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

August 2018

Preliminary and incomplete. Do not cite without permission.

[*] We thank Tony Cookson, Diego Garcia, Gary Gorton, Bryan Kelly, Bonnie Moskowitz, James Pennebacker, 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, and the U.S. Securities and Exchange Commission Division of Economic and Risk Analysis 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

Using a unique dataset of fake stock promotion articles prosecuted by the Securities and Exchange Commission, we examine the impact of fake news. In addition, we use a linguistic algorithm to detect deception in expression for a much larger set of news content using the fake articles as a training sample. We find increased trading activity and temporary price impact from fake news about small firms, but no impact for large firms. Using the SEC investigation as a shock to investor awareness of fake news, we find a marked decrease in reaction to news, particularly content deemed less authentic, but also legitimate news. These findings, including the indirect spillover effects on other news, are most pronounced for small firms with high retail ownership and for the most circulated articles. Understanding the motivation behind the fake articles, we find that small firms engage in corporate actions and insider trading designed to profit from the fake articles, consistent with concerns of coordinated stock price manipulation. No such patterns are observed for large firms. The setting offers a unique opportunity to quantify the direct and indirect impact of fake news.

1. Introduction

False or misleading information can potentially impact social, political, or economic relationships. One prominent recent example is the increased attention "fake news" is receiving. Fake news is a form of disinformation such as hoaxes, frauds, or deceptions designed to mislead consumers of information. With the explosion of (largely unmonitored) shared information platforms, such as social media, blogs, etc. that transmit information, the potential influence of fake and biased news is a growing concern.[1]

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 has primarily been theoretical. For instance, 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. 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.

Debate over the relevance and consequences of fake news is ongoing (Allcott and Gentzkow (2017), Kshetri and Voas (2017), Aymanns et al. (2017)). False content can 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

[1] According to a survey from the Pew Research Center (Gottfried and Shearer (2016)), 62% of American adults get news from a social media site. Allcott and Gentzkow (2017) argue that social media platforms enable content to be disseminated with no significant third party filtering or monitoring, allowing false information to be spread quickly through a vast social network. Vosoughi et al. (2018) find that fake news diffuses faster, deeper, and more broadly than actual news, in part because the fake news is often more extreme and exaggerated in order to increase diffusion. Fake news may have influenced the 2016 U.S. Presidential election (Allcott and Gentzkow (2017), Silverman (2016), Timberg (2016), Silverman and Alexander (2016)), for example, and a study by ReviewMeta (2016) found that fake reviews on Amazon are misleading consumers toward various products (often paid for by the producers of the products).

Mullainathan and Shleifer (2005)). Very little empirical work on fake news exists, however, due to a lack of data, particularly the identification of fake content itself. Indeed, one of the greatest challenges facing shared digital platforms like Amazon, Facebook, Twitter and others today is the ability to detect fake content.

We provide some of the first empirical estimates of the impact of fake news using a unique dataset of false articles in financial markets. The set of identified fake articles come from a Securities and Exchange Commission (SEC) investigation of paid-for false articles on a shared financial news network. An industry "whistle-blower", Rick Pearson, who was a regular contributor on Seeking Alpha, a crowd-sourced content service provider for financial markets, went undercover to investigate fake paid-for articles that he turned over to the SEC. The sample is small, but the identity of fake news is clean – 171 articles by 20 authors covering 47 companies falsely promoting the stock. (We also compare these fake articles to other articles written by the same authors that were not paid-for and presumably not fake.)

The data offer a singular look at identified fake content, overcoming one of the major obstacles in analyzing these issues. However, the sample is small and narrow, making it more difficult to draw general conclusions. To broaden the analysis, we collect articles from Seeking Alpha and another prominent financial crowd-sourced website, 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 over 7,700 publicly traded firms. To identify fake content within this broader set of articles, we appeal to the linguistics literature (Pennebaker et al. (2015), Newman et al. (2003)) and use an algorithm designed to detect deception in expression to assess the authenticity of each article. Importantly, we use the smaller dataset of *known* fake articles from the SEC to validate the algorithm and calibrate a model for measuring the probability of fake news. This is a key and distinct advantage. Absent a set of identifiable fake articles for use as a training set, such endeavors have yielded little success.[2] The algorithm has a type II error on the known fake articles of less than 1% (false

[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 and struggling to do so

positives) and a type I error on the non-fake articles by the same authors of less than 10% (false negatives). The method is conservative and designed to minimize type II errors, where we are likely missing other fake articles, but are confident in the fake news we identify. The prevalence of fake news by our measure is not insignificant and varies meaningfully through time: We classify 2.8% of articles as fake, with the frequency peaking in 2008 at 4.8%.

Our setting is financial markets, and specifically shared-information platforms on financial news and opinions. There are reasons to be both cautious and optimistic on what we can learn about the impact of fake news more broadly from this setting. On the plus side, one of the benefits of financial markets is we can quantify the influence of fake news through prices and trading activity.[3] On the negative side, these information platforms may have little influence on markets either because they are unimportant or due to markets already incorporating the information. Thus, in the backdrop underlying this study is a question of how informationally efficient (Fama (1970)) the market is. Fake news should not matter if markets are perfectly efficient, regardless of what the equilibrium asset pricing model is. In that sense, our setting offers a unique test of market efficiency that circumvents the joint hypothesis problem. Essentially, we run the flip side of the classic event study (Fama et al. (1969)), by examining price and trading responses to a "fake news event." Given competitive arbitrage activity in financial markets, the impact of fake news is likely to be lower than in other settings.

We begin by examining the direct impact of fake news on trading activity. First, we find that abnormal trading volume rises on the days articles on these platforms appear. Second, looking specifically at the SEC sample of known fake articles, we find an even larger trading response to fake news relative to non-fake articles published at the same time on the same platform. This is likely driven by fake articles often being more sensational and diffusing

successfully (Cullan-Jones (2016), Leong (2017), Leathern (2017)).

[3] Arguably, there is little non-pecuniary benefit to consumers of financial news on the platforms. Fake financial news, unlike political or social news, should provide little utility from an entertainment or bias perspective as in Mullainathan and Shleifer (2005). In addition, the costs of fake content here are clear in that if fake news causes less accuracy or erroneous financial decisions, we can directly measure those consequences through trading and price distortions.

more quickly across consumers (Vosoughi, Roy, and Aral (2018)). Turning to the broader set of articles, where we estimate the probability of fake news, we find similar results – abnormal trading volume with less authentic articles. The direct effect on trading is stronger for smaller firms with higher retail ownership and for articles with greater circulation (measured by number of clicks and readers of each article), lending credence to these platforms influencing investor behavior.

We next explore the indirect effects of fake news on trading activity by examining spillover effects from public awareness of the SEC investigation. We exploit the timing of the announced SEC investigation and exposé articles written about the scandal as a shock to investors' awareness of fake news. Do investors react differently to news in general once aware of the existence of fake news? We find that trading volume drops significantly for *any* news article written on these platforms after the event, including legitimate news. The decrease in trading volume is even larger for articles with less authenticity, however. These effects are robust for small, mid, and large-cap stocks, though the effects are strongest for small firms and firms with high retail ownership. In addition, when assessing the comments section to these articles, we find a significant increase in uses of the words "fake" and "fraud" after the scandal, consistent with investors being more concerned or aware of fake news after the event. Importantly, use of these words in the comments has no relation to whether the articles are fake or not, indicating that consumers had no ability to detect fake news, consistent with the difficulty identifying fake content and the response to distrust all news. These findings are consistent with models of fake news such as Allcott and Gentzkow (2017) and Aymanns et al. (2017) where awareness of fake news causes agents to discount all news.

We then turn to pricing effects to see if fake news moves prices in a distortive way. Using the sample of known fake articles from the SEC, we find that the fake promotional articles are able to pump up the stock price for small companies, which subsequently gets fully reversed over the course of a year. Mid-size firms, however, experience a permanent negative price impact when fake articles are written about the firm. Looking at the broader set of articles where we estimate the probability of fake news, we first find that the incidence of fake content is higher for small firms and very low for large firms. We similarly find strong temporary positive price effects for smaller firms, than then fully reverse and turn negative, immediate negative returns for mid-size firms, and no price impact for large firms. These results mirror those from the SEC sample and suggest our methodology for detecting fake news is valid.

We note, too, that 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 framework were not yet known or identified. These results are consistent with the cost of information being greatest for small firms, where in equilibrium paid-for fake content is engaged by small firms, but not by large firms, where there is no price impact. The results for mid-size firms may be consistent with the market not being fooled by paid-for fake content and punishing firms for attempting it.

To investigate further the motivation behind fake news in our setting, to better understand its influence, we begin with the reason Rick Pearson went undercover initially and why the SEC got involved. The original fake articles were part of a promotional pump-and-dump scheme to manipulate the stock price, orchestrated by the firms themselves. For the broader set of probabilistically fake articles we investigate how many are likely motivated by a similar campaign. Another possibility is independent third parties creating a false narrative for their own intentions.

To try and distinguish these motivations, we look at other actions taken by the firm at the time of the article's release. In the week before, during, and after the fake news articles appear, we find that firms are more likely to have press releases and 8-K filings, consistent with a coordinated effort to influence the narrative of news about the firm. Moreover, these actions are clearly present for small firms and, to a lesser extent, mid-size firms, but do not accompany fake news for large firms. Furthermore, we find strong evidence of insiders positioning themselves to benefit from the subsequent price movement in small firms. For large firms, however, we find no evidence of unusual insider trading activity. We also find that the price response to fake news is even greater when insiders trade as well. These results are consistent with a deliberate campaign by smaller firms to manipulate the stock price and take advantage of any price impact. Large firms, however, do not exhibit any of these patterns, suggesting that fake news about large firms may be written by authors with no ties to the company.

Our results provide some of the first empirical estimates of the impact of fake news. Our findings have implications for theories about fake news and news media more generally. The prevalence of fake articles on these information-shared platforms and its impact on trading activity and prices (for small firms) may be consistent with fake news being tailored to consumer's priors as suggested by Allcott and Gentzkow (2017), and more broadly, Gentzkow and Shapiro (2005) and Gentzkow et al. (2015), who argue that biased reporting, of which fake news is one aspect, will arise in equilibrium when verifying authenticity is costly and news is deemed higher quality if closer to a consumer's priors. In addition, the decline in trading activity to all news, including legitimate news, following the public's awareness of fake news from the SEC investigation is consistent with Aymanns et al. (2017) and Allcott and Gentzkow (2017), who argue fake news may increase distrust of media in general.[4] The spillover effect we find on investors' reaction to other, non-fake news may also be related more generally to the economics of norms and institutions like trust and social capital (Guiso et al. (2004), GUISO, SAPIENZA, and ZINGALES (GUISO et al.), Guiso et al. (2010), Sapienza and Zingales (Sapienza and Zingales)).

Our findings also have implications for the informational efficiency of markets, where the price impact we find for small stocks suggests their cost of information is sufficiently high, and hence why small firms may attempt price manipulation in the first place.[5] The

[4]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.

[5]The marginal cost of information determines how informationally efficient financial markets are (Grossman and Stiglitz (1980)). The cost of information can be both a direct cost of gathering, processing, and analyzing information, as well as the indirect costs of misperceiving or misreacting to information stemming

subsequent price reversal is also consistent with fake news producers sacrificing longer-term reputational capital in lieu of short-term gains (Allcott and Gentzkow (2017)). For large cap firms, the lack of any price reaction is also consistent with large firms not attempting any price manipulation, since markets are more efficient for these firms.

Finally, our study provides evidence on the prevalence and effect of fake news on crowd-sourced platforms that continue to grow and gain attention. The results are broadly consistent with other findings suggesting that crowd-sourced services can impact markets (Hu, Chen, De, and Hwang (2014)). If fake news can impact U.S. equity markets, where there is competition for information and arbitrage activity exists, then it may have even greater influence in settings where information costs are high and the ability to correct misinformation is more limited, such as online consumer, marketing, political, and social media networks.

The rest of the paper is organized as follows. Section 2 details our sample of fake news articles obtained from Rick Pearson and the SEC, the broader set of articles from the shared-informaton platforms, and our methodology for assessing the probability of fake news. Section 3 examines a case study of Galena Biopharma that launched the SEC prosecutions to illustrate the issues we investigate more broadly. Section 4 examines investor's response to fake news through trading activity, including spillover effects on non-fake news. Section 5 analyzes the price impact of fake news and Section 6 seeks to understand the motivation behind fake news by looking at coordinated corporate actions and insider trading around the fake articles. Section 7 concludes.

2. Data and Identifying Fake News

We describe our sample of fake articles, the broader sample of articles with unknown authenticity from the same media platforms, and our methodology for identifying probable fake content from the broader sample. Before proceeding, we provide some background on shared-financial news platforms.

from psychological or behavioral biases. Allcott and Gentzkow (2017) suggest that information costs are necessary for fake news production.

2.1. *Shared Financial News Platforms*

We draw our sample of articles 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 through a shared-knowledge platform. As described below, 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 cites contribute and are compensated for their articles.

The popularity of these sites has grown exponentially over the fifteen years of their existence. For example, 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, such as pump-and-dump schemes, since they are virtually unregulated, frequented predominantly by retail investors, and authors on these platforms can use pseudonyms instead of writing under their real names (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 on these platforms face the following legal restrictions. First, it is legal for an author to talk up or down a stock that she is long or short, provided she discloses any positions she has in the stock in a disclaimer that accompanies the article. Failure to disclose can have legal ramifications and although many authors add such disclaimers to their articles, the platforms do not actually verify them. 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.[6]

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

2.2. "For-Sure" Fake Articles

Promotional articles, fraud, and pump-and-dump schemes can be hard to identify and even harder to prove intent to deceive. Our analysis starts with a unique dataset of articles whose authors received payment to write, where the authors illegally did not disclose payment. These unique articles were obtained from an industry insider, Rick Pearson, who as a regular contributor to Seeking Alpha, was approached by a public relations firm to promote stocks by writing fake articles for a fee without disclosing the payment. Non-disclosure of payment not only violates the terms of Seeking Alpha but also SEC regulation Section 17b. Instead, Mr. Pearson decided to go undercover to investigate how rampant this practice was on these platforms 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 two lawsuits: on October 31, 2014 and in 2017 against authors of fake articles and the promotion firms who were paying the authors to generate the articles.[7]

Mr. Pearson kindly shared with us the articles that he has determined to be fake, providing us with 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. During the investigation, the SEC lawyers were able to identify more articles that were paid for by stock promotion firms and deemed to be paid-for fake content.[8] We also contacted Seeking Alpha, and they kindly shared 147 of those articles with us. Of those, we were able to match 60 with Center for Research in Security Prices (CRSP) data that are publicly traded on U.S. exchanges, where the rest of the articles pertain to firms traded over the counter.

---

compensation had been paid.

[7]See filing documents at: http://securities.stanford.edu/filings-documents/1051/GBI00_01/20141031_r01c_14CV00367.pdf and https://www.sec.gov/litigation/complaints/2017/comp23802-lidingo.pdf.

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

Our final dataset of for-sure fake articles consists of 171 articles written by 20 authors about 47 firms.[9]

It is important to define what we mean by *fake* articles. In this smaller 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 some element of false information. How false or wrong that information was 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. Articles may be fake and (mostly) right, as well as fake and (very) wrong. Some of our analysis on the language used in the articles and on their impact on stock prices will help distinguish between these two cases, where we will conclude that most of the articles perpetuated false information. Ultimately, however, it is exceedingly difficult to assess *how* false the articles are. We focus instead on the set of articles with a known intent to deceive, which we call "fake."

We also obtain a sample of other articles written by the same 20 authors now under investigation that were *not* paid for by a PR firm, totaling 334 additional articles about 171 companies published on Seeking Alpha. We use this set of non paid-for articles by the same authors to provide a clean comparison to the fake (paid-for) articles written by those authors, which controls for any author characteristic or heterogeneity in writing style. It is notable that these other non-paid for articles are often written about larger firms, which as we will show, are much less likely to engage in stock promotion schemes. Furthermore, authors may need to establish credibility and a reputation by writing non-fake articles before they can write effective promotional articles. Hence, we refer to these non-paid for articles as

[9]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.

"non-fake" following our definition above and make no statement about the accuracy of the articles themselves. In summary, we focus on authenticity and not accuracy, though some of our analysis may help distinguish between them.

2.3. Further Identifying Fake Articles – LIWC and the Authenticity Score

Our unique data of fake articles provides a sample of unambiguous fake content, overcoming one of the major challenges to studying this issue. However, the sample is small and therefore may make it difficult to draw more general conclusions. To complement these data, we manually download all articles published on Seeking Alpha, as well as a competitor site Motley Fool, representing two of the most prominent financial crowd-sourced platforms. We obtain 203,545 articles from Seeking Alpha over the period 2005 to 2015 and 147,916 articles from Motley Fool from 2009 to 2014. The universe of articles allows us to examine the impact of these platforms, and fake content that might emanate from them, more broadly.

The downside of this much larger dataset of articles is that the articles are of unknown authenticity. We therefore develop a probability function for detecting fake content using an objective and scalable measure that captures the authenticity of the article. Appealing to the linguistics literature, we use a linguistic algorithm designed to detect deception in expression. Specifically, we use the Linguistic Inquiry Word Count model (LIWC2015) from Pennebaker et al. (2015), which is a linguistic tool that focuses on individuals' writing or speech style, and appears to be uniquely adept at measuring individuals' cognitive and emotional states across domains. For instance, Newman et al. (2003) use an experimental setting to develop an authenticity score based on expression style components.[10] 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

[10]These techniques are often used by the Central Intelligence Agency and Federal Bureau of Investigation to assess authenticity in speech or writing.

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. 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 to validate the linguistic algorithm and calibrate the authenticity score into a probability fake news. 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.

Using our unique sample of 171 fake articles and 334 non-fake articles written by the same authors, we test and validate the linguistic algorithm. We compare the LIWC authenticity score, which is normalized between 0 and 100, for the two samples and control for author fixed effects to capture any heterogeneity in author style, content, or reputation, and any selection issues of authors being matched to fake/promotional 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 the two samples' authenticity scores in Figure 1, Panel A highlights the differences, where again we are controlling for author heterogeneity since we examine fake and non-fake articles within the *same author*.

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 the same 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.

While the exact composition of the authenticity score is proprietary, we provide several language characteristics associated with authenticity described in Pennebaker (2011). Panel A of Table 1 reports summary statistics on those characteristics for the for-sure fake and non-fake articles written by the same authors that contribute to the for-sure fake articles' total authenticity score being about half that of the 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 tend to use fewer self-referencing words, fewer words per sentence, fewer insight and relativity words, and more discrepancy verbs. The results in the table line up well with those findings: fake articles' self-referencing score is about half of non-fake articles, have lower insight, lower relativity scores, 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, an impossible task without the for-sure fake articles from Rick Pearson and the SEC.

2.4. Probability of Being Fake

The sample of for-sure fake and non-fake articles also allows us to calibrate the authenticity scores into a probability of fake content. While the LIWC authenticity score is statistically different between fake and non-fake articles, (Panel A, Table 1), 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 from which to calibrate probabilities. 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 larger sample of Seeking Alpha and Motley Fool articles. We use the known fake and non-fake articles to map authenticity scores into a conditional probability of being fake.

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 which ones are T in order to measure these 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 to the fixed point problem can be found assuming that $Prob(F)$ in the sample is representative of $Prob(F)$ in the overall population.

We first apply this approach to the entire sample of Seeking Alpha articles published between 2005 and 2015, covering over 203,000 articles, pertaining to over 7,700 firms. Figure 2 plots the mapping of LIWC authenticity scores (S) into the conditional probability of being fake ($Prob(F|S)$). An authenticity score of 33 (the average for the non-fake articles) corresponds to a conditional probability of being fake of near zero, while an authenticity score of 19 (the average for the fake articles) corresponds to a significant probability of being fake of 3.6%. The relation between the LIWC authenticity score and the probability of being fake is highly nonlinear. Specifically, the sharp increase in probability in the very low authenticity range suggests that articles may be more efficiently and better classified into fake and non-fake using a probability cutoff. We use a cutoff of $Prob(F) > 0.20$ to classify articles as being fake and classify articles with $Prob(F) < 0.01$ as being non-fake, with the rest (articles with $0.01 \leq Prob(F) \leq 0.20$) being classified as ambiguous or "other."[11]

We first examine how accurate our method is at identifying fake news from our specialized small sample of 505 articles (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 of the 505 articles as being fake ($Prob(Fake) > 0.20$),

[11] 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 beginning of a steep increase in probability as shown in Figure 2.

of which 17 are actually fake, indicating that the Type II error rate is very low – one false positive. Our method is conservative, however, since it misses a lot of fake articles. Our algorithm identifies 165 articles (out of 505) as being non-fake. Of those, 17 are actually fake, implying a Type I error of about 10%, which is quite low considering our methodology is designed to minimize type II errors. We exclude articles with $0.01 \leq Prob(Fake) \leq 0.20$ from our analysis, since both Type I and Type II errors will be larger for these articles.

Table 1 Panel A shows summary statistics for the *Fake*, *Non Fake*, and *Other* articles identified by our algorithm on all Seeking Alpha articles published between 2005 and 2015 (203,545 articles). The number of articles in each category, the mean of the *Authenticity* measure that we use to construct the probabilities of being fake, 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 were sorted based on the probabilities, but the magnitude of the difference is interesting and suggests substantial differences in authenticity between the two groups of articles.

Using our methodology, how pervasive are fake articles on financial crowd-sourced platforms? The unconditional probability of a Seeking Alpha article being fake is 2.8% over the entire sample period, peaking at 4.8% in 2008 and dropping to a low of 1.6% in 2013. Figure **??** in Appendix C reports variation in the authenticity scores across the linguistic cues from our algorithm for the average article over time. Dissecting the time variation in scores over time, it appears that the spike in 2008 is driven by more *** words and the low in 2013 seems to come from *** fewer words. Overall, however, the components that make up the authenticity scores seem to move up and down together (the average time-series correlation of the six components we track is 0.**).

We also apply our methodology for identifying fake articles to another sample of articles from another crowd-sourced financial news platform – Motley Fool (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* as we did for Seeking Alpha. 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. Looking at the rest of the components of the authenticity score, the algorithm does a similar job on both samples of articles.

Finally, as another validation exercise we analyze only those 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. Since Mr. Jayson works directly for Motley Fool, it is unlikely he has written fake articles on their platform and unlikely promotional firms would even approach him. Hence, we use his articles as a placebo test of our classification methodology. Using our methodology on Mr. Jayson's articles, we classify 18,361 as reliably non-fake and only 2 of his articles as probabilistically fake (the rest being indeterminate). That is, we classify 0.006% of his articles as fake, suggesting that our algorithm works quite well, since the number of his articles that are fake should be essentially zero.

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 USD millions) for each article group. For-sure fake articles tend to cover firms with a higher fraction of retail investors, and tend to concentrate on smaller firms with low analyst coverage. The probabilistically determined fake and non-fake articles from the broader Seeking Alpha and Motley Fool articles 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 Appendix C examines whether fake articles tend to cluster in specific industries. We separate articles into one of the 12 Fama-French industries that the firms 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 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 identified by Rick Pearson and the SEC, with the majority coming from *Business Equipment*, *Finance*, and *Healthcare* industries.

2.5. Supplemental Datasets

To investigate the motivation behind fake articles, including the hypothesis that these campaigns are ordered by firms and orchestrated by a PR agency, we obtain a dataset of press releases from RavenPack from 2001 to 2015, 8-K disclosure filings from the SEC's Edgar database, stock price data from CRSP, firms' financial information from COMPUSTAT, executive compensation data from Execucomp, and insider trades from Form 4 from Thomson Reuters.

3. A Case Study and a Shock to Investor Awareness of Fake News

To illustrate the motivation behind and the impact of fake articles we aim to examine more broadly, we first dissect a case study of Galena Biopharma Inc., one of the companies that hired a PR firm to solicit paid-for fake articles about its stock. Galena was the first company prosecuted by the SEC for stock price manipulation on these knowledge-sharing platforms. We start by documenting the pump-and-dump scheme orchestrated by Galena and it's unraveling, and later examine how public awareness of the scheme impacted the market's reaction to news more generally.

3.1. Case Study: Galena Biopharma Inc.

On October 31, 2014 the SEC filed a lawsuit in the United States District Court (Case 3:14-cv-00558-SI)[12] on behalf of all persons who bought Galena's common stock between August 6, 2013 and May 14, 2014. The timeline of events (summarized from the lawsuit document) is presented in Figure 3. The figure 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 in some cases falsely claimed specifically *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. If one searches for this fake article today, Seeking Alpha displays a message saying "This author's articles have been removed from Seeking Alpha due to a Terms of Use violation."

Figure 3 shows that over this time, Galena's share price rose from about \$2 to \$7.48 between the summer of 2013 and January of 2014. The publication 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. As the graph shows, Galena's share price increased drastically during the publication of the fake articles, more than tripling in four months.[13] A natural question is why companies paid for these fake articles. We examine data on insider sales, equity offerings, and stock option grants and executive compensation around the release of the fake articles. Galena insiders seemed to take advantage of the price rise from the fake articles through corporate actions and their own personal trading. On September 18, 2013 in an SEO Galena sold 17,500,000 units of stock for net proceeds of

[12] http://securities.stanford.edu/filings-documents/1051/GBI00_01/20141031_r01c_14CV00367.pdf

[13] Four of the paid-for articles published towards the end of 2013 and early 2014 were all written by the same author, John Mylant, who had been an active contributor to Seeking Alpha since 2009. Since the lawsuit, all of his articles have been taken down by Seeking Alpha.

$32.6 million. Then, on November 22, 2013, Galena held a board meeting and granted stock options to executives and directors with a strike price of $3.88. Starting January 17, 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 a combined total of more than $16 million. These events are highlighted in Figure 3.

As the news of insider sales broke, the stock price started to decline dramatically as depicted in Figure 3. Furthermore, in February and early March 2014, several investigative journalists, including Matt Gravitt, Adam Feuerstein, and Richard Pearson, started publishing exposé articles on Seeking Alpha documenting the fraud. While these journalists were uncovering events that were linked directly to Galena, in February and March, 2014, several articles appeared on Barron's and in Fortune,[14] discussing Seeking Alpha's anonymous contributors policy and the fact that Seeking Alpha and other websites had to remove over 100 articles from their site that had been used in stock promotion schemes. Finally, 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 back down to $2.

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.

[14]***CITES HERE***

3.2. A Shock to Awareness of Fake News

The SEC investigation ultimately provided us with a sample of known fake articles from which we validate and calibrate our framework for detecting fake news more broadly. In addition, however, the public revelation of the investigation and subsequent media attention around it also provides us with a unique shock to investor awareness of fake news. We exploit the timing of the announcement of the investigation and revelation of fraudulent articles to test various implications of consumer awareness of fake news.

Fake news can be costly to society in several ways. In addition to the potential costs of individuals believing and acting upon false content, fake news can be costly to society if it damages people's trust in news generally and 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 non-fake news. Using the revelation of the SEC investigation and subsequent exposé articles, we examine whether investors behaved any differently before versus after this event, when the presence of fake news on knowledge sharing platforms suddenly became salient to many consumers on these platforms.

4. Impact on Trading Activity

Financial markets provide an intriguing setting to examine the impact of fake news. On the one hand, they provide a wealth of outcome variables, measured at high frequencies, such as trading activity and market prices, that allow us to measure the impact of fake news. On the other hand, the efficiency of markets may make detection of these effects in prices difficult. Finally, there is also the possibility that these shared knowledge platforms have no influence on investors or markets.

4.1. Direct Effect of Articles on Trading

We begin by addressing whether articles posted on these social platforms have any influence on market participants. We start by examining abnormal trading volume around the publication of all articles, fake and not fake. We focus on volume because we are interested in whether investors who read these articles "react" to them. We examine abnormal trading volume, which is trading volume relative to an expected level of volume in the stock. Alternatively, examining the articles' effect on price volatility addresses whether there is information in the articles not already accounted for by markets (an issue we examine later), but not necessarily whether investors acted in response to the articles. Quantities traded can vary significantly with no price movement (Fama (1970), Grossman and Stiglitz (1980)) or trade can be zero with substantial price movement (Milgrom and Stokey (1982)). We look at both quantities and prices separately.

Panel A of Table 2 examines the effect on trading volume from articles published on those sites. We define abnormal trading volume for stock i as $Vol(i,t)/\frac{1}{T}\sum_{k=1}^{251} Vol(i,t-k)$, which is the trading volume for stock i on day t relative to the average daily trading volume in stock i over the last year (250 trading days).[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 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 a 12.1% increase in abnormal trading volume over the three days following publication. This result implies either 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 coincide.

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

The next three columns of Panel A of Table 2 report results separately for small, medium, and large firms. Small firms are defined as all firms traded on the NYSE, AMEX, and Nasdaq that are smaller than the bottom 10^{th} percentile of NYSE firms, mid-size firms are defined as firms that fall in the 20^{th} to 90^{th} size percentile of NYSE firms, and large firms are defined as firms in the top 10^{th} size percentile of NYSE firms. The effect on abnormal trading volume declines with firm size, and is six times larger for small firms than for large firms. This result is consistent with small firms having more retail investor trading and perhaps a more opaque information environment.

The results in Panel A show that increased trading volume coincides with articles published on these platforms, especially for small firms, suggesting investors may be responding to or are influenced by these articles. In Panel B, we examine whether the authenticity of the article has a differential impact on trading volume. Does fake news have the same impact as non-fake news on the same stock?

We start with our sample of for-sure fake and non-fake articles from the SEC. Focusing on days when an article appeared, we regress the log of abnormal trading volume on a dummy for whether the article is fake, and control for year-month fixed effects. The abnormal volume controls for firm-specific average trading activity absent any news since it compares volume on the day the article appears to the moving average of recent daily volume on the stock. The coefficient on the fake article dummy represents the marginal effect on trading volume for a fake article versus a non-fake article in our small sample of articles of known authenticity. The first column of Panel B shows that fake articles have a larger impact on trading volume than non-fake articles, suggesting that the fake articles garner more attention from investors, which may be because the promotional articles are designed to induce investor interest and are often more sensational than non-fake articles. Also, the paid-for fake articles are often part of a broader scheme to influence markets, and hence some of what the regression may be picking up is the effect from this coordinated effort, which we investigate more thoroughly in Section 6.

The next three columns of Panel B examine the broader set of all articles published on Seeking Alpha and Motley Fool, where we use our calibrated probability function for fake news to classify (probable) fake and non-fake articles. We throw out all ambiguous articles, whose probability of being fake is greater than 0.01 but less than 0.20. We also focus only on days when articles are published on the firm.

The second column of Panel B shows that a fake article generates increased abnormal volume that is 3.4% higher than that generated by non-fake articles (with a t-statistic of 3.17). The next two columns of Panel B examine several author and stock characteristics that may affect the impact of fake, relative to non-fake, articles on trading volume and help establish that these articles are influencing trading activity. We examine whether the effect of fake articles on trading volume differs by author impact, firm size, or the fraction of retail investors in the stock. Author impact is measured as the average price response to the author's previously written articles over the day in which the articles appeared, plus the following two days. Price response is measured as the idiosyncratic volatility of the average daily share price over the three-day window following each previous article, where idiosyncratic returns are defined relative to the Fama and French (1993) three-factor model augmented with a momentum factor. Authors whose previous articles were associated with larger price moves, may receive more subsequent attention from investors. If those authors are also more likely to write fake articles, perhaps because promotion firms rationally select more influential authors, then the higher trading volume associated with fake news could partly be confounded with author reputation. As the third column of Panel B shows, past author impact has a very large effect on abnormal trading volume, suggesting these platforms do affect trading activity, but the interaction term between author impact and fake articles is insignificant, suggesting no selection bias of impactful authors being more likely to invite fake articles.

The last column of Panel B interacts firm size and retail investor ownership with the fake article dummy. We use the percentile of the firm's market cap and retail ownership as regressors (to limit the influence of outliers). The larger impact on trading volume from fake news weakens with firm size, and fake articles have a much stronger effect on trading volume, relative to non-fake articles, for firms with higher retail ownership. Both results are intuitive since larger firms are typically more informationally efficient, facing lower information costs than small firms, and retail investors dominate participation on these shared information platforms. The latter result also provides more direct evidence that (retail) investors pay attention to these sites.

Panel C repeats the regressions in Panel B, replacing abnormal volume as the dependent variable with the idiosyncratic price volatility of the stock. We measure idiosyncratic volatility for a stock as the residual volatility of daily returns on the stock relative to the Fama and French (1993) three factor model, augmented with a momentum factor. The dependent variable is the sum of daily idiosyncratic volatility over the day the article is published plus the next two days. These regressions capture whether the 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 markets are informationally efficient, we expect little or no price movement despite the fact that trading volume rises following these articles being posted. However, since these articles contain fake content, finding significant price movement may indicate markets are less than perfectly efficient. As Panel C reports, we find effects similar to those in Panel B when looking at trading volume: price volatility of the stock rises following fake news, and the effect is strongest for smaller firms with higher retail ownership. While the effects on volatility mirror those on trading volume, the effects are weaker, suggesting that prices respond less to these articles than trading activity, which is intuitive given markets are somewhat efficient.

[16] Textual analysis used to derive sentiment (Antweiler and Frank (Antweiler and Frank), Tetlock (Tetlock), Das and Chen (2007), Jegadeesh and Wu (2013), Heston and Sinha (2017), Boudoukh et al. (2018)) is notoriously challenging and noisy.

4.2. More Evidence on Direct Impact

To provide further evidence that these articles may have directly impacted trading volume in the stocks these articles wrote about, Table C2 in Appendix C examines the impact on trading volume from various measures of how widely circulated the articles are. We use the number of comments other users posted on the articles, the number of followers of the article, and the number of emails the article was distributed to, all of which Seeking Alpha records. Consistent with our findings, the better circulated the article, the greater the impact on abnormal trading volume, suggesting that the articles published on these platforms have a direct affect on trading activity. In addition, interacting the circulation measures with the fake news indicator suggests that well-circulated fake articles have an even greater impact on trading volume, consistent with those articles getting more attention.

As a further test of a direct link between articles published on these platforms and trading activity, we obtain a proprietary supplemental dataset from Seeking Alpha on readership of the articles. For each article published during calendar year 2017 about a U.S. publicly traded firm, we observe the daily number of “clicks" (i.e., the number of times a given article was uploaded) and the number of times it was “read" (i.e., instances in which the reader scrolled to the end of the article). In total, the dataset covers 25,596 articles and 3,118 firms.

Table C3 in Appendix C presents the results, where the first four columns report results from regressing abnormal trading volume following the release of the article the readership circulation variables over the first three days after the article is published. The table shows that abnormal trading volume is positively related to the number of clicks and number of times the article was read by consumers. This evidence provides direct support for these articles influencing trading activity in the stocks the articles were about.

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 appears to disseminate faster and more widely and, as a result, impacts investors more in terms of their trading activity. These results are consistent with fake news being more sensational and more persuasive, catering to the biases and priors of their consumers, and propogating more diffusely through the network as suggested by Allcott and Gentzkow (2017) and Vosoughi, Roy, and Aral (2018).

4.3. Indirect Effects on Trading: Spillover Effects from the Scandal

While fake articles seem to have a direct effect on investors' attention and trading activity, in this subsection we examine the indirect effects of fake news on other news generally. Another unique feature of our study is that we can exploit the timing around the promotional articles scandal that broke in February and March 2014 as an exogenous shock to people's awareness of the presence of fake news on these platforms. This shock provides a novel opportunity to examine any spillover effects from the presence of fake news on news generally, as suggested by theory (Allcott and Gentzkow (2017), Aymanns et al. (2017)). We concentrate on the six months prior to and six months after the scandal, to examine whether the prevalence of fake news and the effect of news in general on abnormal trading volume changes after public revelation of the scandal. We define the *before* period as August 2013 to January 2014, and the *post* period as April 2014 to September 2014, excluding February and March 2014 when the exposé articles were published, as the information event.

Panel A of Table 3 first examines whether the propensity of fake news declined after the scandal and whether the effect varies by firm size. The first column reports results for all firms, where we regress the prevalence of fake news on an indicator for the period after the scandal. The coefficient 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. This null average result, however, masks substantial and interesting heterogeneity. In the second column, we examine whether the author's past impact played a difference in the decision to write fake news after the scandal. Before the scandal, authors who were more impactful are more likely to write fake news, but after the scandal they are less likely to write fake content, suggesting that more impactful authors found the cost of publishing fake news to be higher after the public's awareness of the scandal. The next six columns repeat these regressions separately for small, midsize, and large firms (defined as the smallest 10%, middle 80%, and largest 10% of firms, respectively, based on NYSE market cap breakpoints). Interestingly, and intuitively, the scandal had the biggest impact on the prevalence of fake articles about small firms, relative to medium and large firms. The prevalence of fake content about small firms fell significantly (-24.8%) following the scandal. For large firms, the effect is actually slightly positive, though economically small (a 0.3% increase). These results are consistent with small companies engaging in promotional articles before the scandal, where once the SEC became aware and investigated, there was a sharp decline in this activity. Medium sized firms also show a decline in fake articles, but of much smaller magnitude (-3.1%). In addition, author impact is much stronger on fake content propensity for small firms and the decline in fake news post-scandal is largest for small firms and among the highest impact authors. Author impact overall has no detectable effect on the change in fake news pre- and post-scandal.

In Panel B, we perform a similar analysis by separating firms by retail ownership (within size deciles to control for the relation between firm size and retail ownership) and by industry classification. Stocks with higher retail ownership have a stronger response to both the scandal and higher impact authors, consistent with these platforms having a direct effect on retail investors. As Table C1 in Appendix C shows, the majority of companies caught in the scandal were in the healthcare industry. Therefore, the scandal may have been more salient to investors in that industry. The results confirm that intuition. The drop in the propensity of fake articles is much larger for firms in the healthcare industry. However, the results are not confined to the healthcare sector as there is also a significant decline in fake news across all non-healthcare sectors.

Panels C and D of Table 3 examine the impact of published articles on abnormal trading volume before versus after the scandal. The analysis is similar to Panels A and B, except the dependent variable is abnormal trading volume over days t, $t+1$, and $t+2$. The first column of Panel C reports results from a regression of abnormal volume on the post event indicator, the fake article indicator, and their interaction. The positive coefficient on fake articles is our earlier result from Table 2 that before the scandal, fake articles generated more trading volume. The negative interaction term shows, however, that the effect of fake news on trading volume decreases significantly after the scandal. This result is consistent with investors becoming aware of fake content and muting their trading to news in response. Finally, the strong negative coefficient on the post-event dummy indicates that abnormal trading volume declines for non-fake news as well after the scandal. This result suggests that people responded less to news in general on these platforms, including non-fake news, after the scandal and is consistent with consumers having less trust of news once aware of the possibility of fake news, as theory suggests (Alcott and Gentzkow (2017)). The economic magnitude of the effect is large: a 15% drop in trading volume associated with non-fake news articles after the scandal broke and a 19% drop in volume associated with fake (less authentic) articles post-scandal. Following the scandal, articles posted on these platforms generated 15-20% less trading response, with the effect being even larger for articles with less authenticity. The second column of Panel C also shows that more impactful authors had an even larger increase on abnormal trading volume before the scandal, but a much larger decrease in trading volume after the scandal.

Columns 3 through 8 of Panel C report results separately for small, medium, and large firms. Consistent with our previous results, these effects are all much stronger for smaller firms. Post scandal, the abnormal trading volume associated with fake articles drops 77% and for non-fake articles published on these platforms, trading volume drops 61.5% for the smallest firms. Interestingly, even though few fake articles are written about large firms and none of the firms in the SEC probe were large firms, abnormal trading volume still declined by almost 12% 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. Author impact post event is also much stronger among small firms and negligible among larger firms.

Panel D separates firms by retail ownership (within size deciles) and by industry classification. We find that the effect on trading volume from fake articles and from high-impact authors before the scandal is much larger for high retail ownership firms and the decrease in trading volume after the scandal is also larger for firms with high retail ownership. The effects on abnormal volume are similar for firms in the healthcare and non-healthcare industries, and are consistent with our general findings, where post-scandal abnormal trading volume declines for both fake and non-fake articles.

Table C4 in Appendix C examines the abnormal trading volume response daily for the first five days after the article is published. The biggest trading volume impact occurs on the day the article is published – for both the direct effect on fake articles as well as the spillover effect on non-fake articles after the scandal – with significant effects on volume for the next five days that get slightly weaker each day. This timing is consistent with a direct impact on trading activity from the articles themselves and suggests these platforms have some impact on investor activity.

Panels E and F report the results from the same regressions using 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 in general following the scandal, with a slightly stronger reduction for less authentic articles. These findings are consistent with markets discounting the news from these platforms after revelation of the scandal. The price response to the scandal and the probability of the articles being fake are also stronger for more impactful authors and for articles written about smaller firms with more retail ownership. These results are consistent with those on abnormal trading volume in Panels C and D.

4.4. More Evidence of Spillover Effects

The spillover effect from the awareness of fake news to non-fake news is both interesting and consistent with theoretical conjecture (Allcott and Gentzkow (2017)). The result begs the question: How broadly did the awareness of fake news from the scandal affect investors' response to news generally? Or, was the spillover response merely contained to similar articles on Seeking Alpha, where the promotional articles had resided? This question is difficult to answer beyond our setting for several reasons. For instance, looking at other news events, such as media articles (e.g., Wall Street Journal), analyst reports, or earnings announcements pre- versus post-scandal invites a whole host of confounding factors that are nearly impossible to control for. Plus, we do not know whether the investors who pay attention to or consume news on Seeking Alpha are the same as those who consume these other sources of news. Thus, *any* effect we find (positive, zero, or negative) could be attributed to omitted variable bias or investor heterogeneity, with little hope of signing that bias. Even if we could account for all of these issues, we would still have difficulty interpreting the result because while a negative effect on trading volume associated with these other news events post-scandal *might* be consistent with investors discounting all news after the scandal, a zero effect could also be construed as investors discounting news, but perhaps assuming fake news does not pervade these other news outlets as much, or even a positive effect on volume from other news sources after the scandal could indicate substitution in investor attention across different news producers. In other words, such a test, no matter what the outcome, could be interpreted in the same manner as being consistent with our results and does not help us generalize the spillover effect. A better test would be to find fake news from other media sources and test whether the response on those same media sources is the same before versus after the awareness of that fake news. Alas, we only have one setting – the crowd-sourced financial news platforms – where fake news has been successfully identified.

There is, however, one extension of our results we can examine that helps generalize the findings. As another test of the spillover effect from fake news on other news more generally, the last column of Panel D runs the regression of abnormal trading volume for the sample of articles from Motley Fool only. Since Motley Fool was not part of the scandal, and none of its articles were flagged for failing to disclose paid-for articles as part of a promotional scheme, it is interesting to examine the spillover effect from the scandal on Seeking Alpha to the trading volume response for articles published on Motley Fool. Panel G of Table 3 reports the results and shows that abnormal trading volume declined significantly for Motley Fool articles, too, after the scandal. The results mirror those we find in the previous panels – trading volume declines for all news written on Motley Fool post-scandal and that decline is most significant for small, then medium, then large firms, and is significant even for the largest firms. After the scandal, the difference in trading volume response for probabilistically fake or less authentic content is negligible. These results are very similar to our earlier findings and point to the spillover effect from the scandal extending beyond the specific platform where that scandal occurred. Whether or not these spillover effects extend beyond these shared-knowledge social platforms remains an open question.

Panel H examines how long the spillover response lasts by looking at abnormal trading volume six to twelve months after the scandal and exposé articles. For small and medium firms, the negative trading volume response is still significant six months to a year after the scandal, but for large firms the spillover effect on trading disappears after six months.

Finally, we provide some additional and more direct evidence that the decline in trading volume per published article, 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 variable *Fake Words*, which is a dummy 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 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 – fake news is intended to deceive and therefore is often inaccurate, though could be correct, while inaccurate news may have no deceptive intent. In most of our analysis, we cannot distinguish between these two, but the comments section may provide a glimpse of what consumers on these platforms are concerned about.

Table 4 reports results that look at the language used in the comments to these articles. Panel A examines whether the appearance of *Fake Words* or *Wrong Words* is more prevalent for fake versus non-fake articles over the whole 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 is interesting and suggests that participants on these platforms could not identify or differentiate between fake and non-fake articles. This finding is consistent with the difficulty Facebook, Amazon, and the SEC are having in recognizing and identifying fake content, and, again, is one of the key contributions our unique sample of known fake articles provides. In our setting, users of these platforms were deceived by these articles and there appears to be no indication that consumers of these articles were anymore skeptical of fake articles on these platforms.

The second column of Panel A runs the same regression but uses *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 felt the fake articles were more correct (less wrong) than the non-fake articles. Said differently, the fake articles seem to be more convincing of their statements than the non-fake articles, which may be why they generated more trading volume (and may be why they were used in the promotional campaign).

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 news 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 rather then erroneous news.

Combining the results in Panels A and B, 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 non-fake. As a consequence, we see a marked drop in investor trading volume to 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 more generally.

5. Impact on Returns

The previous section shows the market's response to fake articles through trading activity, finding a significant impact on trading from fake articles on these platforms that diminishes significantly after the promotional scandals became public. In this section, we investigate whether market prices respond to the fake articles. Examining market prices allows us to quantify the economic impact of fake news, which is extremely difficult to do in other settings and hence one of the benefits of analyzing financial markets. This analysis is also a novel test of the informational efficiency of markets, where in a perfectly efficient market fake news will have no impact on prices, even if people trade on it.

5.1. Return Reaction

We start by examining the price reaction to the for-sure fake articles provided by Rick Pearson and subpoenaed by the SEC. We separate firms by size into small and mid-size (there are no large firms in this sample) and examine their return response to the release of for-sure fake articles. The cumulative abnormal returns, measured as equal-weighted residuals from the Fama and French (1993) three-factor model augmented with a momentum factor (using equal-weighted versions of these factors), are constructed starting the day after the article was published until 251 trading days later. Using equal-weighted portfolios of the market, size, value, and momentum factors (*RMRF*, *SMB*, *HML*, and *UMD* factors), from Ken French's website, we estimate betas for each stock i for day t using past daily returns from $t-252$ to $t-1$. We then use those betas to calculate the residual (abnormal) cumulative returns, relative to the same four factors, for stock i for days $t+1$ to $t+251$.

Figure 4 plots the cumulative abnormal returns for the for-sure fake articles for small and mid-size firms. Returns for small firms increase after the article is published, reaching as much as 15%, cumulatively, after about 60 days, before giving up all the gains, and ending with a cumulative negative 10% return by the end of the year. The permanent price impact of -10% 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 10% initially, but the fake news temporarily delayed the decline. For mid-size firms, there is no initial gain in share price – the price starts dropping 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 mid-size firms, where the cost of information is lower.[17] The results suggest that for both small and mid-size firms, the fact that management is trying to prop up the stock price by commissioning fake articles is a

[17] Of course, it's also possible that mid-cap firms' pumping scheme actually works if the returns would have been even worse had they not initiated the fake articles. Hence, another interpretation is that the mid-cap firms fool the market, too, but only do so when other bad news about the firm is present. This narrative is less consistent with the data, however, since we find no evidence of other bad news associated with mid-size firms around the time of the articles.

signal of deteriorating underlying performance. Whether the market subsequently discounts the stock for attempted manipulation or simply recognizes the action as a symptom of poor financial health is indistinguishable in the data. What we can say is that fake articles are associated with long-term negative returns about the firm and that, due to larger limits to arbitrage, a less sophisticated investor base, or higher information costs, the fake articles appear successful at temporarily propping up the stock price of very small firms.

While the articles from Rick Pearson and the SEC's subpoenas are clean and identify fake news, they also constitute a small sample. 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) no fake articles, and plots these price responses separately for small, mid-size, and large firms in our sample (that have at least one fake article). Specifically, we form an equal-weighted portfolio of all firms that have a fake article on day t and an equal-weighted portfolio of all firms that have a non-fake article published on day t, and calculate the residual returns (with respect to the Fama and French (1993) three-factor model augmented with a momentum factor, using equal-weighted versions of the factors) of both portfolios from $t-120$ to $t+251$. Plotted in Figure 5 is the difference between the cumulative returns of the fake article portfolio minus that of the non-fake article portfolio. As the figure shows, among small firms returns to fake articles relative to non-fake articles increase for 6 months by about 8% following publication, and then revert back to their original level. The returns for mid-size firms, however, start dropping almost immediately, and come to a steady state of -5% after about 10 months.

These patterns – small firms experiencing temporary positive returns following fake articles that eventually revert and mid-size firms experiencing an immediate price decrease –

are remarkably similar to the return patterns we found for the for-sure fake articles from the smaller SEC sample (Figure 5). The similarity in results is reassuring and corroborates our methodology and ability to identify fake news. Finally, for large firms, nothing very interesting is happening and the abnormal returns are statistically no different from zero throughout the event. This result makes sense since the market for large firms (the largest 10% on the NYSE) is quite efficient. It may also be the case that the articles we identify as fake among large stocks may not be part of a promotional campaign, but rather may be produced by an independent third party, such as one-off rogue authors. Recall, that the for-sure fake promotional articles from the SEC did not contain any firms of this size. Also, prior to the appearance of fake articles for large firms, there is not a significant decline in stock price or accounting performance, unlike what we see for small and mid-cap stocks prior to fake articles written about them. This circumstantial evidence points to different motivations perhaps driving the production of fake articles of large firms, such as third party producers of those articles unaffiliated with the firm. In the next section, we will test more directly for firm involvement in promotional articles and find no corroborating evidence for large firms, while finding strong evidence for small firms.

We formally test whether the patterns in cumulative abnormal returns for different-sized firms we observe in Figure 5 are statistically significant by estimating the following model:

$$AbnRet_{i,(t+1,t+T)} = \alpha + \beta Fake_{i,t} + \varepsilon_{i,(t+1,t+T)}$$

, where $AbnRet_{i,(t+1,t+T)}$ are cumulative abnormal 4-factor returns for firm i, from one day after the fake article is published until T days, where T = 51, 101, 151, 201, and 251. The results are presented in Table 5. For small firms, the returns in the first 100 days following fake articles are more positive than following non-fake articles, and the difference is statistically significant. This difference disappears after about 10 months and reverts back to zero after a year. For mid-size firms, the returns start decreasing immediately following the publication of fake articles, relative to days with non-fake articles, and continue to decrease for about 10 months, before coming to a steady state at around -4%. Finally, for large firms, the difference is negligible and insignificant.

5.2. Fake articles and firm fundamentals

Although Figure 5 and Table 5 show that the presence of fake articles is usually bad news, especially for mid-size firms, it remains unclear whether the poor returns are due to investors' over/under reaction or whether fake articles are a sign of poor fundamental firm performance. Table 6 examines whether the presence of fake articles is associated with worsening fundamental firm performance, as measured by surprise in unexpected earnings, SUE, which is the seasonally-adjusted change in earnings scaled by the standard deviation of seasonally-adjusted change in earnings over the prior eight quarters; the return on assets, ROA, which is the net income of the firm divided by total assets; and the recent quarterly change in ROA, ΔROA. We regress these performance measures on a fake article dummy equal to one if there was at least one fake article (defined as the probability of the article being fake being > 0.20) in the previous 90 days leading up to the earnings announcement, and zero otherwise. We only include firms in this analysis that had at least one fake article in our sample, and include firm and year-month fixed-effects to the regression.

The results in Table 6 show that the presence of at least one fake article during the quarter is associated with a 0.111 lower SUE disclosed at the end of the quarter. This is a 0.1 standard deviation decrease. In the next three columns we examine the effect separately for small, midsize, and large firms. For small and medium firms the effect is negative, highly statistically significant for medium firms, and insignificant for small firms, though the magnitude of the point estimate for small firms is more than twice that of medium firms. The lack of significance for small firms may be due to low power given the smaller number of observations. Economically, the effect is much bigger for small firms than it is for mid-size firms. For large firms, there is no effect, economically or statistically. Similar results are found using ROA and ΔROA. These results mirror the effect of fake articles on abnormal returns in the last section, and suggest that fake articles are a sign of bad firm performance for small and mid-size firms, but not a signal of financial health for large firms. 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 we investigate next.

6. What Motivates Fake News?

Fake news is designed to deceive for personal or financial 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 would motivate fake news and more likely financial motivations are behind it. Indeed, the SEC investigation focused on promotional articles that were part of a pump-and-dump scheme to defraud securities markets. Through a variety of tests, we have shown that these fake articles induce abnormal trading and temporarily drive up the prices of small stocks, whose recent prior performance was deteriorating. These patterns are consistent with a motivation to hire authors to write fake content to prop up the stock price.

In this section, we investigate in the broader sample of articles, what other actions the firm may take to augment the promotion articles, and what incentives managers have to pump up their stock price. In particular, we examine whether companies are more likely to issue press releases or 8-K filings to accompany the fake articles to give authors of the fake articles more material and credibility. We also seek to identify the pump-and-dump schemes, where one acquires shares at a low price, then inflates the price through fake articles, and then sells the stock, by looking at insider trading from the SEC's Form 4.

As described in Section 3, for the small sample of cases we obtained from the SEC, there is evidence of a coordinated promotion campaign, including press releases and insider trading to profit from the promotion, which is what chiefly caught the regulator's attention. For our broader sample of articles, where we probabilistically assess the occurrence of fake news using the linguistic algorithm, we investigate whether we can find corroborating evidence of corporate actions and insider trading consistent with the promotional motivation behind the articles. This examination also serves as a robustness test of our methodology's ability to detect promotional fake articles more broadly.

6.1. Firm Disclosures

Fake articles promoted by the firm are likely to be more credible if accompanied by a press release or filing of material information, which can also provide some facts for the author to write about and potentially embellish. To test this conjecture, we regress whether there were fake articles in a given week, on whether there was a press release or an SEC 8-K filing in the prior week, the week of, and a week after the fake articles are published.[18] Table 7 reports results separately for small, mid-size, and large firms.

We find that small firms are substantially more likely to have fake articles written about them in the week before, the week of, and the week after they issue a press release or file an 8-K form with the SEC. The coefficients become insignificant if we go out further than those weeks. Mid-size firms have an increased probability of having fake articles written about them in the week of the press release or an 8-K filing as well, but there is no effect for large firms. These results are also consistent with the anecdotal evidence that companies often issue press releases to provide some material for the fake articles.

6.2. Insider Trading

We next examine insider trading to see if insiders in the firm are positioning themselves to benefit from the price impact of the promotion campaign. We regress an indicator variable for whether a firm had predominantly fake articles in a given week on whether insiders were net buyers or net sellers in the week before, the week of, and the week after the fake article is published. *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 perform our analysis separately for small, mid-size, and for large firms.

[18] An 8-K form must be filed with the SEC if a material event occurs at the firm within five business days.

The results are presented in Table 8. For small firms and mid-size firms, insiders start buying shares in the week before the fake article appears (prior weeks show no activity), and then start actively buying the week of and the week after the fake articles are published. These findings are similar to the case study of Galena (Figure 3), where insiders started buying the stock around and shortly after the fake articles come out. As in the Galena case, these campaigns comprise a sequence of fake articles that cumulatively affect prices, and insiders appear to be buying more shares as these articles come out. We do not find any insider trading activity for large firms, consistent with there being no price effect for large firms.

6.3. Insider Trading and Returns

Finally, we examine whether the impact of fake articles on returns is even stronger when insiders purchase stock. We separate articles by whether the firm was a net inside buyer in the two weeks leading up to the article being published. We concentrate on small firms, where the activity primarily takes place, and examine fake and non-fake articles separately to difference out the effect of insider buying generally. The results are presented in Figure 6. In Panel A, we examine the effect of fake articles on returns, separated into articles that followed insider buying versus those that did not. The figure shows that fake articles published following heavy insider buying are associated with prior poor stock price performance that seems to temporarily prop up the stock price significantly. However, we do not see a similar pattern for fake articles that did not follow insider purchases. The difference in returns leading up to the publication of fake articles is very different for the two samples. The articles published following insider purchases are preceded by very sharp drops in share price in the month before publication. The fake articles not associated with insider purchases have flat to lightly increasing returns before publication.

After publication, the share price rises significantly for firms whose insiders are buying and then crashes back down after 150 days. For firms with fake articles written about them that do not have insiders buying shares, there is a small price increase that also turns negative at around six months. These patterns suggest that the articles accompanied by insider purchases are more likely to be orchestrated by the firm to prop up or stabilize prices, and thus come at a very specific time, and might be accompanied by other promotional attempts. The fake articles that are not accompanied by insider purchases might be attempts by third parties to manipulate the stock price for their own benefit without assistance from the firm, and hence seem to be less successful.

To address that the results are not just driven by insider trading per se, and that the fake articles themselves have impact, Panel B performs a similar analysis using only the non-fake articles. As the graph for non-fake articles shows, there no difference in returns for non-fake articles with insider buying versus without insider buying. Hence, it is unlikely insider buying per se is causing the differences in returns we see in Panel A for fake articles. Rather, it is the combination of insider buying with fake articles that seems to matter and is indicative of a comprehensive promotional campaign to pump up the stock price.

7. Conclusion

Using a unique dataset of fake paid-for articles obtained from an SEC investigation, we overcome one of the impediments to analyzing the impact of fake news empirically – identifying fake content. Investigating these specific cases as well as applying a linguistic algorithm to a much larger set of news content using the sample of known fake articles to verify the algorithm, we find increases in abnormal trading volume and temporary price impact following fake news for small firms, but no impact for large firms. Following public revelation of the SEC's investigation, we find a significant spillover effect to news generally, where investors react less to *all* news, even legitimate news, following the scandal. These findings represent some of the first documented effects of fake news and are consistent with theory on the potential impact of fake news (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 the ability to take action to correct it is more limited (e.g., online consumer retail, political news, elections, and social media). More broadly, our findings may have implications for news media generally (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 (Sapienza and Zingales)).

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. 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.

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.

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 stroÌĹmberg and joel waldfogel, eds.

Grossman, S. J. and J. E. Stiglitz (1980). On the impossibility of informationally efficient markets. *The American Economic Review 70*(3), 393–408.

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 âĂIJfake newsâĂİ. *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. A trust crisis. *International Review of Finance 12*(2), 123–131.

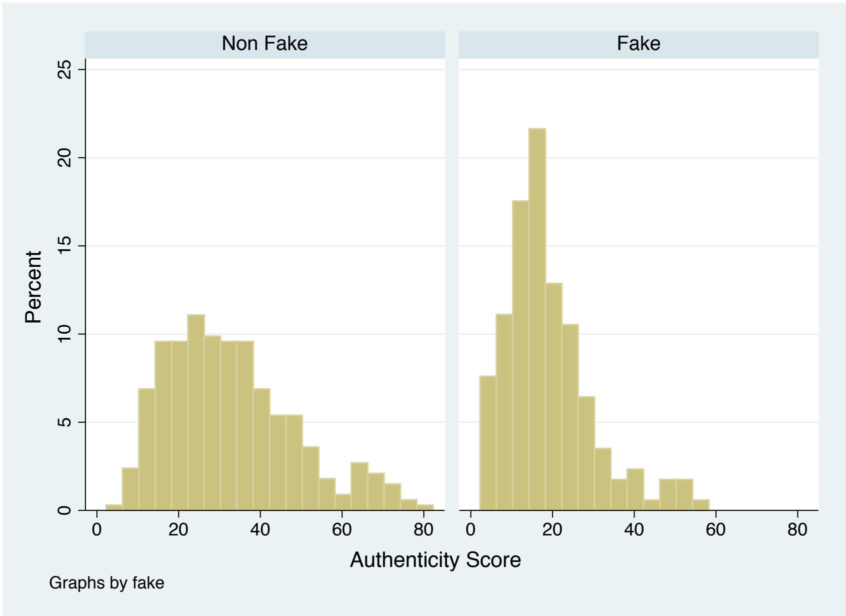
Tetlock, P. C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance 62*(3), 1139–1168.

Vosoughi, S., D. Roy, and S. Aral (2018). 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:



Panel B:

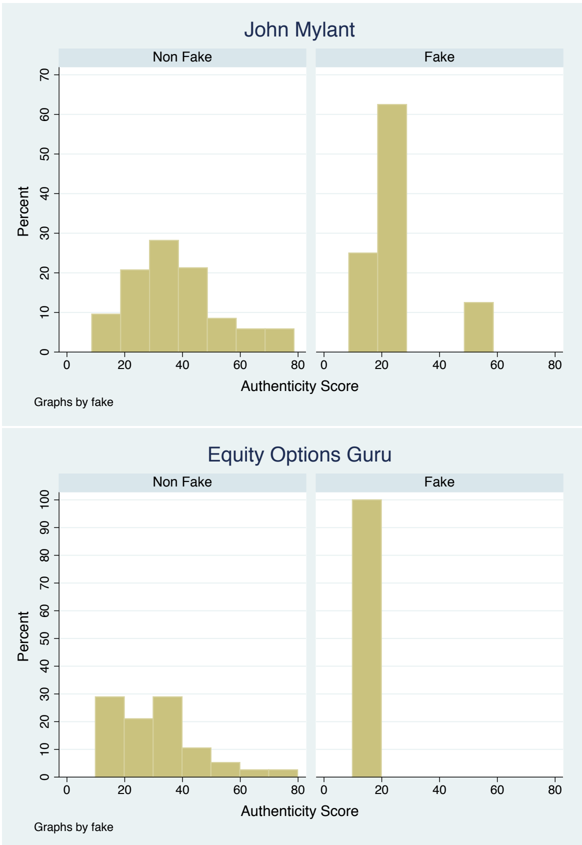


Figure 2. **Authenticity score and the probability of being fake**

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

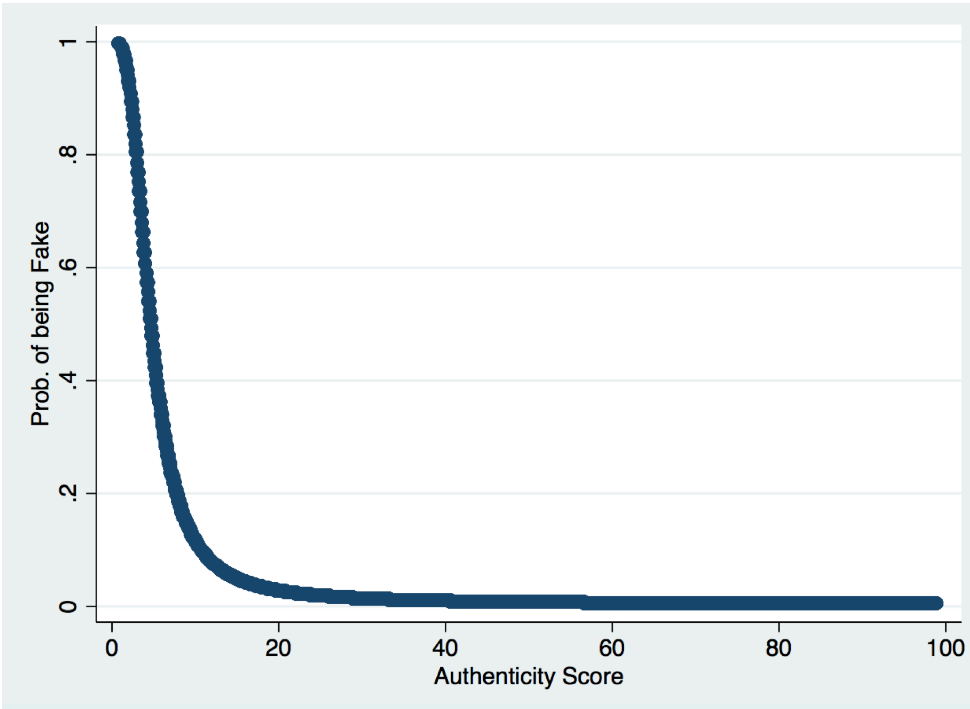


Figure 3. **Example of a Pump-and-Dump Scheme: Galena Biopharma Inc.**

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)[19].
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.

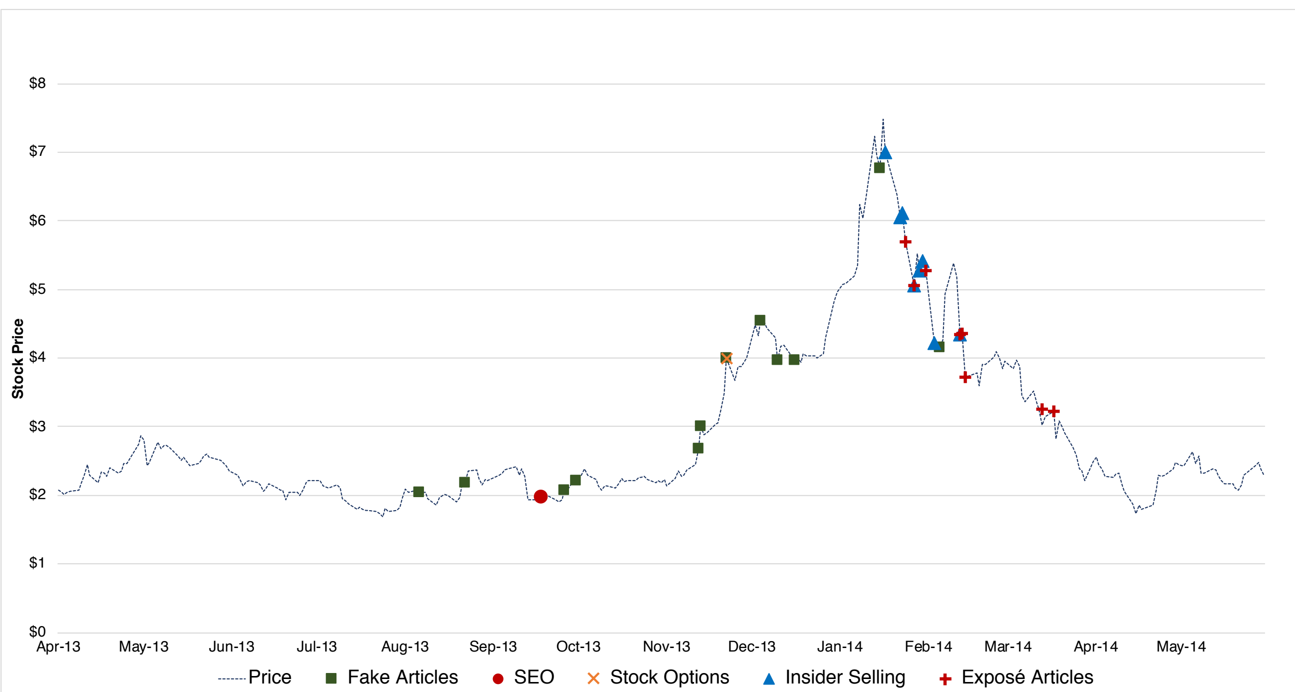


Figure 4. **Abnormal Returns for For-Sure Fake Articles**

The figure depicts cumulative abnormal returns (measured as equal-weighted 4-factor residuals) for for-sure fake articles that were provided to us by Rick Pearson and that were subpoenaed by the SEC. 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 -120 and ending on the day before the article publication. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, medium firms are defined as firms in the 20th-90th percentile of NYSE firms.

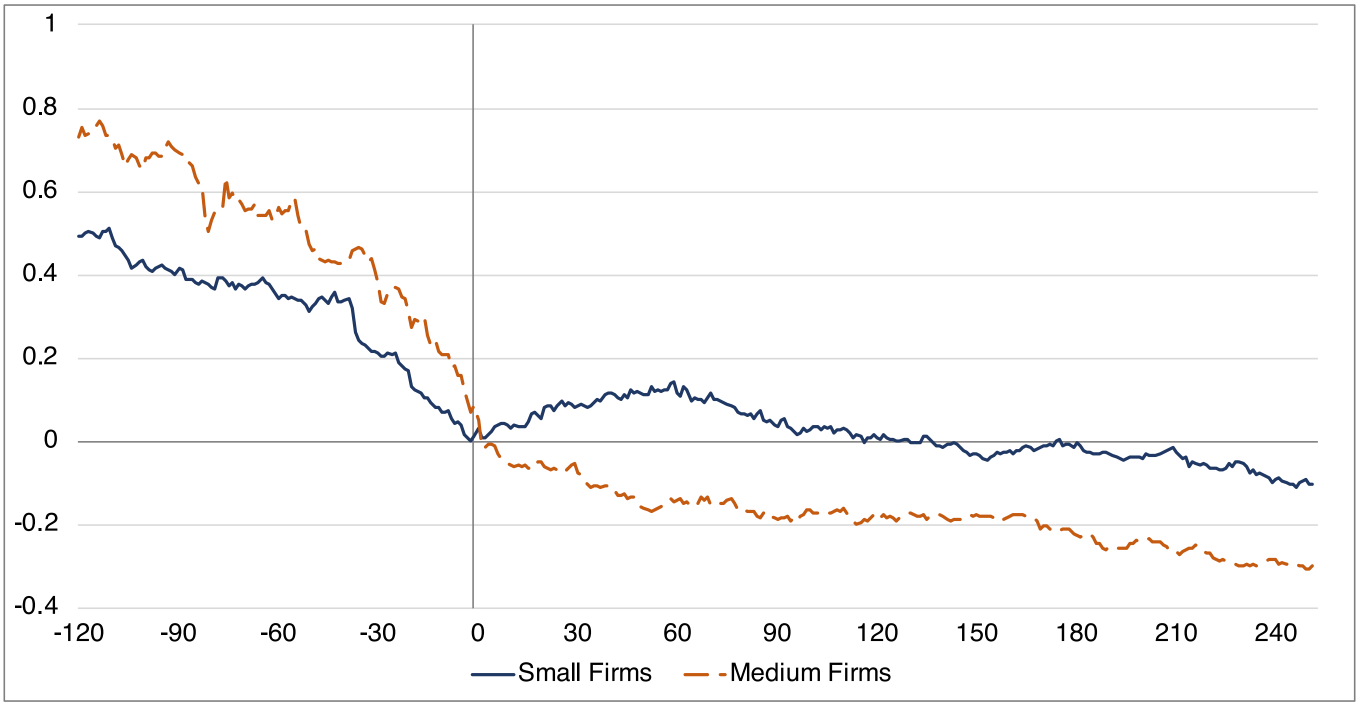


Figure 5. **Abnormal Returns for Fake versus Non-Fake Articles**

The figure depicts the 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, mid-size, and large firms in our sample. 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 -120 and ending on the day before the article publication. Small firms are defined as firms in the bottom 10th percentile of NYSE firms, medium 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.

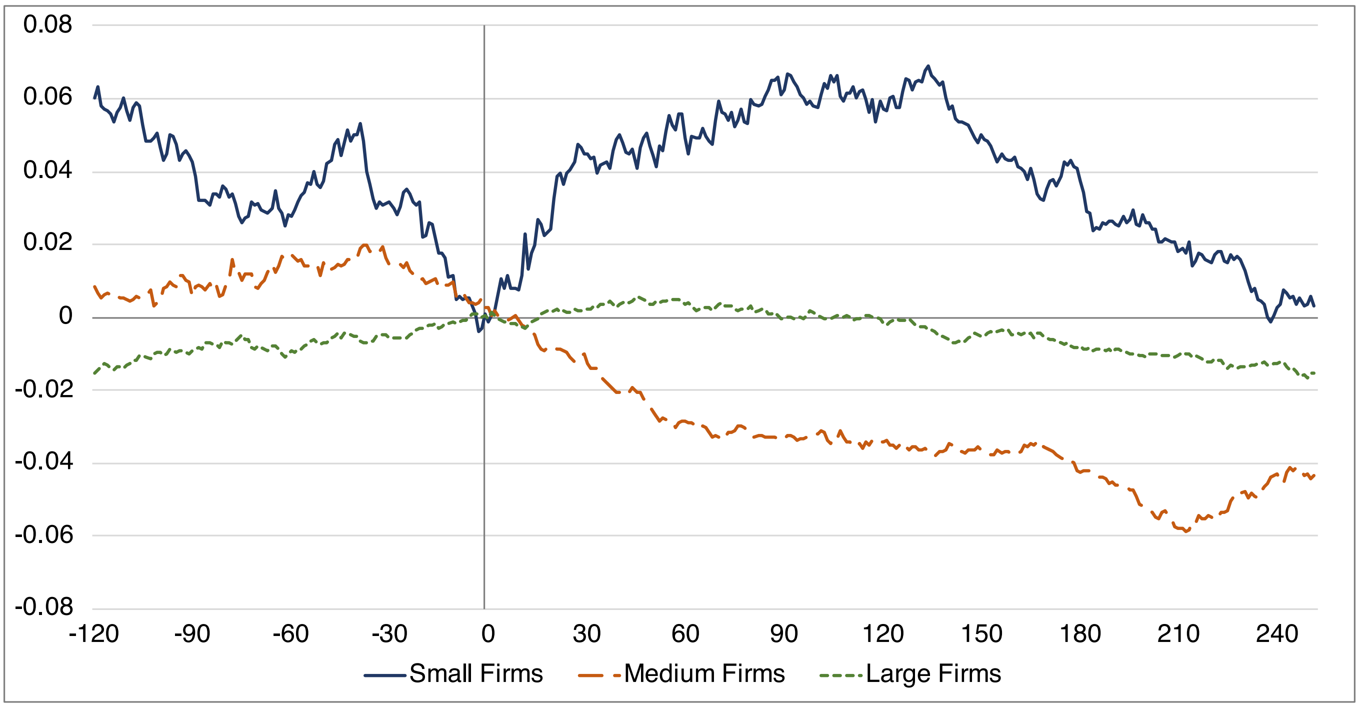


Figure 6. **Abnormal Returns and Insider Trading**

The two figures below depict for small firms the cumulative abnormal returns (measured as equal-weighted 4-factor residuals) for days with articles where the firm was a net insider buyer in the two weeks leading up to the article versus days with articles where the firm was not a net insider buyer. A firm is a net insider buyer if the dollar value of stock bought by insiders is larger than the value of stock sold by insiders. Panel A shows the results for days with fake articles and Panel B shows the results for days with non-fake articles. 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 having a non-fake article, 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 -120 and ending on the day before the article publication. Small firms are defined as firms in the bottom 10th percentile of NYSE firms.

Panel A: Fake Articles with/without Insider Buying for Small Firms

Panel B: Non-Fake Articles with/without Insider Buying for Small Firms

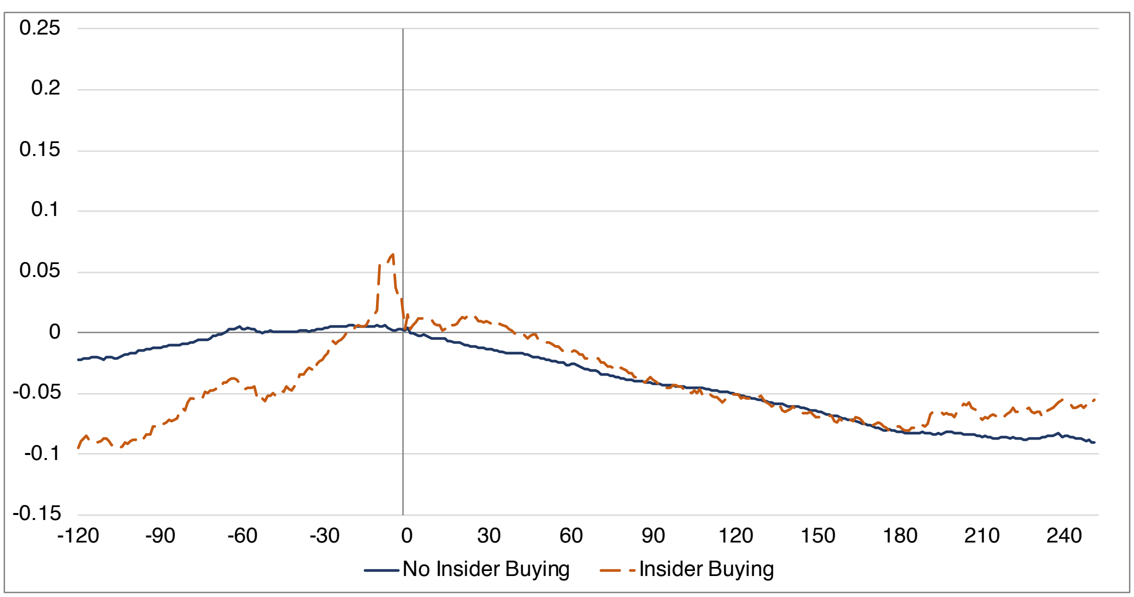


Table 1. **Summary Statistics**

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. We also display the *Clout* measure, which is meant to capture dominance in language. 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
Panel B: Probability of being Fake								
Prob(Fake)	0.08	0.02	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. **Article Impact on Abnormal Volume**

The table examines how investor attention (proxied for by log of abnormal volume on days $t = 0$, $t + 1$, and $t + 2$), responds to fake and non-fake articles as a function of article and author characteristics, abnormal volatility, firm size, and fraction of retail investors. Panel A shows results for all firms that have ever had an article written about them on Seeking Alpha or Motley Fool. Panel B concentrates on days with articles. Abnormal volume is defined as $Vol(t)/AvgVol(t - 250, t - 1)$, summed over days $t = 0$, $t + 1$, and $t + 2$, and then we take the natural log of the sum. *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. *Author Impact* is the average idiosyncratic volatility over days [t, t+2] observed for the author after the release of all prior articles. 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 both firm and year-month fixed effects when firm size or percentage of retail investors are not used as explanatory variables, except for Panel D where we include only year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

Panel A

Dependent variable	$Ln([t, t + 2]$ day abnormal volume)			
		Firm Size		
	All	Small	Medium	Large
Article dummy	0.121***	0.651***	0.183***	0.106***
	(67.73)	(21.8)	(78.96)	(64.78)
Observations	13,445,533	716,815	11,164,898	1,563,820
R^2	0.60	0.46	0.58	0.64

Panel B

Dependent variable	$Ln([t, t + 2]$ day abnormal volume)			
	For-sure Fake		Fake	
Fake News	-0.025	0.034***	0.019	0.159**
	(-0.11)	(3.15)	(0.27)	(2.42)
Author Impact			0.099***	
			(53.50)	
Fake News × Author Impact			-0.000	
			(-0.05)	
ME percentile				-0.022***
				(-256.23)
Fake News × ME percentile				-0.001*
				(-1.69)
Percentile retail investors				0.455***
				(42.55)
Fake News × Percentile retail investors				0.283***
				(3.69)
Observations	363	190,298	185,296	173,455
R-squared	0.92	0.78	0.79	0.31

Panel C

Dependent variable	$ln[(t, t+2]$ idiosyncratic volatility)			
	For-sure Fake		Fake	
Fake News	-1.039	0.037	-0.329**	0.263**
	(-1.50)	(1.57)	(-2.09)	(2.52)
Author Impact			0.234***	
			(56.61)	
Fake News × Author Impact			-0.047**	
			(-2.17)	
ME percentile				-0.022***
				(-164.68)
Fake News × ME percentile				-0.001
				(-1.54)
Percentile retail investors				-0.220***
				(-12.95)
Fake News × Percentile retail investors				0.144
				(1.19)
Observations	205	194,737	189,512	176,845
R-squared	0.74	0.45	0.46	0.17

Panel D

Dependent variable	$Ln([t, t+2]$ day abnormal volume)			
	For-sure Fake		Fake	
Fake News	1.815***	0.212***	0.850***	0.159**
	(8.07)	(9.78)	(6.84)	(2.42)
Author Impact			0.354***	
			(105.01)	
Fake News × Author Impact			0.104***	
			(6.06)	
ME percentile				-0.022***
				(-256.23)
Fake News × ME percentile				-0.001*
				(-1.69)
Percentile retail investors				0.455***
				(42.55)
Fake News × Percentile retail investors				0.283***
				(3.69)
Observations	461	191,559	186,572	173,455
R-squared	0.35	0.02	0.07	0.31

Table 3. **2014 SEC Lawsuit Event Study**

The table examines how fake article intensity and firm-level abnormal volume after the release of articles (days $[t, t+2]$), change around the 2014 exposé articles and SEC lawsuit. The exposé articles and the lawsuit occurred in February and March of 2014. We study the 6-month time periods before and after February and March of 2014. Abnormal volume is defined as $Vol(t)/AvgVol(t-250, t-1)$, summed over days $t=0$, $t+1$, and $t+2$, and then we take the natural log of the sum. *Post Event* is defined as the 6-month time period after February and March, 2014. *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. *Author Impact* is the average idiosyncratic volatility over days [t, t+2] observed for the author after the release of all prior articles. 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. *Retail Ownership* is divided into high and low, depending on whether the fraction of retail owners was above or below the median fraction of retail ownership in the firm's size decile on August, 2013. *Healthcare* industry membership is according to Fama-French's 12 industry classification. Panel G reports results for articles posted on Motley Fool only. Panel H reports results over a longer window period (6-12 months after the event window). We include firm fixed-effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

Panel A

Dependent variable	Fake News							
			Firm Size					
	All Firms		Small		Medium		Large	
Post event	-0.001	-0.093***	-0.041***	-0.287***	-0.002	-0.078***	0.002	-0.030***
	(-1.37)	(-11.17)	(-3.47)	(-4.86)	(-1.27)	(-5.93)	(1.60)	(-2.59)
Author Impact		0.011***		0.040***		0.010***		0.003***
		(12.14)		(7.26)		(6.57)		(2.70)
Post event× Author Impact		-0.012***		-0.036***		-0.010***		-0.004***
		(-11.11)		(-4.19)		(-5.77)		(-2.82)
Observations	33467	32582	1592	1492	13922	13525	17953	17565
R-squared	0.31	0.32	0.69	0.72	0.29	0.30	0.10	0.11

Panel B

Dependent variable	Fake News							
	Retail Ownership				Industry Classification			
	Low Retail		High Retail		Healthcare		Non-Healthcare	
Post event	0.002	-0.032***	-0.001	-0.146***	-0.036***	-0.302***	-0.000	-0.071***
	(1.38)	(-2.64)	(-0.58)	(-11.04)	(-4.04)	(-5.44)	(-0.18)	(-8.62)
Author Impact		0.005***		0.016***		0.039***		0.009***
		(3.86)		(11.24)		(7.23)		(9.92)
Post event× Author Impact		-0.004***		-0.019***		-0.036***		-0.009***
		(-2.84)		(-11.10)		(-4.95)		(-8.67)
Observations	15449	15055	15582	15177	1514	1481	32432	31557
R-squared	0.30	0.30	0.24	0.25	0.51	0.55	0.32	0.33

Panel C

Dependent variable	Log Abnormal Volume [t, t+2]							
	Firm Size							
	All Firms		Small		Medium		Large	
Post event	-0.198***	-0.262***	-0.656***	-1.188***	-0.179***	0.323***	-0.162***	-0.145***
	(-28.32)	(-6.26)	(-9.73)	(-5.25)	(-13.74)	(4.37)	(-19.64)	(-2.68)
Fake article	0.148**		0.333*		-0.007		-0.219**	
	(2.46)		(1.71)		(-0.07)		(-2.32)	
Post event× Fake article	-0.265***		-0.669		-0.079		0.086	
	(-3.49)		(-1.21)		(-0.60)		(0.80)	
Author impact		0.089***		0.114***		0.087***		0.046***
		(19.51)		(4.88)		(10.26)		(8.25)
Post event× Author impact		-0.010*		-0.079**		0.066***		0.000
		(-1.87)		(-2.33)		(6.67)		(0.06)
Observations	32629	53182	1477	2658	13473	21055	17679	29469
R-squared	0.87	0.87	0.84	0.79	0.82	0.82	0.80	0.80

Panel D

Dependent variable	Log Abnormal Volume [t, t+2]							
	Retail Ownership				Industry Classification			
	Low Retail		High Retail		Healthcare		Non-Healthcare	
Post event	-0.220***	0.272***	-0.172***	-0.728***	-0.108***	-0.627***	-0.201***	-0.236***
	(-21.72)	(4.69)	(-15.10)	(-11.18)	(-3.19)	(-4.10)	(-28.24)	(-5.41)
Fake article	-0.099		0.423***		0.341***		0.103	
	(-1.03)		(4.73)		(2.77)		(1.49)	
Post event× Fake article	-0.103		-0.488***		-0.463**		-0.219***	
	(-0.89)		(-4.39)		(-2.07)		(-2.62)	
Author impact		0.066***		0.110***		0.099***		0.088***
		(10.45)		(15.10)		(6.07)		(18.54)
Post event × Author impact		0.063***		-0.074***		-0.060***		-0.007
		(8.30)		(-8.70)		(-2.97)		(-1.17)
Observations	15415	24094	15151	25651	1512	3062	31218	50310
R-squared	0.86	0.86	0.87	0.86	0.89	0.87	0.87	0.87

Panel E

Dependent variable	Log Idiosyncratic Volatility [t, t+2]							
	Firm Size							
	All Firms		Small		Medium		Large	
Post event	-0.317***	0.645***	-0.178	0.169	-0.137***	1.722***	-0.401***	-0.499***
	(-17.26)	(6.01)	(-1.50)	(0.42)	(-4.10)	(9.48)	(-17.08)	(-3.24)
Fake article	-0.120		0.278		-0.188		-0.325	
	(-0.77)		(0.81)		(-0.76)		(-1.23)	
Post event × Fake article	-0.167		-2.243**		-0.223		0.138	
	(-0.84)		(-2.01)		(-0.67)		(0.46)	
Author impact		0.155***		0.061		0.171***		0.183***
		(13.10)		(1.40)		(8.25)		(11.33)
Post event × Author impact		0.119***		0.053		0.246***		-0.019
		(8.45)		(0.88)		(10.09)		(-0.97)
Observations	33034	53843	1567	2856	13698	21425	17769	29562
R-squared	0.51	0.50	0.72	0.66	0.47	0.47	0.32	0.30

Panel F

Dependent variable	Log Idiosyncratic Volatility [t, t+2]							
	Retail Ownership				Industry Classification			
	Low Retail		High Retail		Healthcare		Non-Healthcare	
Post event	-0.231***	1.508***	-0.418***	-0.189	-0.135	-0.268	-0.325***	0.655***
	(-7.64)	(8.80)	(-14.98)	(-1.22)	(-1.50)	(-0.66)	(-17.30)	(5.90)
Fake article	-0.123		0.015		0.370		-0.232	
	(-0.43)		(0.07)		(1.12)		(-1.32)	
Post event × Fake article	-0.399		-0.102		-0.254		-0.085	
	(-1.16)		(-0.37)		(-0.43)		(-0.39)	
Author impact		0.158***		0.149***		0.145***		0.155***
		(8.46)		(8.61)		(3.32)		(12.67)
Post event × Author impact		0.219***		0.023		-0.023		0.121***
		(9.79)		(1.13)		(-0.43)		(8.35)
Observations	15285	23920	15485	26123	1502	3050	31891	51402
R-squared	0.48	0.47	0.55	0.53	0.58	0.55	0.52	0.51

Panel G

Dependent variable	Log Abnormal Volume [t, t+2]							
	Firm Size							
	All Firms		Small		Medium		Large	
Post event	-0.345***	0.018	-1.316***	0.197	-0.329***	0.299*	-0.300***	-0.090
	(-25.76)	(0.21)	(-8.53)	(0.19)	(-11.39)	(1.78)	(-18.93)	(-0.90)
Fake article	-0.737**		-0.275		-0.270		-0.698**	
	(-2.33)		(-0.61)		(-1.51)		(-2.37)	
Post event × Fake article	0.404						0.340	
	(1.26)						(1.13)	
Author impact		0.057***		0.063		0.117***		0.019
		(5.65)		(0.44)		(5.55)		(1.64)
Post event × Author impact		0.044***		0.211		0.085***		0.025*
		(3.99)		(1.44)		(3.81)		(1.95)
Observations	15438	26283	333	460	6044	9284	9061	16539
R-squared	0.89	0.88	0.89	0.89	0.87	0.86	0.83	0.82

Panel H

Dependent variable	Log Abnormal Volume [t, t+2]							
	Firm Size							
	All Firms		Small		Medium		Large	
Post event	-0.072***	-0.675***	-0.761***	-2.181***	-0.096***	-0.317***	0.016	-0.045
	(-8.37)	(-13.35)	(-9.21)	(-8.43)	(-6.29)	(-3.80)	(1.61)	(-0.71)
Fake article	0.112*		0.077		-0.059		-0.144	
	(1.67)		(0.36)		(-0.55)		(-1.43)	
Post event × Fake article	-0.003		-0.550		0.230*		0.246**	
	(-0.04)		(-1.42)		(1.86)		(2.02)	
Author impact		0.101***		0.137***		0.081***		0.049***
		(19.47)		(5.23)		(8.73)		(8.25)
Post event × Author impact		-0.081***		-0.195***		-0.033***		-0.010
		(-12.19)		(-5.15)		(-2.97)		(-1.25)
Observations	22100	35545	1486	2784	9913	15972	10701	16789
R-squared	0.86	0.85	0.85	0.82	0.82	0.81	0.78	0.78

Table 4. **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 textitWrong 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 level respectively. T-statistics are in parentheses.

Panel A

	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

	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 5. **Return Window Regressions – Unconditional**

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. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

	$Ret_{1,51}$	$Ret_{1,101}$	$Ret_{1,151}$	$Ret_{1,201}$	$Ret_{1,251}$
			Small Firms		
Fake Article	0.034	0.063***	0.055*	0.027	0.017
	(1.61)	(2.66)	(1.85)	(0.77)	(0.45)
Constant	-0.022***	-0.045***	-0.064***	-0.078***	-0.086***
	(-8.81)	(-12.62)	(-13.48)	(-13.20)	(-12.25)
Observations	11,622	11,622	11,622	11,622	11,622
R^2	0.000	0.000	0.000	0.000	0.000
			Medium Firms		
Fake Article	-0.017***	-0.020**	-0.028**	-0.045***	-0.038**
	(-3.04)	(-2.45)	(-2.50)	(-3.51)	(-2.50)
Constant	-0.006***	-0.012***	-0.017***	-0.025***	-0.031***
	(-7.76)	(-11.45)	(-12.21)	(-15.07)	(-16.05)
Observations	68,087	68,087	68,087	68,087	68,087
R^2	0.000	0.000	0.000	0.000	0.000
			Large Firms		
Fake Article	0.006*	0.004	0	-0.007	-0.011
	(1.71)	(0.90)	(0.06)	(-0.90)	(-1.33)
Constant	0.001**	0	-0.003***	-0.004***	-0.005***
	(2.26)	(-0.23)	(-2.77)	(-3.96)	(-4.02)
Observations	47,908	47,908	47,908	47,908	47,908
R^2	0.000	0.000	0.000	0.000	0.000

Table 6. **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	-0.111**	-0.373	-0.150**	0.033	-0.002**	-0.008	-0.003***	0.000
	(-2.53)	(-1.52)	(-2.42)	(0.49)	(-2.29)	(-1.00)	(-2.78)	(0.72)
Observations	32,315	5,314	21,064	5,858	31,805	5,170	20,731	5,829
R-squared	0.114	0.196	0.130	0.153	0.594	0.641	0.533	0.460
	ΔROA							
	All Firms	Small	Medium	Large				
Fake Article	-0.002	-0.000	-0.004**	0.001				
	(-1.28)	(-0.02)	(-2.20)	(0.84)				
Observations	30,561	4,797	19,897	5,794				
R-squared	0.058	0.124	0.084	0.080				

Table 7. **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 8. **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

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 $50Ũ$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.

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

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. **Proxies for Article Impact and Abnormal Volume**

The table examines how investor attention (proxied for by log of abnormal volume on days $t = 0$, $t + 1$, and $t + 2$), responds to fake and non-fake articles as well as as to proxies for article impact such as the number of a comments the article received, the number of followers the author had, and the number of users signed up to receive articles about the firm. Abnormal volume is defined as $Vol(t)/AvgVol(t-250, t-1)$, summed over days $t = 0$, $t + 1$, and $t + 2$, and then we take the natural log of the sum. *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. We include both firm and year-month fixed effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

Dependent variable	Log Abnormal Volume [t, t+2]					
ln(num of comments)	0.015***			0.007***		
	(9.35)			(3.47)		
ln(num of followers)		0.013***			0.013***	
		(16.07)			(12.11)	
ln(emailed to)			-0.030***			-0.033***
			(-10.88)			(-9.09)
Fake News				-0.005	0.013	-0.117**
				(-0.30)	(0.31)	(-2.17)
Fake News × impact variable				0.027***	0.002	0.015***
				(2.74)	(0.35)	(2.69)
Observations	188159	189707	187311	110635	111585	110104
R-squared	0.78	0.78	0.78	0.78	0.78	0.78

Table C3. **Individual-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.

Dependent variable	Log Abnormal Volume [t, t+2]				Number of Clicks (ln)	Number of Reads (ln)
Number of clicks (ln)	0.053***		-0.137***			
	(10.68)		(-6.24)			
Number of reads (ln)		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 C4. **Daily Response of Trading Volume to Fake and Non-Fake News in First Three Days**

The table decomposes how the daily response in abnormal firm-level trading volume to the release of articles change around the 2014 exposé articles and SEC lawsuit. The exposé articles and the lawsuit occurred in February and March of 2014. We study the 6-month time periods before and after February and March of 2014. Abnormal volume is defined as $Vol(t)/AvgVol(t-250, t-1)$, where we regress separately the abnormal value of days $t = 0$, $t+1$, $t+2$, $t+3$ on the fake news and post-event dummies. *Post Event* is defined as the 6-month time period after February and March, 2014. *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. We include firm fixed-effects. *, **, and *** indicate statistical significance at the ten, five, and one percent level respectively. T-statistics are in parentheses.

Dependent variable	Log Abnormal Volume					
	t_0	t_1	t_2	t_3	t_4	t_5
Post event	-0.029***	-0.028***	-0.030***	-0.025***	-0.020***	-0.021***
	(-4.74)	(-6.84)	(-7.15)	(-8.12)	(-5.19)	(-8.61)
Fake News	0.286***	0.259***	0.273***	0.057**	0.186***	0.110***
	(5.38)	(7.46)	(7.50)	(2.12)	(5.58)	(5.25)
Post event × Fake Article	-0.298***	-0.273***	-0.277***	-0.067**	-0.185***	-0.107***
	(-4.45)	(-6.23)	(-6.06)	(-2.00)	(-4.43)	(-4.06)
Constant	0.143***	0.112***	0.100***	0.086***	0.076***	0.073***
	(28.12)	(33.86)	(28.80)	(34.02)	(23.97)	(36.52)
Observations	32635	32639	32640	32638	32643	32643
R-squared	0.95	0.96	0.94	0.87	0.45	0.55