THE UNIVERSITY OF CHICAGO

COMPLEXITY AVERSION WHEN SEEKING ALPHA

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

TARIK UMAR

CHICAGO, ILLINOIS

JUNE 2017

I dedicate this dissertation to my wife – Dr. Penny Fang – who encouraged me to apply for the Ph.D. and stuck with me for the five years I was away in Chicago.

Simplicity is attractive.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee of advisors: Lubos Pastor (co-chair), Kelly Shue (co-chair), Amir Sufi, and Richard Thaler. This paper was awarded "best PhD paper" at the 2016 Colorado Finance Summit and was a finalist for the BlackRock Applied Research Award. This work has benefited from feedback at the following seminars: Boston College, Chicago Booth, Dartmouth, Rice University, Securities and Exchange Commission, Texas A&M, University of California San Diego, University of Oklahoma, University of Virginia, Vanderbilt, Washington University-St Louis, and Wharton. The following individuals connected with *Seeking Alpha* provided me with valuable resources: Eli Hoffmann (CEO), Selig Davis (Vice President of Audience and Mobile), and Yoni Madar (Mobile Product Analyst).

ABSTRACT

I provide causal evidence that textual complexity affects investor and market behavior, using two complementary settings. First, examining field data with randomization from *Seeking Alpha*, I find that a standard-deviation increase in headline length (negativity) leads to 12%-fewer (2%-more) views. The effects are larger for less-sophisticated investors. Second, examining firm-earnings releases, I find a market effect by instrumenting title length with company-name length. A standard deviation increase in length leads to 5%-less-announcement turnover, 50-basis-points-tighter-intraday-price ranges, and 40-basis-points-return underreactions, correcting within one month.

CHAPTER 1

COMPLEXITY AVERSION WHEN SEEKING ALPHA

A journalist knows that in order to capture and maintain her readers' limited attention, she must write short, simple, and active reports (Kahneman, 1973). Company disclosures, in contrast, are lengthy and complex. One would hope that investors in a high-stakes market would not be repelled, but prior work suggests complexity matters for investor and market behavior. Studies find that investors underreact to textual content and that longer disclosures are correlated with less trading by individual investors and greater-post-filing volatility.[1] However, omitted variables related to events may drive both investor behavior and readability, as suggested by Bloomfield (2008). Also, longer disclosures may have more subtle information buried in footnotes, and work on investor attention has found that variation in the positioning of content matters for attention.[2] This paper isolates changes in complexity from changes in both the content and positioning of information. I do so using exogenous variation in title length in two distinct, and complementary, field settings. The results establish that complexity is a driver of investor attention and helps explain market reactions to news.

To show textual complexity affects investor attention, I need a measure of investor attention to a particular text, variation in the text unrelated to the event reported, and random assignment of texts to investors. To address these challenges, I analyze field data with ran-

1. See Tetlock (2007); Tetlock et al. (2008); Engelberg (2008); Li (2008); Miller (2010); Loughran and McDonald (2011); Lawrence (2013); Loughran and McDonald (2014); Loughran and Mcdonald (2016).

2. Prior work finds that markets underreact to hard-to-find news that is slow-moving, quiet, and buried in footnotesGiglio and Shue (2014) show investors underreact to the absence of news. Da et al. (2014) finds investors underreact to slow-moving information. Cohen et al. (2015) shows that subtle changes in text of disclosures are informative of returns. DellaVigna and Pollet (2009), Niessner (2014), and Hirshleifer et al. (2009) find investors underreact to news released on days investors are distracted, like Fridays, around holidays, or days with many other announcements. However, there may also be a selection effect that explains the Friday effect (Michaely et al., 2016). Also see (Hirshleifer and Teoh, 2003; Li, 2008; Loughran and McDonald, 2014)

domization from *Seeking Alpha*, a crowdsourced-investment-research firm.[3] On January 3, 2016, *Seeking Alpha* began allowing authors to propose two plausible headlines (titles) for the same stock report. The editor assigned to review the stock report can provide a third title. The two-to-three plausible titles are then randomly sampled on investors who signed up to receive real-time-alert emails about the topic company ("title testing"). The alert emails are identical except for the title, which is randomly assigned from those proposed. None of the body of the stock report is included in the email, and there is only one title in the email. Investors must click a link in the email to read the full stock report. My chief measure of attention to a title is the number of investors who click the link in the email. I use the number of clicks within 30 minutes of the email being sent to focus on attentive and active investors.[4] I also measure the number of investors who scrolled to the end of the report by title.

The *Seeking Alpha*-title-testing data allow me to employ a stock report fixed effect to examine, within-a-report, how differences in title characteristics lead to differences in attention to news. The setting holds fixed the topic firm, author, date, and, uniquely, the event discussed. The randomized assignment of titles to investors tracking the company ensures omitted characteristics of the recipient investors and the news event are not driving the relation between textual attributes and attention.

I find that investors are significantly more attracted to short and simple titles. An increase in title length of one standard deviation leads to 12% fewer page views.[5] To help appreciate the magnitude, consider that company stock reports released on days when the VIX is a standard deviation higher receive 8% more page views. I find a similar negative relation between attention and both average word length and number of words, suggesting

3. *Seeking Alpha* is an online community of 4 million investors who read and write stock reports on specific companies. See Section 1.1 for more details.

4. All of the results hold for other time windows.

5. A title length's mean and standard deviation in character units: $\mu = 49$ and $\sigma = 17$.

that investors prefer simple titles. The magnitudes are surprisingly large given that the subjects are investors in a high-stakes setting and are interested in the news, having signed up to receive alert emails on the company. Also, the subjects are highly-engaged investors, who respond within 30 minutes of an email alert. One might suspect that the magnitudes are even larger for less-engaged, less-interested investors. The relation is similar in magnitude even when comparing pairs of titles with very similar word usage and character length.

The attraction to short and simple titles is stronger when the investors tracking a company are less-sophisticated. I do not have direct data describing the individuals receiving email alerts. However, the public comment section of stock reports reveals individual identities, allowing me to derive measures that characterize the sophistication of investors tracking the firm. The average sophistication of investors who write stock reports is likely to be higher than those who only read reports. Also, more sophisticated investors are likely to be more numerical and write longer comments. Consistent with this reasoning, I find that the sensitivity of attention to title length increases with the fraction of comments from investors who have not written a stock report and decreases with the numerical intensity of comments and with the average length of comments.

Do investors who access reports with longer titles behave more sophisticated? The data includes the number of investors who scroll to the end of the report by title received. The underlying stock report is the same so that differences in the propensity to read to the end must be due to a selection effect. I find that investors who do access stock reports with longer titles are significantly more likely to read the full report. This finding is consistent with the notion that investors with greater cognitive abilities are less sensitive to complexity and also more thorough.

Although the title-testing data with randomization holds the context fixed, I do not write the sampled titles. The analysts craft the plausible titles. Therefore, longer and more positive titles might be more informative, reducing the need to read the stock report.

To resolve this concern, I use the length of a company's legal name to instrument for title length. *Seeking Alpha* titles nearly always include a company's name, and companies with longer names in an industry tend to also have longer titles. The length of a company's legal name likely satisfies the exclusion restriction as legal names are chosen in the past and do not provide investors with new information. Companies with longer names do not include less content in titles, and company-name length is unrelated to firm characteristics after controlling for firm size.

Using the instrument, I find a negative relation between title length and page views. I also find a positive relation between title length and the read-to-end rate. The magnitudes are nearly identical to those determined using the title-testing data alone. Thus, the instrument confirms the direction and magnitude of the effect of non-informative title length on attention. The result suggests investors are complexity averse and not simply reading longer titles less because longer titles are more informative.

The evidence of complexity aversion for investors on *Seeking Alpha* motivates examining whether, in aggregate, investors underreact to company-issued news releases with longer titles. One could argue that while the behavior demonstrated by *Seeking Alpha* investors likely affects individual performance, notably of less-sophisticated investors, the behavior may not matter for market behavior. The market may have relatively more-sophisticated investors present, and the attention of algorithmic traders is likely unaffected by headline length.

I assemble a database of approximately 265,000 company-issued-earnings releases distributed via *PR Newswire* and *Business Wire* during 1992 to 2015. Note that I am no longer using *Seeking Alpha* data but rather earnings-press-release data. Longer titles tend to be more informative, and thus, title length is *positively* related to both turnover and volatility. To isolate variation in title length that is unrelated to the event reported, I instrument title length with the length of the company's legal name. I find a significant *negative* effect of

title length on turnover. A standard deviation increase in a title's length predicts 5% less announcement turnover. Title length also negatively affects announcement volatility, captured by the day's intraday-price range. A standard deviation increase in a title's length predicts a 50-basis-points-tighter-intraday-price range. These relations hold almost exclusively on the announcement day, consistent with title length reducing initial attention to news.

I find evidence of return underreactions for earnings news with longer titles. A standard deviation increase in a title's length predicts a 40-basis-points return underreaction for negative news and a 30-basis-points underreaction to positive news. In other words, I find that the stock price does not rise as much for positive news with non-informatively-longer titles and does not fall as much for negative news with longer titles. I determine positive news by comparing actual earnings to analyst expectations. The underreactions reverse within the next 20 trading days. The reversal of the underreaction is consistent with the instrument – company-name length – capturing non-informative variation in title length. The effects are robust to time splits, firm-size splits, a large set of firm controls, and controlling for the earnings surprise. These findings are consistent with investors paying less attention to news with longer titles and supports the external relevance of the results of the *Seeking Alpha*-field experiment.

The market effects of complexity vary with market volatility and news intensity. On the one hand, on volatile-market days and busy-news days, investors are more time constrained and thus inclined to skip complex news. On the other hand, I find that news demand by *Seeking Alpha* investors increases meaningfully on high VIX days, suggesting that investors perceive higher benefits to engaging with news on such days. The anticipated benefits of reading complex news may be relatively higher on volatile news days and investors may be overall more willing to engage with news regardless of title length. Consistent with the latter reasoning, I find that the sensitivity to title length is greater on low-volatility and slow-news days.

Finally, firms do not appear to strategically choose titles. The earnings surprise is positively related to title length. Firms seem to add positive words to titles and almost never add negative words to titles. Keeping negative words out of titles may be effective as I find that *Seeking Alpha* investors are attracted to negative words. A standard deviation increase in a title's negativity predicts a 2% increase in page views. As sentiment is difficult to measure, especially for short titles, measurement error likely attenuates the magnitude. However, adding positive words leads to less attention and also increases the length of titles, which further reduces the attention the news receives. Firms reporting positive news should reduce non-informative title length. I find no relation between the earnings surprise and usage of "earnings per share" versus "eps", "first quarter" versus "q1", and abbreviated forms of the firm's name. Thus, firms are not managing non-informative title length strategically. For positive news, shortening non-informative length might emphasize the positive content at the end of titles.

In conclusion, this paper offers the first causal evidence that textual complexity matters for investor attention and market reactions to news using two settings – *Seeking Alpha* field data and earnings releases.[6] The evidence that complexity aversion is greater for less-sophisticated investors contributes to the literature relating cognitive abilities and financial decisions.[7] More generally, the documented effect of complexity on attention allocation contributes to efforts to model investor attention allocation.[8] The market effects of title complexity contribute to the empirical evidence that investor attention seems to matter

6. See Engelberg (2008); Li (2008); You and Zhang (2009); Miller (2010); Lehavy et al. (2011); Loughran and McDonald (2011, 2014).

7. See Feng and Seasholes (2005); Grinblatt et al. (2008); Agarwal et al. (2009); Grinblatt et al. (2009, 2011).

8. See Gabaix and Laibson (2005); Gabaix et al. (2006); Peng (2005); Peng and Xiong (2006); Kacperczyk et al. (2009); Bordalo et al. (2013)

for firm outcomes,[9] asset pricing,[10] individual investor trading,[11] announcement effects,[12] and individual investor portfolios.[13] The result that investors pay more attention to more negative headlines contributes to the literature on sentiment and finance.[14] The evidence that firms strategically manage a title's sentiment but not length contributes to the literature on strategic disclosures.[15]

These results motivate policies simplifying reports from businesses, government agencies, and academic institutions (e.g., the SEC's Plain Writing Initiative). I find that even sophisticated academics may be sensitive to title length. Using *Social Science Research Network* data, I find that an author's papers with twice the title length receive 11% fewer views, 13% fewer downloads, and 4% fewer citations.

1.1 *Seeking Alpha* and Title-Testing

Founded in 2004, *Seeking Alpha* has become the leading-crowdsourced-investment-research firm.[16] The platform is highly active, with 4 million registered investors and 85 million page views per month.[17] The community includes over 10,000 contributing analysts writing stock reports and 280,000 commenters. Stock reports cover a broad range of firms, including more

9. See Chemmanur and Yan (2009); Fang and Peress (2009); Ahern and Sosyura (2015).

10. See Corwin and Coughenour (2008); Hou et al. (2009); Bank et al. (2011); Da et al. (2011); Li and Yu (2012); Hillert et al. (2014); Yuan (2015).

11. See Barber et al. (2009); Huddart et al. (2009); Barber and Odean (2012); Loewenstein et al. (2016); Engelberg and Parsons (2016).

12. See Loh (2008); Hirshleifer et al. (2009, 2011); Tetlock (2011); Chakrabarty and Moulton (2012); Drake et al. (2012); Hartzmark and Shue (2015).

13. See French and Poterba (1991); Coval and Moskowitz (1999); Nieuwerburgh and Veldkamp (2009).

14. See Barberis et al. (1998); Antweiler and Frank (2004); Baker and Wurgler (2006); Tetlock (2007); Tetlock et al. (2008); Dougal et al. (2012); Chen et al. (2014); Da et al. (2015).

15. See Patell and Wolfson (1982); Damodaran (1989); Hirshleifer and Teoh (2003); Hirshleifer et al. (2004); Bagnoli et al. (2006); DellaVigna and Pollet (2009); Michaely et al. (2016).

16. David Jackson and venture capital firms Benchmark, Accel, and DAG Ventures own *Seeking Alpha*.

17. See `http://seekingalpha.com/page/who_reads_sa`.

than 4,000 small- and mid-cap stocks in the past year across a variety of sectors.

The audience includes money managers, sell-side analysts, investment bankers, financial advisors, business leaders, entrepreneurs, and retail investors. Over 20% of the audience are financial professionals. Readers tend to be highly active investors. Over 50% of readers purchased stocks in the trailing 30 days. Almost 90% of unique visitors own securities. The readers also tend to be wealthy, with the highest percentage of readers, among any major finance website, managing portfolios with assets greater than $50,000, $100,000, $250,000, $500,000, and $1,000,000.

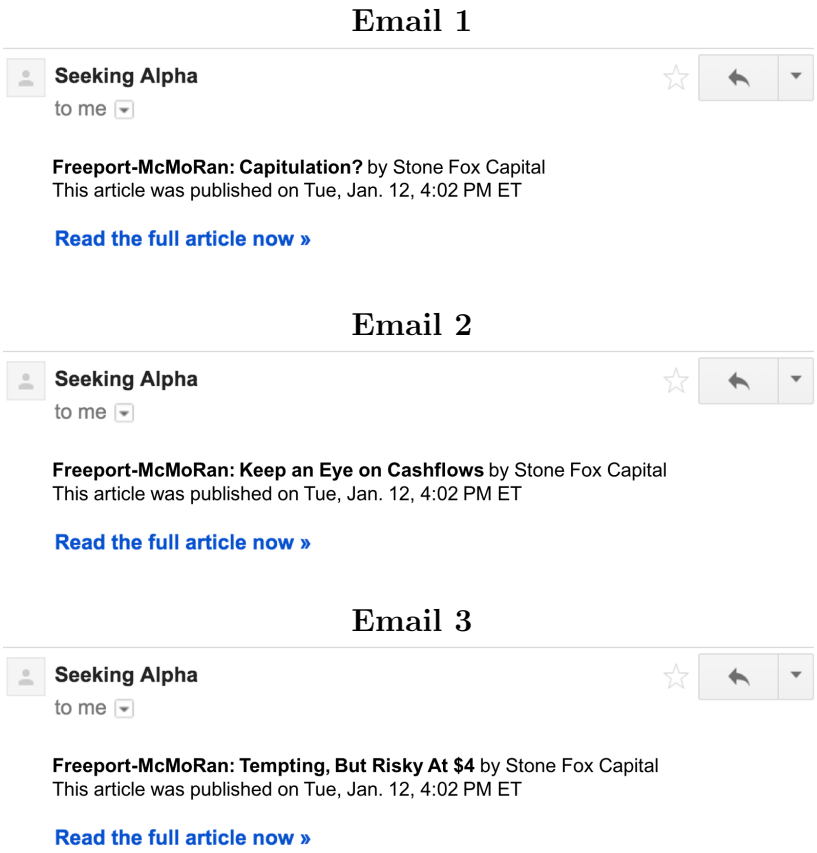
Seeking Alpha hires editors to make sure stock reports are well written and not repetitive of prior published reports. The community generates 600 submissions a day and after editorial review, approximately 200 reports are published daily. The editors provide guaranteed monetary rewards for high-quality content: $35 basic, $150 must-read, and $500 top-idea. Analysts also earn a performance rate of $0.01 per page view.[18] Analysts build a reputation and get feedback via comments to stock reports. The content produced is valuable. Chen et al. (2014) finds the sentiment of the *Seeking Alpha* stock reports and comments predict future returns.

Starting January 3, 2016, *Seeking Alpha* began "title testing." An analyst may propose two titles for a stock report, and the editor assigned to review the stock report may propose a third title. The three titles are sampled with random assignment on investors signed up for alerts on the topic company. Figure 1.1 illustrates three example real-time alert emails sent during the title-testing program for a single report. Each email alert contains the stock report's title in bold, the author's name, a time stamp, and a link to the full stock report. Notice that all elements other than the title are the same across alert emails and that none of the body of the stock report is included in the alert email. Also, there is only one title (news

18. The basic compensation is $35 plus $0.01 per page view. The must-read and top-idea rates guarantee a minimum of $150 and $500 respectively. Analysts can earn more than the minimums if the compensation from page views ($0.01/view) exceeds the minimum.

article) per email. In this example, the proposed titles randomly assigned to investors were "Freeport-McMoRan: Capitulation?," "Freeport-McMoRan: Keep an Eye on Cashflows," and "Freeport-McMoRan: Tempting, But Risky At $4." These alerts were sent to 8,373, 8,274 and 8,289 investors respectively. Within 30 minutes, the titles received 220, 154, and 133 clicks, respectively. *Seeking Alpha*'s title-testing algorithm chooses the title with the most page views after an interval of time as the winning title.

Figure 1.1: Example of email alerts sent during *Seeking Alpha*'s title-testing program. Each of the three alert emails contains one of the three proposed titles for a specific stock report written by Stone Fox Capital on Freeport-McMoRan. The three alert emails shown below were sent to 8,373, 8,274 and 8,289 investors, respectively, with random assignment of title to investor. The recipient investors previously signed up for alerts on Freeport-McMoRan. The titles received 220, 154, and 133 views within 30 minutes, respectively. The chief measure of attention is the number of investors who click the "Read the full article now" link to view the body of the stock report within 30 minutes of the alert email being sent.



Seeking Alpha runs title testing for over 100 stock reports per day. *Seeking Alpha* does not run title testing on stock reports that editors chose as "must-read" or "top-idea," because these are embargoed for paying subscribers for 24 hours.[19] The ultimate audience of the stock report is global. *Seeking Alpha* emails the report in a "Daily Investing Ideas" email to over 500,000 subscribers and publishes the report on major news feeds.

Whereas press releases are primarily distributed before the market opens or immediately following the market's close, the distribution of *Seeking Alpha* stock-report alerts peaks during market hours between 10AM and 1PM EST. The mean click rate in 30 minutes on email alerts is 1.5%. The click rate is meaningfully greater for smaller and less popular companies.

As of December 2016, the *Wall Street Journal* and the *Financial Times* have not implemented title testing, making the *Seeking Alpha* data unique for studying how investors respond to the textual attributes of financial information.

1.2 Data Summary Statistics

1.2.1 Seeking Alpha *Title-Testing Data*

The title-testing program began January 3, 2016 and is ongoing. My sample period ends December 3, 2016, and includes 18,572 unique reports with 41,525 titles covering 3,573 unique securities.

I match each stock report with CRSP data. I discard observations that do not match with CRSP, including all stock reports covering OTC stocks. For this paper, I exclude stock reports about macroeconomic events, ETFs, mutual funds, REITS, currency, retirement, and commodities.[20] I limit the sample to stock reports released on non-holiday weekdays.

19. The population of paying subscribers is too small at the moment to gather enough data to determine a title's relative attractiveness. *Seeking Alpha*'s revenues are primarily advertising based.

20. An attraction to shorter, simpler, and more negative headlines also holds for investors following these

Table 1.1: Summary Statistics

Title-testing data from *Seeking Alpha*, January 4, 2016, to September 5, 2016

Variable	Mean	SD	p25	p50	p75	N
By Company						
Market Capitalization (million)	11333	34497	352	1653	7525	1856
Log Market Capitalization	21.2	2.1	19.7	21.2	22.7	1856
Real Time Alert Subscribers	4429	12254	620	1441	3797	1856
Length Company CRSP Name	21.3	7.8	16	21	27	1856
Log Length CRSP Name	3.0	0.4	2.8	3.0	3.3	1856
By Title						
Title Length (characters)	48.7	17.3	36	46	59	22623
Log Title Length	3.8	0.4	3.6	3.8	4.1	22623
Fraction Title Negative (%)	4.4	9.7	0	0	0	22623
Fraction Title Positive (%)	6.9	11.5	0	0	13	22623
Fraction Title Net (%)	2.6	15.8	0	0	11	22623
% Yield (Views/Email in 30 minutes)	1.5	1.6	0.5	1.0	1.9	22623
By Day						
VIX (%)	16.1	4.0	13.4	14.7	18.2	232

Press-release data from *PR Newswire* and *Business Wire*, 1992-2015

Variable	Mean	SD	p25	p50	p75	N
By Company						
Market Capitalization (million)	3372	14060	107	379	1555	12325
Log Market Capitalization	19.9	1.9	19	20	21	12325
Length Company CRSP Name	19.3	6.5	14	19	24	12325
Log Length CRSP Name	3.0	0.4	2.8	3.0	3.2	12325
By Press Release						
Title Length (characters)	75.1	34.1	52	65	89	264125
Log Title Length	4.2	0.4	4.0	4.2	4.5	264125
Log Turnover	-4.8	1.5	-5.8	-4.8	-3.8	264125
Log Intraday Price Range x 100	7.9	6.8	3.2	5.9	10.2	264125
Log Return(t,t+1) x 100	-0.3	7.6	-3.1	0.0	3.1	264125

This approach reduces the sample to 9,944 unique stock reports with 22,623 titles for 1,856 companies.

For each stock report, the data include the titles associated with the stock report, the number of emails sent by title, the timestamp of the emails (all emails sent at same time), and an identifier for whether the title was from the analyst (original or alternative) or from the editor. The page-view data by title captures the number of clicks on a specific title in the 30-minutes following an email alert. For January and February 2016, I have data on page views for all 30-minute intervals in the 24 hours following the time an email alert is sent. For this limited period, I also have data on the number of email alert subscribers who read the full stock report, measured as the number who scrolled to the end of the report within

other assets.

24 hours.

Each of the two-or-three titles is not equally likely to have the most page views ex-ante. Appendix 1.5.9 Table 1.18 shows the analyst's alternative title is less likely to have the most page views and is slightly longer when compared to the analyst's original title. The editor's title is also less likely to receive the most page views and is generally shorter and more positive in tone than the analyst's original title. I include dummy variables reflecting whether a title is the author's original, author's alternative, or editor's title in all regressions using *Seeking Alpha* data. The results in this paper hold if I discard the editor's title.

I supplement the data by gathering the body of the stock report and all of the comments. The comment data help gauge the sophistication of investors paying attention to the stock. I also collect data on the author (analyst), including the number of years as an author, the number of followers, and the number of published stock reports.

1.2.2 Earnings-Press-Release Data

Using *Seeking Alpha*-title-testing data, I show investors exhibit strong attractions to short, simple, and negative titles. I cannot use the *Seeking Alpha*-title-testing data to identify the effect of title length on market reactions to news. Instead, I transition to a new empirical setting and examine the effect of title length on market reactions to earnings-press releases. I collect press-release titles from *PR Newswire* for the years 1992 to 2015 and from *Business Wire* for the years 1995 to 2012. These two newswires are the primary means of distributing news for public companies. There are approximately 6.2 million releases.

I match press-release titles to CRSP data based on the company's ticker symbol used in the press release or the name in the title. The matched set includes 1.6 million releases. The non-matched press releases tend to come from trade organizations, rating agencies, and non-corporate sources. The remaining titles are for press-releases released directly from public companies.

I further limit the sample to earnings announcements. Earnings announcements occur regularly and provide important information to investors. Announcements are prescheduled, reducing the possibility management time news releases with market or investor behavior. Another advantage of studying responses to earnings announcements is one can measure the sentiment of unexpected news by comparing actual reported earnings with analyst estimates. Also, stock return anomalies around earnings announcements tend to be stronger.[21] I extract earnings news releases by keeping titles with keywords such as "earnings," "first quarter," and "q1." I also find earnings releases by matching the press release dates with earnings-announcement dates in the I/B/E/S detail history file (plus or minus a day). The I/B/E/S data include the latest estimates of earnings by each analyst and the actual reported earnings. I keep firms without analyst coverage. I match the earnings releases with Compustat firm characteristics for the previously announced year.

The final sample contains 264,125 earnings announcements from 12,325 firms for the period from 1992 to 2015. The 25th percentile firm has 8 releases, and the 75th percentile firm has 35 releases. Approximately 50% of observations have analyst earnings estimate data available from I/B/E/S.

1.3 Empirical Results

1.3.1 Complexity and Attention

I measure the complexity of titles using common measures to grade the complexity of scholastic texts: length, the number of words, word length, and word frequency. Title length is measured in character units. Log title length equals the log number of words plus the log average word length. Longer words tend to be more difficult to understand. More words require more synthesis to interpret. The word-frequency measure captures the frequency of

21. Engelberg et al. (2015) finds anomaly returns are seven times higher on earnings announcement days and two times higher on corporate news days. The effects are similar on both the long and short sides.

the least common word in the title relative to word usage in *Seeking Alpha* headlines from 2006 to 2015.[22] I exclude a company's name from headlines when determining the frequency of the least common word.

Table 1.2 shows the following empirical specification: regression of page views by title on measures of complexity by title. Page views are measured over the 30 minutes following an alert email, capturing variation in the attention of highly engaged investors who are also interested enough to sign up for email alerts:

$$\text{Log Views/Emails}_{i,j} \sim \beta \text{ Log Title Complexity}_{i,j} + \alpha_j + \epsilon_{i,j}$$

Because the three titles i for stock report j are randomly assigned, β captures the average effect of textual attributes on attention. The α_j is the stock report fixed effect and holds fixed the firm, analyst, event, and date. Random assignment of titles to investors makes the distribution of omitted characteristics of the audience similar across titles in expectation. I cluster standard errors by stock report, and the significance of the results is highly robust to other forms of clustering.

To illustrate variation in title length within a stock report holding the context fixed, consider the following two titles.

- Bank of America is an attractive long term investment
- Bank of America is an attractive investment if your horizon is longer than a year

The first title received 195 page views in 30 minutes. The second title received 159 page views in 30 minutes.

Table 1.2 regression (1) shows a significant positive relation between title length and attention. However, regression (2), which includes the stock report fixed effect, shows a

22. The results are similar if word frequency is calculated using the frequency of words in 10-K filings from 1994 to 2014.

highly significant, negative relation between title length and page views. The sign flip shows that the unconditional-title-length-attention relation is biased by omitted variables related to the event, author, date, or firm. For example, larger firms tend to have lower response rates and shorter titles, which would lead to a positive OLS correlation between readership yields and title length.

Using the stock-report fixed effect, a standard deviation increase in title length predicts 12%-fewer page views. This magnitude compares to the relation between the VIX and activity on *Seeking Alpha*. A company's stock report released on a day with a standard-deviation-higher-than-normal VIX tends to receive 8% more views. Also, this magnitude characterizes the behavior of investors who have indicated an interest in the news by signing up for alert emails. Less-interested investors may have a higher sensitivity to title length. The within-R^2 is 6%, suggesting title length is an important explanatory variable for attention.[23]

Table 1.2 regression (3) shows a negative relation between attention and both the average length of words in the title and the number of words.[24] This result suggests that investors prefer simpler titles with fewer words and shorter words. Regression (4) shows a negative relation between the frequency of the least common word and attention. The negative relation suggests less-common words attract more attention. After controlling for the least common word in the titles, the coefficients on number of words and word length become more negative. These results suggest unusual information attracts attention, but measures of length repel attention.

The negative relation is not driven only by titles that are very different in length or content. Figure 1.2 shows a clear negative relation in levels between within-stock-report page views and within-stock-report title length. Appendix 1.5.1 confirms that the negative

23. I conduct 10,000 bootstrap out-of-sample tests to evaluate the out-of-sample explanatory power of this relation. The model is estimated using 50% of the stock reports, regressing within-stock-report page views on within-stock-report variation in title length. Testing the model shows 95% of R^2 estimates are between 5% to 7%.

24. The decomposition is log title length = log average word length + log number of words.

relation can be identified even when the sample is restricted to titles that are at most 3 characters different in length. Appendix 1.5.2 shows that the magnitude of the negative relation between title length and attention is nearly identical when the sample is restricted to titles with low and high word overlap, suggesting that differences in information content are not driving the results.

Table 1.2: Regressions of investor attention on title complexity, using *Seeking Alpha*-title-testing-field data. Title length is measured in characters. The frequency of the least common word in the title (excludes firm name) is relative to word usage in *Seeking Alpha* titles from 2006 to 2015. Regressions (2) to (4) include an article fixed effect, which holds the event, firm, author, and date fixed. Standard errors are clustered by article.

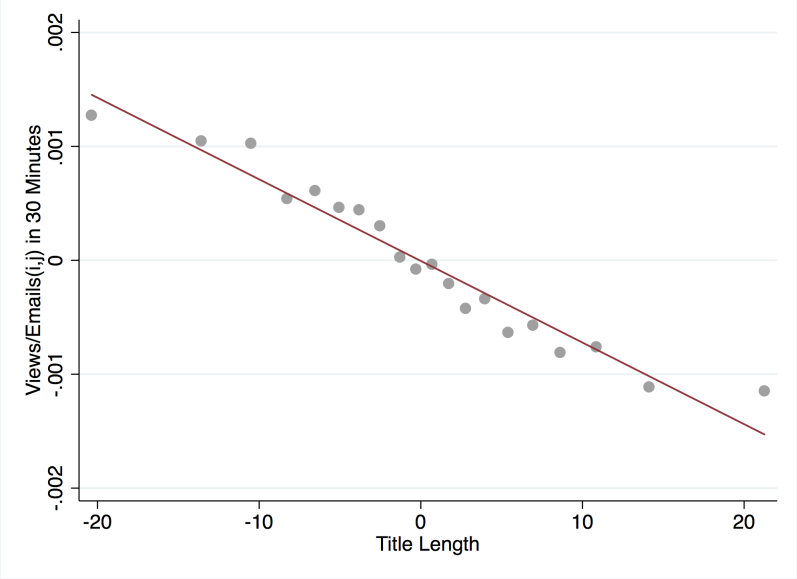
	Log Page Views/Emails			
	(1)	(2)	(3)	(4)
Log Title Length	0.11***	-0.30***		
	(0.02)	(0.01)		
Log Average Word Length			-0.32***	-0.35***
			(0.02)	(0.02)
Log Number of Words			-0.30***	-0.32***
			(0.01)	(0.01)
Log Frequency of Least Common Word in Title				-0.01***
				(0.00)
Article FE	No	Yes	Yes	Yes
Adjusted R^2	0.01	0.91	0.91	0.91
Within R^2	.	0.06	0.06	0.07
Num. Articles	9944	9944	9944	9943
Observations	22623	22623	22623	22607

1.3.2 Cognitive Abilities and Attention

I now examine whether the sensitivity of investor attention to title length is stronger for less-sophisticated investors. I would prefer to use a direct measure of the sophistication of investors signed up for alert emails. However, I do not have these data. Instead, I take advantage of individual data revealed in the actively-used comment sections of stock reports to characterize the sophistication of the topic company's followers.[25] Comments from investors who have contributed a stock report on *Seeking Alpha* previously may be more sophisticated on average than investors who have never contributed a stock report.

25. 60% of comments are made on the day of the report's publication. 20% of comments are made on the following day.

Figure 1.2: Bin scatterplot illustrating the relation between title length in characters and readership (views/emails) in first 30 minutes following alert emails. Title length and yield are demeaned at the stock report level, so that "0" denotes the average yield and average title length for a single stock report. The bin scatterplot divides the sample by title length into equal-sized groups. The mean yield is then determined for each group and plotted as a point.



Also, more-sophisticated investors may be more numerical and write longer comments. More numerical comments have more digits relative to total characters. Table 1.3 regression (3) shows that when more comments come from investors who have never contributed a stock report, the negative sensitivity to title length is stronger. Regressions (4) and (5) show that the sensitivity to title length appears to be weaker when comments are more numerical and longer. I also look at variation in the sensitivity with firm size and popularity since less-sophisticated investors are more likely to be aware of larger and more popular firms ($\rho = 0.7$ between size and popularity) (Barber and Odean, 2012). Regressions (1) and (2) show that when the company is larger and more popular, the audience seems to be more sensitive to title length. These results suggest complexity aversion is lower for more-sophisticated investors.

Another way to gauge whether sophisticated investors have less complexity aversion is to

Table 1.3: Regressions showing how the sensitivity of investor attention to title complexity varies in the cross-section of investor sophistication, using *Seeking Alpha*-title-testing data. Title length is measured in characters. Title length is interacted with firm size, number of followers (number of *Seeking Alpha* investors signed up to receive email alerts for the topic company), the fraction of article comments from non-analysts (*Seeking Alpha* investors who have never written a stock report), the average length of comments, and the fraction of characters in comments that are digits ("numerical comments"). Market capitalization is from CRSP and is calculated as of $t-1$, where t is the email-alert date. Each regression includes an article fixed effect, which holds the event, firm, author, and date fixed. Standard errors are clustered by stock report.

	Log Page Views/Emails					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Title Length	-0.27***	-0.25***	-0.30***	-0.31***	-0.32***	-0.27***
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Log Title Length x Log Market Cap	-0.02***					-0.01
	(0.00)					(0.01)
Log Title Length x Log Followers		-0.03***				-0.01
		(0.01)				(0.01)
Log Title Length x Fraction Comments Non-Analyst			-0.18***			-0.12**
			(0.05)			(0.05)
Log Title Length x Avg. Comment Length				0.04*		0.04**
				(0.02)		(0.02)
Log Title Length x Numerical Comments					0.11**	0.09*
					(0.05)	(0.05)
Article FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.91	0.91	0.91	0.91	0.91	0.91
Within R^2	0.07	0.07	0.07	0.07	0.07	0.07
Num. Articles	9535	9535	9535	9535	9535	9535
Observations	21725	21725	21725	21725	21725	21725

examine the reading intensity of those who click on a relatively-longer title. The underlying report is the same, so differences in read-to-end rates (number who scroll to bottom of the article) predicted by title length suggest a selection effect - a difference in composition of readers. Table 1.4 regression (2) shows the relation between title length and the number of investors who read the full stock report in the 24 hours following the alert email, controlling for the number of investors who click on the stock report. The data is only available for January and February. The coefficient on title length is positive and highly significant. A standard deviation increase in title length predicts a 4% higher read-to-end rate. This finding is consistent with more-sophisticated investors being less sensitive to title length and generally more thorough in information acquisition.

Table 1.4: Regressions showing the relation between the number of investors who read the full stock report (read-to-end rate) and title length, using *Seeking Alpha*-title-testing data. Regression (2) includes an article fixed effect, implying that the underlying stock report is the same, only the titles investors receive differ. Due to data availability, the read-to-end rate sample period is currently January 4, 2016, to February 29, 2016. Standard errors are clustered by stock report.

	Log Read-to-End Rate	
	(1)	(2)
Log Title Length	0.13***	0.11***
	(0.01)	(0.01)
Article FE	No	Yes
Adjusted R^2	0.04	0.72
Within R^2	0.04	0.05
Number of Articles	1686	1686
Observations	4007	4007

1.3.3 Instrumenting Title Length

At this point, I still cannot formally conclude title complexity affects attention. More complex titles might provide more information, reducing the need to click to read the stock report. More informative titles may also lead to better matches between an investor's interests and the content of the report. To shut down the information story, I instrument title length with the length of a company's legal name (not the length of the name in the title).

The relevance condition is satisfied as *Seeking Alpha* headlines generally include the topic company's name and longer names predict longer titles. A monotonicity condition must hold – all companies with longer names have on average longer titles because of their longer names. A very small group of defiers have longer names but use a shorter version (e.g., "International Business Machines" uses "IBM"). Excluding these defiers does not alter the relations.

The exclusion restriction is almost surely satisfied. A company's legal name is chosen in the past, and thus the choice is unrelated to the new information discussed in a stock report. Although a wide variety of firm names is possible, the variety of plausible lengths is smaller. Company-name length may be correlated with the company's industry; for example, "ABC pharmaceuticals" is longer than "XYZ energy." I include SIC4-industry-by-year fixed effects in regressions to control for differences in name length across industries and industry trends.

I examine in Table 1.5 the correlations between company-name length and firm characteristics. I use Compustat's 2015 annual file. I match the Compustat data with data from CRSP and ownership data from Thomson. Regression (1) shows a significant negative relation between a firm's market capitalization and company-name length. Therefore, I orthogonalize company-name length to firm size by regressing name length on a third-degree polynomial of market equity with SIC-4 industry fixed effects and then taking the residuals as "Adjusted Name Length." Regressions (2)-(10) show that adjusted-company-name length is unrelated to other firm characteristics including total assets, book leverage (net debt/total invested capital), revenues, profitability (gross margin and profit margin), age since IPO, the firm's market beta, and ownership. One can see the within-R^2 goes to 0.00 in these regressions. The lack of a significant relation between company-name length and a variety of firm characteristics suggests company-name length provides variation in title length that is unrelated to characteristics of the firm, conditional on a company's size and industry.[26]

Table 1.6 regression (1) shows the first-stage regression of title length on company-name length. The coefficient on company-name length is positive and has a F-statistic of 192, exceeding the threshold of 10 recommended by Stock and Yogo (2005). The first stage result suggests company-name length is highly relevant.

Table 1.6 regression (3) shows the IV relation between title length and attention within an SIC-4 group and date. One cannot use the article fixed effect because I am exploiting variation in company-name length. The coefficient of -0.36 is not meaningfully different from the OLS coefficient of -0.28 in regression (2). The IV magnitude might be more negative if the instrument more cleanly identifies non-informative variation in length. Table 1.6

26. I show, within a firm, that non-informatively longer titles lead to less attention using an experimental Google Survey. I asked 1,000 adults to choose which of two pieces of news on Apple is more interesting. In one survey, the titles are "Apple Inc. sees iPhone sales slump" and "Apple Incorporated launches new smart home." In the second survey, I switch the "Inc." and "Incorporated." The switch changes the relative optical length of the titles, holding the information content and the topic firm fixed. Participants preferred the optically shorter title significantly more in both surveys, consistent with non-informative variation in title length affecting attention.

Table 1.5: Using the Compustat 2015 annual file, this table shows that log company-name length after adjusting for firm size is unrelated to firm characteristics. "Adj Name Length" is determined by taking the residuals from a regression of log company name length on a cubic polynomial of log-firm market capitalization and SIC-4 industry fixed effects. Market beta is estimated using past 5 years of monthly returns. Ownership data is from Thomson. Standard errors are clustered by firm.

	Name Length	Adj Name Length								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log Market Cap	-0.02***									
	(0.00)									
Log Total Assets		-0.00								
		(0.00)								
Book Leverage			0.03							
			(0.03)							
Log Revenue				-0.00						
				(0.00)						
Gross Margin					0.02					
					(0.03)					
EBIT						0.06				
						(0.05)				
Net Income Margin							0.01			
							(0.04)			
Market Beta								0.01		
								(0.01)		
Age since IPO									-0.01	
									(0.01)	
Institutional Ownership										0.00
										(0.00)
SIC-4 FE	SIC-4	SIC-4	SIC-4	SIC-4	SIC-4	SIC-4	SIC-4	SIC-4	SIC-4	SIC-4
Adjusted R^2	0.18	-0.11	-0.11	-0.11	-0.12	-0.14	-0.13	-0.11	-0.08	-0.10
Within R^2	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Observations	4260	4027	4027	3901	3375	2653	3244	3416	2425	3819

regressions (4) and (5) examine the relation between title length and the read-to-end rate measure. The IV coefficient of 0.17 from regression (5) is similar to the 0.09 OLS coefficient from regression (4). This result shows that the observed selection effect is driven by non-informative variation in title length, suggesting that investors who are less sensitive to title length are more thorough and sophisticated. The instrumental-variable results rule out the competing story that the negative-title-length-attention relations are because longer titles are more informative.

Seeking Alpha authors do not seem to be aware of the negative effect title length has on attention. Figure 1.3 shows, from 2009 to 2015, *Seeking Alpha* titles have been steadily increasing in length.

Table 1.6: Regressions showing the relation between instrumented-title length and investor attention, using *Seeking Alpha*-title-testing data. The instrumental variable is the character length of the firm's name. Regression (1) is the first-stage of title length regressed on company-name length. Regression (2) and (3) show that title length affects the number of investors who click to view an email. Regression (3) is the instrumental variable regression. Regressions (4) and (5) show that title length affects the read-to-end rate, or the number of investors who read the full report conditional on viewing the report. Regression (5) is the instrumental regression, showing that investors who click on longer titles are also more thorough. Other controls include a firm's market capitalization and the number of emails sent by title. Standard errors are clustered by stock report.

	(1) Log Title/ Length OLS	(2) Log Views/ Emails OLS	(3) Log Views/ Emails IV	(4) Log Read-to-End Rate OLS	(5) Log Read-to-End Rate IV
Log Company Name Length	0.10***				
	(0.01)				
Log Title Length		-0.28***	-0.36***	0.09***	0.17***
		(0.02)	(0.11)	(0.01)	(0.06)
SIC-4 FE	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes
Firm Controls	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.13	0.58	0.58	0.30	0.29
Within R^2	0.02	0.32	0.32	0.12	0.10
Number of Articles	9943	9944	9943	1686	1686
Observations	22621	22623	22621	4007	4007

1.3.4 Sentiment and Attention

This section examines, using the *Seeking Alpha*-field data, whether the sentiment of headlines matters for investor attention. Understanding the sentiment-attention relation helps test later whether firms manage title length and sentiment.

Using the *Seeking Alpha*-title-testing data, I find a positive relation between the negativity of titles and attention. To measure the sentiment of titles, I count the number of positive and negative words in a title using the lexicons (lists) of words from either the Harvard psychosocial lexicon[27] or Bill McDonald lexicon.[28] The net sentiment of a title is the number of positive words less the number of negative words divided by the number of words in the title. I exclude the firm's name from these calculations.

27. Source: `http://www.wjh.harvard.edu/~inquirer/`.

28. Source: `http://www3.nd.edu/~mcdonald/Word_Lists.html`.

Figure 1.3: Trends in the length of titles.

Figure A: 10-week average title length (characters) of *Seeking Alpha*-stock reports from 2006 to 2015.

Figure B: Median title length (characters) of earnings announcements released via *PR Newswire* and *Business Wire* from 1992 to 2015.

Table 1.7 regression (3) shows more positive titles receive less page views controlling for the length of titles and using the article fixed effect. A one-standard-deviation increase in a title's net sentiment predicts a 2% increase in page views. In regression (4), I decompose net sentiment into the fraction of positive and negative words. Positive words are negatively related to attention, while negative words are positively related to attention. A standard deviation change in positive or negative sentiment predicts a 2% decrease and a 1% increase in attention respectively. The magnitudes are likely attenuated as measuring title sentiment

Table 1.7: Regressions showing relation between title sentiment and investor attention, using *Seeking Alpha*-title-testing data. Title sentiment is measured using the number of words in the title (excluding firm name) that appear in the Bill and McDonald lexicons of positive and negative words. The number of positive and negative words is then scaled by the number of words in the title (excluding firm name). Regressions (2) to (4) include a stock report fixed effect. Standard errors are clustered by stock report.

	Log Views/Emails			
	(1)	(2)	(3)	(4)
Net Sentiment	0.03	-0.15***	-0.12***	
	(0.05)	(0.02)	(0.02)	
Positive Sentiment				-0.14***
				(0.03)
Negative Sentiment				0.09***
				(0.03)
Log Title Length			-0.30***	-0.31***
			(0.01)	(0.01)
Article FE	No	Yes	Yes	Yes
Adjusted R^2	0.01	0.91	0.91	0.91
Within R^2	0.00	0.00	0.07	0.07
Num. Articles	9535	9535	9535	9535
Observations	21725	21725	21725	21725

is difficult as titles are short.

I receive similar results using a machine learning classifier to score the sentiment of headlines. For more details see Appendix 1.5.12.

1.3.5 Title Length Affects Market Behavior

Although I have provided evidence that longer-instrumented titles reduce investor attention using *Seeking Alpha* data, I have not shown this behavior matters for market outcomes. One could argue that while headline attributes matter for the attention of *Seeking Alpha* investors, the results may not hold externally. This section shows that title length does matter for market behavior.

I assemble a database of 265-thousand-earnings-press releases released via *PR Newswire* and *Business Wire* from 1992 to 2015 covering 12.4 thousand firms. The advantage of studying earnings announcements is that earnings events are pre-scheduled, regularly occurring, and significant events. Also, stock return anomalies have been found to be significantly larger

on earnings-announcement dates.[29] Although the press release contains the actual date of the release, I do not have timestamp data to tell whether the release was before market open or after market close. I choose day t to be the first full trading day after the press release is filed.

Since I am no longer using *Seeking Alpha*-title-testing data, but rather company-issued press releases, I need variation in title length that is unrelated to the earnings surprise. I use the instrument – the length of a company's legal name – to identify whether earnings-release-title length affects market outcomes. The first-stage regression using company-name length and press-release title length is highly significant. Appendix 1.5.6 Table 1.4 regressions (1)-(3) show the sensitivity of log title length to the log length of a company's name in CRSP is 0.17. The F-statistic for the hypothesis that the instrument's coefficient is zero is 864, exceeding the threshold of 10 recommended by Stock and Yogo (2005). Regression (4) is in levels rather than logs. The coefficient of 0.67 suggests that a firm with a one character longer name has on average 0.7 more characters in the title. Thus, the number of characters in the name meaningfully impacts the number of characters in the title. The coefficient may be less than one if companies do not always include all of their legal names in titles. For example, firms may not include "Limited" or "LLC." I do not find that firms with longer names have meaningfully less content in titles, and controlling for the non-name content of titles has no effect on results.

The exclusion restriction is likely satisfied as company name length is uncorrelated with earnings surprise measures. Similar to DellaVigna and Pollet (2009), I measure a firm's earnings surprise as the difference between reported EPS and the median analyst forecast of EPS from $t-45$ to $t-3$, where t is the announcement date. I normalize the difference in actual and forecasted EPS by the stock price at $t-3$. Scaling by price reveals the magnitude of the surprise. A \$0.10 surprise is bigger for a \$1 stock than a \$10 stock. I showed previously

29. See Engelberg et al. (2015).

that company name length is unrelated to firm characteristics after controlling for size.

I first examine the relation between instrumented-title length and turnover around earnings announcements. Turnover is the volume traded divided by the number of shares outstanding. If titles are optically longer for non-informative reasons, these titles may attract less attention. Less attention may lead to lower changes in announcement turnover relative to firms with shorter names in the industry. Table 1.8 Panel A regression (1) shows no relation between instrumented title length and turnover at $t-2$. Regression (3) shows a significant negative relation between changes in turnover and instrumented title length on the earnings-announcement day t. A standard-deviation increase in title length predicts 5% less announcement turnover, which is 5% of a standard deviation of the change in turnover on announcement relative to the past 10 days of turnover. Turnover during the days following the announcement does not seem to be related to title length, suggesting that investors do not delay the news until later. Consistent with less volume, I find that the number of trades on the NASDAQ exchange for a stock is negatively related to title length in Panel B.

Volatility and attention tend to be positively correlated on average.[30] Regression (3) in Table 1.8 Panels C and D shows that instrumented title length and volatility are negatively related on earnings announcement days. I measure volatility using the announcement day's intraday price range (high divided by low price) and the absolute close-to-close return. A standard-deviation increase in title length predicts a 50-basis-points decrease in the announcement day's trading range, which is about 10% of a standard deviation of the change in intraday price range on announcement. A standard-deviation increase in title length also predicts a 30-basis point decrease in absolute open-to-close returns, consistent with lower volatility during the day. The relations are concentrated on the announcement day.

While less attention to news due to longer titles should lead to less turnover and less intraday volatility, it is not clear that longer titles should result in a price underreaction.

30. See (Andrei and Hasler, 2014). The attention of less-sophisticated investors proxied for by activity on stock message boards is correlated with greater volatility (Antweiler and Frank, 2004).

Some sophisticated investors are likely to be unaffected by title length. Algorithmic traders are also likely unaffected by title length. However, these more skilled investors may be fewer in number and capital constrained. Alternatively, these skilled investors may trade less aggressively on news if there are fewer noise traders to cloak their information (Kyle, 1985).

Table 1.8 Panel E regression (3) suggests that there is a price underreaction to longer titles for both positive and negative news. I define positive news as a strictly positive earnings surprise relative to median analyst estimates from I/B/E/S. The observation counts are about 50% of the previous panels because only 50% of events have analyst earnings estimates. The coefficient on title length is 1.06, suggesting that for negative news the price does not fall as much. For positive news, the coefficient is -0.64 (-1.70+1.06), suggesting that the price does not rise as much. A one-standard-deviation increase in title length predicts a 40-basis-points underreaction to negative news and a 0.30-basis-points underreaction to positive news.[31] Since the instrument captures non-informative variation in title length, the underreactions should reverse. Summing the coefficients across regressions (4) to (6) shows that a full reversal tends to occur in the next month.

The underreaction is robust. Appendix 1.5.3 Table 1.13 shows that the return effects are fairly consistent across time splits and size splits. Appendix 1.5.4 Table 1.14 regressions (2) through (4) show that, after controlling for firm size, the underreaction is highly robust to adding a variety of firm controls, like the book leverage, revenue, asset value, gross margin, and net income margin, and event controls, notably the earnings surprise.[32]

Table 1.9 shows that investors seem to pay less attention to news with longer titles released on slow-news days (mostly Fridays) and low-VIX days. Using *Seeking Alpha* data, I

31. A standard deviation change in title length is 34 characters, which "is exactly as long as this phrase!"

32. Firm size is negatively related to earnings announcement turnover and returns. Table 1.5 shows company-name length is negatively related to firm size. Thus, omitting controls for firm size would induce a positive bias between name length and changes in turnover and returns.

find in Appendix 1.5.7 that investors read stock reports released on Fridays significantly less. I also find that activity on *Seeking Alpha* is significantly higher when the VIX is higher. While on busy-news days and high-VIX days investors may be more time constrained, investors are also more attentive and complex information may be more useful on those days.

1.3.6 Titles & Earnings Surprises

This section assess whether firms choose title length strategically. Table 1.10 examines what attributes of titles are predicted by the earnings surprise. Regression (1) shows a strong positive relation between earnings surprise and title length. Firms seem to add positive words to titles but avoid adding negative words to titles. Firms seem to believe that formulating titles differently for positive and negative news matters. While negative words grab attention, positive words reduce attention. Putting positive words in the title may have the effect of reducing attention to the news. Also, adding positive words increases title length and reduces attention to news.

Regressions (2) to (4) show that the earnings surprise seems to be unrelated to usage of abbreviations for common terms. For example, "eps" is a well-known abbreviation for "earnings per share." Firms reporting positive news should prefer abbreviating common terms to emphasize the positive words added to titles at the end. In these regressions, there appears to be no relation, suggesting that firms do not strategically alter titles non-informatively.

1.4 Conclusion

Attention is a scarce cognitive resource that investors must allocate. This paper is the first causal evidence that textual complexity is a driver of investor attention and market behavior. These results are also field evidence that complexity more generally matters.

The *Seeking Alpha* title-testing data cleanly show that investors tend to skip over stock

reports with longer titles. I rule out the possibility that longer titles are simply more informative, using company-name length as an instrument for title length. The cognitive abilities of investors matter as less-sophisticated investors are more put-off by complex titles. The large effect title length has on attention suggests that investors are complexity averse. The data also show that investors are more attracted to titles that are more negative in sentiment.

The aversion to title length measured in the *Seeking Alpha* data motivates testing whether investors in aggregate underreact to news with longer titles. One would hope that markets with sophisticated investors and algorithmic traders would not be impacted by title length. Instead, I find that non-informatively-longer-earnings-announcement titles lead to less volume, smaller intraday-price ranges, and return underreactions.

Financial discourse is generally complex, rather than concise and active. The findings motivate efforts by the SEC to simplify firm disclosures. Firms do not seem to be strategic with length, but are strategic with sentiment, writing longer headlines for positive news. These results, measured in a high-stakes setting and with new identification techniques, are also broadly relevant to journalists, marketers, and academics. Academic papers with longer titles receive less abstract views, citations, and downloads (see Appendix 1.5.10). The trend is towards 140 characters (Twitter tweets) and short Facebook posts, suggesting that some media outlets are learning over time that individuals are quite complexity averse, but much discourse in finance is not on trend.

1.5 Appendix

Table 1.8: Regressions showing the causal relation between earnings-announcement-market effects and earnings-press-release-title length. Title length is instrumented with the length of the company's legal name. The dependent variables in Panel A, B, C, D and E are log turnover (x100), log number of trades in a stock on NASDAQ exchange, log intraday price spread (x100), absolute log daily holding return (x100), and log daily holding return (x100). In Panel E, the dependent variables in regressions (4) to (6) are the holding returns for the days designated. The dummy for positive news is one for positive EPS surprises, measured relative to median analyst expectations from $t-3$ to $t-45$. Adjusting returns for the Fama-French three factor model does not noticeably alter the results. All regressions include SIC-4 by year fixed effects to account for industry trends and an announcement day fixed effect. Other controls include a third-degree polynomial of firm market capitalization on day $t-1$ and 10 days of lags of turnover, the intraday spread, absolute returns, and returns. Panel E includes a positive news dummy and interactions of control variables with the positive news dummy. Variables are winsorized at the 1% level. Standard errors are clustered by firm and date.

Panel A: %Δ Log Turnover

	(1) (t-2)	(2) (t-1)	(3) (t)	(4) (t+1)	(5) (t+2)	(6) (t+3)	(7) (t+4)
$\widehat{\text{Log Title Length}}$	-0.71	-4.00	-7.29**	-4.85	-0.91	0.15	5.55*
	(2.34)	(2.67)	(3.49)	(3.05)	(2.84)	(2.80)	(2.83)
Adjusted R^2	0.70	0.72	0.71	0.68	0.64	0.63	0.62
Within R^2	0.63	0.64	0.62	0.59	0.55	0.54	0.53
Observations	257839	257837	257832	257832	257832	257832	256863
Panel B: %Δ Log Number of Trades							
$\widehat{\text{Log Title Length}}$	-2.72	-4.54	-12.17***	-4.04	-0.12	0.92	1.65
	(2.46)	(2.86)	(4.01)	(3.38)	(3.04)	(2.96)	(2.97)
Adjusted R^2	0.92	0.92	0.91	0.91	0.91	0.90	0.90
Within R^2	0.86	0.85	0.83	0.82	0.82	0.82	0.81
Observations	143656	143632	143606	143586	143579	143577	142887
Panel C: %Δ Log Intraday Price Range (High/Low Price)							
$\widehat{\text{Log Title Length}}$	-0.10	-0.28***	-0.92***	-0.44***	-0.17	-0.11	0.01
	(0.09)	(0.10)	(0.20)	(0.11)	(0.11)	(0.11)	(0.10)
Adjusted R^2	0.52	0.49	0.44	0.48	0.47	0.46	0.46
Within R^2	0.36	0.33	0.27	0.31	0.30	0.29	0.29
Observations	261923	260542	259963	259963	259963	259963	259924
Panel D: %Δ Absolute Daily Return							
$\widehat{\text{Log Title Length}}$	-0.11	-0.10	-0.91***	-0.17*	0.03	0.05	0.07
	(0.08)	(0.09)	(0.21)	(0.09)	(0.08)	(0.08)	(0.08)
Adjusted R^2	0.27	0.25	0.20	0.24	0.24	0.24	0.24
Within R^2	0.14	0.13	0.08	0.12	0.11	0.11	0.11
Observations	261923	260542	259963	259963	259963	259963	259924

Panel E: Log Holding Period Return

	(1) (t-2)	(2) (t-1)	(3) (t)	(4) (t+1,t+5)	(5) (t+6,t+10)	(6) (t+10,t+20)
$\widehat{\text{Log Title Length}}$	0.03	-0.00	1.06**	-0.09	-0.19	-0.64
	(0.22)	(0.25)	(0.54)	(0.49)	(0.38)	(0.60)
$\widehat{\text{Log Title Length}}$ x Positive News	-0.02	0.03	-1.70**	0.16	0.42	1.06
	(0.26)	(0.30)	(0.69)	(0.59)	(0.47)	(0.73)
Adjusted R^2	0.16	0.16	0.09	0.16	0.19	0.19
Within R^2	0.01	0.01	0.06	0.01	0.00	0.00
Observations	142501	142333	142361	142361	142361	142361

Table 1.9: Regressions showing how the causal effect of press-release-title length on earnings-announcement-market outcomes varies with the VIX, news intensity, and analyst coverage. The VIX is the log of the ratio of the VIX on day t to the average VIX of 17.8. News intensity is the z-score of news count in a given calendar year. Analyst coverage is 1 if analysts provided earnings estimates ahead of the event and zero otherwise. The instrument for title length is the length of the company's name in CRSP. Earnings-announcement-press-release data are from *PR Newswire* and *Business Wire*. Market data is from CRSP. Log absolute returns and log turnover are multiplied by 100. Regressions include SIC-4 and date fixed effects. Other controls include a polynomial of firm market capitalization on day $t-1$ and 10 days of lags of turnover, intraday spread, squared returns, and returns. All variables are winsorized at the 1% level. Standard errors are clustered by firm and date.

		%Δ Turnover			%Δ Abs Ret			%Δ Spread	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\widehat{\text{Log Title Length}}$	-13.3***	-11.3***	-22.8***	-1.2***	-1.1***	-1.2***	-1.3***	-1.3***	-1.7***
	(4.3)	(4.2)	(5.9)	(0.3)	(0.3)	(0.3)	(0.2)	(0.2)	(0.3)
$\widehat{\text{Log Title Length}}$ x VIX	34.7***			1.5***			0.7		
	(6.9)			(0.5)			(0.5)		
$\widehat{\text{Log Title Length}}$ x News Intensity		9.9***			0.5***			0.5***	
		(2.6)			(0.2)			(0.2)	
$\widehat{\text{Log Title Length}}$ x Analyst			26.5***			0.3			1.1***
			(7.1)			(0.4)			(0.4)
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC-4 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.71	0.71	0.71	0.20	0.20	0.20	0.43	0.43	0.44
Within R^2	0.62	0.62	0.62	0.08	0.08	0.08	0.27	0.27	0.27
Number of Firms	12347	12347	12347	12347	12347	12347	12347	12347	12347
Observations	260062	260062	260062	260062	260062	260062	260062	260062	260062

Table 1.10: Regressions showing that title length is associated with the earnings surprise, but usage of common abbreviations is unrelated to the earnings surprise. For regressions (4) to (6), the dependent variable is 1 if the abbreviation (e.g. "fy" for "fiscal year") is used and 0 if the full form is used. Earnings surprise is the actual earnings per share less median estimated earnings, of those made from $t-45$ to $t-3$. The surprise is scaled by the stock price at $t-3$ and winsorized at the 1% level. Standard errors are clustered by firm.

	Log Title Length	Log # Words in Release	Log Firm Name Length	q or qtr=1 quarter=0	eps=1, earnings per share=0	fy=1 fiscal year=0
	(1)	(2)	(3)	(4)	(5)	(6)
Earnings Surprise	0.45***	-0.08***	-0.00	0.03	-1.18	0.45
	(0.04)	(0.03)	(0.01)	(0.02)	(1.23)	(0.33)
Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.36	0.89	0.98	0.38	0.60	0.71
Within R^2	0.00	0.00	0.00	0.00	0.00	0.00
Number of Firms	9206	9206	9206	8902	2195	1794
Observations	148266	148266	148266	132316	9977	7011

1.5.1 Averse to Small Differences in Title Length

Table 1.11: Regressions showing that the negative relation between title length and attention is identifiable even when the two-to-three titles for a single stock report have very similar character lengths, using *Seeking Alpha*-title-testing data. The sample splits include only titles within a certain difference of lengths (see the "max character difference in length" row). For example, regression (2) only includes titles that are two-or-fewer-characters different in length. All regressions include an article fixed effect, which holds the event, firm, author, and date fixed. Standard errors are clustered by stock report.

	Log Views/Emails			
	(1)	(2)	(3)	(4)
Log Title Length	-0.46	-0.29	-0.32**	-0.43***
	(0.57)	(0.25)	(0.15)	(0.11)
Max Character Difference in Length	1	2	3	4
Article FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.90	0.90	0.90	0.90
Within R^2	0.00	0.00	0.00	0.00
Num. Articles	1137	1797	2447	3056
Observations	2301	3679	5049	6336

1.5.2 Very Similar, Very Different Titles

Table 1.12: Regressions showing that the negative relation between title length and attention is identifiable even when the word overlap of the two-to-three titles per report is high, medium, or low, using *Seeking Alpha*-title-testing data. Title length is measured in characters. Word overlap is measured as the fraction of title length that is shared across the two titles compared. High overlap implies that enough words overlap so that at least 90% of the characters are the same across the two titles. Low overlap titles have less than 10% of characters the same. All regressions include an article fixed effect, which holds the event, firm, author, and date fixed. Standard errors are clustered by stock report.

	Log Views/Emails		
	(1)	(2)	(3)
Log Title Length	-0.34***	-0.31***	-0.31***
	(0.13)	(0.02)	(0.05)
Word Overlap	High	Middle	Low
Article FE	Yes	Yes	Yes
Adjusted R^2	0.92	0.92	0.90
Within R^2	0.02	0.06	0.08
Num. Articles	344	6265	410
Observations	688	12529	820

1.5.3 Robustness: Splits by Year and Firm Size

Table 1.13: Robustness check of Table 1.8 results using sample splits by time and firm size. Panel A regressions split the sample into two time periods, 1992-2003 and 2004-2015. Panel B regressions split the sample into two size groups, firms with market capitalizations less than $1B and firms with capitalizations greater than $1B.

Panel A: Sample Splits by Year

	(1) Ret % (t) 92-03	(2) Ret % (t) 04-15	(3) Turnover % (t) 92-03	(4) Turnover % (t) 04-15	(5) Spread % (t) 92-03	(6) Spread % (t) 04-15	(7) Abs Ret % (t) 92-03	(8) Abs Ret % (t) 04-15
$\widehat{\text{Log Title Length}}$	0.86	1.77**	-8.15	-15.44***	-1.40***	-0.98***	-1.19***	-1.08***
	(0.78)	(0.83)	(4.96)	(5.71)	(0.30)	(0.34)	(0.31)	(0.38)
$\widehat{\text{Log Title Length}}$ x Positive News	-1.83*	-3.06***						
	(1.00)	(1.03)						
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC-4 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.05	0.13	0.67	0.75	0.45	0.43	0.20	0.20
Within R^2	0.03	0.08	0.55	0.66	0.24	0.23	0.07	0.07
Observations	68941	73875	142352	117611	142352	117611	142352	117611

Panel B: Sample Splits by Firm Market Capitalization

	(1) Ret % (t) ≥$1B	(2) Ret % (t) <$1B	(3) Turnover % (t) ≥$1B	(4) Turnover % (t) <$1B	(5) Spread % (t) ≥$1B	(6) Spread % (t) <$1B	(7) Abs Ret % (t) ≥$1B	(8) Abs Ret % (t) <$1B
$\widehat{\text{Log Title Length}}$	0.90	0.66	-10.04*	-12.52***	-0.65*	-1.48***	-1.37***	-1.22***
	(0.99)	(0.77)	(6.01)	(4.81)	(0.35)	(0.30)	(0.43)	(0.31)
$\widehat{\text{Log Title Length}}$ x Positive News	-2.13*	-1.50						
	(1.21)	(1.00)						
Other Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SIC-4 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.10	0.09	0.82	0.68	0.48	0.41	0.25	0.19
Within R^2	0.05	0.06	0.61	0.56	0.17	0.22	0.05	0.07
Observations	51880	90936	71984	187979	71984	187979	71984	187979

1.5.4 Robustness: Adding Firm and Event Controls

Table 1.14: Robustness check of Table 1.8 results by adding firm and event controls. Book leverage is the debt to total assets. Earnings surprise is the difference between reported EPS and the median analyst forecast of EPS from $t-45$ to $t-3$, where t is the announcement date. I normalize the difference in actual and forecasted EPS by the stock price at $t-3$. I restrict the sample to only those firms with all firm and event controls to keep sample size fixed. Requiring analyst-eps estimates substantially reduces the sample size.

	(1)	(2)	(3)	(4)	(5)
	% Ret	% Ret	% Ret	% Ret	% Ret
	(t)	(t)	(t)	(t)	(t)
$\widehat{\text{Log Title Length}}$	0.69	0.99	1.05*	1.04*	1.22*
	(0.64)	(0.63)	(0.63)	(0.63)	(0.63)
$\widehat{\text{Log Title Length}}$ x Positive News	-1.10	-1.92**	-1.89**	-1.89**	-2.15***
	(0.80)	(0.79)	(0.79)	(0.79)	(0.79)
Book Leverage			0.05	0.07	0.08
			(0.12)	(0.12)	(0.12)
Log Revenue			0.14***	0.13***	0.14***
			(0.03)	(0.04)	(0.04)
Log Total Assets			0.26***	0.27***	0.28***
			(0.04)	(0.04)	(0.04)
Gross Margin				0.02	0.05
				(0.15)	(0.15)
Net Income Margin				0.11	0.10
				(0.13)	(0.13)
Earnings Surprise					20.69***
					(1.87)
Log Market Cap (t-1)		-16.09***	-13.22**	-13.21**	-27.00***
		(6.11)	(6.08)	(6.09)	(6.20)
Log Market Cap^2 (t-1)		0.74**	0.60**	0.60**	1.24***
		(0.30)	(0.29)	(0.29)	(0.30)
Log Market Cap^3 (t-1)		-0.01**	-0.01*	-0.01*	-0.02***
		(0.00)	(0.00)	(0.00)	(0.00)
Other Controls	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes
SIC-4 FE	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.09	0.09	0.09	0.09	0.09
Within R^2	0.06	0.06	0.06	0.06	0.06
Observations	130546	130546	130546	130546	130546

1.5.5 Title Attributes Predict Stock-Report Attributes

Table 1.15: Regressions showing the relation between stock-report-title attributes and stock-report-body attributes, using *Seeking Alpha*-title-testing data. Title length and stock report length are measured in characters. Positive and negative sentiment are measured as the fraction of words that are positive or negative according to Bill McDonald's lexicons. Firm names are excluded from the title length and sentiment scoring. The dependent variable in regression (2) is the Gunning fog index. A higher Gunning fog index predicts greater difficulty to read. The fog index depends on the number of words in sentences and the fraction of words with more than three syllables. Each regression includes a firm and date fixed effect. Standard errors are clustered by the topic firm of the stock report.

	(1) Log Stock Report Length	(2) Stock Report's Fog Index	(3) %Positive Words Stock Report	(4) %Negative Words Stock Report
Log Title Length	0.22***	0.76***	-0.11	-0.19***
	(0.04)	(0.14)	(0.09)	(0.07)
Negative Sentiment	-0.12	0.01	0.10	1.36***
	(0.08)	(0.33)	(0.25)	(0.17)
Positive Sentiment	-0.07	0.16	0.89***	-0.06
	(0.08)	(0.35)	(0.23)	(0.13)
Firm FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.18	0.15	0.56	0.47
Within R^2	0.02	0.01	0.01	0.03
Observations	4128	4083	4128	4128

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

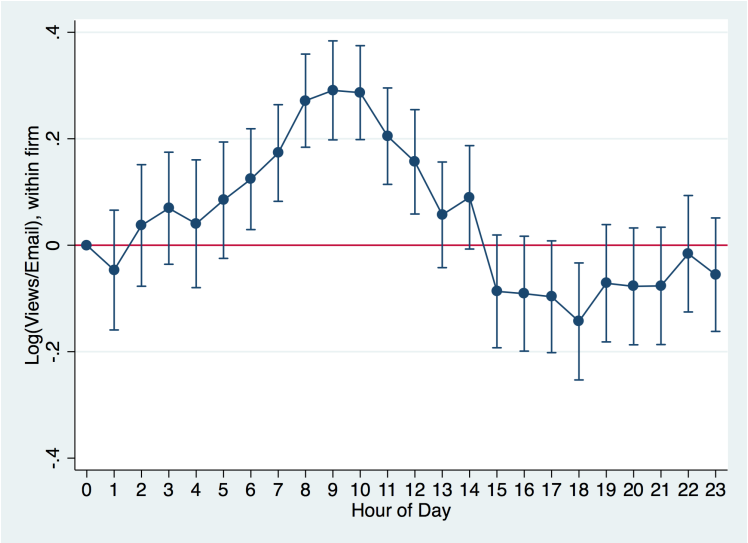
1.5.6 First Stage using Press-Release Data

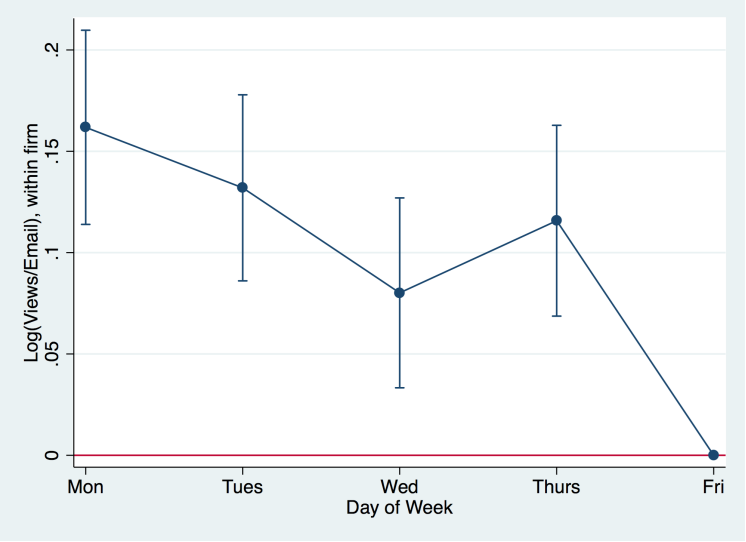
Figure 1.4: Regressions showing the first-stage of title length on firm name length, using earnings-releases from *PR Newswire* and *Business Wire* from 1992 to 2015. Firm name length and title length are measured in characters. The F-statistic for the coefficient on firm name length is 864, exceeding the 10 recommended by Stock and Yogo (2005). Each regression includes an SIC-4 and date fixed effect. Standard errors are clustered by firm.

	(1) Log Title Length	(2) Log Title Length	(3) Log Title Length	(4) Title Length
Log Length Firm Name	0.18***	0.17***	0.17***	
	(0.01)	(0.01)	(0.01)	
Length Firm Name				0.67***
				(0.03)
Log Market Cap			0.01***	1.14***
			(0.00)	(0.11)
Date FE	Yes	Yes	Yes	Yes
SIC-4 FE	Yes	Yes	Yes	Yes
Adjusted R^2	0.08	0.12	0.13	0.11
Within R^2	0.03	0.03	0.03	0.02
Observations	262833	262833	262833	262833

1.5.7 Attention by Hour and Day of Week

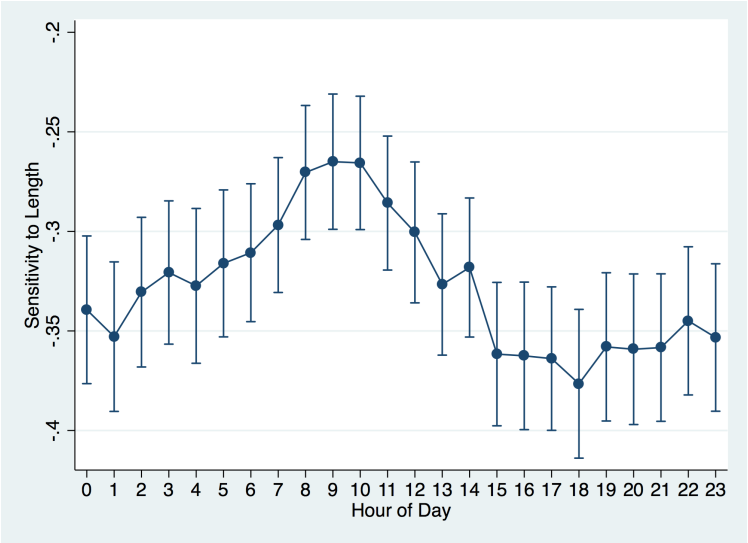
Table 1.16: Figure showing the percentage differences in yield (log views/emails) for stock reports released during the day and week, using *Seeking Alpha*-title-testing data. The "by hour" figure on top includes a firm and date fixed effect. The "by weekday" figure on bottom includes a firm and week fixed effect.

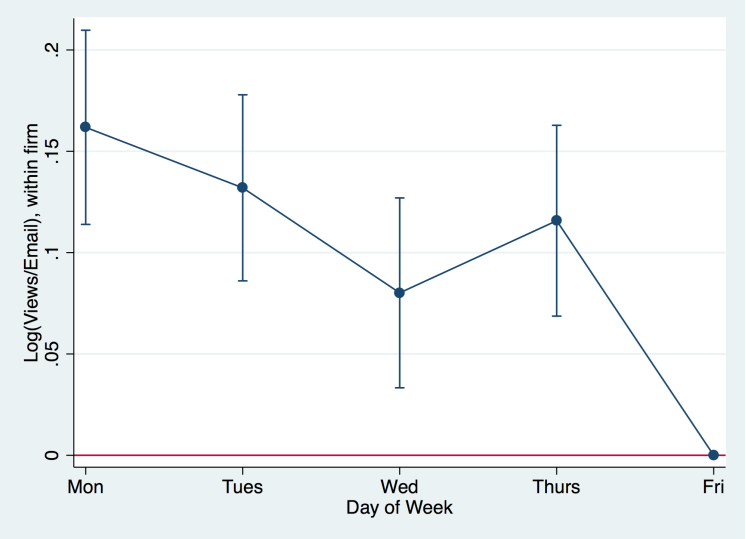




1.5.8 Sensitivity to Title Length by Hour and Day of Week

Table 1.17: Figure showing the percentage differences in yield (log views/emails) for stock reports released during the day and week, using *Seeking Alpha*-title-testing data. The "by hour" figure on top includes a firm and date fixed effect. The "by weekday" figure on bottom includes a firm and week fixed effect.





1.5.9 Differences in Author's and Editor's Titles

This section examines whether the author's original title differs from the author's alternative title and the editor's title.

Regressions (1) and (2) regress whether a title received the most page views of the plausible titles (1 if yes, 0 if no) on a dummy for whether the title was the analyst's original title, the analyst's alternative title, or the editor's title. Controlling for differences in title length, the author's alternative title receives the most page views 4% less often. The editor's proposed title has the most views 10% less. Notice R^2 is negligible in regression (1), suggesting the analyst's order of devising titles is not an important explanatory variable for predicting the winner.

Regression (3) examines whether the titles differ in title length by order. The analyst's alternative title is likely to be 2% longer and the editor's title 16% shorter. Regression (4) examines the negativity of titles by order. The author's alternative title and the editor's title tend to be more positive.

As a result of these findings, I include dummy variables in all regressions using *Seeking Alpha* data to control for whether a title is the author's or editor's and whether a title is the author's original or alternative.

Table 1.18: Regressions showing the differences between author and editor titles, use *Seeking Alpha*-title-testing data. The titles are the author's original, author's alternative, or editor's. Negative sentiment is determined by counting the number of positive and negative words in a title, excluding the firm's name, according to Bill McDonald's lexicons of positive and negative words. Regressions include an article fixed effect, which holds fixed the event, firm, and date.

	(1)	(2)	(3)	(4)
	1(or 0) if Title w/ Most Views	1(or 0) if Title w/ Most Views	Log Title Length	Negative Sentiment
Author Alternative Title(i=2,j)	-0.05***	-0.04***	0.02***	0.00
	(0.01)	(0.01)	(0.00)	(0.00)
Editor Title(i=3,j)	-0.01	-0.10***	-0.17***	0.01***
	(0.01)	(0.01)	(0.01)	(0.00)
Log Title Length		-0.55***		0.02***
		(0.02)		(0.01)
Article FE	Yes	Yes	Yes	Yes
Within R^2	0.00	0.05	0.11	0.00
Num. Clusters	9944	9944	9944	9944
Observations	22624	22624	22624	22624

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.5.10 Title Length and Attention on SSRN

I examine non-experimentally how headline length matters for academics' information-acquisition process. I cycled through academic ID numbers on SSRN from 0 to 700,000. Not all numbers have been assigned to individuals in this interval. The sample includes approximately 50,000 academics. For each academic, I retrieve all papers. Then, for each paper, I collect the abstract, number of abstract views, number of paper downloads, and number of citations.

Table 1.19 provides evidence consistent with the importance of title length. In regression (2), the dependent variable is abstract views, which is comparable to page views in *Seeking Alpha* data. The regression holds the author fixed and controls for the age of the papers. Doubling title length predicts 11% fewer abstract views.

In regression (3), the dependent variable is paper downloads. This measure is comparable to the read-to-end measure in *Seeking Alpha* data. Doubling title length predicts 13% fewer paper downloads. Regression (4) shows that conditional on viewing an abstract with a longer title, longer titles predict academics are more likely to download the paper. Also, longer abstracts predict that academics are more likely to download the paper.

In regression (5), the dependent variable switches to number of citations. About 60% of papers have 0 citations, so the dependent variable is $\log(\text{citations} + 1)$. Doubling title length predicts 4% fewer citations. Conditional on the number of abstract views and number of downloads, longer titles still predict fewer citations.

Table 1.19: This table examines the relation between the titles of SSRN papers of all academics with ID numbers between 1 and 700,000 on SSRN and the paper's abstract views, downloads, and citations.

	(1) Log Abstract Views	(2) Log Abstract Views	(3) Log Paper Downloads	(4) Log Paper Downloads	(5) Log Paper Citations	(6) Log Paper Citations
Log Title Length	-0.22***	-0.11***	-0.13***	0.01*	-0.04***	-0.02***
	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Log Abstract Views				1.22***		0.23***
				(0.00)		(0.01)
Log Downloads						-0.01***
						(0.00)
Log Abstract Length				0.01***		0.05***
				(0.00)		(0.00)
Year-Week FE	No	Yes	Yes	Yes	Yes	Yes
Author FE	No	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.01	0.66	0.55	0.81	0.49	0.51
Within R^2	.	0.00	0.00	0.58	0.00	0.04
Num. Academics	50011	50010	50010	50010	50010	50010
Observations	295390	295349	295349	295349	295349	295349
Clustering SE	Author	Author	Author	Author	Author	Author

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.5.11 Experimental Google Survey

One concern with the company-name-length instrument is that shorter company names may be more familiar or attention grabbing. To provide further support that optical-title length is off putting, I run two national surveys using Google's survey tool. For each survey, I asked 500 adults in the United States between the ages of 35 and 55 the following question: "Which news on Apple is more interesting?" One set of 500 adults received Survey 1, whereas the other set received Survey 2. The news events are the same in both surveys; however, I varied the "Inc." and "Incorporated," which varies the relative optical length of the title non-informatively. In Survey 1, the first title was selected 41% of the time. In Survey 2, the first title was selected 52% of the time. The 95% confidence interval is +/-3%. The shortest title was overall preferred in both surveys even though the information content was held constant. The surveys were conducted at the same time, and Google randomly switched the order of headlines within each survey.

- Survey 1
  - Apple Incorporated Sees iPhone Sales Slump
  - Apple Inc. Launching Smart Home
- Survey 2
  - Apple Inc. Sees iPhone Sales Slump
  - Apple Incorporated Launching Smart Home

1.5.12 Machine Learning Sentiment

Measuring the sentiment using lexicons is less suitable for titles, because titles are short and thus, 75% of titles have no matching words with the lexicons. Also, the lexicon approach equal-weights positive words like "good" and "great" and does not consider combinations of words. Therefore, I check my results using a machine-learning classifier approach.[33] This classifier learns "features" of titles that best predict whether the title is positive or negative. Features are words and combinations of words. The model learns the importance of features from annotated training data. *Seeking Alpha* classifies every stock report as bullish (recommending a purchase) or bearish (recommending a sale or short position). This classification of titles separates positive and negative titles well. The training data are all bullish and bearish titles from 2006 to 2015, a period that predates the title-testing program. The trained model then determines the sentiment of the titles in the title-testing sample. More specifically, the model provides a probability a title is negative. The results using this machine-learning measure of sentiment are very similar.

To illustrate variation in title sentiment within a stock report holding the context fixed, consider the following three titles. The model had a 5.3% confidence the first title is negative, a 31.6% probability that the second title is negative, and a 9.3% confidence that the third title is negative. The model learned that the word "trouble" is more indicative of a "negative" title than worried. The confidence scores are relatively low because the unconditional probability of a title being negative is low since most titles are bullish.

- Prospect Capital: About that Barclays note
- Prospect Capital: Is Barclays borrowing trouble?
- Prospect Capital: Barclays is worried, should you?

33. I use a naïve Bayes classifier, which is a probabilistic classifier based on applying Bayes' theorem with strong (naïve) independence assumptions between features.

REFERENCES

Agarwal, Sumit, John C Driscoll, Xavier Gabaix, and David Laibson, 2009, The age of reason: Financial decisions over the life cycle and implications for regulation, *Brookings Papers on Economic Activity* 2009, 51–101.

Ahern, Kenneth R., and Denis Sosyura, 2015, Rumor has it: Sensationalism in financial media, *Review of Financial Studies* 28, 2050–2093.

Andrei, Daniel, and Michael Hasler, 2014, Investor attention and stock market volatility, *Review of Financial Studies* 28, 33–72.

Antweiler, Werner, and Murray Frank, 2004, Is all that talk just noise? The information content of internet stock message boards, *Journal of Finance* 59, 1259–1294.

Bagnoli, Mark, Michael Clement, and Susan Watts, 2006, Around-the-clock media coverage and the timing of earnings announcements.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *The Journal of Finance* 61, 1645–1680.

Bank, Matthias, Martin Larch, and Georg Peter, 2011, Google search volume and its influence on liquidity and returns of German stocks, *Financial Markets and Portfolio Management* 25, 239–264.

Barber, Brad, and Terrance Odean, 2012, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *The Handbook of News Analytics in Finance* 173–210.

Barber, Brad, Terrance Odean, and Ning Zhu, 2009, Do retail trades move markets?, *Review of Financial Studies* 22, 151–186.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Bloomfield, Robert, 2008, Discussion of "Annual report readability, current earnings, and earnings persistence", *Journal of Accounting and Economics* 45, 221–247.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013, Salience and asset prices, *American Economic Review* 103, 623–628.

Chakrabarty, Bidisha, and Pamela C. Moulton, 2012, Earnings announcements and attention constraints: The role of market design, *Journal of Accounting and Economics* 53, 612–634.

Chemmanur, Thomas, and An Yan, 2009, Product market advertising and new equity issues, *Journal of Financial Economics* 92, 40–65.

Chen, Hailiang, Prabuddha De, Yu Hu, and Byoung Hyoun Hwang, 2014, Wisdom of crowds: The value of stock opinions transmitted through social media, *Review of Financial Studies* 27, 1367–1403.

Cohen, Lauren, Christopher Malloy, and Quoc Nguyen, 2015, Lazy prices.

Corwin, Shane, and Jay Coughenour, 2008, Limited attention and the allocation in securities trading of effort, *Journal of Finance* 63, 3031–3067.

Coval, Joshua, and Tobias Moskowitz, 1999, Home bias at home: Local equity preference in domestic portfolios, *The Journal of Finance* 54, 2045–2073.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, *Journal of Finance* 66, 1461–1499.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2015, The sum of all FEARS investor sentiment and asset prices, *Review of Financial Studies* 28, 1–32.

Da, Zhi, Umit G. Gurun, and Mitch Warachka, 2014, Frog in the pan: Continuous information and momentum, *Review of Financial Studies* 27, 2171–2218.

Damodaran, Aswath, 1989, The weekend effect in information releases: a study of earnings and dividend announcements, *Review of Financial Studies* 2, 607–623.

DellaVigna, Stefano, and Joshua Pollet, 2009, Investor inattention, firm reaction, and friday earnings announcements, *The Journal of Finance* 64, 709–749.

Dougal, Casey, Joseph Engelberg, Diego Garcia, and Christopher Parsons, 2012, Journalists and the stock market, *Review of Financial Studies* 25, 640–679.

Drake, Michael, Darren T. Roulstone, and Jacob Thornock, 2012, Investor information demand: Evidence from google searches around earnings announcements investor information demand, *Journal of Accounting Research* 50, 1001–1040.

Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements.

Engelberg, Joseph, David McLean, and Jeffrey Pontiff, 2015, Anomalies and news.

Engelberg, Joseph, and Christopher Parsons, 2016, Worrying about the stock market: Evidence from hospital admissions, *The Journal of Finance* 71, 1227–1250.

Fang, Lily, and Joel Peress, 2009, Media coverage and the cross-section of stock returns, *The Journal of Finance* 64, 2023–2052.

Feng, Lei, and Mark S. Seasholes, 2005, Do investor sophistication and trading experience eliminate behavioral biases in financial markets?, *Review of Finance* 9, 305–351.

French, Kenneth, and James Poterba, 1991, Investor diversification and international equity markets, *The American Economic Review* 81, 222–226.

Gabaix, Xavier, and David Laibson, 2005, Bounded rationality and directed cognition.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg, 2006, Costly information acquisition: Experimental analysis of a boundedly rational model, *American Economic Review* 96, 1043–1068.

Giglio, Stefano, and Kelly Shue, 2014, No news is news: Do markets underreact to nothing?, *Review of Financial Studies* 27, 3389–3440.

Grinblatt, Mark, S Ikaheimo, and Matti Keloharju, 2008, Are mutual fund fees competitive? What IQ-related behavior tells us.

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa, 2009, Do smart investors outperform dumb investors?

Grinblatt, Mark, Matti Keloharju, and Juhani Linnainmaa, 2011, IQ and stock market participation, *Journal of Finance* 66, 2121–2164.

Hartzmark, Samuel, and Kelly Shue, 2015, A tough act to follow: Contrast effects in financial markets.

Hillert, Alexander, Heiko Jacobs, and Sebastian Müller, 2014, Media makes momentum, *Review of Financial Studies* 27, 3467–3501.

Hirshleifer, David, Sonya Lim, and Siew Teoh, 2004, Disclosure to an audience with limited attention.

Hirshleifer, David, Sonya S Lim, and Siew Hong Teoh, 2011, Limited investor attention and stock market misreactions to accounting information, *Review of Asset Pricing Studies* 1, 35–73.

Hirshleifer, David, Sonya Seongyeon Lim, Siew Hong Teoh, David Hirshleifer, Sonya Seongyeon, and Laura Field, 2009, Driven to distraction: Extraneous events and underreaction to earnings news, *Journal of Finance* 64, 2289–2325.

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics* 36, 337–386.

Hou, Kewei, Lin Peng, and Wei Xiong, 2009, A tale of two anomalies: The implications of investor attention for price and earnings momentum.

Huddart, Steven, Mark Lang, and Michelle Yetman, 2009, Volume and price patterns around stock's 52-week highs and lows: Theory and evidence, *Management Science* 55, 16–31.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, 2009, Attention allocation over the business cycle.

Kahneman, Daniel, 1973, *Attention and Effort* (Prentice-Hall, Inc.).

Kyle, Albert S., 1985, Continuous Auctions and Insider Trading, *Econometrica* 53, 1315–1335.

Lawrence, Alastair, 2013, Individual investors and financial disclosure, *Journal of Accounting and Economics* 56, 130–147.

Lehavy, Reuven, Feng Li, and Kenneth Merkley, 2011, The effect of annual report readability on analyst following and the properties of their earnings forecasts, *Accounting Review* 86, 1087–1115.

Li, Feng, 2008, Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and Economics* 45, 221–247.

Li, Jun, and Jianfeng Yu, 2012, Investor attention, psychological anchors, and stock return predictability, *Journal of Financial Economics* 104, 401–419.

Loewenstein, George, Duane Seppi, Nachum Sicherman, and Stephen Utkus, 2016, Financial attention, *Review of Financial Studies* 29, 863–897.

Loh, Roger, 2008, Investor inattention and the underreaction to recommendations.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? Textual analysis, distionaries, and 10-ks, *Journal of Finance* 66, 35–65.

Loughran, Tim, and Bill McDonald, 2014, Measuring readability in financial disclosures, *Journal of Finance* 69, 1643–1671.

Loughran, Tim, and Bill Mcdonald, 2016, The Use of EDGAR Filings by Investors .

Michaely, Roni, Amir Rubin, and Alexander Vedrashko, 2016, Are Friday announcements special? Overcoming selection bias, *Journal of Financial Economics* 122, 65–85.

Miller, Brian, 2010, The effects of reporting complexity on small and large investor trading, *Accounting Review* 85, 2107–2143.

Niessner, Marina, 2014, Strategic Disclosure Timing and Insider Trading, *SSRN Electronic Journal* 1–54.

Nieuwerburgh, Stijn Van, and Laura Veldkamp, 2009, Information immobility and the home bias puzzle, *The Journal of Finance* 64, 1–22.

Patell, James, and Mark Wolfson, 1982, Good news, bad news, and the intraday timing of corporate disclosures, *The Accounting Review* 57, 509–527.

Peng, Lin, 2005, Learning with information capacity constraints, *The Journal of Financial and Quantitative Analysis* 40, 307–330.

Peng, Lin, and Wei Xiong, 2006, Investor attention, overconfidence and category learning, *Journal of Financial Economics* 80, 563–602.

Stock, James, and Motohiro Yogo, 2005, Testing for weak instruments in linear IV regression., in *Identification and Inference for Econometric Models*, 80–108 (Cambridge University Press).

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

Tetlock, Paul, Maytal Saar-Tsechansky, and Sofus MacSkassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

Tetlock, Paul C., 2011, All the news that's fit to reprint: Do investors react to stale information?, *Review of Financial Studies* 24, 1481–1512.

You, Haifeng, and Xiao-jun Zhang, 2009, Financial reporting complexity and investor underreaction to 10-k information, *Review of Accounting Studies* 14, 559–586.

Yuan, Yu, 2015, Market-wide attention, trading, and stock returns, *Journal of Financial Economics* 116, 548–564.