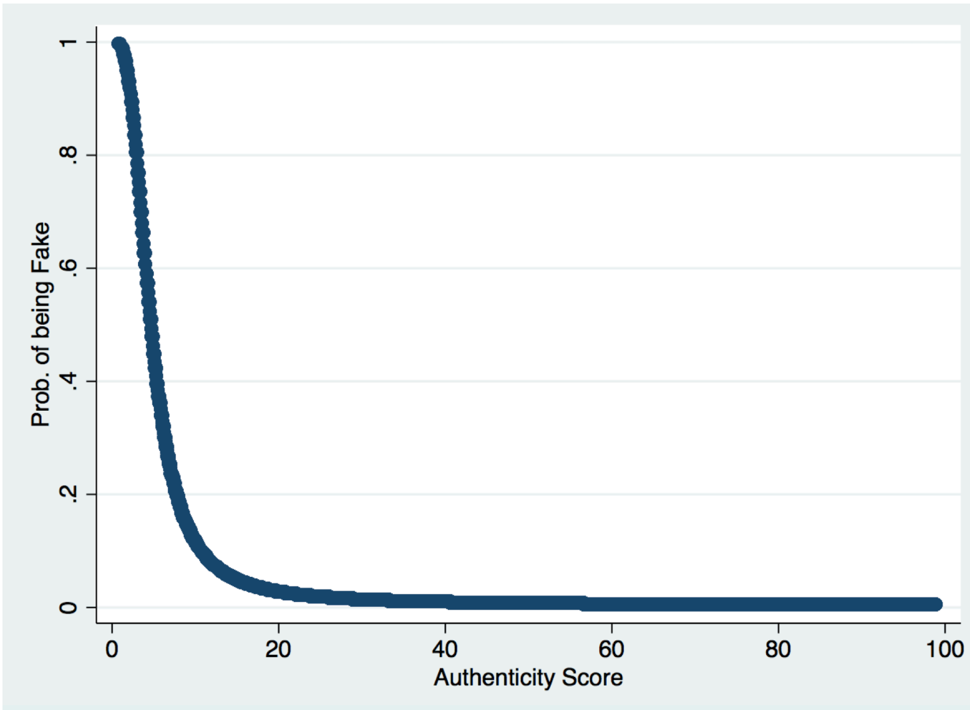


Figure IA2. **Authenticity Score and the Probability of Being Fake**

This figure depicts the relationship between LIWC authenticity scores (S) and the conditional probability of being fake ($Prob(F|S)$) using our validation sample from Rick Pearson and the SEC containing for-sure fake and non-fake articles. We also report the classification error of our methodology that classifies articles as being fake if their probability of being fake is greater than 20% and classifies articles as being non-fake if their probability of being fake is less than 1%, using the mapping of authenticity scores to probabilities. The type I and type II errors for the validation sample are reported.



	Rick Pearson's and SEC articles		
	For-sure fake	Non-fake	Total
Algorithm determined			
Fake	17	1	18
Non-Fake	17	148	165
Not classified/other	137	185	322
Total	171	334	505

Table IA1. **Linguistic Summary Statistics (from the LIWC Model)**

This table presents summary statistics for various LIWC textual measures for our sample of articles on Seeking Alpha and Motley Fool. In the "Rick Pearson & SEC" sample, *For-sure Fake Articles* are articles that have been shared with us by Rick Pearson, or that were subpoenaed by the SEC and shared with us by Seeking Alpha, and *Non Fake* articles are articles by the same authors as the for-sure fake articles that are not fake. *Social Media Articles* are regular articles that we downloaded from Seeking Alpha and Motley Fool. Of those articles, *Fake* articles are articles whose probability of being fake was higher than 20%, *Non Fake* articles are articles with probability of being fake less than 1%, and the rest are classified as *Other*, which are not used in the main analysis. Panel A reports the number of articles as well as the mean *Authenticity* score from LIWC that we use to construct the probabilities of being fake. We also report the means of several of the other variables the comprise the LIWC authenticity score: the average of *1st person singular* mentions (examples: I, me, mine), the *Insight* measure (examples: think, know), *Relativity* measure (examples: area, bend, exit), *Time* measure (examples: end, until, season), *Discrepancy* measure (examples: should, would), and the average number of words per sentence. We also report mean scores from LIWC for other attributes besides authenticity: Clout, Analytic, and Tone as defined by LIWC and described in Table 5. Panel B reports the average probability of being fake using our mapping of the Authenticity score into a probability of being fake, and applying the 20% and 1% cut off probabilities for fake and non-fake identification, respectively. Panel C reports summary statistics on the firms covered in the articles. The average fraction of retail investors, the average number of analysts covering the firm, and the average firm size (in $US millions) for each article group are reported. Differences between measures for Fake and non-Fake articles that are statistically significant at the 5% level are denoted in bold.

	Rick Pearson & SEC		Social Media Articles		
	For-sure Fake	Non Fake	Fake	Non Fake	Other
Panel A: LIWC variables					
Number of articles	171	334	5,301	195,232	150,928
Authentic	**19.09**	**32.79**	**5.51**	**49.11**	22.26
1st pers singular	0.42	0.76	**0.24**	**0.8**	0.40
Words per sentence	57.55	65.23	**25.79**	**20.75**	20.93
Insight	1.52	1.67	**1.48**	**1.88**	1.72
Relativity	**12.92**	**15.11**	**9.72**	**17.04**	13.42
Time	4.97	5.35	**3.38**	**6.42**	4.93
Discrepancy	1.05	1.41	**1.24**	**1.11**	1.17
Clout	**58.25**	**52.31**	**64.71**	**56.07**	60.16
Analytic	**94.65**	**90.97**	**91.94**	**90.94**	91.81
Tone	61.79	57.49	**58.27**	**60.28**	59.04
Panel B: Probability of being Fake					
Prob(Fake)	100%	0%	44.51%	0.58%	3.15%
Panel C: Firm characteristics					
Percent of retail investors	76.66%	50.16%	41.9%	40.09%	42.16%
Number of analysts	6.96	16.76	18.78	18.99	18.45
Firm Size ($Mil)	7.36	58.43	59.18	59.37	61.00