Fake News in Financial Markets

Shimon Kogan*
MIT Sloan School of Management
Interdisciplinary Center Herzliya

Tobias J. Moskowitz
Yale University
School of Management
NBER
AQR

Marina Niessner
Yale University
School of Management

October 2017

Preliminary and incomplete. Do not cite without permission.

Abstract

Using a unique dataset of fake news articles from social media outlets, we conduct a novel test of market efficiency. We examine the impact of false information on prices, which should be zero in an efficient market and thus circumvents the joint hypothesis problem. Our experiment is the flip side of the classic event study tests, where we conduct a series of "non-event" studies. While social media platforms, blogs, and other unmonitored media outlets are becoming a main source of news for many people, they also offer scope for providing misleading or false information. We use two datasets to estimate the prevalence and impact of fake news. The first is a set of paid-for articles obtained from an SEC investigation that are known to be false. The second dataset uses a linguistic algorithm, validated by the known fake articles, that quantifies the probability of ŞfakeŤ news on a much larger set of articles. We find strong temporary price impact and subsequent reversals from the fake news articles for small firms, permanent negative price impact for mid-size firms, and no impact for large firms. We also find that the prevalence of fake articles and their price impact are stronger for more volatile firms that have more retail ownership, less media coverage, and less attention. In addition, we find that insider trading and firm press releases at small and mid-cap

* Emails: skogan@mit.edu, tobias.moskowitz@yale.edu, and marina.niessner@yale.edu

firms coincide with the release of fake news and more strongly predict the price reaction and reversal associated with the fake articles, hinting such firms are possibly engaging in stock price manipulation. No such patterns are found for large firms. The evidence is most consistent with markets being efficient in large and possibly mid-cap stocks, and less efficient in small stocks, where the cost of information is high enough for false news to influence prices temporarily.

1. Introduction

Using a unique dataset of fake news articles, we conduct a novel test of market efficiency. Rather than try to estimate information directly and its impact on prices, which is a product of both the cost of information and its interpretation and value placed by investors, we examine the impact of fake or erroneous information on prices. In a perfectly efficient financial securities market, where the cost of information approaches zero, misinformation will have no impact. As a result, our experiment provides a novel test of market efficiency that avoids the joint hypothesis problem by focusing on a *non-information* event – fake news – where price impact should be precisely zero under any equilibrium model. Our experiment is the flip side of the classic event study test of market efficiency pioneered by Fama, Fisher, Jensen, and Roll (1969), where we conduct a series of "non-event" studies. Instead of measuring the price response to a news event, requiring a benchmark model for prices, we examine the (lack of) price response to false news events, where the price response should be zero under any asset pricing model.

A byproduct of our analysis examines the role new shared information platforms might play in information transmission for financial markets. Social media platforms, blogs, and other unmonitored media outlets are vastly becoming a main source of news for many people.[1] While such platforms can enhance the speed with which information is disseminated and lower the cost of obtaining information, they also offer scope for providing misleading or false information. One prominent example is the proliferation of "fake news," defined as hoaxes, frauds, or deceptions designed to mislead consumers of information.[2] This issue has become important enough that Google, Facebook, Wikipedia, and others are heavily investing to curb the dissemination of fake news.

While the existence of fake news may be problematic in certain settings, it may not matter for financial markets because an efficient market, where the cost of information is

[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.

[2]Facebook defines fake news as "hoaxes shared by spammers" for personal or monetary reasons.

close to zero, should quickly identify the news as fake and have no bearing on prices. On the other hand, finding a significant price response to fake news suggests that markets may not be fully efficient and that the cost of information (for that security) may be significant. 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 from psychological or behavioral biases (e.g., a processing cost that can include psychological barriers to interpreting information such as confirmation bias, inattention, anchoring, framing, and over- and underreaction). While the marginal cost of information is at the heart of determining how informationally efficient financial markets are (Grossman and Stiglitz (1980)), and as a consequence how profitable active investment might be, attempts to estimate such costs are empirically elusive. Our study may perhaps provide a glimpse into the cost of information and speed of price discovery across firms from the price impact of fakes news.

Our sample consists of two datasets to estimate the impact of fake news. The first is a unique dataset of paid-for articles obtained from an industry "whistle-blower", Rick Pearson, a regular contributor on Seeking Alpha, a crowd-sourced content service for financial markets. Mr. Pearson went undercover to investigate other authors on the site and uncovered fake paid-for articles now being investigated by the SEC. The sample size is small, but the identity of fake news is clean – 111 articles by 12 authors covering 46 companies. We compare these to all other articles written by the same authors (336 in total) that were published on the same platform that may not have been fake – covering 171 stocks in total. This first sample represents our cleanest experiment, where there is no ambiguity in identifying fake articles.

While the unique data of fake articles is cleanly identified, it is small and narrow and therefore perhaps difficult to draw more general conclusions. Consequently, we also use a second dataset of hand-collected articles that were published on two (and eventually three) of the most prominent financial crowd-sourced platforms: Seeking Alpha and Motley Fool covering 203,545 articles from 2005 to 2015 for Seeking Alpha, and 147,916 articles from

2009 to 2014 for Motley Fool. Using a linguistic algorithm (based on linguistic science from the literature) that identifies the authenticity of an author's text to probabilistically identify "fake" news, we create a second and larger set of false news events. Importantly, we use our first and smaller dataset of *known* fake articles to validate the algorithm's ability to identify fake news stories. Having an unambiguous sample of fake articles from the undercover sting operation by the SEC is a key advantage, because without it the authenticity algorithm cannot be validated. Echoing the importance of this statement are the challenges Google, Facebook, and Twitter are currently facing trying to identify fake news on their own platforms. For example, Google is currently using human editors to evaluate content in the hopes of training an algorithm to identify false content systematically (Leong (2017)). Absent a set of known and identifiable fake articles, such endeavors have yielded little success. For the same reason, an investor at the time of the fake article's publication could not have constructed such an algorithm either, since the fake articles from our dataset were not yet known or identified.

Using the set of identifiable fake articles from the SEC to train our algorithm and cross-validate it, we find our type II error to be very low – less than 5% of articles are identified as false positives. Hence, our method for identifying fake news in the second dataset is quite conservative, where we classify only 2.8% of articles as being fake in our sample, with the frequency peaking in 2008 at 4.8%, but where we have high confidence that these articles contain false content. Hence, our method is designed to minimize type II errors at the expense of increasing type I errors, where we are likely missing many other fake articles.

Using the samples of truly fake articles and probabilistically fake articles, we investigate whether they impact the market. We find that the incidence of fake articles is slightly higher for small firms, and that the price response in markets is much larger for small and mid-size firms and negligible (precisely zero) for large firms. Small firm prices rise on fake news, which is predominantly positive, by 8% upon its release, which subsequently gets fully reversed over the course of a year. Hence, for small firms the market appears fooled by these articles initially, overpricing small firms with fake news by 8% on average, but then eventually corrects the mispricng. For mid-size firms, the price impact is negative immediately, and there is a permanent 4% discount associated with fake articles written about the firm, suggesting that mid-size firms having fake articles is a bad signal about the firm. For large firms, there is no price impact – initially or long-term – from fake articles written about the firm. These results suggest that the market is efficient with respect to large firms and appears inefficient for small firms, consistent with intuition suggesting that the cost of information (direct and indirect/psychological) is larger for small firms. The evidence for mid-cap firms, however, may be consistent or inconsistent with market efficiency.

To better understand these results, it is useful to consider what motivates the production of fake news about firms? One motivation for the fake articles, which is related to how Rick Pearson went undercover and why the SEC is involved, is that the firms themselves may be orchestrating a promotional pump-and-dump campaign to manipulate the stock price. Another possibility, of course, is that rogue authors wish to create a false narrative about a firm for their own intentions, having no connection to the firm itself. To investigate the first possibility we look at a set of firm actions the firm may be pursuing at the time of the fake articles' release. For example, if these articles are part of paid campaigns by firms orchestrated by a public relations agency, then other actions taken by the firms around these events are likely to be present. We find that the fake news articles are often accompanied with press releases by the firm among small and mid-cap firms, but not among large cap firms. We further find that insider trading in the direction of the fake news (to take advantage of the price impact) is also more likely for small and mid-cap firms, but not large stocks. These results are consistent with a deliberate campaign by the firm to manipulate the stock price and take advantage of the price impact among small firms. For large firms, however, we do not find the same patterns, consistent with fake news about large firms being driven by authors outside of or unassociated with the firm.

We explore what characteristics of firms and articles are associated with the propensity of fake news as well as the magnitude of temporary price impact from the fake news, to better understand the variation in information environments across firms and articles. We find that for small and mid-cap firms, high past volatility and volume are associated with higher propensity of fake news, consistent with more retail investor attention (see Barber and Odean (2007)). Likewise, article email circulation, a proxy for the popularity of the firm among retail investors, is positively related to fake news. In addition, other proxies for attention, such as larger analyst coverage, larger number of stock tweets mentioning the firm, more media coverage, and more readers' comments, are all associated with higher propensity of fake news. We also examine other actions taken by the firm and its insiders prior to the fake news event date, such as share purchases, insider sales, and initiation of press releases. These actions are more prevalent and coincide with the fake news for small and mid-cap firms, but not for large firms.

We also find that these variables are associated with the magnitude of the positive temporary price impact found for small firms, and the permanent negative price impact found for mid-cap firms. The price impact for large firms is non-existent and does not vary with any of these variables. Specifically, for small firms, we find that the price reaction is stronger for firms with higher turnover and volatility, less media coverage, more retail ownership, and more frequently talked about on StockTweets. These characteristics predict both the magnitude of the initial price rise as well as the size of the subsequent reversal. We also find that firms whose articles usually get a lot of comments have a quicker return reaction than firms with fewer comments. Similarly, if the author of the article has many followers, or if many readers subscribe to that firm's articles, the return reaction is much stronger for those firms. Finally, the amount of trading by insiders and the number of press releases issued by the firm affect the differential return reaction, where firms whose management engages in insider buying and press releases around the time of the article have a stronger return reaction.

For mid-size firms, these same characteristics also predict the magnitude of the price reaction. However, the same characteristics associated with small firms having a more positive return reaction are associated with mid-size firm's more negative permanent price reaction. These results suggest that both small and mid-cap firms may be engaging in stock manipulation strategies to pump up the share price and take advantage of the higher price by issuing shares or buying shares before the run-up. In the case of small cap stocks, the market appears to be fooled by this scheme, causing a temporary price increase that subsequently gets reversed. For mid-cap stocks, however, one interpretation is that the market does not get fooled because markets are more efficient for these larger stocks (e.g., the cost of information for these stocks is lower). Instead, the market identifies the news as fake and reacts immediately to the fake news in a negative way by permanently discounting the firms's share price. In this way, the market is efficient for mid-cap firms with respect to these articles. 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 promotional 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.

For large cap firms, they neither seem to be initiating or taking actions in conjuction with the fake news articles and there appears to be no market price reaction of any kind to fake articles written about large firms. These results are consistent with the market being efficient with respect to fake news about large firms and large firms as a result not attempting to engage in a futile effort to manipulate the share price. Rather it appears that the fake articles about large firms are being written by rogue authors unaffiliated with the firms in an effort to increase their own utility.

Our study provides evidence on the prevalence and effect of promotional articles from crowd-sourced financial platforms that continue to grow and gain attention. How important are these platforms and what impact are they having? More specifically, how pervasive are fake articles on financial crowd-sourced platforms and what impact do they have on markets? While these platforms may simply be a side-show for financial markets, our results indicate that there is significant price impact from these promotional fake articles. These results are consistent with other findings suggesting that crowd-sourced services can impact markets (Hu, Chen, De, and Hwang (2014), who show that information content and sentiment from these services predicts returns). Our focus is on the fake news posted on these platforms and their ability to potentially manipulate stock prices.

More broadly, our study provides a unique test of market efficiency. Other studies of market efficiency that seek to circumvent the joint hypothesis test problem appeal to looking at price deviations between equivalent (or near-equivalent) securities, where the same discount rate/equilibrium pricing model differences out between the two securities (e.g., close-end funds (Pontiff (1996)), tech-stock carve outs (Lamont and Thaler (2002), twin shares). A drawback of such studies is that they do not identify an information event and are therefore not able to dissect its impact on the magnitude and duration of any price response. Our unique setting of non-news events allows us to analyze their impact without running into the joint hypothesis problem.

The rest of the paper is organized as follows. Section 2 details our sample on fake news articles and presents our methodology for identifying fake news. Section 3 investigates what motivates the production of fake news across different firms. Section 4 examines the market's response to fake news, including both temporary and permanent price responses, as well as the drivers of differential market responses across different types of firms. Section 5 then examines whether managers engage in insider trading and are more likely to issue press releases around publications of fake articles. Section 6 concludes the paper.

2. Data and Identifying Fake News

We describe the data we obtain on fake articles, our algorithm and it's validation, and provide an example of a specific fake article and its consequences.

2.1. Obtaining Fake Articles

The popularity of financial crowd-sourced platforms such as Seeking Alpha, Motley Fool, TheStreet, etc., has grown exponentially over the last fifteen years. For example, Seeking Alpha went from having two million unique monthly visitors in 2011 to over nine million by 2014. While this innovation has allowed for unprecedented levels of 'democratization' of financial information production, some concerns have been raised about these platforms being susceptible to pump-and-dump schemes, as they are frequented by retail investors, and since many authors on these platforms use pseudonyms instead of writing under their real name.[3] Whereas it is theoretically legal for an author to talk up or down a stock that she is either long or short, she has to disclose any positions she has in the stock in a disclaimer that accompanies the article. While many authors add such disclaimers to their articles, that is not something the platforms actually verify. 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[4]

Even though it does not come as a surprise that such promotions and pump-and-dump schemes exist, it can be hard to identify them, and especially to prove that the authors actually received payment for writing an article. Our analysis starts out with a unique dataset of paid-for articles obtained from an industry "insider," Rick Pearson. Rick, who is a regular contributor to Seeking Alpha, was approached by a PR firm that companies hire to help promote their stock. The PR firm asked him to write articles for a fee without disclosing the payment. Rick went undercover to investigate other authors and has uncovered many fake, paid-for articles where the author didn't disclose the compensation. These fake articles were subsequently taken down by the platforms where they originally appeared, and the firms are now being investigated by the SEC. The SEC has filed two lawsuits in 2014 and in 2017

[3]Even though the platforms claim that they always know the true identity of the author, in case that information is subpoenaed by the SEC.

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

against authors of promotional articles and the PR firms who were paying the authors to generate those articles. Rick has kindly shared with us the articles that he has determined to be fake, providing us with 111 fake articles by 12 authors covering 46 companies. We furthermore were able to obtain all other articles written by the same authors (336 in total) that were published on Seeking Alpha, many of which presumably not paid-for promotional articles, as a baseline comparison for the same authors. These other articles were often written about large firms (171 stocks in total), which as we will show, are less likely to engage in this sort of stock promotion. Furthermore, authors need to establish a reputation writing non-promotional articles before they can write (and get away with) pump-and-dump type articles.

2.2. Example of a Fake Article

To illustrate the process and the impact of fake articles, we provide an example of Galena Biopharma Inc., one of the companies that hired a PR firm to order paid-for promotional articles about its stock. Several very positive articles about Galena appeared on Seeking Alpha and other platforms from 2012 to 2014, which coincided with a substantial run-up and a subsequent drop in Galena's stock price. Figure 1 shows the price of Galena's stock from 2012 to 2015 in light blue, and the appearances of promotional articles in red. Over that time period six identified promotional articles appeared about Galena, with four of them published towards the end of 2013 and early 2014. The four fake articles were all written by the same author, John Mylant. John Mylant had been an active contributor to Seeking Alpha since 2009, even though since this incident all of his articles have been taken down by Seeking Alpha. As the graph shows, Galena's stock prices started to increase drastically, when several fake articles were published, more than tripling in 4 months, before it plummeted back down, once the promotional articles stopped.

One natural question that follows is why the companies pay for these promotional articles. In Figure 1 in dark blue we plot all instances of insider trading between 2012 and 2015 (it's an indicator variable for whether any insider buys/sells were reported to the SEC through

Form 4). The graph shows that insiders executed trades after the stock price almost tripled and right before the stock price crashed again. The SEC brought charges against Galena and its former CEO Mark Ahn "regarding the commissioning of internet publications by outside promotional firms." Mr. Ahn was fired in August 2014 over the controversy, and in December 2016, the SEC, Galena, and Mr. Ahn reached a settlement. The example of a promotional article about Galena is shown in Appendix A, and the 8-K form documenting the settlement in presented in Appendix B. Interestingly, if one were to search for this promotional article now, Seeking Alpha just displays a message saying "This author's articles have been removed from Seeking Alpha due to a Terms of Use violation."

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

While the unique data of fake articles is illustrative of the phenomenon, the sample size is small and it is difficult to draw more general conclusions based on it. The goal of our paper is to estimate the prevalence and the effects of these fake articles on financial markets. In order to do so, we manually download all articles that were published on two of the more prominent financial crowd-sourced platforms: Seeking Alpha and Motley Fool. For Seeking Alpha we obtained 203,668 articles dating from 2005 to 2015 and for Motley Fool we have 147,916 articles dating from 2009 to 2014.

To understand how pervasive the phenomenon is in general and what effect fake articles have on financial markets we develop an objective and scalable measure that captures the authenticity of the article. To that end, we use LIWC2015 (Pennebaker et al. (2015)). LIWC is a linguistic tool that focuses on individuals' writing or speech style, rather than content, and thus appears to be uniquely adept at measuring individuals' cognitive and emotional states across domains. Specific to authenticity, Newman et al. (2003), use an experimental setting to develop an authenticity score based on expression style components. They find that truth-tellers tend to use, for example, more self-reference words and communicate through more complex sentences compared to liars. Intuitively, when people lie, they tend to distance themselves from the story by using fewer "I" or "me"-words. Furthermore, since lying is cognitively taxing, liars tend to use less complicated sentences with fewer details. For example, lying individuals are more likely to say "I walked home," rather than "Usually I take the bus, but it was such a nice day, that I decided to walk home."

2.4. Validation

Given that the LIWC authenticity score was not developed in the context of financial media, one may be skeptical about its ability to distinguish fake from non-fake articles in Finance. After all, financial blogs and articles tend to point to facts, trends, and figures, which are different from narratives. To address this concern, we start out by validating the LIWC authenticity score using the small sample of 111 fake articles and 336 non-fake articles, all written by the same set of authors. That is, we compare the LIWC authenticity score, which is normalized between 0 and 100, for the two samples. The difference in the LIWC authenticity score across the two samples is both economically and statistically large. Relative to an average authenticity score of 33 for non-fake articles, fake articles had a much lower average score of 17 (statistically significant at 1% level). The density plots in Figure 2 illustrate how different the two distributions are. It is important to note that we control for any differences the authors' writing style may have on the authenticity score, as the sample consists of both fake and non-fake articles written by the *same set of authors*.

While the exact composition of the authenticity score is proprietary, several language characteristics are associated with being more or less authentic. In Table 1 we provide a summary of those characteristics for the promotional and non-promotional articles (written by the same authors). *Authentic*, *Clout*, and *Analytics* are proprietary measures meant to measure how authentic, assertive, and analytical the language is, respectively. From the table, we see that the promotional articles' authenticity score is about half the size of non-promotional articles. Promotional articles also use more assertive language, and a slightly more analytical language. We also display the average of the *1st person singular* measure (examples: I, me, mine), *Differentiation* measure (examples: hasn't, but, else), as well as the total number of words in the document, and the average number of words per sentence. According to research by James Pennebacker and co-authors, when people lie they tend to use fewer self-referencing words, and simpler sentences (smaller *Differentiation* measure). The results in the table line up well with those findings: promotional articles' self-referencing score is about half of non-promotional articles, and promotional articles have a smaller *Differentiation* score (use a simpler sentence structure). It's important to note that the promotional articles that we obtain from Rick Pearson are crucial to being able to use LIWC to identify fake articles in Finance, as they provide an out-of-sample test for a methodology that was developed outside of finance.

2.5. Probability of Being Fake

The above validation demonstrates that the LIWC authenticity score has the ability to distinguish between fake and non-fake articles. At the same time, it is not clear how to interpret the cardinal nature of the score – what does a 14 point difference in authenticity score mean? Ideally, we would measure the probability of an article being fake, but the LIWC authenticity score is not a probability. To provide a more direct interpretation of the results and their economic meaning, we develop a mapping of the authenticity score into the probability space. Starting with the validation sample and applying Bayes rule to the overall sample of Seeking Alpha articles, we create a function that maps the authenticity score into a conditional probability of an article being fake.

Specifically, let S be the authenticity score and F (T) denote a fake (true) article. In the validation sample, we know which articles are F and which ones are T, so we can compute $Prob(S|F)$ and $Prob(S|T)$. From Bayes rule, we know that:

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

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 apply this approach to the entire sample of Seeking Alpha articles published between 2005 and 2015, over 203,000 in total, covering over 7,700 firms. There are a number of findings that arise. To start with, we observe the resulting mapping of LIWC authenticity scores (S) into the conditional probability of being fake ($Prob(F|S)$). As Figure 3 depicts, the relation between the two is highly non-linear. As the figure shows, an authenticity score of 31 – the average for the non-fake articles – corresponds to a conditional probability of being fake of close to zero, while an authenticity score of 17 – the average for the fake articles – corresponds to a significant probability of being fake of 3.6%.

This has two important implications. The first is that using the LIWC authenticity score is not equivalent to using the probability. The second is that the sharp increase in probability in the very low authenticity range suggests that articles can be well classified into fake and non-fake ones. Put differently, using a probability cutoff can be an efficient way of separating articles into various types.

Next, we use these results to answer a key question: how pervasive are fake articles on financial crowd-sourced platforms? We find that the unconditional probability of a Seeking Alpha article being fake is 2.8%, peaking at 4.8% in 2008 and dropping to 1.6% in 2013.

We next examine how accurate our method is at identifying fake news. We take the 447 articles (111 promotional and 336 non-promotional articles written by the same authors), generate an authenticity score for them, and calculate their probability of being fake. We then use the cutoff of $Prob(Fake) > 20\%$ to classify articles as being fake[5]. Our algorithm classifies 17 out of the 447 articles as being fake. Out of those 17 articles 16 are actual promotional articles. This suggests that our Type II error rate is very low - we have very few false positives, and our method is very conservative. In other words, while we most likely miss some promotional articles, the ones our algorithm identifies as fake are highly likely to be truly promotional articles.

[5]Our results are not sensitive to the specific cutoffs

We classify articles with $Prob(Fake) < 1\%$ as being non-fake. Our algorithm identifies 156 articles (out of 447) are being non-fake. Of those 8 are actually promotional articles, and the rest are not. Therefore, our Type I error is about 5%, which is quite low. So when we look at articles that our algorithm identifies as being "non-fake," most of them happen to be non-promotional. We exclude articles with $1\% \leq Prob(Fake) \leq 20\%$ from our analysis, as for those articles Type I and Type II errors will be large, and would make our analysis very noisy.

We verify our algorithm in one more way. Seeking Alpha has taken down articles that have been identified as promotional, and periodically removes other articles that violate their terms of use. This can happen for several reasons, including readers or firms pointing out a material error in the article that the author refuses to correct, or if the author didn't write the article him/herself. While Seeking Alpha unfortunately does not store the removed articles, we are able to obtain some of the articles that have been removed using the Wayback machine, a website that keeps archives of the Internet. We examine articles that either "violated the terms of use", or have been removed "due to material error" (suggesting that the author refused to fix the error). Overall, we are able to obtain 12 additional articles from the Wayback machine. While some of those articles may not necessarily be promotional (or fake), the average article is more likely to be and hence these articles probably lie somewhere in between fake and true on average. We examine the authenticity score of the promotional articles that we use in our final dataset, of general Seeking Alpha and Motley Fool articles, and of the articles we were able to obtain using the Wayback machine.

Table 2 shows, for different types of articles on Seeking Alpha and Motley Fool, summary statistics of various LIWC textual measures, the probabilities of being fake, and firm characteristics of the covered firms. *Promotional Articles* are the articles that have been shared with us by Rick Pearson, who went undercover to expose authors who were being paid to write promotional articles for companies, without disclosing the payments. *Wayback Articles* are articles that were taken down by Seeking Alpha because the authors violated the terms of use, that we obtained using the Wayback website. *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 (according to our measure) is higher than 20%, *Non Fake* articles are articles with probability of being fake less than 1%, and the rest are classified as *Other*.

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. The authenticity score is much lower for promotional articles than it is for WayBack articles, which in turn is lower that the authenticity scores of non-fake Seeking Alpha and Motley Fool articles (17.27 vs. 29.56 vs. 50.71 and 46.75, respectively). The differences are statistically significant. The authenticity scores for *Fake* Seeking Alpha and Motley Fool articles are especially low, which is by construction. We also report the means of several other variables provided by LIWC to better understand the textual structure of these articles. In particular we display the means of the two other proprietary measures that LIWC provides: *Clout* and *Analytics*. While LIWC doesn't disclose details about those measures, they are based on research conducted by Kacewicz et al. (2014) and Pennebaker et al. (2014), respectively. *Clout* is meant to measure with how much authority the author speaks, and *Analytics* measures how much analytical thinking the author uses. *Analytics* is closely related to how many numbers are in the text. While, promotional and fake articles have a higher *Clout* score, their *Analytics* scores are quite similar. This suggests that while the authors speak with more authority, when they write fake articles, they use a similar amount of quantitative evidence and analytical thinking. We also display the average use of the *1st person singular* measure (examples: I, me, mine), *Differentiation* measure (examples: hasn't, but, else). Fake articles, as identified by our algorithm, have a much lower fraction of *1st person singular* then the non-fake articles, suggesting that the authors try to distance themselves from the article. The authors use fewer *Differentiation* words, which partially measure how complex a sentence is, which lines up with linguistic research showing that when people are being deceptive, they use a less complicated sentence structure, since lying is cognitively taxing. Interestingly, fake articles (as identified by our algorithm) are shorter but have longer sentences. Promotional articles are much longer, and also have much longer sentences.

In Panel B, we display the average probability of being fake for each of the article categories. For Seeking Alpha and Motley Fool, the difference in magnitudes of the probability of being fake are by construction. For promotional and WayBack articles, we can see that promotional articles are more clearly fake, whereas WayBack articles are more similar to regular, non-fake articles. that we identify.

In Panel C, we display the average fraction of retail investors, the average number of analysts covering the firm, and the average firm size (in Millions of dollars) for the respective article groups. Promotional and WayBack articles tend to cover firms with a much higher fraction of retail investors, whereas the fake versus non-fake articles seem to target firms with similar fraction of retail investors. Similarly, Promotional and WayBack articles tend to concentrate on smaller firms with low analyst coverage, which is not the case for fake and non-fake articles.

2.6. Other Datasets

We investigate the motivation behind these fake articles, where one hypothesis is that these campaigns are ordered by firms and orchestrated by a PR agency. To test this hypothesis, we obtain a dataset of press releases from RavenPack from 2001 to 2015 and collect message volume for a given firm from a Twitter-like platform called StockTwits. If firms are coordinating these articles, they may issue press releases simultaneously to provide material for the promotional articles and may also start Twitter campaigns to reinforce the messages in the fake articles.

We obtain stock price data from CRSP and firms' financial information from COMPUSTAT. We also obtain data on insider trades from Form 4 from Thomson Reuters to see if insiders are trading around these events. Finally, we obtain dates of SEOs from the SDC Platinum database to look at firm equity issuance around these events.

3. What Drives Fake News?

In this section, we explore what firm characteristics are associated with greater temporary price impact from fake news to better understand the variation in information environment across firms and across articles. We look at characteristics associated with the firms such as liquidity (e.g., Amihud Illiquidity, volatility), accounting quality (accruals), and the coverage it receives (analyst coverage, media coverage, tweet volume), its investor base (fraction of retail ownership), characteristics of the blogger/article (num of comments the article received, num of followers the author has, num of individuals on the distribution list), and finally characteristics of the actions taken by the firm and its insiders prior to the event date (share purchases, sales, and initiation of press releases).

We code each characteristic as a dummy variable based on year and size group (small, medium, large) tercile assignment. The characteristic dummy is equal 1(0) if the observation is in the top(bottom) tercile. "Amihud Illiquidity" is measured following Amihud (2002)), "Volatility" is the average volatility during the 12 month period prior to the event date (lagged 12 month daily returns squared variance), "Accruals Quality" is a measure of the quality of the firm accounting (ability of lead/lag cash flow from operations to explain changes in working capital, see Dechow and Dichev (2002)) where higher numbers indicate lower quality, "Analyst Coverage" is a count of the number of analysts covering the firm at the event date, "Retail Own." is the fraction of the firm equity held by retail investors, "# of Tweets" is the count of the number of stock tweets mentioning the firm during the 40 trading days prior to event date, "Media Cov." is the count of the number of articles in the major outlets (e.g., WSJ, NYT) mentioning the firm during the 40 trading days prior to the event date, "# of Comments" is the count of the average number of comments the articles on the event date received, "# of Followers" is the average number of users following the writers of the articles on the event date, "Email Circulation" is the average number of email recipients of the articles on the event date, "Shares Purchased" ("Shares Sold") is a count of the total number of shares purchased (sold) by insiders during the 40 trading days prior to the event date, and "# of Press Releases" is a count of the number of press releases in the 40 trading days prior to the event date. Small firms are defined as firms in the bottom 10th percentile of NYSE firms and mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms.

3.1. Cross-Sectional Firm Characteristics

The cross sectional characteristics proxy for variables that may affect the potential impact and thus incentives of releasing fake news. We start by examining whether the arrival of fake news is more pronounced for certain firms or conditions. Specifically, we regress the fake news dummy on firm characteristics, one at a time, while including month-year fixed effects. We conduct the analysis separately for small and mid-size firms.

Table 3 reports the results. We find that high past volatility and volume are associated with higher propensity of fake news, consistent with both measures proxying for retail investor attention (see Barber and Odean (2007)). Likewise, article email circulation, which is related to the popularity of the firm among retail investors, is positively related to fake news. Many of the other variable are statistically insignificant for small firm, potentially due to the low number of observations. The results for mid-size firms are qualitatively similar to those of small firms but many are also statically significant. In addition to the results above, we find that many proxies for attention are associated with higher propensity of fake news – larger analyst coverage, larger number of tweets mentioning the firm, more media coverage, and more readers' comments.

4. Market Impact

Using the sample of promotional articles as well as the set of probabilistically fake articles from the broader sample, we investigate the market's response to fake news.

4.1. Return Reaction

First we examine the return reaction to the promotional articles that were provided to us by Rick Pearson. We separate the firms into small and mid-size firms and examine the firms' return response to the promotional articles. We classify a firm as small, if its market cap is in the bottom 10th percentile of NYSE stocks, and as medium if it's in the 20th-90th percentile by market cap among NYSE stocks. The cumulative abnormal returns, measured as equal-weighted 4-factor residuals, are constructed starting the day after the article was published until 251 trading days after the article was published. We generate equal-weighted *Mom*, *SMB*, *HML*, and *Mkt* factors, and estimate betas for a given stock i for day t using the window $t - 252$ to $t - 1$. We then use those betas to calculate the residual (abnormal) cumulative returns for stock i for days $t + 1$ to $t + 251$.

Figure 4 plots the cumulative abnormal returns for the promotional articles for small (blue line) and mid-size (red line) firms. Out of the 111 articles that we have, 69 are about small firms, 35 are about mid-size firms, and 7 are about large firms. Returns for small firms increase after the article is published, reaching as much as 20%, cumulatively, after about 60 days, before giving up all the gain, and ending up with a 10% loss towards the end of the year. The permanent price impact of -10% for small firms indicates that once the market figures out the news is fake, investors view this as a bad signal about the firm. Interestingly, for mid-size firms, there is no gain followed by reversal - the price starts dropping after the fake article comes out, and continues to decrease throughout the year. These results suggest that for both small and mid-size firms, the fact that management is trying to prop up the stock price with promotional articles is a signal for deteriorating underlying performance. However, either due to larger limits to arbitrage or to less sophisticated investor base, the promotional articles are successful at temporarily propping up the stock price of small firms, but not of mid-size firms.

The articles we obtained from Rick Pearson are a small sample. We next examine the market response to articles that we classify as fake using LIWC. Since our analysis is at the company-day level, we need to define whether a company had a promotional article on a given day. In order to do so, we calculate the average authenticity score among all articles for a given company/day, and define that a company had fake articles on a given day, if the average authenticity score translates into the probability of being fake of 20% or more, and that a company had no fake news, if the average authenticity score translates into the probability of being fake of less than 1%.

Figure 5 plots the difference between abnormal cumulative returns following days with fake articles, relative to days with no fake articles separately for small, mid-size, and large firms in our sample (that have at least one fake article). As the blue line shows, the returns for small firms increase for 6 months by about 8% following a fake article, relative to a non-fake article, and then revert back to their original level. Whereas the returns for mid-size firms (red line) start dropping almost immediately, and come to a steady state of -5% after about 10 months. It's important to note that small firms experience temporary positive returns following fake articles, whereas mid-size firms see a decrease in returns, which is very similar to return patters following promotional articles that Rick Pearson shared with us (shown in Figure 5), which helps to corroborate the patterns. For large firms (green line) the returns appear to first go up and then decrease, but the magnitudes are quite small – 50 to 100 basis points and are not statistically different from zero. Suggesting that the markets for those firms are quite efficient, and also that the articles that we identify as fake are probably one-off rogue investors, rather than those companies launching promotional campaigns. The promotional articles that we obtained from Rick Pearson only included a few firms in this size category.

Next, we examine whether the patters in cumulative abnormal returns for different-sized firms we observe in Figure 5 are statistically significant. In order to do so, we estimate the following model:

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

where $Ret_{i,(t+1,t+T)}$ are cumulative abnormal 4-factor returns for firm i, from 1 day after the fake article was published until T days, where T is either 51, 101, 151, 201, or 251. The results are presented in Table 4. As we saw in Figure 5, for small firms, the returns in the first 6 to 7 months following fake articles, are more positive than following non-fake articles, and the difference is statistically significant. This gain disappears after about 10 months and is basically 0 after a year. For mid-size firms, the returns start decreasing 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 very small (60 basis points after 3 months) and is barely statistically significant.

4.2. What Drives the Market Reaction?

So far, we have examined differential return reactions to fake articles relative to non-fake articles for small, mid-size, and large firms. Next, we dig deeper into whether the market reacts differently to fake articles depending on the type of firm that the article is about. In particular, for small and mid-size firms, we examine whether the firms' liquidity, investor base, monitoring, and insiders' actions affect the return reaction (for definitions of the cross-sectional variables, see Section 3). For each firm characteristic, we compare the differential return reaction for fake articles relative to non-fake articles for firms in the bottom tercile ($Fake - Low$) to firms in the top tercile ($Fake - High$) of that characteristic.

For small firms the results are presented in Table 5. We find that the return reaction is stronger for firms that are more liquid (lower Amihud measure) and for firms with higher return volatility. Accrual quality seems to not have any effect on the return reaction, whereas firms with a higher analyst coverage have a stronger return reaction than firms with no analyst coverage. However, firms with low media coverage have a stronger return reaction, and so do firms that are often talked about on StockTweets. Fraction of retail ownership seems to have a positive effect on return reaction - firms with more retail owners have a stronger reaction than firms with low retail ownership.

Next, we examine whether the proxy for attention that the firms' articles usually get has an effect on the differential return reaction. We find that firms whose articles usually get a lot of comments have a quicker return reaction than firms with fewer comments. Similarly, if the author has many followers, or if many readers subscribe to that firm's articles, the return reaction is much stronger for those firms. Finally, we examine how the amount of trading by insiders and the number of press releases issued by the firm affect the differential return reaction. We find that firms whose managers, on average, engage more in insider buying and sell fewer shares have a stronger return reaction. Furthermore, firms that issue more press releases have a quicker and a slightly larger return reaction.

We perform a similar analysis for mid-size firms, and the results are presented in Table 6. We find that similar characteristics that are associated with small firms having a more positive return reaction are associated with firm's more negative return reaction for mid-size firms.

5. Insider Trading and Press Releases

So far we have provided evidence that fake articles can influence asset prices. Next we examine what incentives managers have to try to pump up their stock price, and whether they take any actions to facilitate the promotion of the articles. In particular, we look at whether managers engage in what's called "pump-and-dump" schemes, where one acquires shares at a low price, then inflates the price through fake articles, and then sells the stock. While we cannot observe trades for regular investors, this is something we can observe for managers, as they have to report their insider trades to the SEC using Form 4. We further examine whether companies are more likely to issue press releases around the time of fake articles, to give the authors of the fake articles some material to write about.

First, at the monthly level, we regress an indicator variable for whether a firm had predominantly fake articles in a given month on whether the firm was a net insider buyer or a net insider seller in the previous month and in the contemporaneous month, as well as whether the firm issued at least one press release in one of those two months. A firm is a net buyer (seller) if insiders bought more shares, in dollar value, than they sold in a given month (sold more shares than they bought). We define an indicator variable, *Fake Article*, to be 1 if the probability of being fake associated with the average authenticity score for articles written about the firm in the given month is great than 20%. We perform our analysis separately for small, mid-size, and for 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.

We do not tabulate the results for brevity. We find that for small and mid-size firms, insider buying and issuing of press releases is strongly associated with the prevalence of fake articles in the same month. Whereas there is no association between insider trading and fake articles in the previous and the following months. We further find, that for large firms neither insider trading nor press releases have any connection with the prevalence of fake articles.

The above analysis shows that insiders buy stock and issue press releases in the same month as fake articles come out. Next, we zoom in on the weeks around fake articles and examine the timing in more detail. We run similar regressions as we did at the monthly level, except now everything is defined at the weekly level. Therefore, 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 news came out. *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 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 month is great than 20%. *PR* is an indicator variable for whether the firm issued at least one press release in a given week. We perform our analysis separately for small, mid-size and for large firms.

The results are presented in Table 7. For small firms and mid-size firms, insiders don't trade in the weeks leading up to fake articles, and then start actively buying the week of and the week after the fake articles come out. These findings are consistent with Figure 1, where insiders start buying Galena's stock around/after fake articles come out. We do not find a similar result for large firms. Furthermore, we find that small firms are more likely to have fake articles the same week and the week after they issue a press release, and mid-size firms are more likely to have fake articles the week of the press release. This is consistent with the anecdotal evidence that companies often issue press releases to provide some material for the fake articles. Again, large firms do not seem to engage in this behavior. These results are consistent with a deliberate campaign by the firm to manipulate the stock price and take advantage of the price impact.

Further evidence on the link between insiders trading and fake news can be obtained from Table 5. Firms with abnormally high insider purchase activity are associated with significantly more pronounced positive response to the release of fake news. That is, firms with insider purchases prior to fake news releases observe abnormal returns of up to 22.5% in the first month after the release compared to only 1.3% for firms with fake news but no insider purchases. Consistent patterns are observed when analyzing insider sells: firms with abnormally low insider sells observe much higher return response following the release of fake news. While it is hard to determine a causal relation, it is possible that insiders time their buying and selling decisions in anticipation of fake news releases. Interestingly, in both cases, the price impact of fake news is transitory as abnormal return at the end of the year is not statistically different from zero. For medium size firms (see Table 5) the results suggest a more negative response to fake news among firms with abnormally high insider sell activity. For neither small nor medium size firms we find that the propensity of press releases has a discernible impact on the effect of fake articles on prices.

6. Conclusion

Examining the impact of false information on prices using our unique datasets of fake articles, our novel test of market efficiency finds that markets respond to erroneous information in small stocks, possibly leading to potential price manipulation. The "non-event" studies we conduct find strong temporary price impact and subsequent reversals from fake news for small firms that coincide with insider trading and firm press releases, that predict the magnitude of the price reaction and reversal. We find similar results for mid-size firms except there is no temporary price increase and only a permanent price decrease associated with fake news, especially when coordinated with insider trades and press releases by the firm. Large cap stocks exhibit none of these patterns nor any price impact from the fake articles.

The evidence suggests that markets are efficient with respect to fake news for large and possibly mid-cap firms, but is inefficient for small cap stocks, consistent with information costs being greater for smaller firms. Small firms therefore engage in possible price manipulation that temporarily props up the share price and eventually reverses over the course of the year. Mid-size firms seem to engage in similar behavior, but the market isn't fooled and applies an immediate permanent price discount on those firms. Large firms do not engage in this behavior, consistent with its share prices being immune from fake news and the cost of information low enough in large firms that prices remain efficient.

Further research seeks to understand why firms may or may not engage in price manipulation, how these social media platforms can amplify or hinder the price discovery process, and potentially provide a way to measure the cost of information across firms.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of financial markets 5*(1), 31–56.

Dechow, P. M. and I. D. Dichev (2002). The quality of accruals and earnings: The role of accrual estimation errors. *The accounting review 77*(s-1), 35–59.

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

Kacewicz, E., J. W. Pennebaker, M. Davis, M. Jeon, and A. C. Graesser (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology 33*(2), 125–143.

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., C. K. Chung, J. Frazee, G. M. Lavergne, and D. I. Beaver (2014). When small words foretell academic success: The case of college admissions essays. *PloS one 9*(12), e115844.

APPENDIX

Appendix A: An example of a promotional article about Galena Biopharma Inc.

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

Figure 1. Example of a pump-and-dump scheme

This figure depicts the stock price of Galena Biopharma Inc. from 2012 - 2015, as well as occurrences of fake articles being published on Seeking Alpha and instances of trading by insiders at Galena.

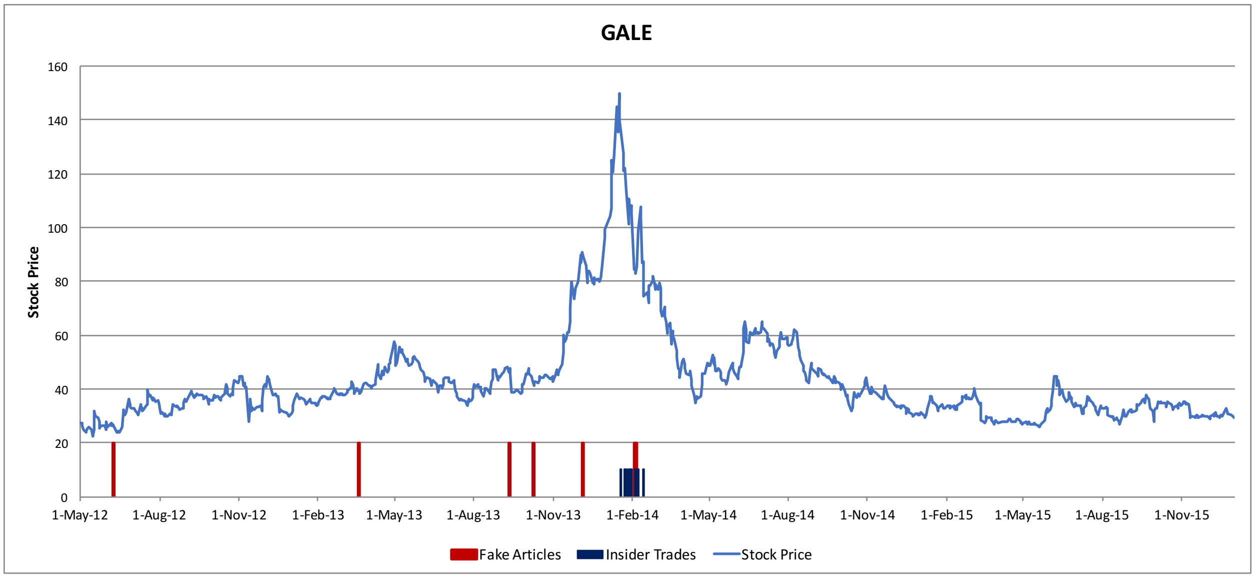

Figure 2. Authenticity Scores

This figure depicts the distribution of authenticity scores for fake and non-fake articles in our validation sample of 111 fake and 336 non-fake articles.

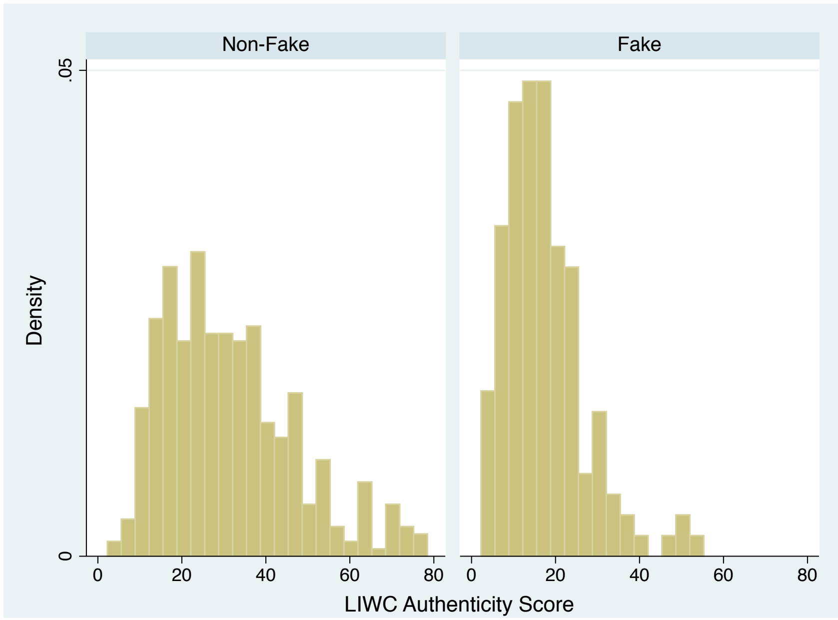

Figure 3. Authenticity score and the probability of being fake

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

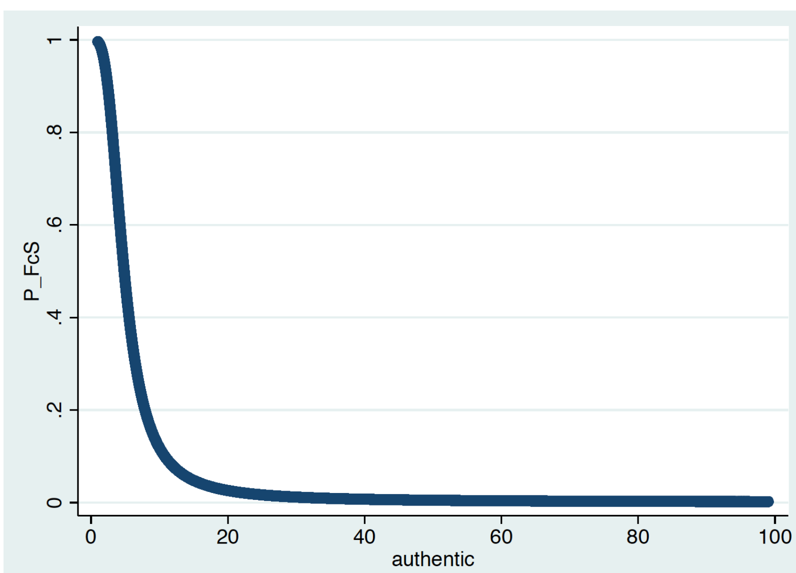

Figure 4. Event Study – 4-factor CAR (promotional articles)

The figure depicts the progression of cumulative abnormal returns (measured as equal-weighted 4-factor residuals) for promotional articles provided to us by Rick Pearson. The cumulative returns are measured starting with the day after the article was published until the 251 trading days after the article was published. 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.

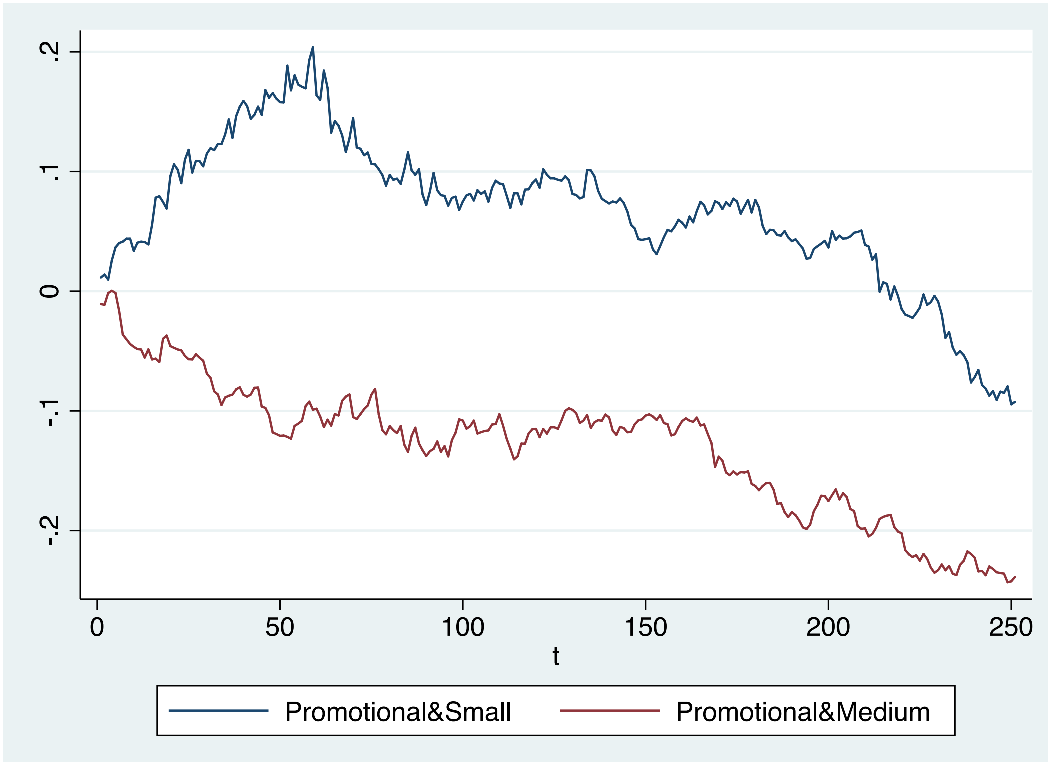

Figure 5. Event Study – 4-factor CAR

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. 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.

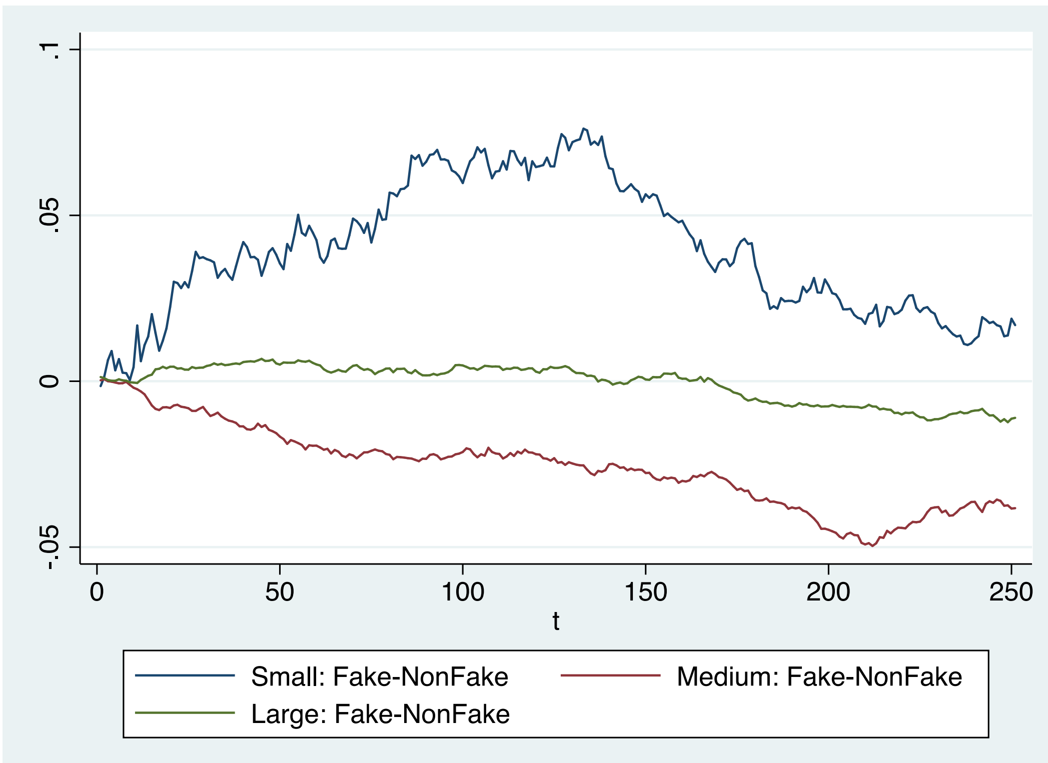

Table 1. LIWC and Promotional Articles

This table presents the summary statistics for various LIWC textual measures for promotional articles that were shared with us byr Rick Pearson, and non-promotional articles that were written by the same authors. *Promotional* articles are articles that have been shared with us by Rick Pearson, who went undercover to expose authors who were being paid to write promotional articles for companies, without disclosing the payments. *Non-Promotional* articles, are articles that were written by the same authors, but about larger firms, that are unlikely to be promotional. We display the number of articles in each category as well as the mean of the *Authenticity* measure from LIWC. We also report the means of several other variables provided by LIWC to better understand the textual structure. In particular we display the means of the *Clout* and *Analytics* measures. Both are proprietary LIWC measures, meant to measure authority and analytical content of the text, respectively. We also display the average of the *1st person singular* measure (examples: I, me, mine), *Differentiation* measure (examples: hasn't, but, else), as well as the total number of words in the document, and the average number of words per sentence.

	Promotional	Non-promotional
Number of articles	111	336
Authentic	17.27	32.77
Clout	61.04	52.33
Analytics	94.69	90.97
1st pers singular	0.37	0.76
Differentiation	1.48	2.02
Word Count	1,611	1,161
Words per sentence	46	65

Table 2. 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. *Promotional Articles* are articles that have been shared with us by Rick Pearson, who went undercover to expose authors who were being paid to write promotional articles for companies, without disclosing the payments. *Wayback Articles* are articles that were taken down by Seeking Alpha because the author violated the terms of use, that we obtained using the Wayback website. *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 better understand the textual structure. In particular we display the means of the *Clout* and *Analytics* measures. Both are proprietary LIWC measures, meant to measure authority and analytical content of the text, respectively. We also display the average of the *1st person singular* measure (examples: I, me, mine), *Differentiation* measure (examples: hasn't, but, else), as well as the total number of words in the document, and the average number of words per sentence. In Panel B, we display the average probability of being fake, for each of the article categories. In Panel C, for the firms that are covered in the respective article groups, we provide the average fraction of retail investors, the average number of analysts covering the firm, and the average firm size (in Millions of dollars).

			Seeking Alpha			Motley Fool		
	Promotional	WayBack	Non Fake	Fake	Other	Non Fake	Fake	Other
Panel A: LIWC variables								
Number of articles	111	12	116,289	3,933	83,323	78,943	1,368	67,605
Authentic	17.27	29.56	50.71	5.44	22.51	46.75	5.71	21.96
Clout	61.04	55.32	52.84	62.04	57.06	60.83	72.40	63.99
Analytics	94.69	93.25	91.85	93.01	92.72	89.60	88.86	90.69
1st pers singular	0.37	0.35	0.98	0.25	0.54	0.53	0.20	0.23
Differentiation	1.48	1.72	2.32	1.53	2.05	2.47	1.34	2.10
Word Count	1,611	1,237	841	609	934	692	523	635
Words Per sentence	46	23	22	24	22	19	31	19
Panel B: Probability of being Fake								
Prob(Fake)	0.11	0.02	0.01	0.45	0.03	0.01	0.42	0.03
Panel C: Firm characteristics								
Percent of retail investors	70.22%	68.84%	42.46%	42.32%	44.96%	36.78%	40.88%	38.99%
Numer of Analysts	8.09	4.14	18.33	16.83	16.67	19.84	23.21	20.34
Firm Size ($Mil)	11.27	1.34	51.72	44.12	45.17	70.58	101.97	80.40

Table 3. Fake News and Firm Characteristics

The table reports results from regressing a dummy variable for whether an article was fake on various characteristics. For definitions of the cross-sectional variables, see Section 3. The top panel reports results for small firms (firms that are in the bottom 10th percentile of NYSE firms) and the bottom panel reports results for mid-size firms (firms in the 20th-90th percentile of NYSE firms). Regressions include month-year fixed effects.

	Amihud Illiquidity	Volatility	Abn Volume	Accruals Quality	Analyst Coverage	Retail Own.	# of Tweets	Media Cov.	# of Comments	# of Followers	Email Circulation
						Small Firms					
Coeff.	0.0052	0.0194**	0.0095*	-0.0078	0.0013	0.0011	0.003	0.0097	0.0062	-0.0005	0.0161**
	(1.17)	(2.23)	(1.93)	(-1.39)	(0.28)	(0.26)	(0.67)	(1.17)	(0.88)	(-0.08)	(2.32)
Obs.	7,054	5,540	6,841	3,635	11,621	6,030	8,889	11,621	3,055	2,963	2,922
R^2	0.040	0.040	0.040	0.050	0.020	0.050	0.030	0.020	0.100	0.100	0.090
						Medium Firms					
Coeff.	0.0005	0.0099***	0.0077***	0.0015	0.0072***	0.0028	0.0030*	0.0084***	0.0054***	-0.0009	0.002
	(0.29)	(2.89)	(4.10)	(0.81)	(3.58)	(1.54)	(1.74)	(4.39)	(2.75)	(-0.46)	(0.96)
Obs.	44,680	17,381	43,494	20,737	68,087	40,820	49,420	67,127	21,197	19,865	20,014
R^2	0.030	0.030	0.030	0.030	0.030	0.020	0.030	0.030	0.050	0.050	0.050

Table 4. 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.

	$Ret_{1,51}$	$Ret_{1,101}$	$Ret_{1,151}$	$Ret_{1,201}$	$Ret_{1,251}$
			Small Firms		
Fake	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	-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	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 5. Return Window Regressions – Small Firms

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. We estimate the effect of fake news separately for each return window and characteristic group (high/low). Small firms are defined as firms in the bottom 10th percentile of NYSE firms. For definitions of the cross-sectional variables, see Section 3.

		$Ret_{1,51}$	$Ret_{1,101}$	$Ret_{1,151}$	$Ret_{1,201}$	$Ret_{1,251}$
Amihud Illiquidity	Fake - Low	0.008	0.041	0.103**	0.057	0.063
		(0.31)	(1.10)	(2.32)	(1.24)	(1.14)
	Fake - High	-0.008	0.04	0.03	-0.001	-0.019
		(-0.19)	(0.86)	(0.50)	(-0.01)	(-0.27)
Volatility	Fake - Low	0.024**	0.022	0.054	0.047	0.069
		(2.05)	(0.68)	(1.55)	(1.24)	(1.35)
	Fake - High	0.075*	0.131***	0.102*	0.084	0.049
		(1.67)	(2.82)	(1.82)	(1.21)	(0.65)
Accruals Quality	Fake - Low	0.074	0.095	0.019	-0.067	-0.062
		(1.62)	(1.53)	(0.41)	(-1.19)	(-0.91)
	Fake - High	0.053	0.101	0.114	0.103	0.076
		(1.28)	(1.63)	(1.20)	(1.18)	(0.86)
Analyst Coverage	Fake - Low	0.019	0.028	0.038	0.041	0.015
		(0.57)	(0.92)	(0.84)	(0.74)	(0.25)
	Fake - High	0.047*	0.094***	0.072*	0.02	0.025
		(1.82)	(2.71)	(1.83)	(0.46)	(0.53)
Retail Ownership	Fake - Low	-0.01	-0.018	-0.038	-0.053	-0.007
		(-0.31)	(-0.45)	(-0.79)	(-0.95)	(-0.10)
	Fake - High	0.064	0.081	0.051	0.041	0.025
		(1.34)	(1.57)	(0.94)	(0.63)	(0.34)
Tweets	Fake - Low	0.038	0.060**	0.055	0.016	-0.003
		(1.33)	(2.03)	(1.41)	(0.35)	(-0.06)
	Fake - High	0.055	0.104*	0.071	0.042	0.077
		(1.49)	(1.80)	(1.20)	(0.69)	(0.98)
Media Coverage	Fake - Low	0.041*	0.072***	0.063**	0.034	0.02
		(1.92)	(3.03)	(2.08)	(0.98)	(0.55)
	Fake - High	-0.1	-0.096	-0.082	-0.122	-0.067
		(-1.06)	(-0.72)	(-0.49)	(-0.64)	(-0.29)
Num of Comments	Fake - Low	-0.02	0.107*	0.161**	0.031	0.015
		(-0.45)	(1.77)	(2.37)	(0.49)	(0.19)
	Fake - High	0.069	0.142*	0.1	0.091	0.07
		(1.45)	(1.78)	(1.33)	(1.33)	(0.95)
Num of Followers	Fake - Low	-0.006	0.068	0.084	0.026	-0.003
		(-0.13)	(1.39)	(1.24)	(0.37)	(-0.04)
	Fake - High	0.027	0.181***	0.210***	0.117	0.017
		(0.61)	(2.67)	(2.74)	(1.60)	(0.25)
Email Circulation	Fake - Low	-0.078	0.039	0.09	0.033	0.005
		(-1.50)	(0.51)	(1.01)	(0.36)	(0.05)
	Fake - High	0.057	0.156**	0.178**	0.073	-0.009
		(1.51)	(2.29)	(2.40)	(0.87)	(-0.09)
Shares Puchased	Fake - Low	0.013	0.051**	0.044	0.014	0.001
		(0.62)	(2.04)	(1.41)	(0.39)	(0.02)
	Fake - High	0.225**	0.177**	0.156*	0.141	0.163
		(2.27)	(2.23)	(1.71)	(1.16)	(1.34)
Shares Sold	Fake - Low	0.041*	0.069***	0.066**	0.042	0.032
		(1.83)	(2.76)	(2.11)	(1.14)	(0.81)
	Fake - High	-0.079*	-0.036	-0.114	-0.204***	-0.222**
		(-1.96)	(-0.73)	(-1.50)	(-3.15)	(-2.44)
Press Releases	Fake - Low	0.032	0.062**	0.068*	0.057	0.039
		(1.22)	(2.19)	(1.73)	(1.27)	(0.80)
	Fake - High	0.052	0.083*	0.04	-0.018	-0.004
		(1.29)	(1.69)	(0.81)	(-0.31)	(-0.07)

Table 6. Return Window Regressions – Mid Size Firms

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. We estimate the effect of fake news separately for each return window and characteristic group (high/low). Mid-size firms are defined as firms in the 20th-90th percentile of NYSE firms. For definitions of the cross-sectional variables, see Section 3.

		$Ret_{1,51}$	$Ret_{1,101}$	$Ret_{1,151}$	$Ret_{1,201}$	$Ret_{1,251}$
Amihud Illiquidity	Fake - Low	-0.007	-0.012	-0.030*	-0.059***	-0.061***
		(-0.77)	(-1.03)	(-1.93)	(-3.18)	(-2.85)
	Fake - High	-0.032***	-0.042***	-0.032	-0.050**	-0.037
		(-2.81)	(-2.62)	(-1.45)	(-2.02)	(-1.20)
Volatility	Fake - Low	-0.01	-0.007	-0.015	-0.030**	-0.039**
		(-1.50)	(-0.77)	(-1.39)	(-2.17)	(-2.51)
	Fake - High	-0.025	-0.016	0.003	-0.023	0.032
		(-1.25)	(-0.56)	(0.08)	(-0.50)	(0.57)
Accruals Quality	Fake - Low	-0.024*	-0.012	-0.038	-0.054	-0.054
		(-1.69)	(-0.50)	(-1.28)	(-1.50)	(-1.26)
	Fake - High	-0.008	-0.011	-0.031	-0.067**	-0.058
		(-0.58)	(-0.58)	(-1.30)	(-2.26)	(-1.64)
Analyst Coverage	Fake - Low	-0.009	0.001	0.032	-0.004	0.004
		(-0.55)	(0.06)	(0.92)	(-0.11)	(0.09)
	Fake - High	-0.019***	-0.025***	-0.040***	-0.054***	-0.047***
		(-3.19)	(-2.84)	(-3.62)	(-4.13)	(-3.02)
Retail Ownership	Fake - Low	-0.003	0	-0.015	-0.026	-0.025
		(-0.31)	(0.04)	(-0.89)	(-1.24)	(-1.07)
	Fake - High	-0.028**	-0.032**	-0.03	-0.057**	-0.048*
		(-2.44)	(-2.25)	(-1.54)	(-2.55)	(-1.75)
Tweets	Fake - Low	-0.025***	-0.033***	-0.048***	-0.069***	-0.072***
		(-3.09)	(-2.81)	(-3.21)	(-3.95)	(-3.63)
	Fake - High	0.002	0.006	0.008	0.011	0.027
		(0.14)	(0.36)	(0.34)	(0.41)	(0.80)
Media Coverage	Fake - Low	-0.019***	-0.022**	-0.019	-0.033**	-0.022
		(-3.24)	(-2.38)	(-1.48)	(-2.15)	(-1.22)
	Fake - High	-0.026*	-0.037**	-0.067***	-0.107***	-0.108***
		(-1.89)	(-2.05)	(-2.94)	(-4.09)	(-3.58)
Num of Comments	Fake - Low	-0.001	-0.005	-0.028	-0.051	-0.046
		(-0.08)	(-0.22)	(-0.92)	(-1.58)	(-1.22)
	Fake - High	-0.048**	-0.072***	-0.109***	-0.149***	-0.144***
		(-2.51)	(-2.97)	(-3.27)	(-3.99)	(-3.26)
Num of Followers	Fake - Low	-0.039***	-0.051**	-0.061*	-0.075**	-0.037
		(-2.77)	(-2.48)	(-1.79)	(-1.98)	(-0.74)
	Fake - High	-0.014	-0.013	-0.029	-0.071*	-0.124***
		(-0.56)	(-0.46)	(-0.80)	(-1.91)	(-3.14)
Email Circulation	Fake - Low	-0.027*	-0.048**	-0.088***	-0.093***	-0.098***
		(-1.88)	(-2.06)	(-3.11)	(-2.80)	(-2.86)
	Fake - High	-0.051***	-0.070***	-0.108***	-0.141***	-0.150***
		(-2.61)	(-3.43)	(-3.58)	(-4.07)	(-3.71)
Shares Purchased	Fake - Low	-0.017***	-0.019**	-0.023*	-0.041***	-0.036**
		(-2.84)	(-2.15)	(-1.94)	(-3.00)	(-2.25)
	Fake - High	-0.019	-0.038	-0.088***	-0.100***	-0.063
		(-1.22)	(-1.51)	(-3.02)	(-3.07)	(-1.55)
Shares Sold	Fake - Low	-0.015**	-0.012	-0.013	-0.030**	-0.025
		(-2.19)	(-1.27)	(-0.97)	(-1.96)	(-1.41)
	Fake - High	-0.025**	-0.043***	-0.072***	-0.091***	-0.076***
		(-2.37)	(-2.86)	(-3.70)	(-3.75)	(-2.63)
Press Releases	Fake - Low	-0.012	-0.013	-0.013	-0.044**	-0.047**
		(-1.29)	(-0.96)	(-0.68)	(-2.17)	(-2.00)
	Fake - High	-0.018*	-0.024*	-0.037*	-0.047**	-0.028
		(-1.84)	(-1.65)	(-1.95)	(-2.01)	(-0.99)

Table 7. Insider Trading and Fake News (Weekly Level)

In this table, we examine whether there are more likely to be fake news 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 news 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.

	Fake Article								
	Small Firms			Mid-size Firms			Large Firms		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Buyer (week-1)	-0.0002			0.0006			-0.0002		
	(-0.23)			(0.66)			(-0.39)		
Seller (week-1)	0.0007			0.0003			-0.0007		
	(0.68)			(0.84)			(-0.90)		
PR (week-1)	0.0013**			0.0002			-0.0010		
	(2.52)			(0.82)			(-1.63)		
Buyer (week=0)		0.0050***			0.0039***			-0.0005	
		(2.96)			(4.13)			(-0.63)	
Seller (week=0)		0.0016			0.0004			-0.0013	
		(0.91)			(1.02)			(-1.53)	
PR (week=0)		0.0028***			0.0023***			0.0002	
		(4.16)			(6.84)			(1.04)	
Buyer (week+1)			0.0058***			0.0022*			-0.0013
			(3.30)			(1.94)			(-0.56)
Seller (week+1)			0.0010			0.0006			0.0005
			(1.01)			(1.63)			(0.54)
PR (week+1)			0.0004			0.0003			0.0002
			(0.61)			(0.90)			(0.25)
Observations	137,560	137,998	137,719	406,508	407,379	406,593	86,956	87,104	86,946
R-squared	0.010	0.012	0.011	0.007	0.008	0.007	0.013	0.013	0.013
Year-month, Firm FEs	X	X	X	X	X	X	X	X	X