Fake News, Investor Attention, and Market Reaction

Jonathan Clarke[1], Hailiang Chen[2], Ding Du[3], and Yu (Jeffrey) Hu[4]

June 2018

Abstract

Does fake news in financial markets attract more investor attention and have a significant impact on stock prices? We use the SEC crackdown of stock promotion schemes in April 2017 to examine investor attention and the stock price reaction to fake news articles. Using data from Seeking Alpha, we find that fake news stories generate significantly more attention than a control sample of legitimate articles. We find no evidence that article commenters can detect fake news. Seeking Alpha editors have only modest ability to detect fake news. The broader stock market appears to price fake news correctly. The stock price reaction to the release of fake news is not significantly different than a matched control sample over short and longer-term windows. We conclude by presenting a machine learning algorithm that is successful in identifying fake news articles.

Keywords: Fake news, investor attention, financial technology, textual analysis, social media

[1] Georgia Institute of Technology, Scheller College of Business, email:jonathan.clarke@scheller.gatech.edu
[2] City University of Hong Kong, Department of Information Systems, email: hailchen@cityu.edu.hk
[3] Carnegie Mellon University, email: dingd@andrew.cmu.edu
[4] Georgia Institute of Technology, Scheller College of Business, email: Jeffrey.hu@scheller.gatech.edu

1. Introduction

Fake news has received considerable press attention in recent years. Facebook, Google, and other social media platforms have received negative attention for promoting articles that later proved to be false. Based on anecdotal evidence, the reach of the fake news stories is staggering. Facebook estimated that fake news stories about the 2016 election reached over 126 million Americans. Allcott and Gentzkow (2017) show that education, age, and total media consumption are correlated with more accurate beliefs about the credibility of news headlines.

The real impact of fake news is less well understood. For example, it is not clear that fake news, despite its reach, had a significant impact on the outcome of the 2016 election. In this paper, we first explore how fake news articles differ from legitimate news articles in the domain of financial markets. We then study the real impact of fake news articles by examining both investor attention and the stock price reaction to fake news articles. Our sample of fake news articles is obtained from the SEC crackdown of several stock promotion schemes in April 2017. The SEC enforcement action covered 27 stock promoters who generated bullish articles about public companies under the guise of impartiality and independence. Undisclosed to market participants, the stock promoters were being compensated by companies to write the fake news stories. Many of the fake news articles were distributed on the website Seeking Alpha, one of the largest social media websites for financial markets. Of the 27 stock promoters initially identified by the SEC, 17 agreed to pay settlements ranging from $2,200 to $3 million.

As an example of how these schemes worked, consider the case of Lidingo Holdings. Between 2011 and 2014, Lidingo paid writers to generate dozens of optimistic articles for public company clients and failed to disclose to investors that these were paid promotions. The articles typically appeared on Seeking Alpha and were touted as impartial and independent advice. Over the time period covered, Lidingo received in excess of $1 million dollars in compensation for the articles.

We identify all promotional articles listed in the SEC complaint that appeared on the website Seeking Alpha between August 1, 2011 and December 31, 2013. This sample consists of 383 fake news articles. We also have information on the 157,253 legitimate news articles published during the same time period on Seeking Alpha. For articles published between August 1, 2012 to March 31, 2013, we collect a proprietary dataset from the company Seeking Alpha containing a number of data items relating to investor attention, including the number of times the article was viewed (i.e., page views), the number of distinct machine cookies that viewed the page, the number of times the article was read to the end, and the number of page views for which the comments were read. We also collect data on the number of comments received by each article and the content of each comment in our sample.

We have four primary results. First, relative to a control sample, fake news stories attract substantial investor attention. Page views, unique pages views, and the number of users who read the article to the end are all significantly higher for the sample of fake news articles. The results are economically meaningful. Fake news articles generate 82% more page views on average than legitimate news articles.

Second, we find that article commenters are unable to identify fake news stories. Fake news articles do not receive a significantly different number of comments compared with legitimate news articles. The level of contradiction between the sentiments revealed in the Seeking Alpha article and the comments it receives is not significantly greater for fake news articles relative to a control sample. We next examine whether Seeking Alpha editors are able to identify the fake news. Editors assign scores (on a four-point scale) across three dimensions to each article. The dimensions capture how convincing the article is (double weighted); how actionable the article is; and how well presented the article is. We find no economically meaningful evidence that editors assign lower scores to fake news articles. However, editors are 8.6% less likely to choose a fake news article as an editor pick.

Third, we show that the stock market appears to discount fake news articles. We find that the stock price reaction to fake news articles is not statistically different than a matched control sample over both short and long-run windows surrounding the release of the fake news.

Finally, we develop a machine learning algorithm based on linguistic characteristics for identifying fake news stories. Building on the literature that relies on linguistic styles to detect deception (Newman et al. 2003), we first demonstrate that there are statistically significant differences in 65 out of the total 94 output variables defined by the Linguistic Inquiry and Word Count (LIWC 2015) dictionary (Francis et al. 1993; Pennebaker et al. 2015) between the sample of fake news articles and its matched control sample of legitimate news articles. We then train a logistic regression classifier based on the output variables generated from the LIWC 2015 dictionary. Our classifier is able to achieve on average an F1 score of 93% on out-of-sample predictions.

Our findings are important to social media platforms, regulators, and investors. Our machine learning algorithm provides a method for social media platforms and regulators to identify attempts to manipulate the market by releasing fake news stories. For investors, our results suggest that fake news articles account for a small minority of articles on Seeking Alpha. Nonetheless, the fake news stories attract significant investor attention. Article commenters and Seeking Alpha editors have limited ability to detect fake news articles. However, the market in aggregate is able to discern fake news. The stock price reaction to fake news articles is not statistically different than a matched control sample over short and long horizons.

The remainder of the paper proceeds as follows. Section 2 reviews prior literature related with fake news. Section 3 discusses our data and defines the key variables used in the analyses. Section 4 presents our primary results. Section 5 concludes.

2. Literature Review

Our paper contributes to several strands of research across the information systems, accounting, and finance literatures. In the information systems literature, for example, several papers examine the incidence and impact of fake product reviews. Luca and Zervas (2016) find that approximately 16% of reviews on Yelp, a crowd-sourced review platform, are fake. Hu et al. (2012) and Mukherjee et al. (2012) use textual analysis to identify fake reviews. Lappas et al. (2016) show that in certain markets as little as 50 fake reviews is sufficient for an attacker to surpass the visibility of its competitors. This paper extends the literature from the study of fake product reviews to the study of fake news.

Our paper also adds to the growing literature on using machine learning and textual analysis to detect fraudulent or deceptive activity. Larker and Zakolyukina (2012) use textual analysis to predict deceptive discussions in earnings conference calls. The authors' prediction model outperforms a random guessing strategy by 6 to 16%. Hobson et al. (2011) find that CEO speech patterns can also be used to detect accounting restatements. Abbasi et al. (2012) develop a meta-learning model for detecting financial fraud. Purda and Skillicorn (2014) also provide a tool for identifying fraudulent reports in the management discussion and analysis sections of annual reports. We add to this literature by developing a machine learning algorithm for detecting fake news in financial markets.

In addition to contributing to the nascent literature on fake news, our paper adds to the finance literature on stock price manipulation. Aggrawal and Wu (2006) use data from SEC enforcement actions to show that prices increase through the manipulation period and fall post manipulation. Sabherwal et al. (2011) provide evidence that stock message boards can be used to manipulate stock prices. Leuz et al. (2017) find that 6% of active investors participate in at least one "pump-and-dump scheme" between 2002 and 2015. Investor losses are large, averaging nearly 30%. Our paper adds to this literature by providing unique evidence on the investor attention associated with stock manipulation schemes.

We also add to the evidence on company sponsored research. Many of the fake news articles in our sample were paid for by company management. While Kirk (2011) finds that company sponsored research is generally informative, especially when research firms have established policies that reduce conflicts of interest, our research highlights the dark side of company sponsored research. Publicly traded firms can work with stock promoters to temporarily manipulate stock prices higher.

Finally, this paper adds to the growing accounting and finance literatures on social media and the stock market. Antweiler and Frank (2004) show that the effect of messages posted on Internet stock message boards such as Yahoo! Finance and Raging Bull on stock returns is statistically significant but economically small. Chen et al. (2014) show that articles and comments on Seeking Alpha are informative and can be used to predict future stock returns. Jame et al. (2016) show that crowdsourced earnings forecasts on Estimize are incrementally useful in predicting firm earnings and measuring the market's expectations. Bartov et al. (2018) show that tweets on the Twitter platform can help predict a firm's quarterly earnings and the announcement returns associated with the earnings announcement.

3. Data and Variables

3.1 Fake News

We use the SEC's enforcement action issued on April 10th, 2017 (Reuters 2017; SEC 2017) to identify fake news stories that were distributed on financial websites and part of stock promotion schemes. A list of 494 fake stock news articles is made publicly available by SEC.[5] These articles were published from August 16, 2011 to March 10, 2014 on 13 different financial websites, among which Seeking Alpha (SA) had the largest number of fake news articles (412 out of 492). Seeking Alpha is one of the largest investment related social media sites in the world. It has over ten million unique visitors in 2018 (https://seekingalpha.com/page/who_reads_sa). The site relies on a crowdsourced contributor network to publish opinion and analysis articles on a wide range of stocks.

We obtain a proprietary dataset from Seeking Alpha on 157,636 opinion and analysis articles written by contributors on Seeking Alpha and their commentaries between August 1, 2011 and December 31, 2013. Within this time period, we find 383 fake news articles covered by the SEC complaint. For our analysis, we assume all the remaining 157,253 articles are legitimate ones. *Fake* is an indicator variable taking the value of one if the article was identified by the SEC as fake news and zero otherwise. Table 1 presents the summary statistics of all main variables for the groups of fake news articles (Panel A) and legitimate news articles (Panel B).

---

[5] A copy of this list is available at https://ftalphaville-cdn.ft.com/wp-content/uploads/2017/04/10231526/Stock-promoters.pdf, last accessed May 24, 2018.

For each article, we have the following data: article id, title, main text, date of publication, author name, and list of stock tickers covered in the article. *Length* is the total number of words in an article. To measure the sentiment revealed in an article, we adopt the dictionary-based approach widely applied in the finance literature (e.g., Tetlock 2007, Tetlock et al. 2008, Chen et al. 2014) and use the negative word list developed by Loughran and McDonald (2011) specifically for financial contexts. *%NegWord* is the fraction of negative words in an article in percentage points. *Premium* is an indicator variable denoting whether the author of the article receives monetary compensation from Seeking Alpha, which is determined mainly by the page views of the article. As shown in Table 1, compared with legitimate news articles, fake news articles are on average longer in length, less negative (i.e., more positive) in sentiment, and less likely to participate in Seeking Alpha's premium partnership program to receive monetary compensation.

3.2 Investor Attention

We also obtain a proprietary dataset from Seeking Alpha on the attention received by all articles published from August 1, 2012 to March 31, 2013. This large dataset contains article-hourly level attention measures for 41,948 articles, 156 of which are fake news articles. The measures of attention include page views (*PVs*), the number of unique visitors (*Uniques*), the number of times an article was read to the end (*ReadToEnd*), and the number of page views for which the comments were read (*ReadComment*). Most of the page views occur within the first few days of article publication, and the attention fades away very fast. Among the first week, the proportions of page views generated on each day for an average article are 57.6%, 24.6%, 7.3%, 4.2%, 2.7%, 2.0%, and 1.6%, respectively. Comparing between fake news and legitimate news articles, we find that on average fake news articles generate more page views, have a larger number of unique visitors, and are read to end more often. However, the comments of fake news articles are read less often.

3.3 Commenter and Editor Reactions

To examine how SA commenters and editors may react to fake news and legitimate articles differently and in particular whether SA commenters and editors would be able to detect fake news, we construct several measures to assess their reactions to SA articles.

Comment is the total number of comments received by an article. Although the number of comments is often positively correlated with the number of page views, the number of comments reveals the level of discussions around an article by the community. For the articles that receive at least one comment, *%Contradiction* is defined as the absolute difference between the fraction of negative words in the article and the average fraction of negative words across all comments received by the article (in percentage points). This variable is intended to measure whether commenters have different views from those expressed in an article.

EditorPick is an indicator variable denoting whether the article is featured in the Editors' Picks section. Starting from June 20, 2012, editors at Seeking Alpha assign scores (1-4 scale) to each article along several dimensions. *ConvincingScore* is designed to capture whether the writer in the article can share pertinent information about the stock, sector or style of investing. *ActionableScore* is designed to measure whether the article provided new information about a sector or security. *WellPresentedScore* measures whether the article was well-written and leveraged images, charts, and data sources to craft a logical, easy-to-understand investment thesis. *AggregateScore* is a composite of these three scores, among which the *Convincing Score* is double weighted. These scores are available for 268 fake news articles and 91,392 legitimate news articles. The variables related with editor reactions represent the SA editors' opinion about the underlying quality of an article. If an article does not report facts or real news, then it should be rated lower from an editorial point of view.

3.4 Abnormal Returns

We capture the market impact of a Seeking Alpha article by focusing on its primary ticker. To identify the main focus of an article, either the author or a Seeking Alpha editor tags an article with a primary stock ticker. When there is no primary ticker for an article, the article either does not cover any specific stock ticker or discusses several stock tickers all together. For all fake news articles, we consider the stock ticker identified by the SEC complaint as the primary ticker of an article. 12 distinct stock tickers are associated with fake news articles. For legitimate news articles, we perform market return analysis only for the subsect of articles that contain a primary ticker.

Stock prices and returns data are collected mainly from CRSP. We require the firms to have a CRSP share code of 10 or 11. This excluded ADRs, SBIs, closed-end funds, and REITS. Because the stocks covered by fake news articles are very small firms, the CRSP data is relatively more incomplete for such firms. We therefore supplement the CRSP data with manually collected data from Yahoo Finance and Google Finance for fake news articles. In total, we have returns data for 346 fake news articles and 51,051 legitimate news articles in our sample.

We use standard event study methods to estimate the excess return for firm *i* on date *t* as:

$$ARet_{i,t} = Ret_{i,t} - E\left(Ret_{i,t}\right)$$

where $ARet_{i,t}$ is the abnormal return, $Ret_{i,t}$ is the actual return, and $E(Ret_{i,t})$ is the expected return. In order to have more observations for our analysis, we use the market adjusted returns as the abnormal returns. We are interested in both short-term and long-term market performances. The short-term performances are measured over the [0, +1] and [0, +2] windows. The long-term performances are measured over two windows from 6 months to a year: [+3, +120] and [+3, +242].

We also calculate the abnormal return over the period from [-30,-1] relative to the publication date to control for the run-up in the stock price prior to publication.

3.5 Firm Characteristics

For all articles that contain a primary ticker, we also consider the firm's characteristics including market capitalization (*Size*); market-to-book ratio (*Market-to-Book*), return on assets (*ROA*), and leverage (*LEV*). These variables are measured at the fiscal year end of the year prior to the article being published. Compared to legitimate news articles, fake news stories tend to target small firms (based on market value). The size difference is quite large. The average market value of firms that are the subject of fake news stories is $41.3 million, while legitimate news articles is $57.3 billion. The average market capitalization of firms that are the subject of fake news articles would place them in the micro-cap universe of stocks. Fake news firms have a significantly lower profitability as measured by ROA and higher leverage than legitimate news firms.

4. Analyses and Results

Our analysis of fake news proceeds as follows. In Sections 4.1, we identify a subset of legitimate news articles on firms with similar characteristics as those covered by fake news articles. In Sections 4.2 to 4.5, we compare fake news articles with a matched group of legitimate news articles in terms of investor attention, and reactions from SA commenters and editors, and stock returns, respectively. In section 4.6, we present a machine learning algorithm for detecting fake news in financial markets based on linguistic characteristics.

4.1 Propensity Score Matching

The summary statistics in Table 1 reveal that the sample of fake news articles labelled by SEC are different from the population of legitimate news articles in many aspects. Given that the fake news articles mainly target micro-cap firms, we first conduct propensity score matching (Caliendo and Kopeining 2008; Dehejia and Wahba, 2002; Rosenbaum and Rubin 1983) to find a matched group of legitimate news articles that cover a set of firms similar as those targeted by fake news articles. For this purpose, we conduct a Probit regression on the full sample of both fake and legitimate news articles, for which we have all the firm characteristics variables including *Size*, *Market-to-Ratio*, *ROA*, and *LEV*. These firm characteristics are measured as of the fiscal year end of the year prior to article publication. We also include the abnormal return over the previous month prior to article publication in the regression. Both sector and quarter dummies are included too. Table 2 presents the regression results of the Probit regression. Column (1) presents the full sample result. We find that market value, market-to-book ratio, and past return are negatively associated with the probability of being a fake news article, while return on assets and leverage are positively associated with the probability of being a fake news article.

We perform the one-on-one nearest neighbor matching method to find a matched legitimate news article for each fake news article in our sample. With a caliper of 0.02 and common support, we find a matched sample of 496 articles in total, among which 248 are fake news articles. Column (2) of Table 2 presents the result of the Probit regression on this matched sample. All the firm characteristics and past return are statistically insignificant at the 10% level, indicating that the firms covered by the matched legitimate news articles share similar characteristics as those of fake news articles.

4.2 Investor attention surrounding fake news articles

In this section, we examine investor attention to fake news articles relative to the matched group of legitimate news articles identified in Section 4.1. Before making the comparison in a regression framework, we first visually inspect and compare the trends of investor attention to these two types of articles over time. As mentioned in the data section, the attention dataset we obtain from Seeking Alpha only covers part of the study period, so the number of fake news articles used for this analysis is 86.

Figure 1 plots the mean of each of the four attention measures for fake and legitimate news for the first seven days after the release of the article. Among these four attention measures, fake news stories generate higher levels of investor attention than legitimate news articles in page views, the number of unique viewers, and the number of views that reach the end of an article (Panels A, B, and C). The difference is particularly large in the first two days. Attention dissipates quickly after the first two days. In addition, it appears that there is no difference between these two types of articles in terms of the number of page views that also read the comments (Panel D).

We further examine the investor attention of fake news articles in a regression framework. Table 3 presents the results of the regressions. The dependent variable is one of the four measures of attention measured over the first three days after the release of the article.[6] *Fake* is the main independent variable of interest. We control for a host of article characteristics and firm characteristics. Controls for article characteristics include the article length (*Log(Length)*), fraction of negative words (*%NegWord*), whether the author was paid to write the article (*Premium*), whether the article was an editor-pick (*EditorPick*), and the number of comments (*Log(Comment))*.

[6] Our results remain qualitatively the same when the attention variables are measured over a different time window, e.g., first seven days after article publication.

Controls for firm characteristics include market value (*Size*), market to book ratio (*Market-to-Book*), return on assets (*ROA*), and leverage (*LEV*). One-month past return ($ARet_{-30,-1}$) is included too. Both industry sector fixed effects and calendar quarter fixed effects are also included in the regressions.

For the first three measures of investor attention, we find that fake news articles generate more investor attention than legitimate news stories. The coefficient on *Fake* is positive and statistically significant at the 1% level in Columns (1) to (3) of Table 3. The results are also economically meaningful. For example, fake news articles on average generate 82.0% more page views than legitimate news articles. However, we find no statistically significant difference in the number of page views that also read article comments between fake news articles and legitimate news articles. Overall, the regression results are quite consistent with the insights revealed in Figure 1.

4.3 Commenter reaction to fake news

In this section, we examine whether and how individuals commenting on Seeking Alpha articles may react differently to fake news. We regress *Comment* and %*Contradiction* on the *Fake* indicator variable and include controls for various article characteristics and firm characteristics.

The results are presented in Table 4. In Column (1), the relation between the *Fake* indicator variable and *Comment* is insignificant. Thus, fake news articles do not generate a statistically different amount of comments than the matched sample of legitimate news articles. Together with the result for page views presented in the previous section, we can infer that fake news articles only attract more views but do not necessarily incite more discussions.

In Column (2), we regress %*Contradiction* on the *Fake* news indicator and the same set of control variables as in Column (1). If commenters were able to detect fake news, we would observe a higher level of disagreement between the sentiment by commenters and the sentiment revealed from the article. If commenters agreed more often with the article author for whatever reasons (e.g., commenters are hired by the author to promote the article by making comments, or commenters genuinely believe in the viewpoints shared by the author), then we would observe a lower level of disagreement or contradiction. The coefficient on *Fake* in Column (2) turns out to be statistically insignificant at the 10% level. This suggests that fake news articles generate neither more nor less contradiction in terms of sentiments revealed from the comments and the original article than the matched sample of legitimate news articles.

Overall, the results in Table 4 suggest that fake news articles do not generate more comments or more disagreement than legitimate news articles. This provides some initial evidence that commenters are not able to detect fake news stories.

4.4 Editor reaction to fake news

We next examine whether Seeking Alpha editors score fake news articles lower than legitimate articles. Starting from June 20th, 2012, editorial board at Seeking Alpha reviews all articles submitted to Seeking Alpha. Articles are given scores (1-4 scale) along three dimensions: a *convincing score*, an *actionable score*, and a *well-presented score*. *Convincing score* is designed to capture whether the writer in the article can share pertinent information about the stock, sector or style of investing. *Actionable Score* is designed to measure whether the article provided new information about a sector or security. *Well-presented score* measures whether the article was well-written and leveraged images, charts, and data sources to craft a logical, easy-to-understand investment thesis. An *Aggregate score* is computed as a composite of the *Actionable, Well-presented*, and *Convincing* scores, where the *Convincing Score* receives double weight. In addition to these scores, we also track whether an article was an editor pick. According to Seeking

Alpha, editor picks are highly discretionary and are based on the perceived timeliness and impact of the article.

Table 5 presents regression results. In Column (1), fake news articles are associated with an 8.60% lower likelihood of being an editor pick. Columns (2) through (4) show that fake news is negatively related to editor scores for convincing, actionable, and well-presented. At first glance, the results appear to suggest that editors have an ability to discern fake news from legitimate news. However, the results are not economically meaningful, because scores are at most 0.176 points lower for fake news articles on a 4 point scale. Column (5) examines the aggregate editor score. While the coefficient on the *Fake* indicator is also negative and statistically significant at the 1% level, the magnitude of the result is again not economically meaningful.

Overall, the evidence in Sections 4.3 and 4.4 suggests that neither editors nor commenters can reliably detect fake news. In the next section, we examine the market reaction to fake news articles.

4.5 Stock returns around fake news stories

In this section, we examine the stock price reaction to fake news articles. There are two possible hypotheses related to the stock price reaction to fake news. In an informationally efficient market, the abnormal performance surrounding the release of fake news should be a non-event. Consequently, there should be no abnormal performance surrounding the release of the fake new. If the market is fooled by the release of fake news, however, then there could be an initial reaction to the news story followed by return reversal. Aggarwal and Wu (2006) find evidence consistent with latter hypothesis in their analysis of stock market manipulations investigated by the SEC.

Our results are presented in Table 6. We examine returns over four windows: [0,+1]; [0,+2]; [+3, +120]; and [+3, +242]. For the former two short term return windows, we consider the return performance on the day of article publication instead of skipping the first day, because we are more interested in how the market reacts to fake news from the very beginning but less interested in predicting future returns. For the latter two long term return windows, we control for the initial market reaction. In all columns, we control for article characteristics and firm characteristics and include sector and quarter fixed effects. The key variable of interest is the *Fake* indicator. Across all four specifications, the stock price reaction to fake news articles is not statistically different than the control sample. In other words, the market reacts to fake news articles in a similar way as it reacts to the matched control sample of legitimate news articles in both short and long terms. This provides some evidence that the market is not fooled by the release of fake news.

4.6 Detecting Fake News Based on Linguistic Characteristics

In this last analysis, we explore whether it is feasible to detect fake news by conducting textual analysis and extracting features related with linguistic styles. Prior studies in the literature have shown the possibility of relying on linguistic styles to detect deception (e.g., Newman et al. 2003). We conjecture that the authors of fake news stories would make involuntarily different word choices than the authors of legitimate news stories despite the various efforts made by fake news stories' authors to disguise and hide their true intentions. We test this conjecture by analyzing the 93 output variables generated by the Linguistic Inquiry and Word Count (LIWC 2015) software (Francis et al. 1993; Pennebaker et al. 2015) for the matched sample of 248 fake news articles and 248 legitimate news articles. More specifically, we perform paired t-tests on the 93 output variables from LIWC 2015 and present the results in Table 7. We find that 65 out of these 93 linguistic characteristics are significantly different between fake news articles and legitimate news articles. This implies that linguistic styles can be potentially helpful in detecting fake news.

To further investigate how much linguistic styles can contribute to the detection of fake news, we develop a logistic regression classifier based on the 93 output variables from the LIWC 2015 software. Because the number of fake news articles in our sample is much smaller than the number of legitimate news article, we have the class imbalance problem. We adopt the following approach to address this problem. First, we randomly select 383 articles from the population of 157,253 legitimate news articles. Second, together with the 383 fake news articles we have, we split the sample of 766 articles into a training dataset and a testing dataset with a ratio of 7:3. This way we ensure that both fake news and legitimate news articles are equally represented in the training and testing datasets. Third, we train a logistic regression classifier on the training dataset and then perform prediction on the testing dataset to obtain the performance measures such as precision, recall, and F1 score. Finally, to utilize more information about legitimate news articles that are available in our sample, we repeat the previous three steps for 100 times and assess the prediction accuracy across these 100 experiments.

Table 8 summarizes the results of these 100 experiments. We find that the simple logistic regression classifier based on linguistic styles is effective in detecting fake news stories. On average, this classifier has achieved a precision of 92%, a recall of 94% and an F1 score of 93%. While neither commenters nor editors are able to reliably identify fake news, our results from machine learning and textual analysis imply that linguistic styles can be particularly informative in detecting fake news. We acknowledge that one challenge to implement such machine learning techniques in practice is the lack of sufficient and accurate data on fake news.

5. Conclusion

This paper examines investor attention and the stock price reaction to articles identified by the SEC as fake news. Many of the articles identified as fake news appeared on the website

Seeking Alpha, which allows to examine whether the fake news articles captured investor attention and the impact of the fake news on stock prices.

We find that investor attention, as measured by page views and the number of times an article was read to the end, is significantly higher for fake news articles than a control sample of legitimate news articles. The results are both economically and statistically significant. Fake news articles generate about 82% more page views than a matched sample of legitimate articles.

We find no evidence that the stock price reaction to fake news articles follows a "pump and dump pattern." Rather, the market appears to discount for fake news articles and we find no significant abnormal returns relative to a matched control sample.

We find little evidence that article commenters or Seeking Alpha editors can successfully identify fake news articles. Therefore, we explore whether machine learning techniques can reliably be used to identify fake news articles. Through a textual analysis of both fake news articles and legitimate news articles that specifically looks at linguistic styles, we find that fake news articles differ from legitimate news articles in word choices. A logistic regression classifier solely based on linguistic characteristics is able to successfully identify fake news with high confidence.

Given the proliferation of fake news articles on social media and the attempts to use fake news to influence everything from stock prices to political elections, we believe these results have widespread implications.

Our results have important implications for investors and regulators. Our findings suggest that the market discounts fake news. For regulators, however, our results highlight the difficulty with detecting fake news. Neither article commenters nor editors were able to detect fake news. Our machine learning algorithm, however, shows promise in detecting fake news.

One limitation of our study is that we focus only on cases where a stock promoter uses fake news to manipulate a stock price higher and is identified by the SEC. We cannot identify cases where a short-seller might use fake news to drive stock prices lower. Our sample of fake news articles is also skewed heavily towards micro-cap stocks. Thus, we cannot easily address whether fake news can be used to manipulate stock prices for large companies that likely have significant institutional and analyst following. We leave these topics for future work.

References

Abbasi A., Albrecht C., Vance A., Hansen J. 2012. MetaFraud: A meta-learning framework for detecting financial fraud. *MIS* Quarterly 36: 1293-327.

Allcott, H, Gentzkow M. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31: 211-236.

Antweiler, W. and M. Z. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59: 1259-1294.

Aggarwal, R. K. and G. Wu. 2006. Stock market manipulations. *Journal of Business* 79, 1915-1953.

Bartov, E., Faurel L., and Mohanram P. 2018. Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review*, (93:3), pp. 25-57.

Caliendo, M., Kopeining, S. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys,* (22:1), pp. 31-72.

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

Dehejia, R.H. and Wahba, S. 2002. "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*, (84:1), pp. 151-161.

Francis, J. W. P. M. E., & Booth, R. J. 1993. Linguistic Inquiry and Word Count. Technical Report. Technical Report, Dallas, TX: Southern Methodist University.

Hobson, J., Mayew, W., Venkatachalam, M. 2012. Analyzing speech to detect financial misreporting. *Journal of Accounting Research* 50, 349-392.

Hu, N, Bose I, Koh NS, Liu L. 2012. Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems* 52 (3): 674-684.

Jame, R., Johnston, R., Markov, S., and Wolfe, M. C. 2016. The Value of Crowdsourced Earnings Forecasts. *Journal of Accounting Research* (54:4) pp. 1077-1110.

Kirk, Marcus. 2011. Research for sale: determinants and consequences of paid-for analyst research. *Journal of Financial Economics* 100, 182-200.

Lappas, T, Sabnis G., Valkanas G. 2018. The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research* 27: 940-961.

Larcker, D., Zakolyukina, A. 2012. Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50, 495-540.

Leuz, C., Meyer S., Muhn M., Soltes E, Hackethal A. 2017. Who falls prey to the Wolf of Wall Street? Investor participation in market manipulation. *NBER Working Paper.*

Luca, M., Zervas, G. 2016. Fake it till you make it: Reputation, competition, and Yelp review fraud. Management Science (62:12), pp. 3412-3427.

Loughran, T. and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-ks. *Journal of Finance* 66, 35-65.

Mukherjee, A., Liu, B., Glance, N. 2012. Spotting fake review groups in consumer reivews. In *Proceedings of the 21st International Conference on World Wide Web*: 191-200.

Newman, M., J. Pennebaker, D. Berry, and J. Richards. 2003. Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin* 29, 665-675.

Pennebaker, J. W., Booth, R. J., Boyd, R. L., Francis, M. E. 2015. Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Purda, L., Skillicorn D. 2015. Accounting variables, deception, and a bag of words: Assessing the tools of fraud detection. *Contemporary Accounting Research* 32, 1193-1223.

Reuters. 2017. SEC targets fake stock news on financial websites. https://www.reuters.com/article/us-sec-fakenews-idUSKBN17C1YP

Rosenbaum, P. R., and Rubin, D. B. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* (70:1), pp. 41–55.

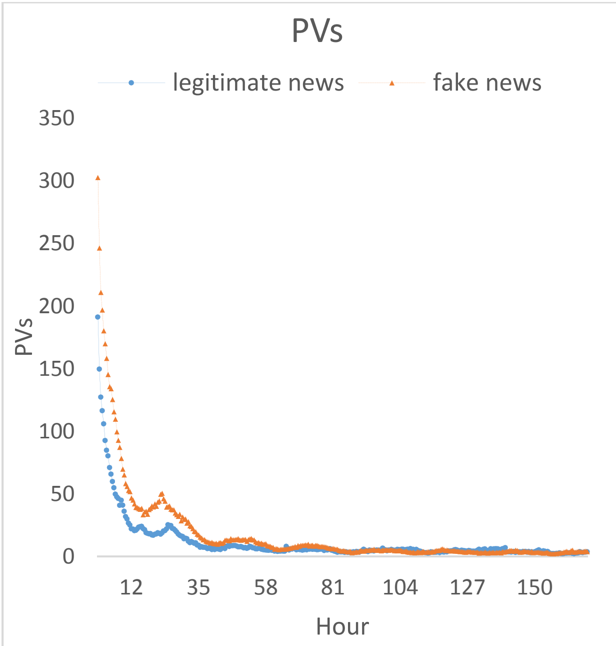
Sabherwal, S., S. Sarkar, and Y. Zhang. 2011. Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news. *Journal of Business, Finance, and Accounting* 38, 1209-1237.

SEC. 2017. SEC: Payments for Bullish Articles on Stocks Must Be Disclosed to Investors. https://www.sec.gov/news/press-release/2017-79

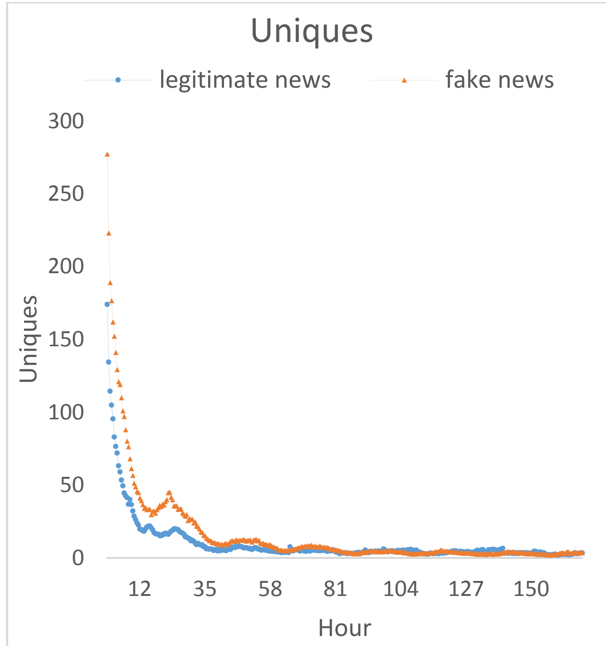
Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62: 1139-68.

Tetlock P., Saar-Tsechansky M, Macskassy S. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63: 1437-1467.

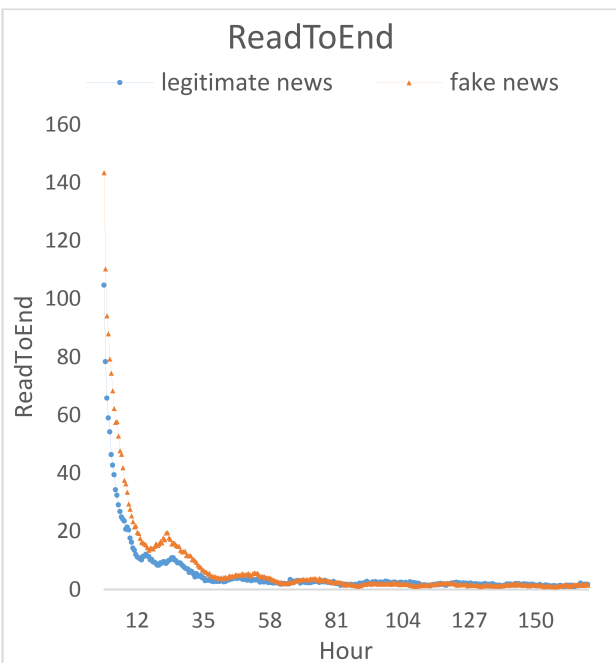
Panel A. *PVs*



Panel B. *Uniques*



Panel C. *ReadToEnd*



Panel D. *ReadComment*

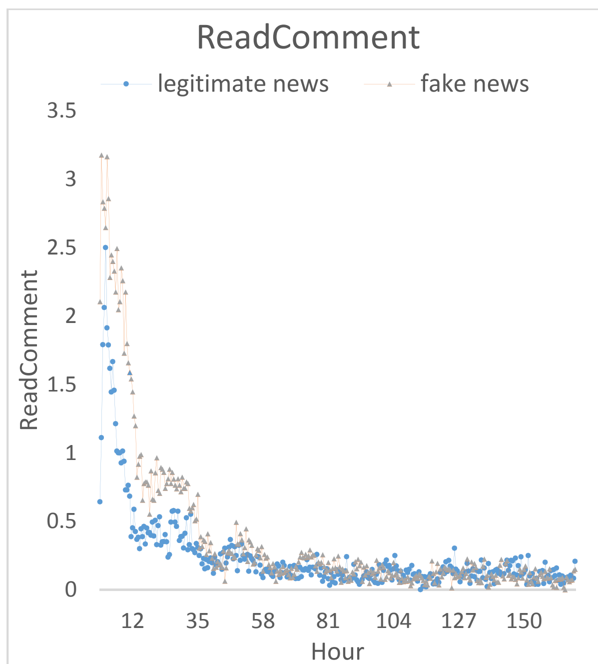


Figure 1. Investor attention on fake news

Table 1. Descriptive statistics

Variable	#Obs	Median	Mean	Std. Dev.	Min	Max
Panel A. Fake news articles						
Length	383	1750	1753.4	625.0	423	4521
%NegWord	383	0.99	1.07	0.51	0	4.28
Premium	383	0	0.41	0.49	0	1
PVs	156	3607.5	4041.4	2670.3	377	13875
Uniques	156	3296.5	3595.4	2358.3	338	12517
ReadToEnd	156	1355.5	1667.0	1135.9	112	5963
ReadComment	156	0	79.5	159.2	0	869
EditorPick	383	0	0.050	0.22	0	1
ConvincingScore	268	3	3.13	0.33	3	4
ActionableScore	268	3	3.17	0.38	3	4
WellPresentedScore	268	3	3.19	0.40	2	4
AggregateScore	268	3	3.18	0.38	3	4
Comment	383	7	11.2	11.9	0	79
%Contradiction	358	0.54	0.69	0.68	0	5.84
$ARet_{0,1}$	346	0.0078	0.022	0.10	-0.23	0.65
$ARet_{0,2}$	346	0.0034	0.027	0.13	-0.25	0.72
$ARet_{3,120}$	346	-0.062	0.018	0.45	-0.77	1.89
$ARet_{3,242}$	346	-0.14	-0.0055	0.55	-0.85	2.06
$ARet_{-30,-1}$	346	0.031	0.12	0.35	-0.49	1.49
Size	357	35.6	41.3	29.8	1.66	126.9
Market-to-Book	357	2.19	-20.3	83.5	-335.7	19.5
ROA	357	-0.66	-3.47	15.8	-114.1	-0.19
LEV	357	0.56	8.90	54.2	0.27	391.3
Panel B. Legitimate news articles						
Length	157253	861	1008.1	668.1	100	16325
%NegWord	157253	1.27	1.48	1.05	0	14.7
Premium	157253	1	0.65	0.48	0	1
PVs	41792	1886	2954.7	4167.2	1	447869
Uniques	41792	1706	2646.8	3771.1	1	432365
ReadToEnd	41792	885	1399.8	1887.5	0	193814
ReadComment	41792	0	116.7	348.9	0	31022
EditorPick	157253	0	0.045	0.21	0	1
ConvincingScore	91375	3	3.11	0.32	1	4
ActionableScore	91375	3	3.13	0.34	1	4
WellPresentedScore	91375	3	3.22	0.42	1	4
AggregateScore	91375	3	3.14	0.36	1	4
Comment	157253	4	14.6	33.5	0	1202
%Contradiction	124760	0.69	0.96	1.19	0	99.0
$ARet_{0,1}$	51051	0.00011	0.0011	0.056	-0.85	2.46
$ARet_{0,2}$	51051	-0.00017	0.00099	0.064	-0.86	2.13
$ARet_{3,120}$	51051	-0.023	-0.010	0.32	-1.11	16.3
$ARet_{3,242}$	51051	-0.047	-0.016	0.47	-1.30	8.02
$ARet_{-30,-1}$	51051	-0.0049	0.0026	0.18	-0.91	5.85
Size	54746	9273.8	57682.9	110235.6	0.0055	626550.4
Market-to-Book	54746	2.64	6.41	61.9	-4027.2	3141.5
ROA	54746	0.058	0.037	3.85	-182.2	198.7
LEV	54746	0.53	0.60	3.36	0	391.3

Table 2. Propensity Score Matching

VARIABLES	(1) *Fake* Before Matching	(2) *Fake* After Matching
Log(Size)	-0.818***	0.106
	(-18.22)	(1.28)
Market-to-Book	-0.001**	0.000
	(-2.14)	(1.25)
ROA	0.085**	0.053
	(2.23)	(0.86)
LEV	0.051**	0.015
	(1.98)	(0.21)
$ARet_{-30,-1}$	-0.212**	0.013
	(-2.38)	(0.11)
Constant	3.366***	-0.742
	(5.82)	(-0.91)
Sector FE	Yes	Yes
Quarter FE	Yes	Yes
Observations	14,518	496
Pseudo R^2	0.616	0.018
Log Likelihood	-592.8	-337.6

Table 3. Investor attention to fake news

VARIABLES	(1) *Log(PVs)*	(2) *Log(Uniques)*	(3) *Log(ReadToEnd)*	(4) *Log(ReadComment)*
Fake	0.820***	0.843***	0.842***	-0.298
	(6.24)	(6.57)	(6.39)	(-0.79)
Log(Length)	0.199*	0.196*	-0.082	0.270
	(1.75)	(1.71)	(-0.68)	(0.86)
%NegWord	0.166*	0.156	0.160	0.021
	(1.77)	(1.61)	(1.60)	(0.07)
Premium	-0.028	-0.022	0.002	-0.171
	(-0.28)	(-0.22)	(0.02)	(-0.57)
EditorPick	0.032	-0.030	-0.081	-1.083***
	(0.19)	(-0.17)	(-0.43)	(-2.83)
Log(Comment)	0.316***	0.306***	0.346***	0.471***
	(6.50)	(6.34)	(7.00)	(3.58)
Log(Size)	0.056	0.068	0.072	-0.306
	(0.74)	(0.90)	(0.95)	(-1.63)
Market-to-Book	0.002***	0.002***	0.002**	0.002
	(2.73)	(2.77)	(2.51)	(0.77)
ROA	-0.131	-0.141*	-0.136*	0.434**
	(-1.65)	(-1.77)	(-1.72)	(2.34)
LEV	-0.185***	-0.190***	-0.167**	0.097
	(-2.76)	(-2.80)	(-2.52)	(0.67)
$ARet_{-30,-1}$	-0.022	-0.009	-0.038	-0.780**
	(-0.22)	(-0.09)	(-0.38)	(-2.60)
Constant	5.262***	5.168***	6.324***	-0.813
	(5.91)	(5.79)	(6.83)	(-0.33)
Sector FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Observations	172	172	172	172
Adj. R^2	0.502	0.506	0.486	0.428

Notes: *, **, and *** indicate p<0.1, p<0.05, and p<0.01, respectively. Robust t-statistics are presented in parentheses.

Table 4. Commenter reaction to fake news

VARIABLES	(1) *Comment*	(2) *%Contradiction*
Fake	-3.247	-0.001
	(-1.43)	(-0.02)
Log(Length)	8.838**	-0.165**
	(2.09)	(-2.54)
%NegWord	2.476*	0.118*
	(1.88)	(1.95)
Premium	5.303***	0.222***
	(3.38)	(3.82)
EditorPick	3.293	0.014
	(0.71)	(0.13)
Log(Comment)		-0.157***
		(-4.62)
Log(CommentWord)		-0.092
		(-1.52)
Log(Size)	0.817	-0.011
	(0.85)	(-0.29)
Market-to-Book	0.003	0.000
	(1.43)	(0.15)
ROA	-2.531	-0.027
	(-1.46)	(-0.70)
LEV	-2.704	0.019
	(-1.64)	(0.55)
$ARet_{-30,-1}$	2.083	-0.027
	(0.99)	(-0.54)
Constant	-69.984**	2.165***
	(-2.09)	(3.15)
Sector FE	Yes	Yes
Quarter FE	Yes	Yes
Observations	496	457
Adjusted R^2	0.132	0.094

Notes: *, **, and *** indicate p<0.1, p<0.05, and p<0.01, respectively. Robust t-statistics are presented in parentheses.

Table 5. Editor reaction to fake news

VARIABLES	(1) *EditorPick*	(2) *Convincing Score*	(3) *Actionable Score*	(4) *WellPresented Score*	(5) *Aggregate Score*
Fake	-0.086***	-0.167***	-0.176***	-0.112**	-0.176***
	(-2.96)	(-3.87)	(-3.46)	(-2.10)	(-3.61)
Log(Length)	0.168***	0.297***	0.211***	0.210***	0.308***
	(5.17)	(6.49)	(4.36)	(3.88)	(6.31)
%NegWord	0.032	0.057**	-0.026	-0.014	0.037
	(1.52)	(1.97)	(-0.78)	(-0.39)	(1.17)
Premium	-0.018	-0.035	-0.010	0.021	-0.033
	(-0.68)	(-0.86)	(-0.22)	(0.42)	(-0.72)
EditorPick	-0.002	-0.005	0.026	-0.014	-0.011
	(-0.14)	(-0.22)	(1.20)	(-0.54)	(-0.48)
Log(Comment)	0.015	0.010	0.001	0.073**	0.004
	(0.94)	(0.46)	(0.04)	(2.23)	(0.16)
Log(Size)	-0.000	-0.001	-0.000	0.001	-0.001
	(-0.39)	(-0.97)	(-0.37)	(1.19)	(-0.91)
Market-to-Book	0.007	-0.030	-0.067	-0.017	-0.045
	(0.37)	(-0.77)	(-1.55)	(-0.34)	(-0.98)
ROA	-0.009	0.011	-0.026	0.010	0.012
	(-0.65)	(0.54)	(-1.24)	(0.39)	(0.61)
LEV	-0.086***	-0.167***	-0.176***	-0.112**	-0.176***
	(-2.96)	(-3.87)	(-3.46)	(-2.10)	(-3.61)
$ARet_{-30,-1}$	-0.005	0.013	0.007	-0.001	0.017
	(-0.43)	(0.67)	(0.26)	(-0.04)	(0.80)
Constant	-1.307***	1.085**	1.673***	1.925***	1.036**
	(-4.92)	(2.23)	(4.39)	(3.78)	(2.03)
Sector FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
Observations	496	350	350	350	350
Adj. R^2	0.106	0.160	0.041	0.069	0.119

Notes: *, **, and *** indicate p<0.1, p<0.05, and p<0.01, respectively. Robust t-statistics are presented in parentheses.

Table 6. Market reaction to fake news

VARIABLES	(1) $ARet_{0,1}$	(1) $ARet_{0,2}$	(2) $ARet_{3,120}$	(3) $ARet_{3,242}$
Fake	-0.011	-0.016	-0.010	0.032
	(-0.86)	(-1.22)	(-0.19)	(0.45)
$ARet_{0,2}$			-0.137	-0.328
			(-0.75)	(-1.44)
Log(Length)	-0.002	0.009	0.062	0.023
	(-0.11)	(0.60)	(1.19)	(0.29)
%NegWord	-0.009	-0.016	-0.098***	-0.089*
	(-0.95)	(-1.63)	(-2.63)	(-1.68)
Premium	0.017	0.012	-0.019	-0.007
	(1.39)	(0.99)	(-0.36)	(-0.10)
EditorPick	-0.007	0.014	0.035	-0.024
	(-0.35)	(0.61)	(0.38)	(-0.23)
Log(Comment)	-0.002	0.003	-0.010	0.006
	(-0.31)	(0.57)	(-0.35)	(0.15)
Log(Size)	0.007	-0.004	0.001	-0.061
	(0.56)	(-0.37)	(0.04)	(-1.11)
Market-to-Book	0.000	0.000	0.000***	0.000**
	(1.10)	(0.31)	(2.74)	(2.29)
ROA	-0.012	-0.002	0.060**	0.111***
	(-0.92)	(-0.40)	(2.25)	(2.79)
LEV	-0.004	-0.002	0.078**	0.109**
	(-0.69)	(-0.21)	(2.27)	(2.31)
$ARet_{-30,-1}$	-0.001	-0.017	-0.207***	-0.229***
	(-0.21)	(-1.32)	(-5.12)	(-5.23)
Constant	-0.060	-0.078	-0.747*	-0.536
	(-0.41)	(-0.56)	(-1.69)	(-0.85)
Sector FE	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes
Observations	496	496	496	496
Adj. R^2	-0.019	-0.019	0.163	0.109

Notes: *, **, and *** indicate p<0.1, p<0.05, and p<0.01, respectively. Robust t-statistics are presented in parentheses.

Table 7. Linguistic characteristics of fake news

	Fake news (N=248)		Legitimate news (N=248)		t-test	
LIWC variables	mean	stdev	mean	stdev	t-stat	p-value
Word count	1644.1	581.8	1333.7	1067.9	4.02	0.000
Analytical thinking	92.4	5.14	93.3	5.36	-1.85	0.065
Clout	50.5	6.13	54.5	8.50	-6.01	0.000
Authentic	29.1	11.9	28.5	14.2	0.47	0.639
Emotional tone	63.5	14.3	56.8	18.2	4.59	0.000
Words per sentence	25.2	3.73	22.7	3.93	7.31	0.000
Words>6 letters	27.9	3.40	27.7	3.69	0.36	0.718
Dictionary words	77.4	4.14	75.5	4.68	4.65	0.000
Function Words	43.7	3.28	42.4	3.95	4.14	0.000
Total pronouns	6.41	1.78	6.39	1.97	0.15	0.884
Personal pronouns	1.40	0.84	2.05	1.22	-6.93	0.000
1st pers singular	0.74	0.53	0.91	0.75	-2.93	0.004
1st pers plural	0.23	0.30	0.48	0.67	-5.30	0.000
2nd person	0.14	0.26	0.16	0.27	-0.69	0.493
3rd pers singular	0.054	0.11	0.11	0.27	-2.91	0.004
3rd pers plural	0.24	0.26	0.40	0.41	-5.28	0.000
Impersonal pronouns	5.01	1.31	4.34	1.31	5.74	0.000
Articles	8.55	1.12	8.09	1.24	4.31	0.000
Prepositions	14.7	1.20	14.9	1.33	-2.07	0.039
Auxiliary verbs	6.88	1.24	6.53	1.47	2.88	0.004
Common adverbs	3.10	0.88	2.76	1.07	3.87	0.000
Conjunctions	5.35	0.84	4.66	0.90	8.83	0.000
Negations	0.71	0.33	0.86	0.42	-4.18	0.000
Regular verbs	10.4	1.77	10.0	2.21	2.11	0.035
Adjectives	5.09	0.95	4.61	1.02	5.44	0.000
Comparatives	2.91	0.71	2.55	0.77	5.48	0.000
Interrogatives	0.87	0.34	0.90	0.44	-0.93	0.354
Numbers	4.04	1.63	4.86	2.31	-4.56	0.000
Quantifiers	2.26	0.56	2.05	0.61	3.97	0.000
Affect Words	4.01	0.74	3.96	1.03	0.52	0.605
Positive emotion	3.01	0.63	2.81	0.80	3.12	0.002
Negative emotion	0.97	0.44	1.12	0.64	-3.15	0.002
Anxiety	0.21	0.17	0.29	0.25	-3.87	0.000
Anger	0.20	0.18	0.15	0.20	2.71	0.007
Sadness	0.31	0.25	0.34	0.26	-1.12	0.262
Social Words	3.53	1.09	4.37	1.40	-7.54	0.000
Family	0.010	0.031	0.020	0.059	-2.15	0.032
Friends	0.12	0.18	0.15	0.18	-2.32	0.021

Female referents	0.038	0.11	0.039	0.13	-0.10	0.923
Male referents	0.047	0.11	0.13	0.30	-4.16	0.000
Cognitive Processes	10.1	1.86	9.43	1.83	3.88	0.000
Insight	1.88	0.64	1.85	0.70	0.56	0.577
Cause	2.47	0.69	2.30	0.71	2.59	0.010
Discrepancies	1.29	0.46	1.14	0.54	3.45	0.001
Tentativeness	2.41	0.64	2.24	0.78	2.71	0.007
Certainty	1.03	0.40	0.97	0.44	1.43	0.154
Differentiation	2.32	0.73	2.06	0.71	3.88	0.000
Perceptual Processes	1.07	0.45	1.14	0.61	-1.54	0.125
Seeing	0.64	0.31	0.59	0.36	1.41	0.159
Hearing	0.13	0.13	0.17	0.21	-2.42	0.016
Feeling	0.18	0.17	0.28	0.36	-3.79	0.000
Biological Processes	3.71	1.77	3.15	1.94	3.31	0.001
Body	0.59	0.46	0.52	0.58	1.53	0.128
Health/illness	2.92	1.53	2.48	1.66	3.07	0.002
Sexuality	0.16	0.28	0.089	0.22	3.20	0.001
Ingesting	0.22	0.53	0.19	0.26	0.70	0.484
Drives and Needs	6.68	1.08	7.06	1.56	-3.16	0.002
Affiliation	0.83	0.49	1.34	0.78	-8.71	0.000
Achievement	2.36	0.70	1.97	0.65	6.44	0.000
Power	2.69	0.65	2.83	0.91	-2.05	0.041
Reward focus	1.57	0.51	1.26	0.56	6.41	0.000
Risk focus	0.56	0.32	0.67	0.43	-3.42	0.001
Past focus	2.12	0.74	2.21	0.95	-1.28	0.202
Present focus	7.12	1.37	7.21	1.75	-0.67	0.503
Future focus	1.68	0.50	1.45	0.61	4.68	0.000
Relativity	13.9	2.06	13.8	2.41	0.68	0.496
Motion	1.72	0.59	1.68	0.62	0.63	0.529
Space	7.01	0.97	7.24	1.38	-2.16	0.031
Time	5.40	1.31	5.08	1.39	2.65	0.008
Work	6.32	1.57	5.97	1.58	2.46	0.014
Leisure	0.38	0.29	0.35	0.27	1.32	0.187
Home	0.040	0.061	0.066	0.10	-3.46	0.001
Money	2.79	1.20	3.30	1.73	-3.82	0.000
Religion	0.044	0.16	0.022	0.057	2.05	0.041
Death	0.066	0.093	0.061	0.11	0.59	0.558
Informal Speech	0.15	0.13	0.29	0.29	-6.79	0.000
Swear words	0.00028	0.0044	0.0098	0.098	-1.53	0.127
Netspeak	0.029	0.071	0.14	0.26	-6.39	0.000
Assent	0.015	0.035	0.035	0.074	-3.85	0.000
Nonfluencies	0.11	0.11	0.14	0.15	-2.27	0.024
Fillers	0.00016	0.0025	0	0	1.00	0.318

All Punctuation	15.7	2.22	16.7	2.94	-4.48	0.000
Periods	4.26	0.59	4.78	0.83	-8.09	0.000
Commas	4.90	1.17	4.40	1.29	4.49	0.000
Colons	0.20	0.14	0.42	0.37	-8.63	0.000
Semicolons	0.086	0.11	0.075	0.14	1.01	0.313
Question marks	0.068	0.13	0.10	0.21	-2.05	0.041
Exclamation marks	0.0096	0.034	0.017	0.10	-1.08	0.280
Dashes	1.50	0.68	1.55	0.91	-0.70	0.484
Quotation marks	0.37	0.39	0.40	0.50	-0.61	0.545
Apostrophes	1.09	0.43	1.06	0.67	0.54	0.589
Parentheses (pairs)	1.59	0.75	1.99	1.20	-4.34	0.000
Other punctuation	1.58	0.83	1.91	1.36	-3.32	0.001

Table 8. Performance of the Machine Learning Algorithm for Detecting Fake News

	Median	Mean	Std. Dev.	Min	Max
Precision	0.93	0.92	0.03	0.86	0.97
Recall	0.95	0.94	0.02	0.88	0.99
F1	0.93	0.93	0.02	0.88	0.97

Notes: Results are based on 100 experiments.