Fake News: Evidence from Financial Markets

Shimon Kogan*
MIT Sloan School of Management
Interdisciplinary Center Herzliya

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

Marina Niessner
AQR Capital Management

April 2019

* We thank Tony Cookson, 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, and AQR Capital Management 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@mit.edu, tobias.moskowitz@yale.edu, and marina.niessner@aqr.com

Abstract

We examine fake news in financial markets, a laboratory that offers an opportunity to quantify its direct and indirect impact. We study three experimental settings. The first is a unique dataset of unambiguous fake articles on financial news platforms prosecuted by the Securities and Exchange Commission. The second applies a linguistic algorithm to detect deception in expression on the universe of articles on these platforms, using the first sample to validate and calibrate the algorithm. The third is an event study exploiting the SEC investigation as a public shock to investor awareness of fake news. We find that trading activity and price volatility rise with fake news about the firms mentioned in the articles. Following public revelation of the existence of fake news, we find an immediate decrease in reaction to all news, including legitimate news, on these platforms, consistent with indirect spillover effects of fake news conjectured by theory. These findings are predominant among small firms with high retail ownership, and are stronger for more circulated articles. Our results are consistent with economic theory on media bias and its application to fake news.

1. Introduction

False or misleading information can potentially impact social, political, and economic relationships. A recent and prominent example is the increased attention "fake news" is receiving. Fake news is a form of disinformation, including hoaxes, frauds, or deceptions, designed to mislead consumers of news. The economics of fake news is an interesting and young area of study. What motivates fake news? What impact does it have? What are the welfare costs and benefits of monitoring it? What policy prescriptions should be considered?

Analysis of these issues in Economics has primarily been theoretical, and even outside of Economics there is a paucity of evidence studying the causal impact of fake news on behavior outside of laboratory settings.[1] False content may impose private and public costs by making it more difficult for consumers to infer the truth, reduce positive social externalities from shared-information platforms, increase skepticism and distrust of legitimate news, and potentially cause resource misallocation. On the other hand, consumers may derive utility from fake news (as entertainment or if slanted toward their biases, as in Mullainathan and Shleifer (2005)). Little empirical evidence on these issues exists, however, due to lack of data, particularly the identification of unambiguous fake content.[2] With the explosion of largely unmonitored shared information platforms, such as social media, blogs, and other crowd-sourced content, the potential influence of fake and biased news is a growing concern.[3]

[1] Allcott and Gentzkow (2017) model fake news as an extension of Gentzkow and Shapiro (2005) and Gentzkow et al. (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, with some debate over the relevance and consequences of fake news. 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.

[2] For example, Amazon, Google, Twitter, and Facebook are currently using human editors to evaluate content in the hopes of training an algorithm to identify false content systematically with limited success (Cullan-Jones (2016), Leong (2017), Leathern (2017)).

[3] 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. (2018a) 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. Fake news has been studied with regard to the 2016 U.S. Presidential election (Allcott and Gentzkow (2017), Silverman (2016), Timberg (2016), Silverman and

We provide some of the first empirical estimates of the direct and indirect impact of fake news using three empirical settings. The first is a dataset of identified fake articles from a Securities and Exchange Commission (SEC) investigation into paid-for false content on shared financial news networks. The sample is small, but the identity of fake news is clean – stemming from an undercover investigation by an industry whistle blower, Rick Pearson, resulting in 171 articles by 20 authors about 47 companies that knowingly provided false information about the stock. The data offer a singular look at identified fake content.

While the first setting provides identified fake content, the sample is small and narrow. To broaden the analysis, and perhaps draw more general conclusions, we collect all articles from two prominent financial crowd-sourced websites – Seeking Alpha and Motley Fool – obtaining 203,545 articles from 2005 to 2015 for Seeking Alpha, and 147,916 articles from 2009 to 2014 for Motley Fool, covering 7,700 publicly traded firms. We then attempt to identify fake content within this broader set of articles using a linguistic algorithm (Pennebaker et al. (2015), Newman et al. (2003)) designed to detect deception in expression to assess the authenticity of each article. Using an "off-the-shelf" algorithm has pros and cons. On the one hand, it allows us to avoid in-sample overfitting, since the algorithm was developed in laboratory settings. On the other hand, it is unclear whether this linguistic method will work for a shared information setting about financial news and topics, whose content and structure could be fundamentally different. Therefore, we use the first empirical setting of known fake articles from the SEC investigation to validate the algorithm. This serves a dual purpose because it also allows us to calibrate a model for the probability of fake news to examine a broader set of content.

Even with our identified set of fake articles from the SEC, we show how difficult it is to detect fake content systematically. Our calibrated model classifies news as fake, non-fake, and ambiguous with the objective of minimizing classification error so that we can confidently identify fake and non-fake content. The algorithm is successful on some dimensions, but

Alexander (2016)), and ReviewMeta (2016) examines fake reviews on Amazon.

performs poorly on others. For instance, it has a type II error on the known fake articles of less than 1% (false positives) and a type I error on the non-fake articles by the same authors of less than 10% (false negatives). However, the low classification error comes at a cost because most articles cannot be confidently classified, highlighting one of the major challenges and tradeoffs in attempting to detect and quantify fake news systematically using linguistic tools and more generally. This exercise highlights well the tradeoff of precision versus breadth of fake content identification. The linguistic tool when calibrated with a stringent threshold can deliver precise classification for the most extreme articles, but misses a lot of other fake content because of its inability to classify most articles. For our purposes, we focus on precision of the classification and hence adopt a very conservative measure. Despite the conservatism, the prevalence of fake news we identify in the broader sample is significant (2.8% of articles).

Another virtue of using and dissecting the linguistic algorithm is that it can shed light on important characteristics of articles that signal false content or intent to deceive. In addition to investigating and highlighting those features, we also examine the many other diagnostics produced by the linguistic tool to examine what other characteristics the articles might differ on. The analysis may help future detection models, but can also show whether fake articles are designed to attract more attention, engender a greater response, and increase influence, which improves our understanding of their impact.

Our third empirical setting does not require identification of fake news at all. Rather, we exploit the public revelation of the SEC's investigation on these platforms (that ultimately led to the dataset for our first experiment) as a shock to the market's awareness of fake news. We first show that the market seemed largely unaware of fake content before the announcement, and then examine the market's response to news before versus after the event. We use this setting to test another implication from theory that fake news imposes externalities on other news. We examine the shock from the public's awareness of fake news on the market's reaction to news in general, including legitimate news.

We first examine the direct impact of fake versus non-fake news on trading activity. Using the first empirical setting of known fake articles from the SEC, we find a larger trading response to fake news relative to non-fake articles published at the same time on the same platform, and controlling for other factors influencing trading volume. Abnormal trading volume rises by more than 50% over the three days following a fake article relative to a non-fake article. This effect is concentrated in the smallest ten percent of firms on public stock exchanges and is not significant among the largest firms. The effect is also bigger for stocks with higher retail investor ownership, where a ten percent increase in retail ownership results in a 7% increase in the trading volume response to fake news relative to non-fake news. The stronger impact on trading activity is likely driven by fake articles being more sensational and diffusing more quickly across consumers (Vosoughi et al. (2018b)). Further corroborating that story, we find that fake articles generate more "clicks" and more "reads." The larger influence of fake articles likely stems from fake articles, by design, being crafted to attract more attention and influence. The linguistic algorithm shows these articles have a stronger and more authoritative tone and a clearer hierarchical structure. Since the goal of fake articles is to influence, it is not surprising that they are designed and written in a way that seeks to attract attention, and they appear successful at that aim.

Turning to the broader set of articles in the second experiment, where we estimate the probability of fake news, we find similar patterns but more muted effects. We find that the direct effect on trading activity from fake news is stronger for smaller firms with higher retail ownership and for articles with greater circulation (measured by number of clicks and readers), lending credence to these platforms influencing investor behavior.

Exploring the indirect effects of fake news on trading activity using the third empirical setting from the SEC announcement event, we find that trading volume drops significantly for *any* news article written on these platforms after the public became aware of fake content. Comparing trading volume before versus after the SEC announced investigation, trading volume drops by 5.2% for all news on these platforms following the information shock, with the drop being even larger for small firms with high retail ownership. We also find decreases in trading volume on the Motley Fool platform, which is a competitor platform that was not part of the SEC investigation, indicating that awareness of fake news caused a spillover effect in trading volume for other news platforms as well. These findings are consistent with theory (Allcott and Gentzkow (2017) and Aymanns et al. (2017)) arguing that awareness of fake news causes agents to discount all news. Looking at the the comments section to these articles using natural language processing (NLP), we find a significant increase in uses of the words "fake" and "fraud" after the SEC investigation came to light, consistent with participants being more concerned and aware of fake news after the SEC announcement. Importantly, the frequency of these words in the comments has no predictive power to detect fake articles, indicating that readers had no ability to identify fake news, but were just more aware of its existence, consistent with their response to distrust all news.

To ensure these results are not driven by trends in news or response to news or to unobservable effects that happen to coincide with the SEC event, we conduct several additional tests. First, we show no pre trends in news or response to news prior to the SEC announcement. Second, we run several "placebo" or falsification tests designed to pick up general effects associated with news and the market's response to news, but unlikely associated with fake news from these platforms. We look at news media more generally outside of these platforms, specifically newspaper articles from the Wall Street Journal and New York Times and examine whether there are any spillover effects from the SEC investigation more broadly. It is unlikely that awareness of fake news on these social media platforms would extend to the WSJ or NYT, yet general trends in news and the market's response to it should be evident in other types of media as well. Consistent with the SEC event providing a shock to fake news awareness on these platforms, and not other trends in news or responses to it, we do not find a commensurate decline in trading volume in response to news from these newspapers. This indicates that investors' distrust of news from the social media platforms did not extend to more prominent types of media, and the lack of change in trading activity in response to newspaper articles provides a useful falsification test ruling out trends in response to news or unobservable effects on trading volume that coincide with the event.[4]

We conduct similar falsification exercises with corporate filings and press releases as other sources of news that are also unlikely to be polluted by fake content or at least interpreted differently after the SEC investigation of the social media platforms. Once again, we find no economically or statistically significant change in market response to these news sources before versus after the SEC event, consistent with other news trends or unobservables not likely confounded with our event study. These results also suggest, however, that the spillover effects from investor awareness of fake news are limited to the social media platforms.

We also examine pricing effects to see if fake news moves prices in a distortive way. If markets are informationally efficient, then despite the changes in trading volume, prices will be unaffected. In this case, fake news may distort attention and trading, but not firm values. Using the sample of known fake articles from the SEC, we find that the fake promotional articles increase idiosyncratic stock volatility by roughly 40 percent relative to non-fake articles over the three days after publication, with the effects concentrated among small firms with high retail ownership. Looking at the direction of price movement, the average fake article is positive and pushes up stock prices for the smallest decile of companies on the NYSE by an average 8% over the next six months, which subsequently gets fully reversed over the course of a year and eventually becomes cumulatively negative at −2.5%. Looking at the broader set of articles where we estimate the probability of fake news, we find similar but much weaker patterns of temporary positive price effects for the smallest firms that get fully reversed and turn negative. For large firms, we find no price impact.[5]

The results are consistent with markets being less efficient for small firms, where the cost of information is higher and the average investor is smaller and perhaps less sophisticated.[6]

[4]This test assumes general news trends affect all media sources and hence cannot completely rule out trends in social media news that are different from other media potentially affecting the results.

[5]An investor at the time of the article's publication could not have constructed or used a similar methodology to detect the probability of false content since the fake articles from which we calibrate our model were not yet known or identified.

[6]The cost of information can be both a direct cost of gathering, processing, and analyzing information,

Consistent with this notion, paid-for fake content from the SEC investigation was exclusively engaged by small firms, and not by large firms, as expected in equilibrium. We also find that small firms are more likely to issue press releases and 8-K filings coinciding with the fake articles, consistent with a coordinated effort to influence the narrative of news about the firm, and find evidence of insiders positioning themselves to benefit from the price movement.[7] These findings speak to the motivation for fake news in our setting.

Our results provide some of the first empirical estimates of the direct and indirect impacts of fake news, which have implications for theory. The prevalence of fake content and its impact on trading activity is consistent with fake news being tailored to consumer's priors as suggested by Allcott and Gentzkow (2017), and more broadly consistent with media bias (Gentzkow and Shapiro (2005) and Gentzkow et al. (2015)). The price patterns we find for small firms may also be consistent with fake 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 fake news from 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.[8] The results may also be related, more generally, to the economics of norms and institutions such as trust and social capital (Guiso et al. (2004), Guiso, Sapienza, and Zingales (Guiso et al.), Guiso et al. (2010), Sapienza and Zingales (2012)).

Finally, given our setting is financial markets, specifically shared-information platforms on financial news, there are reasons to be both cautious and optimistic on what we can learn about the impact of fake news more broadly. One of the benefits of financial markets is the ability to quantify effects through prices and trading activity. On the other hand, these

as well as the indirect costs of misperceiving or misreacting to information from behavioral biases. Allcott and Gentzkow (2017) argue information costs are necessary for fake news production.

[7]The reason Rick Pearson went undercover initially and why the SEC got involved was because many fake articles were tied to promotional pump-and-dump schemes to manipulate the stock price.

[8]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.

specific information platforms may have little influence on markets either because they are unimportant or due to markets already incorporating the information. While a 2015 study by Greenwich Associates found that 48% of institutional investors use social media to "read timely news," this news may not matter if markets are efficient (Fama (1970)). Fake news, in particular, should have no influence in an informationally efficient market, regardless of the equilibrium asset pricing model. Hence, our setting offers a unique test of market efficiency that circumvents the joint hypothesis problem. We run the flip side of the classic event study (Fama et al. (1969)) by examining market responses to fake news events and find (for small stocks) that trading activity and prices are affected.

Our results provide more evidence on the growing impact of crowd-sourced information platforms (Chen et al. (2014)). If fake news can impact U.S. equity markets, where competition for information is intense, markets are liquid, and arbitrage activity exists, then it could have even greater influence in settings where information costs are high and the ability to correct misinformation is more limited, such as online consumer, political, and socially shared-information networks.

The rest of the paper is organized as follows. Section 2 details our first two empirical settings: the sample of fake news articles obtained from the SEC and the broader set of all articles from the shared-information platforms that we apply a linguistic algorithm to assess fake content. Section 3 analyzes the direct impact of fake news through investor trading activity and price impact. Section 4 describes the third empirical setting – an event study of the SEC's investigation that provides a shock to public awareness of fake news – to measure the indirect effects of fake news on the market's reaction to news generally. Section 5 considers the motivation of fake news in our context. Section 6 concludes.

2. Identifying Fake News

We detail our first two empirical settings using articles from knowledge-sharing financial platforms. We first describe these platforms and the data we obtain. We then describe our first sample of fake articles from the SEC. Using this sample, we validate and calibrate a linguistic model for identifying probable fake content. We then apply this model to the broader sample of articles with unknown authenticity from the same media platforms that generates our second empirical setting.

2.1. Knowledge Sharing Platforms

Our sample of articles comes from the two largest financial crowd-sourced platforms: Seeking Alpha and Motley Fool. Seeking Alpha is an online news service provider for financial markets, whose content is provided by independent contributors. The company has had 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 also through a shared-knowledge platform. We obtain the articles posted on these platforms, including their content, authorship, and in the case of Seeking Alpha, commentary from other users. Appendix A details how authors on these sites contribute and are compensated for their articles.

The popularity of these sites has grown considerably. Seeking Alpha grew from two million unique monthly visitors in 2011 to over nine million in 2014, generating 40 million visits per month. While these platforms allow for the 'democratization' of financial information production, concerns have been raised about their susceptibility to fraud, since they are virtually unregulated, frequented predominantly by retail investors, and authors on these platforms can use pseudonyms (though the platforms claim they know the true identity of each author, in case that information is subpoenaed by the SEC, which it was in the cases we examine below). 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 can have legal ramifications. What is *illegal*, according to Section 17b of the securities code, is to fail to disclose any direct or indirect compensation that the author received from the company, a broker-dealer, or from an underwriter.[9] We exploit a

[9]In June 2012, Seeking Alpha announced it would no longer permit publication of articles for which

subset of fraudulent articles on these platforms, identified by an undercover investigation of paid-for false content and eventual SEC prosecution for our first empirical setting.

2.2. "For-Sure" Fake Articles

We examine a unique dataset of articles whose authors were paid to write fake content, and where the authors illegally did not disclose payment. The articles are obtained from an industry insider, Rick Pearson, who, as a regular contributor to Seeking Alpha, was approached by a public relations firm to promote certain stocks by writing articles with false information for a fee without disclosing the payment. Mr. Pearson instead went undercover to investigate how rampant this practice was and uncovered more than one hundred fake, paid-for articles by other authors who did not disclose their compensation. He turned the evidence over to the SEC, who investigated each of these cases. The fake articles were subsequently taken down by the platforms once the SEC informed them of the investigations. The SEC filed its first lawsuit pertaining to these articles on October 31, 2014, prosecuting the authors, the promotion firms who paid them, and in some cases the companies and their executives who hired the promotion firms.[10]

Mr. Pearson kindly provided the articles to us that he determined to be fake: 111 fake articles by 12 authors covering 46 publicly traded companies. We also obtained a second set of known or, as we will refer to them, "for-sure" fake articles, which the SEC identified during the investigation containing paid-for fake content.[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 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

compensation had been paid.

[10]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.

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

of "for-sure" fake articles consists of 171 articles written by 20 different authors about 47 publicly traded firms.[12]

It is important to define what we mean by *fake* articles. In the sample from Rick Pearson and the SEC, the fake articles are those that were paid for by a promotional firm and not disclosed, and many of the authors admitted that the articles were written to deceive the market and manipulate the stock price. Consequently, these articles contained false information that authors knew to be incorrect at the time. How false or wrong that information turned out to be is difficult to assess. For example, an article could intend to deceive by embellishing the prospects of the firm, but could turn out to be mostly correct in that assessment ex-post. In other instances, the deception may be grossly off. Hence, our fake articles are about *intent* to deceive and not necessarily about whether they are right or wrong ex-post. Put differently, the articles contain information that the authors know to be wrong ex ante, at the time they write the articles, but that information could turn out to be closer to the truth ex post. For instance, stating falsely that earnings will rise by 1% when the evidence suggests a fall in earnings, but ex post earnings end up rising by 0.5% due to unexpected positive news. We focus on authenticity and not accuracy. Some of our analysis on the language used in the articles and the impact on stock prices helps assess how wrong the information is.

To provide some insight into the content of these fake articles, we highlight a recent example from our sample that was prosecuted by the SEC in September 2018. One of the fake, 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:

[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 fake articles about the same firms, with only one new firm identified.

> *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 promotion article. Many of the fake, paid-for articles from the SEC involve similar issues and often coincide with a public relations campaign orchestrated by the firm to artificially prop up the stock price, including press releases, filings, and corporate actions. We investigate these issues more below and attempt to control for these actions when assessing outcomes on trading activity and prices.

We also compare the fake articles to a set of non-fake articles by the same authors to difference out unobservable heterogeneity in author style and reputation that may help us better identify fake content in the broader sample as well as more precisely measure the impact of fake articles controlling for other effects that could also impact investor attention. Specifically, we obtain a sample of other articles written by the same 20 authors that are under the SEC's investigation, but that were *not* paid for by a PR firm and have no evidence of being false. These are 334 additional articles from the same set of authors covering 171 companies published on the same platform that are not fake. We refer to these articles as "non-fake" following our definition above.

2.3. Assessing Authenticity in the Broader Set of Articles

Our second empirical setting uses the broader set of all articles written on these platforms and attempts to assess their probability of being fake. While our unique sample of fake articles from the SEC provides unambiguous fake news, the sample is small and may raise external validity concerns for drawing general conclusions. To complement these data, we manually download all articles published on Seeking Alpha and Motley Fool, obtaining 203,545 articles from Seeking Alpha over the period 2005 to 2015 and 147,916 articles from Motley Fool from 2009 to 2014. This exercise also provides an assessment of how difficult it is to detect fake content and some linguistic cues that may help achieve this aim.

2.3.1 How Do You Tell if Someone Is Lying?

We develop a probability function for detecting fake content using an objective and scalable measure based on quantifiable research from linguistics. Specifically, we use a linguistic algorithm designed to detect deception in expression – the Linguistic Inquiry Word Count model (LIWC2015) from Pennebaker et al. (2015) – which focuses on individuals' writing or speech style, and appears somewhat successful at measuring 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, one of which is the authenticity score that detects deception in expression. While the exact formula for the authenticity score is proprietary, Pennebaker (2011) describes which linguistic traits are associated with honesty. In particular, truth-tellers tend to use more self-reference words and communicate through longer sentences compared to liars. When people lie, they tend to distance themselves from the story by using fewer "I" or "me"-words. 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 have used similar methods to assess authenticity in speech or writing.

For example, consider the two statements of 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 to the public.

The algorithm uses a combination of these linguistic traits to generate the authenticity measure. A unique and critical advantage of our study is that we use the for-sure fake articles from Rick Pearson and the SEC (our first setting) to validate the linguistic algorithm and calibrate the authenticity score into a probability of fake news. One of the major challenges to studying this issue is the lack of known fake content. 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 context. Financial blogs and articles tend to point to facts, trends, and figures, which may be decidedly different from narratives that were used to develop the linguistic algorithm, controlling for author-specific features and heterogeneity. Our unique sample of 171 for-sure fake articles and 334 non-fake articles written by the same authors, provides a validation and test of the generalizability of the linguistic algorithm. We compare the LIWC authenticity score, which is normalized between 0 and 100 (where a high score denotes a higher level of authenticity), for the two samples and control for author fixed effects to capture heterogeneity in author style, content, or reputation, and any matching of authors to types of articles.

Panel A of Table 1 reports the difference in the LIWC authenticity scores for the fake and non-fake samples. 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 1, Panel A highlights the differences, controlling for author heterogeneity since we examine fake and non-fake articles for the same authors. Panel B of Figure 1 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, most of the fake articles have very low authenticity scores.

Panel A of Table 1 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, having fewer self-referencing, lower insight, lower relativity, and higher discrepancy scores on average. These findings provide an out-of-sample test of the LIWC algorithm that validates it in a unique setting, which would not be possible without the sample of known fake articles from the SEC. These differences suggest that fake articles are written differently and hence are potentially detectable based on linguistic cues, though we evaluate below the efficiency of such techniques. In addition, the LIWC model also produces scores for other attributes besides authenticity, including "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), though the latter should not be confused with "sentiment" as defined in the Finance literature (Tetlock (2007)), which relates to whether the news on prices is expected to be positive. As Panel A of Table 1 shows, fake articles are different along these other dimensions as well, though the differences are smaller than the authenticity differences. As we argue below, it is not surprising that fake articles differ along multiple dimensions since they are designed to mislead and influence.

2.3.2 Probability of Being Fake

The sample of for-sure fake and non-fake articles allows us to calibrate and quantify the authenticity scores into a probability of fake content. 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. Again, 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 the broader sample of Seeking Alpha and Motley Fool articles to convert authenticity scores into a conditional probability of fake news.

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, 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 population.[13]

Figure 2 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

[13]While we could estimate $P(F|S)$ directly in our smaller sample and then integrate over the distribution of S, the concern is that our smaller sample is highly selected by the SEC (see Section 5) and, therefore, will not give an accurate picture of the frequency of fake news.

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 $Prob(F) > 0.20$ to classify articles as being fake and classify articles with $Prob(F) < 0.01$ as being non-fake, with the remaining articles ($0.01 \leq Prob(F) \leq 0.20$) being classified as ambiguous or "other."[14] 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. As mentioned previously, this conservative cutoff comes at a cost, where many articles will fail to be identified as the "ambiguous" region is wide.

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. We generate an authenticity score for each article, and calculate its probability of being fake. 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 – 154 fake articles are not classified as fake. This result illustrates one of the major difficulties in identifying fake news. The linguistic algorithms are generally poor at identifying the majority of fake content. In this case, missing 154 out of 171 (90%) fake articles. However, among the 18 articles it does identify as fake, there is only one false positive. Hence, the LIWC algorithm appears useful at identifying the most extreme content, where the linguistic cues are clear and discerning. For our purposes, the aim is to identify accurately some fake news, and not all fake news. For this task, the linguistic algorithm with a stringent threshold achieves that aim quite accurately. If the objective had been to identify all fake news – a more challenging task for sure – these algorithms may not work particularly well and will have large classification errors.

Our algorithm also identifies 165 articles as being non-fake. Of those, 17 are actually fake,

[14] 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 2.

implying an error of about 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 186 (334 - 165 + 17) non-fake articles not being classified. Articles falling into the ambiguous classification region (with $0.01 \leq Prob(Fake) \leq 0.20$) comprise 64% of the sample. Yet, for the 36% of articles we do classify, we are confident in their classification as either fake or not fake. There is a tradeoff in how confident we wish to be in our classification versus how many articles we wish to classify. This tradeoff highlights the limitations of the linguistic algorithm and echoes some of the challenges facing social media platforms in flagging fake content. In our case, we choose to minimize the number of falsely identified fake articles and falsely identified non-fake articles, so that when we look at their differential impact, we are confident that we are measuring the difference between fake and authentic news.

Using our fake news probability model, calibrated to the sample of known, for-sure fake and non-fake articles, we apply our methodology to the broader sample of all articles. Table 1 Panel A shows 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 function: the articles we identify as fake have an average 0.45 $Prob(F)$ based on their authenticity score, while the average probability for articles we identify as non-fake is less than 0.01. Obviously, the articles are sorted based on the probabilities, but the magnitude of the difference is interesting and suggests substantial differences in authenticity scores between the two groups of articles, which Table 1 reports are 5.4 versus 50.7, respectively.

We also apply our methodology for identifying fake articles to another sample of articles from another crowd-sourced financial news platform – Motley Fool, where we have 147,916 articles from 2009 to 2014. Applying the LIWC algorithm, we obtain similar differences in authenticity scores and probabilities in classifying Motley Fool articles into *Fake* and

Non-Fake. The unconditional probability of fake news on the Motley Fool sample is 2.7%, almost identical to the 2.8% we found for Seeking Alpha. This indicates that the fraction of articles that exhibit strong linguistic cues that identify with false content is the same on both social media platforms. Looking at the rest of the components of the authenticity score, the algorithm does a similar job on the Motley Fool sample.

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 for a long time. Hence, it is unlikely he has written fake articles on their platform and equally unlikely that promotional firms would approach him to do so. We, use Mr. Jayson's articles as a placebo test of our classification methodology, where intuitively we believe none of his articles to be fake. Using our methodology, we classify 18,361 of Mr. Jayson's articles as reliably (99% probable) non-fake and only 2 of his articles as probabilistically fake. In other words, we classify 0.006% of his articles as fake, which is consistent with our prior that he wrote zero fake articles and suggests our conservative classification methodology works well at identification. Though, once again, the tradeoff of our conservative methodology is that many of Mr. Jayson's articles (38.8%) cannot be classified at all.

Finally, Panel C of Table 1 reports 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. For-sure fake articles tend to cover small firms with a high fraction of retail investor ownership and low analyst coverage. The probabilistically determined fake and non-fake articles from the broader Seeking Alpha and Motley Fool samples exhibit more muted differences. Notably, the Motley Fool articles are written about significantly larger firms than Seeking Alpha, and the for-sure fake articles identified by Rick Pearson and the SEC are about tiny firms whose average market capitalization is only $7.4 million.

Table C1 in the Internet Appendix also examines whether fake articles tend to cluster in specific industries. We separate articles into one of the 12 Fama-French industries that the firms mentioned in the articles belong to. For the for-sure fake articles provided to us by Rick Pearson and the SEC, 81% are about firms in the *Healthcare* industry. This finding is not too surprising as these articles came from authors who were hired primarily by two PR firms that concentrated on the healthcare industry. For the non-fake articles, the majority of firms belong to *Business Equipment*, *Healthcare*, *Finance*, and *Manufacturing* industries. The industry composition of *Fake* and *Non-Fake* articles we identify on Seeking Alpha and Motley Fool using our algorithm is similar to the Non-Fake articles' industry composition from the smaller sample of articles from the SEC.

3. Trading Activity and Price Impact

In this section, we examine the impact of fake news on trading activity and prices. Financial markets provide a useful setting to examine the impact of fake news 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. Price impact measures whether the news affects prices. It is very reasonable to find that (fake) news causes trading, but that trading has no impact on prices due to markets being informationally efficient with respect to these articles (Fama (1970)). In this case, fake news could cause excess trading but no price distortion. On the other hand, trade can be zero with substantial price movement (Milgrom and Stokey (1982)). We examine both.

3.1. Do Articles on the Platforms Have an Impact?

Before looking at fake articles, we first address whether articles posted in general (fake or non-fake) on these social platforms have any influence on market participants or markets. We start by examining abnormal trading volume around the publication of all articles. We focus on abnormal volume because we are interested in whether investors "react" to these articles. Of course, it is also quite plausible the reverse is true, that articles react to trading activity. To try and establish some causal interpretation, we examine *abnormal* or unexpected trading volume changes in the future. Abnormal trading volume is trading volume relative to expected volume in the stock, which we proxy for using the recent average daily volume in the stock (defined below). We look at future changes in abnormal volume in the days following the article's publication. A reverse causal story implies that authors are writing articles in anticipation of future *unexpected* trading activity. If true, we would call that "news." Nevertheless, we also control for lagged abnormal volume from the previous trading day and other news coming from the firm such as recent SEC filings and press releases about the firm.

Panel A of Table 2 examines the effect on abnormal trading volume from articles published on these sites. We define abnormal trading volume for stock i as $Vol(i,t)/\frac{1}{T}\sum_{k=20}^{140} Vol(i,t-k)$, which is the trading volume for stock i on day t relative to the average daily trading volume in stock i over the last 6 months (skipping a month).[15] We sum abnormal volume over days $t = 0$, $t + 1$, and $t + 2$, where $t = 0$ is the date the article appears on the website and then regress the natural logarithm of abnormal volume on an indicator variable for whether there is any article on these sites about the firm on a given day, regardless of its authenticity. We also 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.

As the first column of Panel A of Table 2 shows, an article published on Seeking Alpha or Motley Fool is associated with an 86% increase in abnormal trading volume over the three days following publication. This result implies that investors are trading in direct response to the articles or, more generally, are trading in response to whatever news is coming out that day that these articles are discussing. While the increase seems large, this first regression controls for no other variables, except time fixed effects. Moreover, as we show below, the bulk of the effect is concentrated in very small and illiquid firms, where trading volume changes (as a percentage) can be enormous, with outliers in the thousands of percents.

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

As we will show, articles on knowledge sharing platforms in the SEC-prosecuted cases are often written following press releases or SEC filings. In the second column of Panel A, we control for whether there is an SEC filing (10-K, 10-Q, or 8-K) or a company-issued press release in the three days leading up to the article. Furthermore, to control for serial correlation in abnormal trading volume, we include lagged abnormal trading volume on day $t-1$ as a regressor, which also captures other events we may be missing that could affect trading activity. After the controls, the effect on abnormal trading volume declines to 37%. To make sure these results are not all coming from the day the news is released, Table C2 in the internet appendix reports the effect on trading volume separately for the same day and for one and two days after the article's publication. Of the 37% rise in abnormal trading volume, 15.5% occurs on the day the article is published, 12.1% the following day, and 10.1% two days later. The abnormal volume following an article increases for about two weeks.

The next three columns of Panel A of Table 2 report results separately for small, medium, and large firms. Small firms are defined as smaller than the bottom 10^{th} percentile of NYSE firms, mid-size firms fall in the 20^{th} to 90^{th} size percentile of NYSE firms, and large firms are in the top 10^{th} size percentile of NYSE firms. The effect on abnormal trading volume declines strongly with firm size, with the effect six times larger for small firms than for large firms (80.9% increase versus 8.2% increase). This result is consistent with small firms having less volume and liquidity, less active large investors, and a more opaque information environment. In the last two columns, we separate firms into high and low retail ownership (above or below median retail ownership last month), since retail investors largely participate on these social platforms, and find that the effect on abnormal trading volume is twice as large for firms with high retail ownership.

The internet appendix reports some robustness tests of these results. First, Table C3 includes firm fixed effects to difference out any unobservable firm heterogeneity over the sample period – the results are nearly identical. Second, Table C4 controls for some outliers in percentage change in volume by winsorizing the most extreme five percent of abnormal volume observations. Since abnormal volume is the dependent variable, it is always questionable to winsorize, unless we think these extreme observations are data errors. The effects are obviously more muted from winsorizing but the patterns stay the same: a 33.3% increase in abnormal volume over the next three days that is larger for small firms (54.1%) and high retail ownership firms (41.7%). Our findings are not driven by a few extreme observations. The point estimate for the full sample is roughly the same, but the effect from winsorizing on the smallest decile of firms, where percentage volume changes are more extreme, reduces the effect from 80.9% to 54.1%. Given the similar patterns and the fact that we do not believe the extreme volume changes are errors, we do not winsorize observations.

Panel B of Table 2 repeats the regressions in Panel A, replacing abnormal volume as the dependent variable with the idiosyncratic price volatility of the stock. 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), which forms 125 equal-weighted portfolios based on $5 \times 5 \times 5$ sorts of stocks using size, value, and momentum characteristics that are related to expected returns. The dependent variable is the sum of daily idiosyncratic volatility on the day the article is published plus the next two days. This analysis captures whether articles moved prices around the days they were published. We examine price volatility as opposed to returns because it is exceedingly difficult to sign the direction of the content of the articles.[16] Hence, looking at volatility or the absolute value of returns captures whether prices moved significantly in relation to the articles published on that day. If the market has already incorporated the news, then the expected absolute return change should be zero.

[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. In unreported results, we attempted to use sentiment models to sign the news in the articles and found very unreliable effects. In Section 5 we look at signed returns for the SEC sample, which were largely fake articles positively promoting the stock, and hence know the sign/direction of the news.

As Panel B reports, we find effects similar to those for trading volume in Panel A – daily price volatility or the absolute return of the stock rises following articles published on these platforms, even after controlling for recent SEC filings, firm press releases, and return volatility in the days leading up to the article. The effect is strongest for smaller firms with higher retail ownership. 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 6.8% 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)).

In summary, the results in Table 2 show that the articles written on these platforms coincide with increased trading volume and larger price movement for the stocks mentioned in the articles several days after the article's publication. This evidence suggests that the articles are capturing investor attention and influencing behavior or that the articles coincide with other news that is influencing investor behavior and prices. In the next subsection, we investigate whether fake articles, which presumably are not informative, have a similar, weaker, or stronger effect on investor trading behavior and market prices.

3.2. Impact of Fake Articles

Do fake articles have an impact on trading volume and volatility? Since fake articles convey false information, informed investors should ignore them. Moreover, fake articles may be less likely to coincide with other real news. On the other hand, fake articles may be deployed precisely to embellish real news and mislead consumers. Panel A of Table 3 reports results from the same regressions as Table 2, but includes a dummy variable for whether the article is "for-sure fake" from our SEC sample. As the first column of Panel A shows, the for-sure fake articles are associated with significant increases in abnormal trading volume. The impact of fake articles on trading volume is significant, suggesting that fake articles, too, influence investor behavior. The coefficient on the Fake dummy represents the marginal effect on trading volume of a fake relative to a non-fake article, which indicates another 50% increase in trading volume over the next three days. This seems large, but is reasonable considering that the SEC ex post selected cases that had large impact and is in line with excerpts from the most recent SEC lawsuits.[17] The larger impact of fake articles may also indicate that they are different than non-fake articles along other dimensions as well that might also affect trading volume. Table 1 shows that fake articles differ on other linguistic characteristics (e.g., clout) that may also influence investor reaction. The second and third columns of Panel A interact the fake article dummy with the market equity decile of the company the article is written about and the retail ownership percentage of the firm. The impact on trading volume is larger for smaller firms and for firms with higher retail ownership. These cross-sectional firm results are consistent with where the articles are expected to have more impact (e.g., on retail investors) and hence are less likely to be driven by unobservable effects of news in general or trends in the market's response to news in the economy.

The first three columns of Panel B of Table 3 report results from the same regressions using idiosyncratic volatility as the dependent variable. Consistent with the abnormal trading volume results, fake articles have an additional significant impact on price movements relative to non-fake articles. The magnitude is large, too, which is not surprising since this sample is based off of the SEC investigations, which are more likely to focus on cases where the articles had price impact.

3.3. Using Probabilistically Fake Articles

The last three columns of Panels A and B of Table 3 report results using a dummy variable for whether the article is probabilistically fake using our calibrated probability function for fake news. As Panel A shows, the coefficient on the LIWC Fake dummy is not reliably

[17]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%."*

different from zero, indicating that (probabilistically) fake articles (as estimated by our algorithm) have the same impact on trading volume as all other articles. To be precise, this result implies that authenticity of the articles, as measured by linguistic cues, has no differential impact on volume. Contrasting this result with the bigger impact impact on volume we find for the for-sure fake articles, the difference is likely due to the SEC articles being a selected sample or to other attributes associated with the for-sure fake articles – sensational, dramatic, or other styles – that are not being modeled or captured by the LIWC authenticity score, which only looks at one particular aspect of linguistic style.

The next two columns of Panel A interact the LIWC Fake dummy with size deciles and retail ownership, which indicate significant effects on trading volume for the smallest firms with the highest retail ownership, consistent with the results hose from the narrower set of for-sure fake articles. Panel B shows that the probabilistically fake articles have a more muted impact on stock price movement, producing the same signed coefficient we get from the for-sure fake sample, but where nothing is statistically significant, suggesting that LIWC fake articles have similar impact on price movement. On the other hand, the same sign but lack of significance could be due to LIWC being a noisy measure of fake content and the difficulty in identifying fake news.

3.4. More Evidence on Direct Impact

To further test a direct link between articles published on these platforms and trading activity, we obtain a proprietary supplemental dataset from Seeking Alpha on 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 uploaded) and the number of times the article is "read," which is the instances in which a reader scrolled to the end of the article.[18] In total, the dataset covers 25,596 articles about 3,118 publicly traded firms.

Table C5 in the internet appendix presents results from regressing abnormal trading

[18]We obviously do not know if it was actually read, but scrolling through the article implies that some time was spent on it.

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 trading volume is positively related to the number of clicks and number of times the article is read by consumers, suggesting articles that influence circulation and readership are also associated with more trading activity in the stock. While causality is difficult to determine, and indeed both readership and trading activity may be driven by the importance of the news about the firm, this evidence more directly links the articles to future abnormal trading activity in the stocks the articles discuss.

The last two columns report results from regressions of the readership circulation variables on the fake article dummy to examine whether readership is affected by article authenticity. We find that fake articles are clicked more heavily and read more heavily, consistent with those articles also affecting trading volume more. Fake news seems to disseminate faster and more widely and impacts trading activity more. The larger influence of fake articles is likely due to other attributes also associated with fake news, such as being more sensational, more persuasive, and catering to the biases and priors of consumers. Indeed, Table 1 showed that fake articles may be different on multiple dimensions. The results are consistent with fake news propagating more diffusely through the network as suggested by Allcott and Gentzkow (2017) and Vosoughi et al. (2018b).

4. A Shock to Investor Awareness of Fake News

Our third empirical setting examines a different aspect of fake news to test another theoretical implication. We use the announcement of the initial SEC investigation into the promotional articles that comprise our first sample as an exogenous shock to the public awareness of fake news and examine the market's response to news before and after this shock. This exercise does not require being able to detect fake content.

4.1. Galena Biopharma Inc.

We begin with Galena Biopharma Inc., which was the first case prosecuted by the SEC for stock price manipulation on knowledge-sharing platforms. Galena encompasses the "event" which made the public aware of the existence of fake news. It also provides a micro-study of the direct impact these articles have on the stock's trading activity and prices as well as the motivation behind fake articles.

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.[19] Figure 3 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 PR companies Lidingo and DreamTeam to publish a series of promotional articles on third-party websites, like Seeking Alpha, that Galena paid for. The articles did not disclose the payments that the authors received, which violated the terms of Seeking Alpha and SEC regulation, and in some cases falsely claimed *not* to have received any payment. The lawsuit documents at least twelve promotional articles of this type. Appendix B contains an example of one of the fake articles written about Galena.[20]

Figure 3 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 fake articles are highlighted on the graph by the green boxes and often coincide with a bump in stock price on that day and a steady increase in price several days after. The motivation behind the scheme seems to have been a pump-and-dump campaign, 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

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

[20] This article and others like it that are part of the SEC investigation have been removed from Seeking Alpha. Searching for this fake article today, Seeking Alpha displays a message stating: "This author's articles have been removed from Seeking Alpha due to a Terms of Use violation."

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 3, where as news of insider sales broke, the stock price declined dramatically.

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 A reports the 8-K form documenting the settlement. By that point Galena's stock price had dropped to $2 a share.[21]

4.2. A Shock to Awareness of Fake News

The public revelation of the SEC's investigation and subsequent media attention around it provides a unique shock to investor awareness of fake news. We exploit the timing of the announcement to test additional implications of fake news.

In addition to the direct costs of individuals believing and acting upon false content, fake news can be costly if it damages people's trust in news generally and causes them to discount legitimate news (Allcott and Gentzkow (2017), Kshetri and Voas (2017), and Aymanns et al. (2017)). Our unique setting provides an opportunity to measure the potential spillover effects of fake news on people's trust in news. Using the revelation of the SEC investigation, we examine whether investors behaved any differently before versus after the event, when the presence of fake news on knowledge sharing platforms suddenly became salient to many consumers on these platforms.

[21] 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., among others. Furthermore, according to the SEC lawsuit, more than a hundred market makers facilitated trading in the company's stock.

4.3. Spillover Effects from Fake News

We use the period from February to March 2014 as the event that provides a shock to people's awareness of fake news. We examine the propensity of fake news and abnormal trading activity associated with articles six months prior to and six months after the event (August 2013 to January 2014 and April 2014 to September 2014, respectively).

Panel A of Table 4 first examines whether the propensity of fake news declines after the scandal. We regress a dummy variable of whether the article was probabilistically fake, on a dummy for 6 months after the SEC announcement event, controlling for SEC filings, firm press releases, and lagged abnormal volume in the days leading up to the article's publication. In addition, we include the number of news articles about the firm from the NYT and WSJ (obtained from Factiva). These controls help capture the amount of news occurring at the same time that are covered by the media. The coefficient on the post-scandal period is indistinguishable from zero, indicating that the prevalence of fake news, or more precisely the authenticity score of the fake articles, is similar before and after the scandal. However, this average result masks substantial heterogeneity. The next three columns separately report results for small, midsize, and large firms. The prevalence of fake articles about small firms falls significantly by 1.2% following the scandal. These results are consistent with small companies, who engage or were willing to engage in promotional articles before the scandal, ceasing or decreasing this activity after the SEC announcement.

Panel B of Table 4 examines the impact of all published articles on abnormal trading volume before versus after the scandal. The first column of Panel B reports results from a regression of abnormal volume on an article indicator, the 6-month post event indicator, and their interaction. The positive coefficient on articles confirms our earlier result from Table 2 that articles are associated with larger trading volume in the three days after they are published. The negative interaction term with the post-event dummy shows, however, that the effect of articles on trading volume decreases significantly after the scandal. This result is consistent with investors becoming aware of fake content and muting their trading response to news in general on these platforms. The strong negative coefficient on the post-event dummy indicates that abnormal trading volume declines by 17.9% after the scandal, and the interaction term with articles published on the platform is associated with another 4.1% decline in trading volume after the scandal. The results suggest that investors respond less to news on these platforms, including legitimate news, after the scandal and is consistent with consumers having less trust in the news once aware of the existence of fake news, as theory suggests (Allcott and Gentzkow (2017)). The economic magnitude of the effect is large – a 4.1% drop in trading volume associated with news articles after the scandal relative to before the event.

An alternative explanation for this spillover effect from the scandal is that news or reaction to news is simply lower in the post-scandal period for other reasons. In other words, the scandal happened to coincide with a time when news became lighter or investor reaction to news was more muted. To address this issue, we conduct several tests. First, we control for other news sources in the regression above to account for the amount of news. Second, we examine pre-trends in news and trading volume below and find no significant trends. Third, we conduct several placebo or falsification tests to rule out alternative explanations. Finally, we show cross-section effects across firms and article characteristics that are consistent with the awareness of fake news causing the decline in trading volume in response to news and less likely driven by trends in news or trading activity.

Figure 4 examines the daily abnormal trading volume response for one week before and four trading weeks after the article is published. We estimate the following model:

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

and plot the coefficient β_1 at the daily level, with 95% confidence error bars. The graph displays the average trading volume reaction to all articles after the scandal, and shows significant trading decreases on the day the article is published, and for the next two trading weeks, before eventually returning to pre-scandal levels. We see no trend in trading volume response to news before the event. These results suggest that investors' reaction to articles on these platforms decreases after the scandal and as a result of the scandal.

While the results in the first column of Panel B control for the level of SEC filings, press releases, other media (e.g., WSJ, NYT articles), and lagged abnormal trading volume in the days leading up to the article's publication, in the second column, we also interact the frequency of SEC filings, firm press releases, other news media, and changes in abnormal trading volume with the post-scandal dummy. The interaction terms serve as falsification exercises or placebo tests of the market responding to news on these platforms and the shock of fake news awareness on the platforms. In particular, an alternative explanation for the decline in trading volume in response to news after the scandal is that there is less information content, less news, or less firm activity in the post-event period that happened by chance to coincide with the timing of the SEC announcement. If so, then interactions between corporate filings, press releases, and other media news should be negative as well. As the table shows, however, the interaction terms are negligible and insignificant, and two out of three have the wrong sign to be consistent with this alternative story. The magnitudes of these interactions with the post-event dummy are trivially small – 0.2% increase in trading volume response to SEC filings, 0.3% decrease to press releases, and 0.6% increase to articles in the NYT and WSJ – none of which are reliably different from zero. We find no discernible difference in firm news or activity before versus after the scandal and no reliable difference between the trading volume response to WSJ or NYT articles before versus after the scandal, despite the fact that all three of these media sources themselves do have a significant impact on trading activity generally. For example, SEC filings, press releases, and newspaper articles increase abnormal trading volume by 13%, 29%, and 12.5%, respectively, yet after the scandal we find no difference in response to these other sources of news.

The drop in trading volume associated with published articles on these social media platforms is likely a reduced response from investors to news specifically coming from these platforms, and not press releases, other public filings, or other media, and not any market trends in information production or lower trading activity. The evidence is most consistent with investors discounting all news on these platforms, even legitimate news, after the scandal due to increasing distrust of content from these platforms after the SEC revealed the existence of some fake articles. The magnitude of the drop in abnormal volume is even larger and more significant after accounting for the other activity post-scandal, decreasing volume by an additional 7.5% per article after the event. These findings provide some of the first evidence on the indirect spillover effects of fake news on news in general, as conjectured by theory (Allcott and Gentzkow (2017)). As we will show below, consumers were largely unaware of and unable to detect fake news, consistent with their response to discount all news on the platforms following their awareness of fake content.

Columns 3 through 7 of Panel B report results separately for small, medium, and large firms, as well as for firms with high and low retail ownership. Consistent with our previous results, these effects are all much stronger for smaller firms, and firms with high retail ownership. Post scandal, the abnormal trading volume associated with articles published on these platforms drops by 35% for the smallest firms. Interestingly, even though few fake articles are written about large firms and none of the articles in the SEC probe pertained to large firms, abnormal trading volume still declines by 11.7% for each published article about large firms that appeared on these platforms after the scandal, despite nearly all of these articles being authentic. This result provides further evidence of a spillover effect from fake news to other legitimate news content. Stocks with high retail ownership have a 30.3% drop in abnormal volume post-scandal compared to only a 9.5% drop for low retail ownership stocks. Since retail investors tend to dominate participation on these sites, this result provides a more direct link to these platforms influencing trading activity. These cross-sectional results are also less consistent with alternative explanations of trends in news or investor activity, which would have to be a more complicated story to accommodate these facts.

4.4. Generalizing Spillover Effects

The spillover effect from the awareness of fake news to all news, including legitimate news, is interesting and consistent with theory (Allcott and Gentzkow (2017)). The result begs the question: How broadly does the awareness of fake news from the scandal affect investors' response to news generally? Was the spillover response merely contained to similar articles on Seeking Alpha, where many of the promotional articles the SEC investigated resided, or did it impact news from other sources? We find a significant decline in response to news for all articles published on the social platforms where fake news originated, but do not find commensurate decrease in reaction to other news sources such as the WJS or NYT. The evidence is consistent with investors discounting all news on the social platforms, but recognizing or believing press releases, the WSJ, and NYT, are less subject to fake news. Alternatively, the average investor who trades on press releases may be different than the average investor who trades on news from these social platforms.[22]

While the indirect effects of fake news on these platforms do not seem to spillover to press releases or the WSJ or NYT, we find that they do spillover to other similar media outlets that were not part of the scandal. Specifically, as a test to generalize the spillover effect from fake news onto other news more generally, we examine the trading response after the scandal for articles on the Motley Fool platform only. Motley Fool was not part of the fake news scandal, and none of its articles were flagged for failing to disclose paid-for content as part of a promotional scheme or were investigated by the SEC. Hence, we examine whether the spillover effect from the scandal, contained largely on Seeking Alpha, also had an effect on the trading volume response to articles published on Motley Fool, a similar shared-knowledge platform that was not part of the investigation. This analysis helps measure the scope of the spillover effect from the awareness of fake news and whether similar social platforms that were not part of the scandal, but likely share the same readership, were also affected. The

[22]For these reasons, we think the corporate filings, press releases, and newspaper article interactions provide compelling falsification tests that support our main findings.

last column of Panel B of Table 4 reports the results and shows that abnormal trading volume also declines significantly for Motley Fool articles after the scandal. The result points to the spillover effect from the scandal extending beyond the specific platform where the scandal occurred. The awareness of fake news seems to impact other related news sources – in this case a competitor shared-information platform where the scandal did not occur. But, as our previous analysis shows, it does not have an impact on very different news sources such as press releases or newspaper articles. These results make sense if investors simply discount all social news as a result of the scandal but think other news sources are more immune to false content, or if the set of investors who consume social news is simply different from those who consume other news sources.

Panel C of Table 4 reports the results from the same regressions as in Panel B, but uses idiosyncratic volatility as the dependent variable instead of abnormal trading volume. The results are consistent with the trading volume findings, where there is significantly reduced impact on price volatility from articles after the information shock from the scandal, especially for small firms with high retail ownership. We also find opposite-signed price movement for press releases, SEC filings, and other news media after the scandal, suggesting other trends or omitted variables in the market's response to news in general are not driving the results. We further find that price movements from the Motley Fool articles are also consistent with the reaction to articles on Seeking Alpha after the scandal, providing further evidence of a spillover effect to other shared-information platforms. These findings are consistent with markets discounting news from these platforms after revelation and awareness of fake news.

4.5. More Direct Evidence of Spillover Effects

To strengthen the story, we provide some additional evidence that the decline in trading volume response to articles, and spillover decline in volume for non-fake news after the scandal, is due to investors being made aware of fake news. Specifically, we examine the posted comments to the articles published on these sites in the six months before and after the scandal. In the comments section pertaining to each article, we add up the mention of the words "fake" or "fraud" and compute a dummy variables (*Fake Words*) equal to one if readers use these words. We then regress the frequency of *Fake Words* on a dummy for fake articles as well as a dummy for the six-month period after the scandal.

To test an alternative hypothesis, we also compute the frequency of the words "wrong" or "not right" from the comments section and create a dummy variable *Wrong Words*, which is equal to one if readers use these words in their comments. This variable helps distinguish between erroneous or inaccurate information from fraudulent or deceptive information. The distinction is subtle because it relies on intent. The comments section provides a glimpse of what consumers may be concerned about.

Panel A of Table 5 examines whether the appearance of *Fake Words* or *Wrong Words* is more prevalent for fake versus non-fake articles over the entire sample period. We regress the prevalence of *Fake Words* on the fake article dummy in the first column and find that the words "fake" or "fraud" are not used more frequently with fake articles. This null result suggests that participants on these platforms could not identify or differentiate between fake and non-fake articles. In our setting, participants on these platforms were deceived by these articles with no indication that consumers were skeptical or aware of fake content.

The second column of Panel A runs the same regression using *Wrong Words* as the dependent variable. Here, there is a strong negative association between fake articles and use of the words "wrong" or "not right" in the comments section. This result suggests that investors feel the fake articles are more accurate (less wrong) than the non-fake articles. Fake articles seem to be more convincing of their statements than the non-fake articles, which may be why they generate more trading volume.

Panel B runs similar regressions using the *Post Event* dummy instead of the *Fake Article* dummy, where the post-event dummy is the six-month time period after the scandal. Interestingly, after the scandal, the incidence of the words "fake" and "fraud" *increased* significantly (t-statistic of 2.73), implying that participants on these platforms were indeed more concerned with or commented more about false content on these sites after the scandal. This evidence corroborates the decline in trading volume witnessed post-scandal for all articles and suggests general mistrust of news from these platforms. The use of "wrong" words is no more prevalent after versus before the scandal. Hence, after the SEC announced investigation and subsequent exposé articles, participants on these platforms seemed more concerned with fake news.

Combining the results in Panels A and B of Table 5, the evidence paints a picture of investors and consumers of information on these platforms being largely unaware of fake news before the SEC investigation and then suddenly becoming aware after the scandal, but having no ability to differentiate or detect which articles are fake and not fake. As a consequence, we see a marked drop in investor trading volume to all articles published on these sites, regardless of their authenticity, creating a significant spillover effect from the revelation of the existence of fake news on legitimate news.

4.6. Variation in Article Characteristics

To further examine the link between the articles and trading volume, we examine whether authors who have more followers, and have written more articles, have a bigger impact, as well as whether articles that receive more comments lead to greater trading volume. We also analyze whether the trading volume reaction is higher when the article is more quantitative in nature and/or references accounting data, where presumably it is less likely to be false since numbers, such as earnings, can be verified from other sources. We look at the fraction of the article text comprised of numbers, as well as the number of words that have "earn" as part of the word. We regress abnormal trading volume for stock i on the number of followers an author has, log number of comments an article received, the number of past articles the author has written, and the fraction of numbers that appear in the text as well as the fraction of mentions of "earn." We also control for whether there was at least one SEC filing and one firm-issued press-release in the three trading days leading up to the publication date, and control for abnormal trading volume the day before the article's publication. The results are presented in Panel A of Table 6. We find that articles by authors with more influence as well as articles that get more comments are associated with a larger impact on trading volume. Furthermore, articles that seem to be more quantitative, also have a bigger impact.

Next we examine whether these characteristics have an impact on how the SEC scandal affected the trading volume reaction to the article. In particular, we rerun the regressions from Panel A, examining the time period six months before and six months after the SEC scandal. The results are presented in Panel B of Table 6. Similar to earlier analysis, we find that the trading volume response to articles is lower in the post-period. However, the drop is not as large if the author has more followers and has written more articles in the past. This suggests that people's trust in the articles decreases less for authors that have a better reputation. We further find that the decrease is not as large for articles that mention "earn" in the article, suggesting that articles that cover accounting-related and hard information are not discounted as much after the scandal. These results are consistent with the decline in trading volume after the scandal being a response to distrust in news from these platforms. The types of articles that garner more trust (author reputation) or maybe easier to verify (use of hard information) are not discounted as much.

5. What Motivates Fake News?

Finally, we investigate what might be motivating the fake articles on these platforms. Using the Galena case that launched the broader SEC investigation, we examine whether other cases have similar characteristics and motivations to better understand the existence and prevalence of fake news. This analysis serves several purposes: it may help us better quantify the economic impact of fake news, provides another test of the linguistic algorithm's ability to detect fake content, and may help identify other fake news.

5.1. Firm Performance

We start by examining the price reaction to the other for-sure fake articles from the SEC to see if a similar pattern as Figure 3 for Galena exists for the other firms involved in the scandal. We conduct the flip-side of the classic event study in financial economics (Fama et al. (1969)) by examining the return response to false news. This exercise is a novel test of the informational efficiency of markets, where in a perfectly efficient market fake news should have no impact on prices, regardless of the underlying equilibrium asset pricing model. We separate firms by size into small and non-small (there are no large firms in this sample) and examine their return response to the release of for-sure fake articles, by plotting the cumulative abnormal returns, measured as the difference between the return of the stock and a matched-portfolio of similar stocks (one of 125 equal-weighted portfolios based on size, book-to-market equity, and momentum), for days $t + 1$ to $t + 251$ after a fake article appeared about the firm.

Figure 5 plots the difference between cumulative abnormal returns for the for-sure fake articles, relative to days with non-fake articles. Returns for small firms increase after the fake article is published (relative to non-fake articles), reaching as much as 13% cumulatively, after about 60 days, before giving up all the gains and ending with a cumulative negative 5% return after a year. This pattern matches that of Galena in Figure 3. The permanent price impact of -5% for small firms indicates either that once the market figures out the news is fake, investors view this as a bad signal about the firm or that the true price should have dropped by 5% initially, but the fake news temporarily propped up the price and delayed the decline. For non-small firms, the price starts dropping immediately after the fake article comes out and continues to decrease throughout the year. This result could be consistent with the market figuring out the news is fake immediately for larger firms, where the cost of information is lower, or that the returns would have been even worse had the firm not initiated the fake articles.

We next examine the market price response to articles that we classify as probabilistically fake using the linguistic algorithm on the larger universe of all articles on these platforms. Since our analysis is at the firm-day level, we define whether a firm had a fake article on a given day using the average probability of being fake of all articles written about the firm on that day. Figure 5 plots the difference between abnormal cumulative returns following days with (probabilistically) fake articles, relative to days with (probabilistically) non-fake articles, and plots the responses separately for small and non-small firms in our sample (that have at least one fake article). As the figure shows, among small firms, returns following fake articles relative to non-fake articles increase for 6 months by about 5% following publication, and then revert back to their original level. These patterns are remarkably similar to the return patterns we found for the for-sure fake articles from the smaller SEC sample, further supporting our algorithm's ability to identify fake content. The magnitudes are, not surprisingly, much smaller here since, unlike the SEC example, we identify fake news with noise plus the SEC is likely to go after the most extreme cases, so there may be selection bias in the first sample. For larger firms, we find nothing, which makes sense since the market is more efficient for larger firms who also are less likely to engage in promotional campaigns. The lack of results on larger firms is another useful falsification exercise. We formally test whether the patterns in cumulative abnormal returns for fake news articles about different-sized firms over different horizons are statistically significant in Table C6 in the internet appendix. We find statistically significant results for small firms and no impact for larger firms.[23]

5.2. Other Firm Actions

Fake news is designed to deceive for financial or personal gain, including perhaps the utility of fooling people and/or influencing others. In our setting of financial markets, it seems less likely that private utility benefits motivate fake news. The SEC investigation focused on promotional articles as part of pump-and-dump schemes to defraud securities markets. Our findings on the impact on abnormal trading and temporary prices are consistent with

[23]One question is whether the poor long-term returns to small firms that promote fake articles are due to investors' over/under reaction or whether fake articles are a sign of poor fundamental firm performance. Table C7 in the internet appendix shows that the presence of fake articles is associated with worsening fundamental firm performance, as measured by surprise in unexpected earnings, the return on assets, and its recent quarterly change. These findings are consistent with a possible motivation for engaging in promotional campaigns for financially troubled small firms that include hiring fake articles to prop up the stock price.

a motivation to hire authors to write fake content to promote the stock. Consistent with this motive, Table C8 shows that these firms are more likely to issue press releases and 8-K filings within the same week to accompany the fake articles, perhaps to give authors of the fake articles more material and credibility and to influence the narrative of the firm. We also find in Table C9 that insider trading coincides with the fake articles and is positioned to profit from the price movement caused by the promotion.[24] While these actions are rampant among the SEC-prosecuted sample, we also find similar evidence for our broader sample of articles, where we probabilistically assess the occurrence of fake news using the linguistic algorithm. Consistent with our earlier results, we find these effects, too, to be predominantly contained among small firms.

The evidence suggests that an improved method for detecting fake content may involve examining other actions taken by the firm in addition to textual analysis. As one example, when we combine the probability of fake articles with the dual presence of insider trading to benefit from stock promotion, we find sharper price impact patterns. In addition, fake articles published following insider purchases are preceded by very sharp drops in share price in the month before publication, whereas fake articles not associated with insider purchases have flat to lightly increasing returns before publication. However, even for firms with fake articles written about them that do not have insiders buying shares, there is still a small price increase that also turns negative after several months, suggesting the results are not just driven by insider trading. In addition, performing a similar analysis using only the non-fake articles, there is no difference in returns for non-fake articles with insider buying versus without insider buying. Hence, it is not insider buying per se that drives the returns. Rather, it is the combination of insider buying with fake articles that seems to matter most and is indicative of a comprehensive promotional campaign that motivates the production of fake news in our setting.

[24]We obtain data on press releases from RavenPack from 2001 to 2015, 8-K disclosure filings from the SEC's Edgar database, and insider trades from Form 4 from Thomson Reuters.

6. Conclusion

We study three empirical settings to assess the economic impact of fake news: a unique dataset of fake paid-for articles on financial media crowd-sourced platforms prosecuted by the SEC, a broader set of articles on these platforms that we apply a linguistic algorithm to detect fake content, using the first sample of known fake articles to verify and calibrate the algorithm, and the SEC's announced investigation that provides a shock to the public's awareness of fake news. We find that fake news increases abnormal trading volume and imposes temporary price impact on small firms. Following public revelation of the existence of fake news, we find a significant spillover effect to news generally, where investors react less to *all* news, even legitimate news on these platforms. These findings represent some of the first documented direct and indirect effects of fake news that are consistent with theory (Allcott and Gentzkow (2017), Aymanns et al. (2017), and Kshetri and Voas (2017)).

Our study provides evidence on the prevalence and effect of fake news from crowd-sourced information platforms that continue to grow and gain attention. Financial markets may provide a lower bound on the impact of disinformation in other settings, where information costs are higher and where the ability to take action to correct its distortions is more limited (e.g., online consumer retail, political news, elections, and social media). More broadly, our findings may have more general implications for news media (e.g., Gentzkow and Shapiro (2005) and Gentzkow et al. (2015)) and for trust and social capital (e.g., Guiso et al. (2004), Guiso, Sapienza, and Zingales (Guiso et al.), Guiso et al. (2010), and Sapienza and Zingales (2012)).

References

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*.

Boudoukh, J., R. Feldman, S. Kogan, and M. Richardson (2018). Information, trading, and volatility: Evidence from firm-specific news. *The Review of Financial Studies*, hhy083.

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.

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.

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

Fama, E. F., L. Fisher, M. C. Jensen, and R. Roll (1969). The adjustment of stock prices to new information. *International Economic Review 10*(1), 1–21.

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

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.

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

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

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 (2010, March). Civic capital as the missing link. Working Paper 15845, National Bureau of Economic Research.

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

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

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.

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.

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.

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

Silverman, C. (2016). This analysis shows how viral fake election news stories outperformed real news on facebook. *BuzzFeed News 16*.

Silverman, C. and L. Alexander (2016). How teens in the balkans are duping trump supporters with fake news. *Buzzfeed News 3*.

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.

Timberg, C. (2016). Russian propaganda effort helped spread ‘fake news’ during election, experts say. *Washington Post 24*.

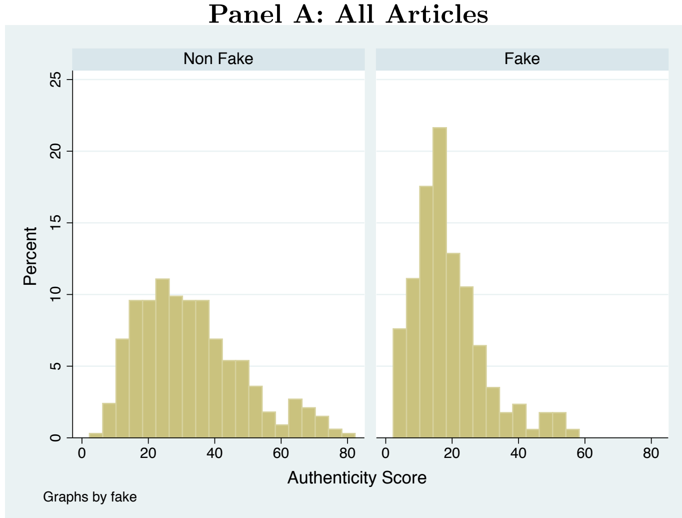
Vosoughi, S., D. Roy, and S. Aral (2018a). The spread of true and false news online. *Science 359*(6380), 1146–1151.

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

Figure 1. **Authenticity Scores**

This figure depicts the distribution of authenticity scores for fake and non-fake articles. In Panel A, we plot authenticity scores for all the articles in our validation sample of 171 fake and 334 non-fake articles. In Panel B, we plot authenticity scores for two authors in our validation sample with the most articles

Panel A: All Articles



Panel B: Author-Specific

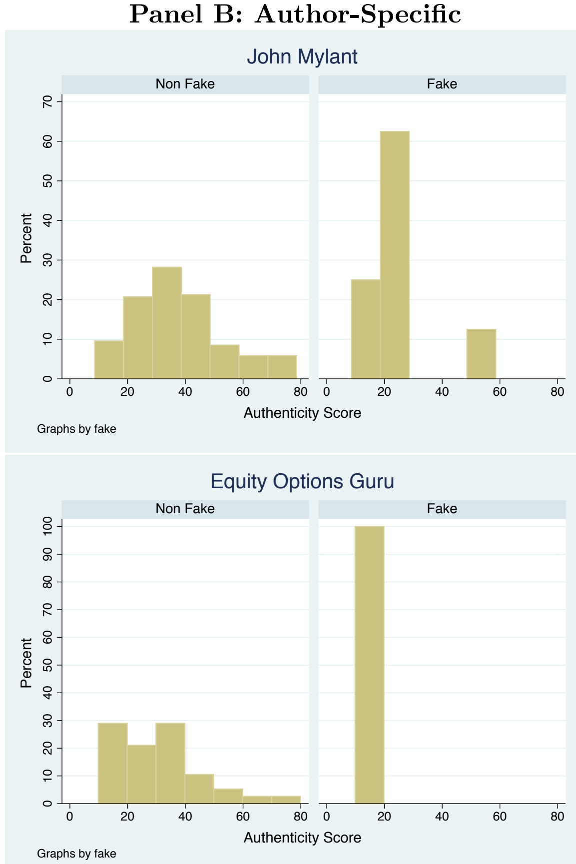


Figure 2. **Authenticity score and the probability of being fake**

This figure depicts the relationship between LIWC authenticity scores (S) and the conditional probability of being fake ($Prob(F|S)$).

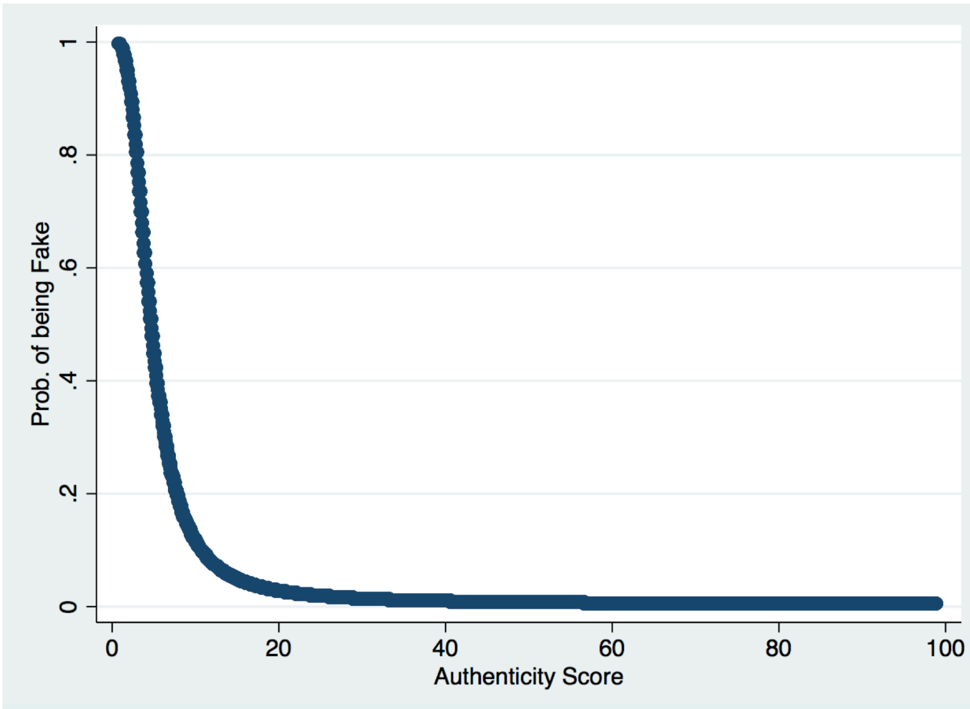


Figure 3. **Example of a Pump-and-Dump Scheme: Galena Biopharma Inc.**

This figure depicts the stock price of Galena Biopharma Inc. from April 2013 - 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 of in early January. During the board meeting on January 16, 2014, where the board reviewed the preliminary 2013 earnings which have 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 has been 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 an SEC probe.

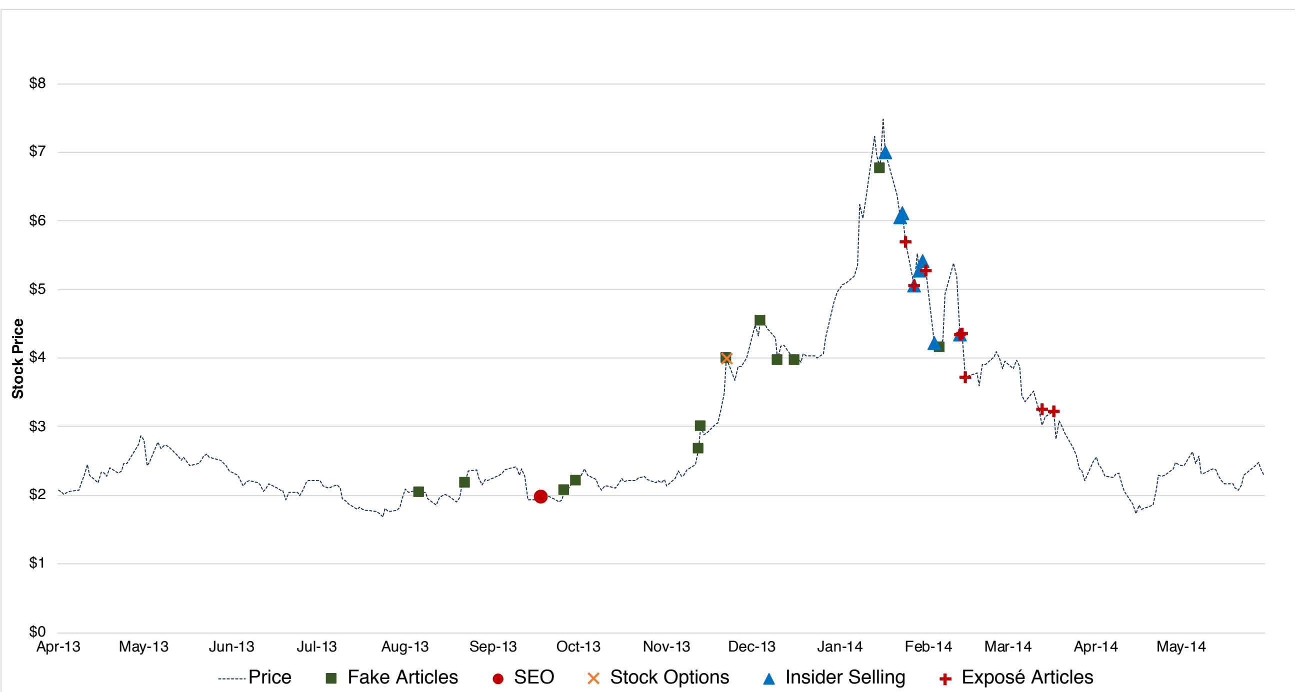


Figure 4. **2014 SEC Lawsuit**

This figure examines whether the salience of the presence of fake news on the platforms, proxied for by the timing of SEC Lawsuit disclosure, had an impact on investors' reaction to the articles (proxied for by log of abnormal volume on days $t = 0$, $t+1$, and $t+2$). We plot the difference in reaction of abnormal volume to articles in the 6 months before and 6 months after the disclosure period. For the graph we estimate the following model:

$$Log(AbVol)_t = \alpha + \beta_1 Article \times 6MonthAfter + \beta_2 Article + \beta_3 Post - Event + Controls + \epsilon$$

where $Log(AbVol)_t$ is the abnormal trading volume 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 1 April - 30 September, 2014. The estimation period is 1 August, 2013 - January 31, 2014, and 1 April - 30 September, 2014. In the figure we graph the estimates of β_1 for four trading weeks after the article was published. The bars represent 95% confidence bands around the point estimates.

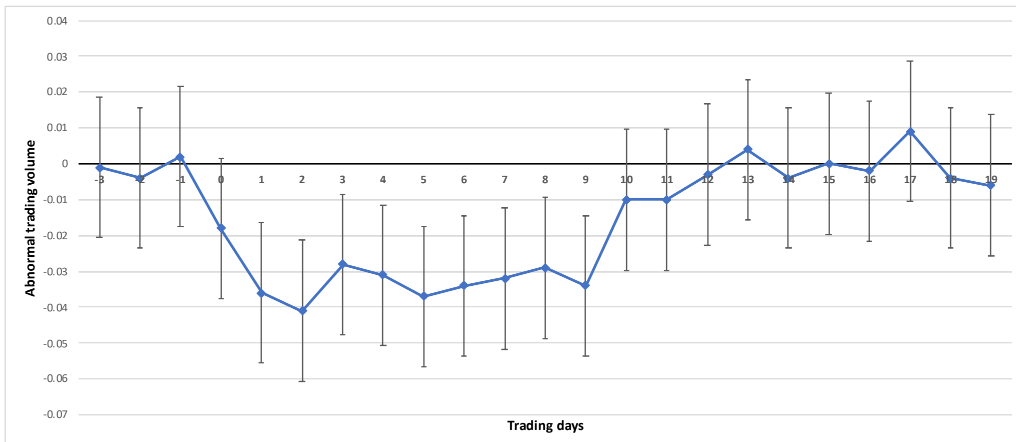
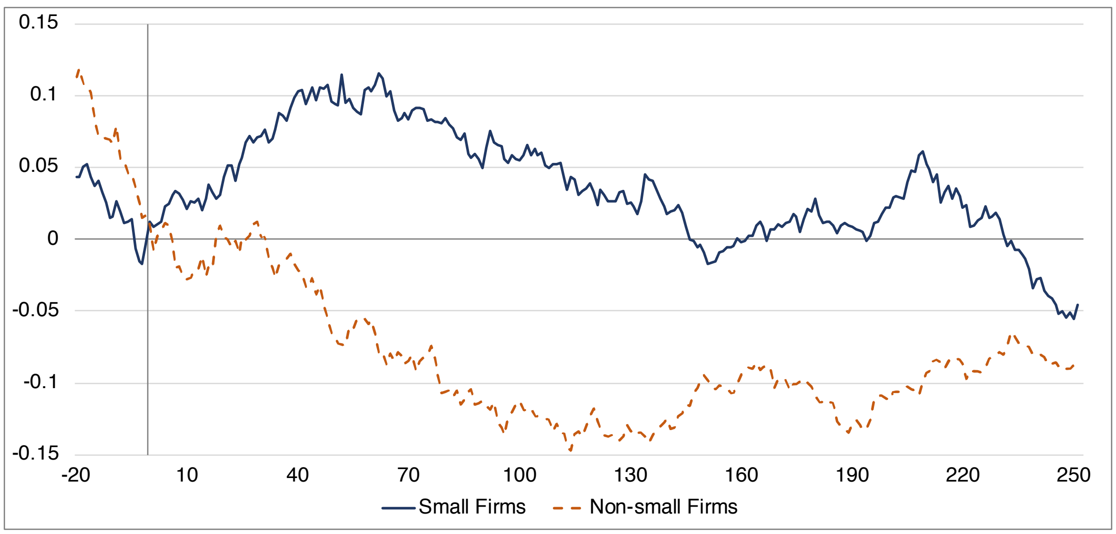


Figure 5. **Abnormal Returns for For-Sure Fake Articles**

The figure depicts difference in cumulative abnormal returns (measured as equal-weighted 4-factor residuals) between days with fake articles and days with non-fake articles separately for small and non-small firms in our sample. In Panel A, fake news are for for-sure fake articles that were provided to us by Rick Pearson and that were subpoenaed by the SEC. In Panel B, we designate a given day t for company i to have a fake article, if the probability of being fake, associated with the average authenticity score for all articles about firm i on day t, is greater than 20%. Similarly, we designate a day t for company i as not having any fake articles, if the probability of being fake, associated with the average authenticity score for all articles about firm i on day t, is less than 1%. The cumulative returns are measured starting with the day after the article was published until the 251 trading days after the article was published. For the time period before the article was published we measure cumulative returns starting with the day -20 and ending on the day before the article publication. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, non-small firms are defined as firms above the 10th percentile of NYSE firms.

Panel A: For-Sure Fake Articles from the SEC



Panel B: Probabilistically Fake Articles from LIWC Algorithm

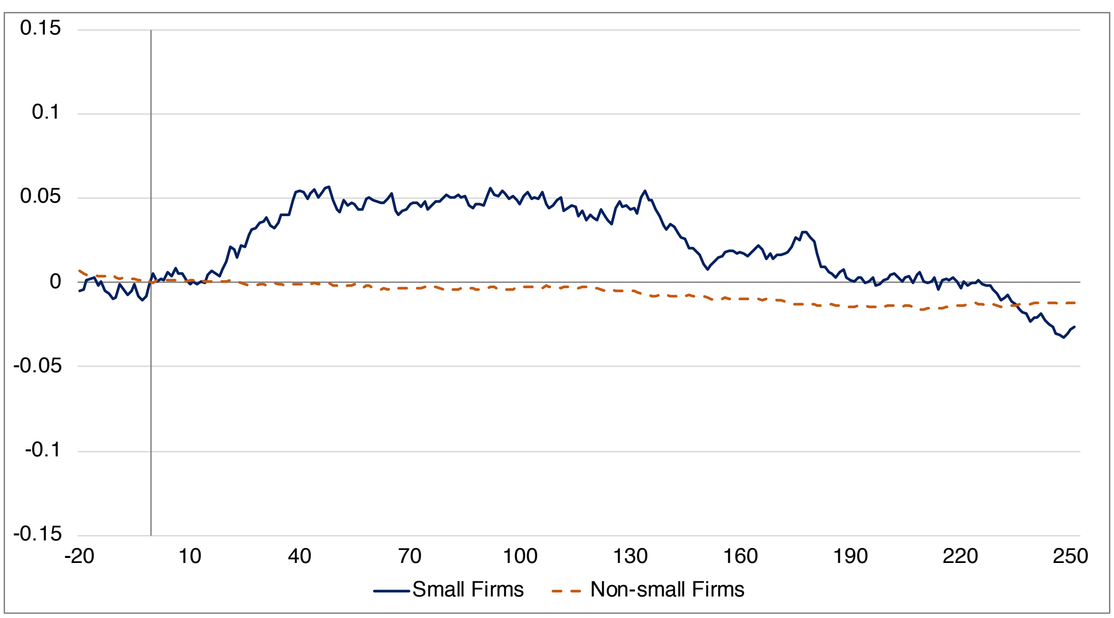


Table 1. **Summary Statistics**

This table presents the summary statistics for various LIWC textual measures and firm characteristics of the covered firms, for different types of articles on Seeking Alpha and Motley Fool. *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. *Seeking Alpha Articles* and *Motley Fool 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 our main analysis. In Panel A, we display the number of articles in each category as well as the mean of the *Authenticity* measure that we use to construct the probabilities of being fake. We also report the means of several other variables provided by LIWC to help better understand the authenticity score. In particular we display the means of the average of the *1st person singular* measure (examples: I, me, mine), *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. In Panel B, we display the average probability of being fake, for each of the article categories. In Panel C, for the firms that are covered in the respective article groups, we provide the average fraction of retail investors, the average number of analysts covering the firm, and the average firm size (in Millions of dollars). The differences between Fake and non-Fake article measures that are statistically significant at the 5% level, when we include author fixed effects, are marked in bold.

	Rick Pearson & SEC		Seeking Alpha			Motley Fool		
	For-sure Fake	Non Fake	Fake	Non Fake	Other	Fake	Non Fake	Other
Panel A: LIWC variables								
Number of articles	171	334	3,933	116,289	83,323	1,368	78,943	67,605
Authentic	**19.09**	**32.79**	**5.44**	**50.71**	22.51	**5.71**	**46.75**	21.96
1st pers singular	0.42	0.76	**0.25**	**0.98**	0.54	**0.20**	**0.53**	0.23
Words per sentence	57.55	65.23	**23.89**	**21.76**	22.18	**31.23**	**19.28**	19.39
Insight	1.52	1.67	**1.43**	**1.75**	1.63	**1.62**	**2.08**	1.84
Relativity	**12.92**	**15.11**	**9.90**	**17.37**	13.53	**9.20**	**16.57**	13.29
Time	4.97	5.35	**3.40**	**6.34**	4.68	**3.34**	**6.54**	5.23
Discrepancy	1.41	1.05	**1.40**	**1.12**	1.22	**0.76**	**1.08**	1.11
Clout	**58.25**	**52.31**	**62.04**	**52.84**	57.06	**72.40**	**60.83**	63.99
Analytic	**94.65**	**90.97**	**93.01**	**91.85**	92.72	**88.86**	**89.60**	90.69
Tone	61.79	57.49	**55.02**	**57.19**	58.13	**75.38**	**59.85**	60.16
Panel B: Probability of being Fake								
Prob(Fake)	1.00	0.00	0.45	0.01	0.03	0.42	0.01	0.03
Panel C: Firm characteristics								
Percent of retail investors	76.66%	50.15%	42.32%	42.46%	44.96%	40.88%	36.78%	38.99%
Numer of Analysts	6.96	16.76	16.83	18.33	16.67	23.21	19.84	20.34
Firm Size ($Mil)	7.36	58.43	44.12	51.72	45.17	101.97	70.58	80.40

Table 2. **Effect of articles on trading volume and volatility**

The table examines whether investors react to articles posted on Seeking Alpha and Motley Fool (proxied for by log of abnormal volume on 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. *Article* is a dummy variable equal to 1 if there was at least one article published about the firm on a given day. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. Firm size is measured in the prior trading month. High retail ownership is defined as being above the median in the prior trading month, and low retail ownership is defined as being below the median. In Panel A abnormal volume is defined as $Vol(t)/AvgVol(t-146, t-20)$, summed over days $t = 0$, $t+1$, and $t+2$, and then we take the natural log of the sum. We control for the natural log of abnormal trading volume on the day $t-1$. In Panel B return volatility is defined as abnormal returns squared, summed over days $t = 0$, $t+1$, and $t+2$. We control for return volatility on days $t-3$ to $t-1$. We include year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively, with t-statistics in parentheses.

Panel A: Effect on Abnormal Trading Volume From All Articles

	Log($[t, t+2]$ day abnormal volume)						
			Firm Size			Retail ownership	
	All	All	Small	Medium	Large	High	Low
Article	0.856***	0.377***	0.809***	0.263***	0.082***	0.511***	0.255***
	(191.71)	(105.42)	(53.56)	(77.12)	(22.80)	(67.00)	(74.87)
SEC Filing$_{t-3,t}$		0.106***	0.104***	0.081***	0.031***	0.055***	0.106***
		(70.11)	(29.65)	(57.70)	(8.84)	(17.84)	(67.41)
Press Release$_{t-3,t}$		0.275***	0.358***	0.174***	0.015***	0.285***	0.207***
		(184.88)	(89.35)	(130.29)	(5.44)	(89.66)	(146.41)
Log(AbVol)$_{t-1}$		1.464***	1.369***	1.487***	1.415***	1.406***	1.385***
		(2730.59)	(1553.37)	(2108.70)	(516.12)	(1669.71)	(1620.83)
Observations	13,951,256	13,890,419	5,213,107	8,114,607	562,705	5,690,068	5,726,556
R-squared	0.024	0.370	0.337	0.392	0.438	0.348	0.360
Year-Month FE	X	X	X	X	X	X	X

Panel B: Effect on Return Volatility From All Articles

	Volatility on days [t,t+2]						
			Firm Size			Retail ownership	
	All	All	Small	Medium	Large	High	Low
Article	0.076***	0.068***	0.997***	0.208***	0.023***	0.162***	0.079***
	(3.91)	(3.50)	(8.65)	(25.50)	(21.42)	(4.62)	(14.56)
SEC Filing$_{t-3,t}$		0.057***	0.138***	0.031***	0.004***	0.103***	0.040***
		(7.01)	(5.60)	(9.44)	(3.94)	(7.61)	(15.90)
Press Release$_{t-3,t}$		0.013*	0.183***	0.056***	-0.003***	0.187***	0.063***
		(1.68)	(6.56)	(17.98)	(-3.64)	(13.26)	(28.26)
Volatility$_{t-1}$		0.652***	0.550***	2.336***	23.879***	0.499***	3.924***
		(35.16)	(17.58)	(59.69)	(176.32)	(26.97)	(90.07)
Observations	10,617,750	10,617,750	3,566,554	6,517,737	533,459	3,801,741	5,260,212
R-squared	0.001	0.001	0.002	0.004	0.136	0.002	0.010
Year-Month FE	X	X	X	X	X	X	X

Table 3. **Effect of fake versus non-fake articles on trading volume and volatility**

The table examines whether investors react differently to fake articles posted on Seeking Alpha and Motley Fool (proxied for by log of abnormal volume on 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. *Article* is a dummy variable equal to 1 if there was at least one article published about the firm on a given day. *For Sure Fake* is a dummy equal to 1 if the article is one of the 171 articles identified by Rick Pearson or the SEC as fake. *LIWC Fake* is a dummy equal to 1 if the probability of the average article about the firm on a given day being fake is $> 20\%$, 0 if the probability of an average article being fake is $< 1\%$, and missing otherwise. We also include controls for SEC filings, press releases, and abnormal volume on day $t-1$. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. Firm size is measured in the prior trading month. High retail ownership is defined as being above the median in the prior trading month, and low retail ownership is defined as being below the median. In Panel A abnormal volume is defined as $Vol(t)/AvgVol(t-146, t-20)$, summed over days $t = 0$, $t+1$, and $t+2$, and then we take the natural log of the sum. We control for the natural log of abnormal trading volume on the day $t-1$. In Panel B return volatility is defined as abnormal returns squared, summed over days $t = 0$, $t+1$, and $t+2$. We control for return volatility on days $t-3$ to $t-1$. We include year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively, with t-statistics in parentheses.

Panel A: Effect on Abnormal Trading Volume from Fake versus Non-Fake Articles

	Log$[t, t+2]$ day abnormal volume					
	For Sure Fake			LIWC Fake		
Article	0.376***	0.176***	0.340***	0.377***	0.177***	0.340***
	(105.28)	(49.14)	(94.07)	(104.75)	(48.99)	(93.63)
Fake	0.502***	1.051***	0.594	-0.004	0.448***	-0.097*
	(3.78)	(6.09)	(1.12)	(-0.12)	(6.44)	(-1.67)
ME decile		0.007***			0.007***	
		(393.51)			(393.54)	
Fake × ME decile		-0.012**			-0.007***	
		(-2.15)			(-7.17)	
% Retail Inv.			-0.362***			-0.362***
			(-222.17)			(-222.16)
Fake × % Retail Inv.			0.701			0.260**
			(1.08)			(2.23)
Observations	13,890,419	13,890,419	11,416,624	13,890,419	13,890,419	11,416,624
R-squared	0.370	0.377	0.358	0.370	0.377	0.358
Year-Month FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X

Panel B: Effect on Return Volatility from Fake versus Non-Fake Articles

	Volatility on days [t,t+2]					
	For Sure Fake			LIWC Fake		
Article	0.066***	0.282***	0.092***	0.068***	0.284***	0.094***
	(3.40)	(14.32)	(7.37)	(3.48)	(14.32)	(7.49)
Fake	2.968***	2.739***	9.553***	-0.010	0.470	-0.153
	(3.97)	(2.80)	(5.22)	(-0.06)	(1.14)	(-0.77)
ME decile		-0.007***			-0.007***	
		(-74.59)			(-74.58)	
Fake × ME decile		-0.005			-0.007	
		(-0.17)			(-1.26)	
% Retail Inv.			0.462***			0.462***
			(77.73)			(77.73)
Fake × % Retail Inv.			-6.980***			0.372
			(-3.08)			(0.91)
Observations	10,617,750	10,617,750	9,061,953	10,617,750	10,617,750	9,061,953
R-squared	0.001	0.002	0.003	0.001	0.002	0.003
Year-Month FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X

Table 4. **2014 SEC Lawsuit**

In Panel A of this table we examine whether the frequency of fake articles has changed after the SEC lawsuit. In Panel B we examine whether the salience of the presence of fake news on the platforms had an impact on investors' reaction to the articles (proxied for by log of abnormal volume on days $t = 0$, $t + 1$, and $t + 2$). In Panel C we examine whether there was an impact on return volatility following the publication of articles. We examine the time period of six months before and six months after February/March 2014 period. 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. *LIWC Fake* is a dummy equal to 1 if the probability of the average article about the firm on a given day being fake is $> 20\%$, 0 if the probability of an average article being fake is $< 1\%$, and missing otherwise. Abnormal volume is defined as $Vol(t)/AvgVol(t-146, t-20)$, summed over days $t = 0$, $t + 1$, and $t + 2$, and then we take the natural log of the sum. We control for the natural log of abnormal trading volume on the day $t - 1$. Return volatility is defined as abnormal returns squared, summed over days $t = 0$, $t + 1$, and $t + 2$. For the volatility regressions, we control for return volatility on days $t - 3$ to $t - 1$. *Article* is a dummy variable equal to 1 if there was at least one article published about the firm on a given day. *6 months after* is a dummy equal to 1 if the time period is 1 April - 30 September, 2014 . We also include controls for SEC filings, press releases, and abnormal volume on day $t - 1$. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. *Other Media (t-3 to t)* is a dummy variable if there was at least one WSJ or NYT article about the firm in the past 3 trading days. *Post-Event* is defined as the 6-month time period after February and March, 2014. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. Firm size is measured in the prior trading month. High retail ownership is defined as being above the median in the prior trading month, and low retail ownership is defined as being below the median. We include year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

Panel A: Propensity of Fake News After SEC Investigation

	LIWC Fake			
		Firm Size		
	All	Small	Medium	Large
Post-Event	0.001	-0.012**	0.000	0.004**
	(0.88)	(-2.15)	(0.19)	(2.40)
Observations	33,498	1,706	16,580	15,212
R-squared	0.000	0.003	0.001	0.000
Controls	X	X	X	X

Panel B: Effect on Abnormal Trading Volume After SEC Investigation

	$Ln([t, t+2]$ day abnormal volume)							
			Firm Size			Retail Ownership		
	All	All	Small	Medium	Large	High	Low	MF Articles
Article	0.389***	0.407***	1.233***	0.289***	0.170***	0.584***	0.242***	0.730***
	(24.32)	(25.21)	(21.08)	(18.11)	(10.68)	(19.13)	(15.24)	(13.40)
Post-Event	-0.179***	-0.170***	-0.349***	-0.080***	-0.035***	-0.302***	-0.095***	-0.179***
	(-61.97)	(-50.29)	(-58.39)	(-28.28)	(-4.00)	(-61.04)	(-28.84)	(-62.70)
Article × Post-Event	-0.041**	-0.075***	-0.357***	-0.068***	-0.117***	-0.066	-0.031	-0.335***
	(-2.05)	(-3.66)	(-4.44)	(-3.45)	(-5.78)	(-1.64)	(-1.64)	(-5.87)
SEC Filing$_{t-3,t}$	0.131***	0.129***	0.135***	0.104***	0.079***	0.106***	0.116***	0.133***
	(30.69)	(20.69)	(14.00)	(25.94)	(7.71)	(12.84)	(27.33)	(31.09)
SEC Filing$_{t-3,t}$ × Post-Event		0.002						
		(0.24)						
Press Release$_{t-3,t}$	0.285***	0.287***	0.371***	0.221***	0.016*	0.312***	0.231***	0.289***
	(58.23)	(41.29)	(29.25)	(49.67)	(1.65)	(30.56)	(50.92)	(58.95)
Press Release$_{t-3,t}$ × Post-Event		-0.003						
		(-0.33)						
Other Media$_{t-3,t}$	0.070***	0.125***	0.331***	0.015	-0.038***	0.159***	-0.017*	0.104***
	(6.82)	(6.07)	(4.91)	(1.49)	(-3.86)	(7.42)	(-1.74)	(10.26)
Other Media$_{t-3,t}$ × Post-Event		0.006						
		(0.40)						
Log(AbVol)$_{t-1}$	1.479***	1.461***	1.432***	1.485***	1.476***	1.437***	1.404***	1.480***
	(876.45)	(612.48)	(531.22)	(659.81)	(163.03)	(559.18)	(534.90)	(877.35)
Log(AbVol)$_{t-1}$ × Post-Event		0.036***						
		(10.65)						
Observations	1,401,509	1,401,509	549,047	799,318	53,144	607,521	610,819	1,401,509
R-squared	0.367	0.367	0.354	0.366	0.346	0.355	0.336	0.367

Panel C: Effect on Return Volatility After SEC Investigation

	Volatility on days [t,t+2]							
			Firm Size			Retail Ownership		
	All	All	Small	Medium	Large	High	Low	MF Articles
Article	0.195***	0.201***	1.830***	0.156***	0.017***	0.388***	0.077**	0.076
	(5.72)	(5.84)	(13.29)	(4.12)	(10.04)	(5.10)	(2.57)	(0.68)
Post-Event	-0.055***	-0.063***	-0.123***	-0.015**	-0.003***	-0.084***	0.002	-0.058***
	(-8.08)	(-8.22)	(-8.06)	(-2.11)	(-2.67)	(-5.96)	(0.29)	(-8.59)
Article × Post-Event	-0.076*	-0.091**	-1.209***	0.073	-0.007***	-0.303***	0.076**	0.112
	(-1.81)	(-2.10)	(-6.35)	(1.56)	(-3.07)	(-3.00)	(2.10)	(0.95)
SEC Filing$_{t-3,t}$	0.035***	0.016	0.040*	0.038***	0.006***	0.047**	0.052***	0.036***
	(3.82)	(1.20)	(1.87)	(4.06)	(5.17)	(2.28)	(6.42)	(3.90)
SEC Filing$_{t-3,t}$ × Post-Event		0.035*						
		(1.94)						
Press Release$_{t-3,t}$	0.042***	0.007	0.122***	0.076***	-0.003***	0.120***	0.080***	0.043***
	(4.00)	(0.49)	(4.32)	(7.27)	(-2.68)	(4.61)	(9.33)	(4.14)
Press Release$_{t-3,t}$ × Post-Event		0.071***						
		(3.41)						
Other Media$_{t-3,t}$	-0.174***	-0.196***	-0.095	-0.034	-0.000	-0.284***	-0.082***	-0.160***
	(-8.21)	(-6.52)	(-0.65)	(-1.44)	(-0.40)	(-5.58)	(-4.62)	(-7.62)
Other Media$_{t-3,t}$ × Post-Event		0.045						
		(1.06)						
Log(AbVol)$_{t-1}$	4.654***	6.079***	6.859***	1.516***	6.855***	4.354***	0.860***	4.662***
	(44.20)	(43.70)	(36.16)	(11.90)	(14.96)	(24.82)	(6.34)	(44.27)
Log(AbVol)$_{t-1}$ × Post-Event		-3.339***						
		(-15.69)						
Observations	1,005,279	1,005,279	336,165	619,729	49,385	386,117	547,402	1,005,279
R-squared	0.002	0.002	0.005	0.001	0.009	0.002	0.000	0.002

Table 5. **Language in Comments around the 2014 SEC Lawsuit**

In this table we examine whether readers are more likely to mention words like "fake," or "wrong" in the comments to the articles. In particular *Fake Words*, is a dummy equal to 1 if the readers used the words "fake" or "fraud" in their comments. *Wrong Words* is a dummy equal to 1 if the readers used the words "wrong" or "not right" in their comments. We study the 6-month time periods before and after February and March of 2014. In Panel A, we examine whether the appearance of *Fake Words* or *Wrong Words* is different for fake versus non-fake articles. In Panel B, *Post-Event* is defined as the 6-month time period after February and March, 2014. We include firm fixed-effects. *, **, and *** indicate statistical significance at the ten, five, and one percent levels, respectively, with t-statistics in parentheses.

Panel A: Use of "Fake" and "Wrong" Words in Comments Section of Fake Versus Non-Fake Articles

	Fake Words	Wrong Words
Fake Article	-0.004	-0.070**
	(-0.27)	(-2.27)
Constant	0.072***	0.348***
	(35.27)	(92.70)
Observations	16,332	16,332
R-squared	0.000	0.000

Panel B: Use of "Fake" and "Wrong" Words in Comments Section After SEC Investigation

	Fake Words	Wrong Words
Post-Event	0.007***	0.000
	(2.73)	(0.06)
Constant	0.069***	0.306***
	(39.17)	(97.79)
Observations	46,172	46,172
R-squared	0.000	0.000

Table 6. **Article-Level Analysis**

In this table, we examine whether author and article characteristics have an impact on investors' reaction to the articles (proxied for by log of abnormal volume on days $t = 0$, $t + 1$, and $t + 2$). In Panel A we examine our entire time period, and in Panel B we concentrate on the time period of six months before and six months after February/March 2014 period. In Panel B, 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 is defined as $Vol(t)/AvgVol(t-146, t-20)$, summed over days $t = 0$, $t + 1$, and $t + 2$, and then we take the natural log of the sum. *Log(Num Followers)* is defined as the natural log of the number of followers the author of the article has, *Log(Num Comments)* is the natural log of one plus the number of comments that the article received. *Log(Num Past Articles)* is defined as the natural log of one plus the number of articles that the author has written prior of writing the article that we are examining. *Frac Numbers in Text* is the standardized faction of how often the author uses numbers in the article. *Frac "earn" mentions* is the standardized faction of how often the words that have a stem "earn." *Post-Event* is a dummy equal to 1 if the time period is 1 April - 30 September, 2014 . We also include controls for SEC filings, press releases, and abnormal volume on day $t-1$. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. We control for the natural log of abnormal trading volume on the day $t-1$. We include year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

Panel A: Full Sample Period

	Log(Abnormal Volume [t, t+2])				
Log(Num followers)	0.012***				
	(7.83)				
Log(Num comments)		0.019***			
		(7.48)			
Log(Num past articles)			0.016***		
			(15.58)		
Frac Numbers in Text				0.046***	
				(20.35)	
Frac "earn" mentions					0.081***
					(35.11)
Observations	198,629	198,629	345,708	345,704	345,704
R-squared	0.467	0.467	0.440	0.440	0.442
Year-Month FE	X	X	X	X	X
Controls	X	X	X	X	X

Panel B: Comparing 6-Months Before Versus After SEC Investigation

	Ln(Abnormal Volume [t, t+2])				
Log(Num followers)	-0.010				
	(-1.55)				
Log(Num followers) × Post-Event	0.024***				
	(2.73)				
Log(Num comments)		0.034***			
		(3.85)			
Log(Num comments) × Post-Event		0.007			
		(0.56)			
Log(Num past articles)			-0.000		
			(-0.00)		
Log(Num past articles) × Post-Event			0.065***		
			(10.72)		
Frac Numbers in Text				0.002	
				(0.25)	
Frac Numbers in Text × Post-Event				0.009	
				(0.75)	
Frac "earn" mentions					0.023*
					(1.92)
Frac "earn" mentions × Post-Event					0.061***
					(4.43)
Post-Event	-0.328***	-0.179***	-0.482***	-0.154***	-0.148***
	(-5.33)	(-5.31)	(-16.33)	(-11.92)	(-11.48)
Observations	29,265	29,265	56,292	56,291	56,291
R-squared	0.435	0.436	0.443	0.439	0.440
Controls	X	X	X	X	X

APPENDIX

Appendix A: Contributors and compensation for authorship on shared-knowlege 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.

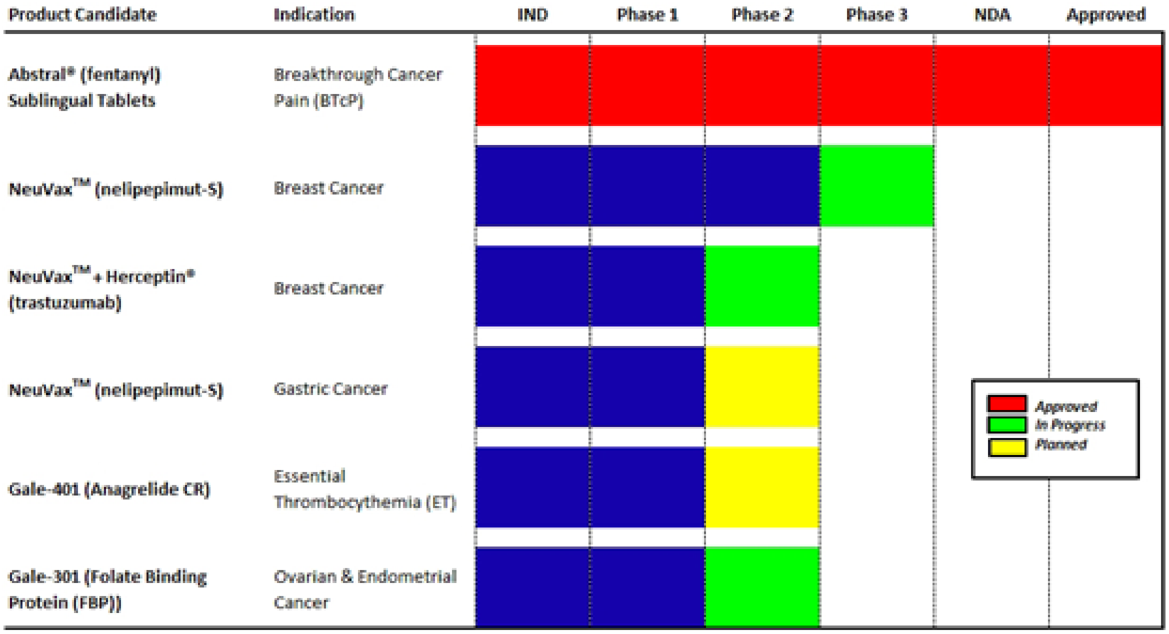
Presently, the company has 105.2 million outstanding shares of stock trading at $4.22, which gives it a market cap of $441.10 million.

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

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 for "Fake News: Evidence from Financial Markets"

Table C1. **Fake Articles and Industries**

This table presents the distribution of articles by Fama-French 12 industries, for different types of articles on Seeking Alpha and Motley Fool. *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. *Seeking Alpha Articles* and *Motley Fool 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 our main analysis.

	Rick Pearson & SEC		Seeking Alpha			Motley Fool		
Industry	For-sure Fake	Non-Fake	Fake	Non-Fake	Others	Fake	Non-Fake	Others
Consumer NonDurables	-	2.45%	2.57%	5.19%	4.53%	5.67%	5.19%	5.19%
Consumer Durables	-	4.49%	3.13%	3.52%	3.37%	6.66%	5.04%	4.04%
Manufacturing	2.30%	12.65%	4.55%	7.26%	5.82%	8.05%	9.98%	8.09%
Energy	-	8.16%	4.9%	6.52%	6.17%	5.26%	5.66%	6.68%
Chemicals	1.15%	1.22%	1.46%	1.79%	1.78%	1.97%	2.44%	2.34%
Business Equipment	4.60%	27.35%	28.13%	23.66%	25.91%	26.87%	26.22%	25.39%
Telecom	-	2.86%	6.39%	4.77%	4.72%	4.35%	3.61%	3.87%
Utilities	-	-	1.11%	0.99%	1.46%	1.23%	1.66%	2.1%
Shops	-	2.86%	6.84%	12.19%	9.21%	13.72%	13.69%	11.62%
Healthcare	81.61%	17.14%	10.63%	5.38%	9.6%	7.81%	7.92%	10.4%
Finance	-	13.06%	22.2%	16.67%	16.42%	10.85%	6.49%	8.9%
Other	10.34%	7.76%	8.09%	12.06%	11.03%	7.56%	12.11%	11.38%

Table C2. **Effect of articles on trading volume - Daily Reaction**

The table examines whether investors react to articles posted on Seeking Alpha and Motley Fool (proxied for by log of abnormal volume on 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. *Article* is a dummy variable equal to 1 if there was at least one article published about the firm on a given day. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. Firm size is measured in the prior trading month. High retail ownership is defined as being above the median in the prior trading month, and low retail ownership is defined as being below the median. In Panel A abnormal volume is defined as $\text{Log}(Vol(t)/AvgVol(t-146, t-20))$. We control for the natural log of abnormal trading volume on the day $t - 1$. We include year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively, with t-statistics in parentheses.

	Log(AbVol)(t=0)	Log(AbVol)(t+1)	Log(AbVol)(t+2)
Article	0.155***	0.121***	0.101***
	(104.15)	(77.24)	(62.88)
SEC Filing (t-3 to t)	0.054***	0.037***	0.015***
	(86.07)	(56.18)	(21.59)
PR (t-3 to t)	0.112***	0.099***	0.064***
	(181.36)	(150.96)	(96.34)
Log(AbVol) (t-1)	0.552***	0.475***	0.438***
	(2,476.84)	(2,015.37)	(1,818.76)
Observations	13,890,419	13,890,419	13,890,419
R-squared	0.325	0.244	0.208
Year-Month FE	X	X	X

Table C3. **Effect of articles on trading volume and volatility - Firm Fixed Effects**

The table examines whether investors react to articles posted on Seeking Alpha and Motley Fool (proxied for by log of abnormal volume on 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. *Article* is a dummy variable equal to 1 if there was at least one article published about the firm on a given day. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. Firm size is measured in the prior trading month. High retail ownership is defined as being above the median in the prior trading month, and low retail ownership is defined as being below the median. In Panel A abnormal volume is defined as $Vol(t)/AvgVol(t - 146, t - 20)$, summed over days $t = 0$, $t + 1$, and $t + 2$, and then we take the natural log of the sum. We control for the natural log of abnormal trading volume on the day $t - 1$. In Panel B return volatility is defined as abnormal returns squared, summed over days $t = 0$, $t + 1$, and $t + 2$. We control for return volatility on days $t - 3$ to $t - 1$. We include year-month and firm fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively, with t-statistics in parentheses.

Panel A: Effect on Abnormal Trading Volume From All Articles (Firm FEs)

	$Ln([t, t + 2]$ day abnormal volume)						
			Firm Size			Retail ownership	
	All	All	Small	Medium	Large	High	Low
Article	0.540***	0.270***	0.764***	0.262***	0.116***	0.354***	0.232***
	(119.67)	(72.70)	(51.09)	(75.38)	(29.18)	(44.91)	(64.78)
SEC Filing (t-3 to t)		0.112***	0.154***	0.090***	0.038***	0.141***	0.102***
		(72.74)	(42.99)	(62.41)	(10.66)	(45.13)	(65.13)
PR (t-3 to t)		0.228***	0.356***	0.190***	0.035***	0.319***	0.203***
		(146.58)	(86.73)	(136.19)	(11.01)	(97.67)	(139.72)
Log(AbVol) (t-1)		1.363***	1.287***	1.435***	1.389***	1.279***	1.308***
		(2,509.81)	(1,448.44)	(2,018.01)	(504.21)	(1,498.63)	(1,517.78)
Observations	13,951,252	13,890,416	5,213,103	8,114,607	562,705	5,690,068	5,726,556
R-squared	0.117	0.395	0.359	0.403	0.443	0.381	0.379
Year-Month FE	X	X	X	X	X	X	X
Firm FE	X	X	X	X	X	X	X

Panel B: Effect on Return Volatility From All Articles (Firm FEs)

	Volatility on days [t,t+2]						
			Firm Size			Retail ownership	
	All	All	Small	Medium	Large	High	Low
Article	0.232***	0.217***	0.940***	0.190***	0.029***	0.336***	0.146***
	(11.23)	(10.49)	(8.12)	(22.66)	(25.18)	(9.08)	(25.28)
SEC Filing (t-3 to t)		0.066***	0.119***	0.035***	0.007***	0.079***	0.041***
		(7.83)	(4.73)	(10.51)	(6.49)	(5.72)	(16.62)
PR (t-3 to t)		0.125***	0.292***	0.079***	0.009***	0.222***	0.091***
		(14.87)	(10.13)	(24.23)	(9.55)	(15.17)	(39.56)
Volatility (t-3 to t-1)		0.500***	0.420***	1.877***	21.636***	0.352***	1.792***
		(26.91)	(13.37)	(47.92)	(158.83)	(19.02)	(41.08)
Observations	10,617,750	10,617,750	3,566,554	6,517,737	533,459	3,801,741	5,260,212
R-squared	0.005	0.005	0.006	0.009	0.151	0.013	0.030
Year-Month FE	X	X	X	X	X	X	X
Firm FE	X	X	X	X	X	X	X

Table C4. **Effect of articles on trading volume and volatility - Winzorised Measures**

The table examines whether investors react to articles posted on Seeking Alpha and Motley Fool (proxied for by log of abnormal volume on 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. *Article* is a dummy variable equal to 1 if there was at least one article published about the firm on a given day. *SEC Filing (t-3 to t)* is a dummy variable if there was at least one SEC filling (10K, 10Q, or 8K) over the past 3 trading days, and *PR (t-3 to t)* is a dummy variable if there was at least one press release issued by the firm over the past three trading days. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. Firm size is measured in the prior trading month. High retail ownership is defined as being above the median in the prior trading month, and low retail ownership is defined as being below the median. In Panel A abnormal volume is defined as $Vol(t)/AvgVol(t-146, t-20)$, summed over days $t = 0$, $t+1$, and $t+2$, and then we take the natural log of the sum. We control for the natural log of abnormal trading volume on the day $t-1$. In Panel B return volatility is defined as abnormal returns squared, summed over days $t = 0$, $t+1$, and $t+2$. We control for return volatility on days $t-3$ to $t-1$. Both measures are winzorisized at the 5% level. We include year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively, with t-statistics in parentheses.

Panel A: Effect on Abnormal Trading Volume From All Articles (Winzorised)

	$Ln([t, t+2]$ day abnormal volume$)$						
			Firm Size			Retail ownership	
	All	All	Small	Medium	Large	High	Low
Article	0.717***	0.337***	0.529***	0.212***	0.071***	0.421***	0.229***
	(199.00)	(114.11)	(45.49)	(67.92)	(20.84)	(68.42)	(74.08)
SEC Filing (t-3 to t)		0.105***	0.082***	0.078***	0.034***	0.042***	0.098***
		(83.54)	(30.25)	(60.21)	(10.27)	(16.88)	(68.92)
PR (t-3 to t)		0.257***	0.285***	0.157***	0.021***	0.225***	0.184***
		(208.19)	(92.46)	(128.15)	(8.20)	(87.70)	(143.93)
Log(AbVol) (t-1)		1.135***	0.971***	1.301***	1.309***	1.065***	1.228***
		(2,555.93)	(1,430.65)	(2,017.65)	(503.02)	(1,568.75)	(1,582.73)
Observations	13,951,256	13,890,419	5,213,107	8,114,607	562,705	5,690,068	5,726,556
R-squared	0.029	0.346	0.304	0.377	0.435	0.324	0.355
Year-Month FE	X	X	X	X	X	X	X

Panel B: Effect on Return Volatility From All Articles (Winzorised)

	Volatility on days [t,t+2]						
			Firm Size			Retail ownership	
	All	All	Small	Medium	Large	High	Low
Article	0.001	0.000	0.114***	0.080***	0.014***	-0.005***	0.023***
	(1.62)	(0.13)	(46.47)	(138.78)	(31.07)	(-3.69)	(41.60)
SEC Filing (t-3 to t)		0.013***	0.024***	0.013***	0.003***	0.017***	0.017***
		(55.73)	(46.45)	(54.97)	(5.78)	(36.69)	(67.71)
PR (t-3 to t)		-0.002***	0.031***	0.023***	-0.004***	0.034***	0.023***
		(-9.04)	(51.85)	(106.39)	(-11.85)	(68.88)	(104.67)
Volatility (t-3 to t-1)		0.047***	0.028***	0.336***	8.558***	0.027***	0.829***
		(87.58)	(41.78)	(121.90)	(145.55)	(40.58)	(189.95)
Observations	10,617,750	10,617,750	3,566,554	6,517,737	533,459	3,801,741	5,260,212
R-squared	0.110	0.111	0.122	0.154	0.234	0.105	0.149
Year-Month FE	X	X	X	X	X	X	X

Table C5. **Article-Level Analysis**

In this table, we examine the relation between Seeking Alpha readership and abnormal firm-level volume and how readership is related to articles being fake. The analysis is at the firm/article level, including date and firm fixed effects. *Fake Article* is a dummy equal to 1 if the probability of an article being fake is > 20%, 0 if the probability of an article being fake is < 1%, and missing otherwise. The number of clicks and the number of reads are measured over days 0-2 (logged). *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

	Log(Abnormal Volume [t, t+2])				Log(Number of Clicks)	Log(Number of Reads)
Log(Number of clicks)	0.053***		-0.137***			
	(10.68)		(-6.24)			
Log(Number of reads)		0.060***	0.191***			
		(12.43)	(8.89)			
Fraction of reads				0.460***		
				(8.51)		
Fake article					0.163***	0.121**
					(2.91)	(2.12)
						0.80
Observations	14567	14567	14567	14567	15093	15093
R-squared	0.89	0.89	0.89	0.89	0.81	0.80

Table C6. **Return Window Regressions**

The table reports results from regressing 4-factor cumulative abnormal returns $Ret_{1,51}$, $Ret_{1,101}$, $Ret_{1,151}$, $Ret_{1,201}$, $Ret_{1,251}$ on a dummy variable for whether an article was fake. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. The estimation sample are the articlesl*, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

	$AbRet_{1,51}$	$AbRet_{1,101}$	$AbRet_{1,151}$	$AbRet_{1,201}$	$AbRet_{1,251}$
			Small Firms		
Fake Article	0.042*	0.052**	0.009	0.007	-0.023
	(1.95)	(2.01)	(0.32)	(0.21)	(-0.66)
SEC (t-3 to t)	-0.007	-0.005	-0.023	-0.045**	-0.037
	(-0.78)	(-0.37)	(-1.40)	(-2.09)	(-1.46)
PR (t-3 to t)	0.044***	0.046***	0.053**	0.069***	0.103***
	(4.46)	(2.90)	(2.10)	(2.70)	(3.80)
Observations	6,838	6,838	6,838	6,838	6,838
R-squared	0.005	0.003	0.001	0.002	0.003
			Non-Small Firms		
Fake Article	-0.002	-0.002	-0.009	-0.013	-0.012
	(-0.52)	(-0.46)	(-1.30)	(-1.61)	(-1.28)
SEC (t-3 to t)	0.000	0.001	-0.001	0.001	0.000
	(0.40)	(0.47)	(-0.54)	(0.27)	(0.13)
PR (t-3 to t)	0.003***	0.007***	0.010***	0.014***	0.019***
	(3.38)	(5.30)	(5.40)	(6.59)	(7.75)
Observations	104,859	104,859	104,859	104,859	104,859
R-squared	0.000	0.000	0.000	0.000	0.001

Table C7. **Fake Articles and Fundamental Performance**

This table examines whether the presence of fake articles during a quarter is associated with deteriorating fundamental performance. We measure fundamental performance in several ways. As *SUE*, which is defined as the seasonally-adjusted change in earnings scaled by the standard deviation of seasonally-adjusted change over the prior eight quarters. Also, as *ROA*, defined as the firm?s return on assets defined as net income scaled by beginning-of-quarter total assets, as well as ΔROA, defined as same-quarter annual change in ROA. *Fake Article* is a dummy equal to 1 there was at least one fake article in the 90 days leading up to earnings announcements, and 0 otherwise. We define an article as being fake if the probability of the article being fake is > 20%. We only include firms in this analysis that had at least one fake article in our sample. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and Large firms are defined as the top 10th percentile of the NYSE firms. We include firm and year-month fixed-effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

	SUE				ROA			
	All Firms	Small	Medium	Large	All Firms	Small	Medium	Large
Fake Article	-1.094*	-3.546***	0.912	-1.244	0.007	-0.024	0.038	-0.008
	(-1.92)	(-2.99)	(1.02)	(-1.28)	(0.53)	(-0.60)	-1.47	(-0.86)
Observations	32,315	5,237	20,971	5,926	31,803	5,128	20,693	5,903
R-squared	0.114	0.180	0.129	0.145	0.594	0.641	0.522	0.461

	ΔROA			
	All Firms	Small	Medium	Large
Fake Article	-0.018	-0.117**	0.030	-0.007
	(-1.10)	(-2.21)	(1.22)	(-0.51)
Observations	30,554	4,754	19,863	5,865
R-squared	0.058	0.099	0.084	0.086

Table C8. **Fake News and Firm Announcements (Weekly Level)**

In this table, we examine whether there are more likely to be fake articles in the weeks around and contemporaneous with insider trading. At the weekly level, we regress a dummy variable for whether a firm had predominantly fake articles in a given week (w = 0) on whether the firm was a net buyer or net a seller in the previous week (w-1), the contemporaneous week (w=0), and the following week (w=1), and a dummy variable for whether the firm issued a press release in weeks w-1, w=0, or w+1. *Net Buyer* (*Net Seller*) is an indicator for whether insiders bought more shares in dollar value than they sold in a given week (sold more shares than they bought). We define a dummy variable (*Fake Article*) for whether a firm had predominantly fake articles in a given week as 1 if the probability of being fake associated with the average authenticity score for articles written about the firm in the given week is great than 20%. PR is an indicator variable for whether the firm issues at least one press release in a given week. We perform our analysis separately for small, mid-size, and large firms. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and large firms are defined as firms above the 90th percentile of NYSE firms. Standard errors are double-clustered at the year-month and firm level. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

	Fake Article								
	Small Firms			Mid-size Firms			Large Firms		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Press Release (week-1)	0.0011**			0.0002			-0.0010*		
	(2.07)			(0.85)			(-1.75)		
8K filing (week-1)	0.0013**			0.0001			0.0003		
	(2.50)			(0.29)			(0.49)		
Press Release (week=0)		0.0026***			0.0020***			0.0008	
		(3.78)			(6.36)			(0.76)	
8K filing (week=0)		0.0018***			0.0017***			0.0002	
		(2.92)			(4.57)			(1.07)	
Press Release (week+1)			0.0002			0.0002			0.0001
			(0.36)			(0.71)			(0.18)
8K filing (week+1)			0.0014**			0.0004			0.0006
			(2.49)			(1.51)			(0.88)
Observations	137,560	137,998	137,719	406,508	407,379	406,593	86,956	87,104	86,946
R-squared	0.010	0.011	0.011	0.007	0.008	0.007	0.013	0.013	0.013

Table C9. **Insider Trading and Fake News (Weekly Level)**

In this table, we examine whether there are more likely to be fake articles in the weeks around and contemporaneous with insider trading. At the weekly level, we regress a dummy variable for whether a firm had predominantly fake articles in a given week (w = 0) on whether the firm was a net buyer or net a seller in the previous week (w-1), the contemporaneous week (w=0), and the following week (w=1), and a dummy variable for whether the firm issued a press release in weeks w-1, w=0, or w+1. *Net Buyer* (*Net Seller*) is an indicator for whether insiders bought more shares in dollar value than they sold in a given week (sold more shares than they bought). We define a dummy variable (*Fake Article*) for whether a firm had predominantly fake articles in a given week as 1 if the probability of being fake associated with the average authenticity score for articles written about the firm in the given week is great than 20%. *PR* is an indicator variable for whether the firm issues at least one press release in a given week. We perform our analysis separately for small, mid-size, and large firms. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms, and large firms are defined as firms above the 90th percentile of NYSE firms. Standard errors are double-clustered at the year-month and firm level. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

	Fake Article								
	Small Firms			Mid-size Firms			Large Firms		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Buyer (week-1)	0.0025*			0.0020*			-0.0062		
	(1.92)			(1.81)			(-1.41)		
Seller (week-1)	0.0008			0.0004			-0.0008		
	(0.72)			(0.90)			(-0.91)		
Buyer (week=0)		0.0051***			0.0040***			-0.005	
		(3.03)			(4.15)			(-0.62)	
Seller (week=0)		0.0017			0.0004			-0.0013	
		(1.56)			(1.05)			(-1.50)	
Buyer (week+1)			0.0058***			0.0022*			-0.0013
			(3.27)			(1.94)			(-0.55)
Seller (week+1)			0.0010			0.0006			0.0005
			(1.04)			(1.64)			(0.53)
Observations	137,593	137,998	137,721	406,575	407,379	406,595	86,959	87,104	86,946
R-squared	0.010	0.011	0.011	0.007	0.007	0.007	0.013	0.013	0.013