INVESTOR ATTENTION, TEXTUAL STYLE, AND INSIDER TRADING PLANS

By

Phillip Baeza

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Business Administration – Finance – Doctor of Philosophy

2022

ABSTRACT

INVESTOR ATTENTION, TEXTUAL STYLE, AND INSIDER TRADING PLANS

By

Phillip Baeza

The first chapter of this dissertation analyzes how the textual style of firm disclosures affects investors' information acquisition patterns. Using the SEC EDGAR server logs, I show that investors seek more information in the firm's previous filings when the 10-K is difficult to read and more negative in sentiment. This sensitivity is stronger for small firms and those with few analysts following, suggesting that a weak information environment helps motivate investors to broaden their research. Moreover, I find that owners of the company's stock are far less sensitive to textual attributes than non-owners, but they are more likely to increase their holdings when they do extra research. This chapter is the first to directly analyze how heterogeneity in text style affects those who read disclosures.

The second chapter of this dissertation examines abuse of insider stock trading plans made under the SEC's Rule 10b5-1. These plans are meant to defend against allegations of trading on private information. Since the rule was enacted in 2000, however, the plans have been shrouded in secrecy with the vast majority being unannounced until the first trade is made. This chapter studies 10b5-1 plan announcements in 8-K filings to see how many shares the insider plans on selling and compare this with how many shares the insider ends up selling. We find that insiders sell the proposed number of shares in only 24% of announced 10b5-1 plans. We then investigate the firm characteristics that predict following through on 10b5-1 plans. This chapter also updates previous literature's findings on 10b5-1 trades and plan announcements.

This dissertation is dedicated to my wife, Samantha Baeza, and my son, Ellison Baeza.

ACKNOWLEDGEMENTS

I thank my dissertation committee - Ryan Israelsen, Zsuzsanna Fluck, Zoran Ivkovic, and Hao Jiang - for their guidance throughout my time at Michigan State. I owe special thanks to Ryan Israelsen, the chair of my committee, for incredible mentorship.

I also thank the rest of the finance department faculty, especially Andrei Simonov, Morad Zekhnini, Dmitriy Muravyev, Nuri Ersahin, Kirt Butler, Mark Schroder, Antoinette Tessmer, and Miriam Schwartz-Ziv for helpful comments and job market advice.

Finally, I thank my family for endless love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. Textual Style and Investor Follow-up Attention

1. INTRODUCTION

Finance literature has shown that the style of text in firm disclosures predicts real outcomes. For example, the 10-K's readability and sentiment correlate with future stock returns (Feldman et al 2010), earnings persistence (Li 2008), volatility, and even the likelihood of lawsuits (Loughran and McDonald 2014)[1]. These papers have largely ignored the behavior of the individuals who consume and process text data. On the other hand, a separate strand of literature describes how investor attention is necessary to incorporate news into stock prices (recently, Da et al 2011 and Ben-Rephael et al 2017). In this paper, I combine these two areas and show how textual style affects subsequent investor learning. The results help illuminate the black box of investor information acquisition and suggest that firms may be able to alter their disclosures' textual style to draw attention to or away from themselves (i.e., if firms have negative events that they do not want investors being reminded of, they may want to write their 10-K in a certain way).

The Securities and Exchange Commission (SEC) provides a perfect setting to study this problem with their public database of EDGAR (Electronic Data Gathering, Analysis, and Retrieval) server logs. This data reveals which company disclosures are downloaded from which IP address[2] at what time. The server logs have been used to measure attention at the firm level and to gauge interest in specific filings (Chen et al 2020, Li et al 2019, Crane et al 2020, and others). There has been little analysis, however, on the specific patterns of research by individual EDGAR users. This paper opens this topic and sheds light on what attributes of text data motivate investors to learn more about the company.

---

[1] Earnings announcements (Bushee et al 2018), mutual fund prospectuses (Tucker et al 2020), and M&A announcements (Pely and Schock 2020) are other recently popular settings to study the effects of textual style.
[2] The last octet of the IP addresses is anonymized. Chen et al 2020 develop a cipher to decode the last octet using web traffic from another website.

In particular, I analyze how investors conduct follow-up research after viewing recent 10-K filings. Investors read 10-Ks for a summary of the firm's fiscal year, but the format and style of these disclosures vary widely. In 2017, for example, the raw file size of 10-Ks ranged between 43 kB and 256,646 kB. In this paper, I analyze how heterogeneity in the textual style of these filings affects whether EDGAR users choose to seek more information about the company. To measure this activity, I first identify the "interested" investors that download an annual report within six trading days of its public release. Out of these users, I mark as "determined" those who subsequently view another of the company's SEC filings within the six-trading day period. My primary variable of study, *PercentDetermined*, is the percentage of "interested" users that are "determined" to learn more. The goal of this paper is to explain which traits of disclosure text result in higher *PercentDetermined* and what impact this has on the company.

In all my regressions, I include common control variables that might also affect investor attention. Most of these are significant predictors of *PercentDetermined* in logical directions. Past year stock price volatility, absolute filing day return, number of Compustat Business Segments, and number of SEC disclosures around the 10-K filing date have positive effects on *PercentDetermined*. These variables control for uncertainty and complexity about the firm that might compel investors to conduct further research. Past year alpha works in the opposite direction; EDGAR users tend to follow up on a firm more often when the stock performed poorly over the past year.

My main results describe how readability of the 10-K affects *PercentDetermined*. Does a difficult-to-read filing cause investors to give up on the firm, or do they persist and seek more information? I test this by regressing *PercentDetermined* on a few readability measures, including Bonsall et al 2017's Bog Index, and controls. I find that when the 10-K is more difficult to read, *PercentDetermined* is higher as EDGAR users seek to supplement their information set from the firm's other SEC filings. If a firm wants to discourage investors from going back and scrutinizing their recent SEC filings, then they should write a clear 10-K. Investors are even more sensitive to readability for small firms and those with less analyst coverage. For these companies, the information environment is already weaker, and a poorly written annual report exacerbates the need for more research. In Grossman and Stiglitz 1973 terms, the observable signal of the future stock return is more volatile for these firms, so the benefit from becoming informed is greater.

Other traits of the 10-K's text also affect EDGAR users' tendency to conduct follow-up research. Investors seek more information when the filing contains a higher percentage of negative, weak modal, and strong modal words and a lower percentage of positive words. The results on sentiment suggest that investors follow up on a stock when they suspect that its price might drop. This might be an opportunity to dig deeper on the company and predict whether the dip will reverse or not. This explanation also fits with the results on readability; when an investor looks at a poorly written 10-K, they might see an opportunity to gain an informational advantage over other market participants. Weak modal words, such as "could" and "might", represent uncertainty and encourage follow up research much like readability does. The significant effect of strong modal words, like "always" and "will", is harder to interpret. Loughran and McDonald 2014 show that these terms predict future disclosure of material weakness. Perhaps investors sense this signal and scrutinize other SEC filings to uncover the truth.

In an effort to better understand this effect, I study an exogenous shock to the way investors use electronic 10-K filings on EDGAR: the introduction of XBRL. To make filings easier to use for investors, XBRL (eXtensible Business Reporting Language) standardizes the tags for accounting items. I find that the implementation of XBRL reduces the need for investors to conduct follow-up attention, but it does not have a consistent effect on *PercentDetermined*'s sensitivity to textual attributes. It has an insignificant impact on readability's effect, it dampens the effect of negative words, and it accentuates the effect of positive words.

Prior literature has established that low readability in the 10-K leads to future uncertainty about the firm, such as higher stock price volatility (Lehavy et al 2011, Loughran and McDonald 2014, and Bonsall et al 2017). I document that investors tend to fight against this effect by searching for more information on the firm. However, their extra effort does not reduce uncertainty caused by unreadability. In fact, the interaction of P*ercentDetermined* * *Unreadability*[3] predicts higher stock price volatility, analyst dispersion, and earnings surprise. This is likely due to uncertainty around the firm that correlates with low readability and high *PercentDetermined* but is not fully controlled for in my model. If I could eliminate this explanation, it would be entertaining to conclude that follow-up research only serves to confuse investors further.

Next, I examine whether professional investors behave differently. This requires identifying the organization that owns each IP address for each day in the EDGAR server logs. Using the American Registry for Internet Numbers (ARIN)'s WhoWas database, I merge in organization names for about 70% of the requests in the EDGAR server logs. Then, I use regular expressions along with manual searching to identify a subset of these that are likely financial institutions. I use activity from these organizations to construct *PercentDetermined* again.

My first hypothesis is that financial professionals will be more sensitive to textual attributes than the full sample. They should be more skilled at gleaning information from the firm's other SEC filings and, thus, more willing to try doing so. However, the regression results for this group are very similar to those of the full sample. My second hypothesis is that follow-up research from

---

[3] For the readability measures I use, high values represent low readability.

professional investors will be more effective at reducing the uncertainty caused by low readability. There is some evidence of this. While the coefficients for *PercentDetermined* * *Unreadability* are positive and significant for the full sample, these coefficients are not significantly different from zero for professional investors. It is possible that while follow-up research from these sophisticated users positively correlates with uncertainty, their specialized scrutiny reduces the confusion effect of readability to near zero. However, it is difficult to separate the two competing effects.

Finally, I analyze how information acquisition patterns change when EDGAR users own the stock. Using the ARIN data, I match IP addresses from the log files to investors that file Form 13-F. I merge in their quarterly holdings data from the WRDS SEC Analytics Suite 13-F database to see if they act differently for stocks they own. I find that institutional investors are more likely to conduct follow-up research if they own the stock. However, owners are much less sensitive to text attributes than non-owners. For example, an increase of one in the Bog Index raises the probability of following up by 0.92 percentage points for non-owners and only 0.16 percentage points for owners. I also find that owners that follow up on the 10-K are 10% more likely to increase their holdings than if they do not. Taken together, these results highlight the importance of textual style for firms wishing to attract institutional investors. Public companies must be deliberate in writing their disclosures and should consider the impact of readability and sentiment on the breadth of investor information gathering.

The rest of the paper is organized as follows: Section 2 discusses the literature in text analysis and information acquisition; Section 3 describes the data and sample construction; Section 4 presents the results for the main sample; Section 5 reports the results for financial professionals; Section 6 examines the impact of ownership in the stock; and Section 7 concludes.

2. LITERATURE REVIEW

2.1. Textual Analysis

Finance and accounting literature has recently begun to study the effects of textual style of firm disclosures. Document readability has long been a core concern of the SEC. In 1998, they issued “A Plain English Handbook: How to create clear SEC disclosure documents”, which provides a detailed guide on how to clearly communicate financial information. Finance and accounting literature has since studied how to measure readability and identify its effect on firms. Most early papers use the Fog Index, which is equal to $0.4 * (\# \, of \, words \, per \, sentence + \% \, of \, words \, longer \, than \, 2 \, syllables)$. Like most other measures of readability, high values indicate low readability. Li et al 2008, Biddle et al 2009, and Lehavy et al 2011 are just a few papers that use the Fog Index to measure 10-K readability. However, Loughran and McDonald 2014 show that the Fog Index is ill-suited for financial documents as the most frequent words longer than 2 syllables are common business terms, like “financial” and “company”.

A host of alternatives to the Fog Index have been proposed, and three seem to have risen to the top: file size, number of words, and the Bog Index. Loughran and McDonald suggest using the size, in bytes, of the 10-K filing as a measure of confusing text. Ertugrul et al 2020 and other recent papers trust this simple metric to identify readability. The reasoning is that most elements of unreadability mentioned in the Plain English Handbook contribute positively to file size. Less readable documents are said to contain run-on sentences, unwieldy legalese, and technical jargon. In their paper, the authors show that larger file size predicts high stock price volatility, earnings surprise, and analyst dispersion. Number of words has the same intuition behind its use; the more words a 10-K uses, the more opaque it appears to investors. Another motivation for these two measures is their ease of observation. There is very little arbitrary parsing required, except in identifying the main body of text to count number of words. In contrast, Bonsall et al 2017 introduce the Bog Index, which is produced by running the 10-K's text through a program called StyleWriter. The program identifies violations of plain English and sentence length. The recent literature has adopted this index as a more scientific metric of readability. Balachandran et al 2020 uses the Bog Index to analyze 10-K readability's effect on various M&A outcomes. Out of the three measures I use, Bog Index is certainly the most rigorous as it analyzes specific text patterns.

This paper also studies the sentiment of text. Similar to the readability literature, investor sentiment research has grappled with how to measure their variable of interest. The early methods rely on equilibrium trading outcomes. Baker and Wurgler 2006 construct an index of pre-existing measures which are based on mutual fund prices, stock trading volume, equity issuances, and dividend premiums. Textual analysis entered the fray with Tetlock 2007 using counts of negative words in Wall Street Journal columns to predict Dow Jones Index returns. Since then, textual analysis has been the leading way to measure sentiment and is used widely for firm disclosures. Loughran and McDonald 2011 develop word lists specifically for business settings and show their predictive power for stock returns, trading volume, litigation, etc. Bodnaruk et al 2015 shows how the frequency of constraining words in 10-K filings predicts liquidity events in the future.

2.2 Investor Attention

A wide literature studies how investors allocate their limited attention Kahneman 1973 and how this is necessary for information transmission. DellaVigna and Pollet 2009 show that earnings announcements (EAs) made on Fridays have slower responses in stock prices. They also find that firms know this and strategically schedule EAs with bad results on Fridays to dampen the reaction of investors. Similarly, Hirschleifer et al 2009 find that investors respond more slowly to EAs when there are many other firms announcing on the same day. Da et al 2011 are the first to directly measure investor attention. They use the Google Trends Search Volume Index to show that high abnormal attention accentuates IPO filing day returns and subsequent price reversal. Ben-Rephael et al 2017 propose using Bloomberg data to identify abnormal attention from sophisticated investors. They demonstrate that price drifts after EAs and analyst recommendation changes are attenuated by high attention.

To my knowledge, there is only one database that describes the timestamped information-gathering activities by specific users: the SEC EDGAR log files. So far, most papers have only used the EDGAR log files to study levels of attention to firms or to specific filings. Drake et al 2014 conduct a wide survey of the database, and their main tests rely on aggregating attention at the firm level. They do study individuals' frequency of use, but they do not examine the sequence of activity. Drake et al 2017 summarize EDGAR users' demographics by identifying the ZIP codes of their IP addresses.

A few of these papers exploit the timing of EDGAR requests, but they do not study the specific research patterns by individual users. For example, Li et al 2019 show that when investors study a firm's SEC filings before an earnings report, the post earnings announcement drift is weaker. They aggregate EDGAR attention by investor in the 30 days leading up to earnings announcement. Gibbons et al 2021 look at how analysts use EDGAR immediately prior to their forecasts and recommendations. They find that analysts that conduct research on EDGAR have more accurate and informative reports.

2.3. The Relationship between Text and Attention

There are a few crossover papers that combine text analysis with information acquisition.

With a similar spirit to this paper, Lehavy et al 2011 show that less readable 10-Ks cause analysts to take more time with their reports and increase the information content of the analysis. This supports my conclusion that low readability in firm disclosures causes EDGAR users to seek more information. Cohen et al 2020 show the negative impact of changes in 10-K filings on firm performance and stock returns. The effects are decreased when investors download both the recent 10-K and the previous year's report. Chen et al 2020 examine how exogenous shocks to the information environment of a firm affect how hedge funds conduct research on EDGAR. They find that hedge funds tend to ramp up their information acquisition and earn higher abnormal returns on stocks affected by brokerage closings. Their final takeaway is that sophisticated investors can substitute analyst reports with research on public information.

Cao et al 2020 discuss the reverse of my research question: how does investor attention affect readability? They analyze how attention from automated EDGAR users causes companies to improve the "machine readability" of the subsequent year's 10-K. In contrast, I study how text attributes affect investor attention for human users. This paper complements the papers in this subsection to help illustrate the relationship between text readability, investor attention, and real firm outcomes.

3. DATA AND SAMPLE CREATION

The sample for this paper starts with all 139,851 10-K filings made between January 1st, 2003 and March 31st, 2017. Following Loughran and McDonald 2014, I remove firm-year duplicates, filings made less than 180 days before the firm's last 10-K, and those with fewer than 2,000 words. To maximize the usefulness of *PercentDetermined*, I also remove filings with fewer than five IP addresses viewing within six trading days of its release. Then, I require that each observation has enough data from CRSP, Compustat, and I/B/E/S to create my model's variables.[4] After these reductions, the final sample contains 33,793 filings. Table 1 displays the summary of this sample.

To identify readability of each 10-K, I select three of the most popular measures from recent literature: Bog Index, size of the filing in bytes, and number of words. The Bog Index is the result of running the 10-K's text through a program called StyleWriter. The software identifies plain English violations, such as passive language, legalese, and run-on sentences. The output is an index that ranges from 54 to 131 in my final sample. I retrieve the data for this measure from Samuel Bonsall's website[5]. File size is a simpler measure widely used in recent literature. The idea is that larger files are tougher for investors to understand and may, whether intentionally or not, make some information difficult to find. Number of words in the 10-K has a similar reasoning, except that it excludes tables, graphics, HTML and XML tags, and other non-text data. The data for file size and number of words are retrieved from Bill McDonald's website.[6]

Other 10-K sentiment measures come from counts of Loughran and McDonald (LM) 2011's word lists, also available at Bill McDonald's website. In regressions, I use the simple percentages of negative, positive, strong modal, and weak modal words in the 10-K's text. Other weighting schemes, like term frequency inverse document frequency (tf.idf), reduce the weight of words that appear in many documents, like "loss" or "cost". The literature has shown that tf.idf enhances the predictive power of the LM word lists. In this paper, however, I show that simple proportions affect investor attention.

This paper exploits the only dataset that details information acquisition by user: the

[4] See appendix for detail on each variable.
[5] https://sites.psu.edu/sambonsall/
[6] https://sraf.nd.edu/

EDGAR server logs. The database is published on the SEC website[7] for January 1st, 2003 to June 30th, 2017[8]. Each observation describes a server request with the variables: IP address, datetime, filing ID, company ID (CIK), server response code, and an indicator for self-described web crawlers. I focus on server requests with response codes between 200-299, which indicates a successful download. Then, I follow Lee et al 2015's method to eliminate "robot downloads"; I remove activity from IP addresses that download more than 50 unique companies' filings in one day. I also remove observations with user agents that self-identified as web crawlers. It is important to filter this activity out because I want to identify when a human studies a 10-K and reacts to it.

My specific interest is investors' EDGAR activity after looking at recent 10-K filings. First, I identify when a user downloads a 10-K within 6 trading days (about 8 calendar days) of its filing. I mark those IP addresses as interested. Then, I record which filings of the same company the user downloads within the 6-trading day period. If a user looks at any of the company's other SEC filings, I mark them as determined. Figure 1 illustrates the timeline of what "determined" investors do, an activity that I call "follow-up research" throughout the paper. I choose this time period because investors are less likely to have seen other news about the company, either from the firm's own disclosures or outsiders analyzing the annual report. The exact threshold of 6 trading days is not important, and my results are robust to time periods ranging from the day of the filing to two weeks after. Note that the server logs include weekends and holidays, which have lower EDGAR activity but are still represented in this paper.

In some regressions, I study the logged number of "determined" IP addresses. This is more

---

[7] https://www.sec.gov/dera/data/edgar-log-file-data-set.html

[8] The log files are missing for September 24, 2005 through May 10, 2006. This omission would only be significant for this paper if the behavior of EDGAR users in this time period were sufficiently different to affect my estimations. There is no evidence or logical reason that this might be the case.

of a measure of raw attention in the company. In contrast, the percentage of "determined" out of the "interested" investors, which I call *PercentDetermined*, normalizes by the level of attention and represents a response by human EDGAR users who were exposed to the 10-K. Higher *PercentDetermined* identifies when a large portion of investors read the 10-K and then choose to gather more information. The purpose of this paper is to explain why they feel the need to study the company more closely.

There are a few limitations of this approach to studying investor information acquisition. First, it is possible that EDGAR users are significantly different than other investors, and my results might only apply to them. Many investors access 10-K filings through a secondary source, like FactSet or Bloomberg. Others simply read analyst reports to see the main points. It is easy to imagine that EDGAR users could be more traditional or might not have enough money for a paid subscription that displays the 10-K in a different format. It is possible that my results are only significant for EDGAR users and do not represent the behavior of the entire population of investors. Second, it is also possible that EDGAR users do not make the decision to conduct follow-up research based on the 10-K's text. They might remember that the company issued an important 8-K a few weeks ago, and the decision to look at it has nothing to do with the 10-K. This phenomenon would simply add noise to my tests, biasing the coefficients towards zero. If I still detect predictive power of the 10-K's text on *PercentDetermined*, then the statistical significance of the true effect should be larger.

Finally, the data used to match IP addresses by day come from the American Registry of Internet Numbers (ARIN). I use Chen et al 2020's cipher to obtain the fourth octet of each IP address. I use a Python script to access ARIN's historical WhoWas database to request all IP addresses that downloaded at least 100 10-K filings from 2003 to 2017. About 70% of these searches produce a result from ARIN, which provides a name, phone number, and physical address for the organization that owns that IP address on each day. I use these pieces of information to match activity on the log files to specific entities. For 13F filers, I retrieve the names, addresses, and phone numbers from EDGAR. Data on 13F filer holdings comes from the WRDS SEC Analytics Suite.

4. MAIN SAMPLE

First, I analyze *PercentDetermined* in the main sample. This includes all human EDGAR users: retail investors, institutional investors, analysts, government agencies, etc. When investors view recent 10-K filings, which textual attributes predict higher follow-up attention?

4.1. Textual Style Affects Follow-Up Attention

Table 2 shows how each measure of 10-K readability affects investors’ tendency to follow up. Each column represents an OLS panel regression with standard errors clustered by industry and year. In columns 1-5, the dependent variable is the logged number of IP addresses determined, *Log(IPs determined)*. This is the log of the numerator of *PercentDetermined*. In these regressions, Gross File Size and Number of Words are both positive and significant predictors of *Log(IPs determined)*. The coefficient for Bog Index is positive and statistically significant without control variables (untabulated here) but loses significance when they are included. This insignificance may be due to noise in the non-text portion of Gross File Size (pictures, tables, HTML tags, etc.) In columns 6-10, the dependent variable is the percentage of determined IP addresses out of those who are interested, *PercentDetermined*. In these tests, Bog Index and Number of Words are positive predictors and significant at the 1% level, while Gross File Size is negative and insignificant. An increase of one standard deviation in the Bog Index (6.86 points) correlates with a 0.56 percentage point increase in *PercentDetermined*. Overall, it seems that low readability in firm disclosures encourages investors to seek information elsewhere.

The main concern with these results is that low readability might simply coincide with uncertainty around the firm that also drives higher *PercentDetermined*. To show this point, Table 3 displays strong evidence that firms specifically design unreadable 10-Ks when their stock performed poorly over the past year and was highly volatile. This result complements those of Cao et al 2020, who show that companies adapt the machine readability of their 10-Ks after the level of robot downloads is high. Low readability is also positively correlated with absolute filing day return and number of SEC filings in the past year, which both suggest investor uncertainty. I address this issue by including these variables as controls in Table 2. The same variables that predict readability also predict *Log(IPs Determined)* and *PercentDetermined* in the same direction. While it is still possible that other factors are the driving force of investor follow-up research, my controls do account for some of this.

Table 4 shows the sensitivity of *PercentDetermined* to other textual attributes of the 10-K. I include Bog Index as an independent variable to compare with the proportional counts from the LM word lists: Negative, Positive, Weak Modal, and Strong Modal. Columns 1-6 display *Log(IPs Determined)* as the dependent variable, while columns 7-12 use *PercentDetermined*. The proportions of negative and strong modal words increase both the level of follow-up research and the normalized value, *PercentDetermined*. Negative sentiment might excite EDGAR users who suspect an upcoming dip in the stock price. They might seek more information to try to predict whether the dip will reverse or not. Strong modal words are known to predict future reports of material weakness. Perhaps their use in the 10-K sets off investors to this potential, and they attempt to uncover the truth. The presence of positive words predicts lower values of the two dependent variables. When firms use fewer positive words that may be expected in an annual report, investors grow concerned. More weak modal words predict fewer determined investors overall but a higher proportion of *PercentDetermined*. It seems that overall attention is lower for disclosures with less certain diction, but those that do pay attention are likely to pursue more information. This is difficult to understand intuitively, and there may be strong unobserved factors that affect the amount of weak modal words and the two dependent variables.

4.2. XBRL Introduction

As another way of addressing this concern, I consider the effect of XBRL introduction as a quasi-exogenous shock to how sensitive investors should be to textual attributes. XBRL (eXtensible Business Reporting Language) formatting enables accurate identification of accounting items by human and automated users. Its required adoption was staggered from 2011-2013, but many firms in my sample start using XBRL earlier or later than required. My hypothesis is that when firms start using XBRL formatting, the readability and sentiment of their 10-K's text should have a weaker effect on PercentDetermined. Table 5 shows that XBRL introduction reduces PercentDetermined by 3.9 percentage points. However, after companies start issuing 10-Ks with XBRL, the three measures of readability have about the same effect on PercentDetermined (only Bog Index is shown here, but the coefficients are similar). On the other hand, it seems that XBRL accentuates investors' sensitivity to positive sentiment and dampens their responsiveness to negative sentiment. XBRL was meant to improve standardization and ease of use of firm disclosures. My results suggest that XBRL serves to highlight some of the differences in textual style.

Next, I separate the sample into large and small firms, based on the monthly NYSE size breakpoints from Ken French's website. Small firms are below the 30th percentile of Market Equity, and large firms are above the 70th percentile. I run the same OLS regressions on these two samples and compare their coefficients in the two panels of Table 6. The results show that readability, measured by Bog Index, has a similar impact for each sample, but sentiment affects these firms differently. For small firms, investors are less sensitive to positive text in the 10-K. However, they are much more likely to pursue follow-up research when small firms use negative words compared to large firms. In fact, the coefficient for N_Negative is negative in the sample of large firms. This result suggests that when a small firm uses negative words in disclosures, investors are much more interested in learning more about the company. They might suspect that the firm's stock price will drop, and they seek more information to predict if the dip will reverse. This opportunity should be more profitable for small firms, which have a weaker information environment. In Grossman and Stiglitz terms, the observable portion of stock returns has a higher volatility for these firms. Investors earn a higher utility reward for learning more about these companies.

The same should be true for firms with few analysts following. With fewer individuals helping to process the firm's information, the benefit to becoming informed should be greater. In untabulated results, I separately analyze the top and bottom terciles of analyst following. The results for this breakdown are exactly analogous to those using firm size except for the coefficient on readability. For firms with fewer analysts, *PercentDetermined* is more sensitive to readability. This suggests that when the information environment is weaker, investors have a stronger incentive to follow up on a confusing 10-K.

4.3. Follow-Up Attention and Confusion measures

Next, I examine how follow-up research affects firm outcomes. Specifically, I study if increased follow-up attention mitigates the confusion caused by low readability in the 10-K. Loughran and McDonald 2014 and Bonsall et al 2017 test their measures of readability on stock price volatility, analyst accuracy, and analyst dispersion. My hypothesis is that when investors conduct more follow-up research (*PercentDetermined* is higher), the effect of readability on these "confusion" outcomes will be attenuated. I estimate the following equation using OLS:

$$Confusion_{it} = \beta_0 + \beta_1 * Readability_{it} + \beta_2 * PercentDetermined_{it} + \beta_3 * Readability * PercentDetermined + \beta_4 * Controls_{it} + FEs_{it} + \epsilon_{it},$$

where $Confusion_{it}$ is either stock price RMSE, absolute unexpected earnings, or analyst dispersion, and $Readability_{it}$ is either Bog Index, Log(Gross File Size), or Log(Words). The coefficient of interest is β_3, representing the interaction of readability and *PercentDetermined*. I expect this coefficient to be negative to reflect that the effect low readability (higher values of readability measures) can be mitigated by greater follow-up investor attention. Columns 3, 5, and 7 in Table 7 confirm the results from the literature that low readability predicts high confusion in the future. That is, β_1 is positive and statistically significant in the above equation for all nine pairs of Confusion and Readability measures. However, β_3 is positive for eight of the nine pairings and significantly so (at least at the 10% level) for six of these. It is likely that both readability and *PercentDetermined* correlate with another dimension of firm uncertainty that is not captured by control variables, like absolute filing day return. Investors follow up more often for these disclosures and the filing tends to be more difficult to read. If I could control for this mystery uncertainty factor and still saw the same results for β_3, then I might conclude that follow up research actually amplifies the effect of low readability. That is, investors and analysts become even more confused by searching the firm's other SEC filings. For now, isolating the interaction of readability and *PercentDetermined* remains an unsolved empirical issue.

5. FINANCIAL PROFESSIONALS

In this section and the next, I study how the heterogeneity of investors affects how they respond to textual style. I design a regular expression to identify EDGAR users who are likely financial firms. The script identifies words such as "financial", "bank", and "advisor". I also manually code in large financial companies like Wells Fargo and Goldman Sachs. As a result, I mark 6.2% of the EGDAR activity with ARIN data as "finance professionals".

In Table 8, I use this sample to retest the effect of 10-K text style on *PercentDetermined*. My hypothesis is that these sophisticated investors will be more sensitive to text attributes. They should have a greater ability to sift through the company's other SEC filings and understand the truth behind the firm's condition. Low readability should make them even more interested in learning about the firm because their cost may be lower and their signal of the stock's return may be more precise. However, the results are mostly indistinguishable from the main sample. This may reflect that typical EDGAR users conduct follow-up research in competent and effective ways that mirror true professionals. On the other hand, average investors may overestimate their own ability to uncover the truth about firms.

My second hypothesis is that follow-up research from the most sophisticated EDGAR users will be more effective at mitigating the effects of low readability. I estimate the same regression equation used with the full sample. In seven of the nine tests shown in Table 9, the interaction of readability and *PercentDetermined* has a positive effect on confusion measures. However, the coefficients are generally smaller in magnitude than those for the full sample, and none of them are statistically distinguishable from zero. This could reflect that the professional investors that I identify do have some mitigating effect on low readability. The evidence is quite weak, however. At the least, my results do not show that profession investors make confusion worse, as the full sample does.

6. OWNERS VERSUS NON-OWNERS

Are owners of a company's stock more sensitive to readability and sentiment in firm disclosures? I match IP addresses in the EDGAR log files to 13-F filers using the ARIN WhoWas data. First, I deanonymize the fourth octet of each IP address using Chen et al 2020's cipher table. Then, I search the full IP addresses that have downloaded at least 100 10-Ks in ARIN's WhoWas database. The search results provide the name, address, and phone number of each organization. I match some of these to the information on 13-F filers on EDGAR, manually verifying the names. The majority of my matches were made using the Python package fuzzywuzzy on the organization names. I use this package for each 13-F filer to identify the five ARIN organizations with the most similar names. Then, I manually selected the fuzzywuzzy matches that were obviously correct. Overall, I found ARIN matches for 379 unique companies, identified by CIK in the WRDS database. 111 of these companies had multiple organization matches in ARIN's WhoWas data. Finally, for each of the investor-10-K pairs, I merge in each 13-F filer's holdings data at the most recent quarter-end before the 10-K filing and two quarter-ends after. In summary, I end up with 27,019 investor-10-K observations which include an indicator equal to one if they followed up, textual attributes of the 10-K, and other control variables used in previous regressions.

I examine why these 379 companies choose to follow up after viewing a 10-K or not. I find that 29.4% of the owners follow up, while only 21.6% of non-owners follow up, but what accounts for the variation within these groups? Do they make the decision to follow up on the 10-K in similar ways? I estimate the following model using logit:

$$1(FollowedUp_{ift}) \ = \beta_0 * Constant \ + \beta_1 * TextAttributes_{ft} \ + \beta_2 * Controls_{ft} \ + \epsilon_{ft}$$

where $1(FollowedUp_{ift})$ is an indicator equal to one if investor *i* pursued further information for firm *f* at time *t*, $TextAttributes_{ft}$ is either Bog Index, N_Negative, N_Positive, N_ModalWeak, or N_ModalStrong for firm *f* at time *t*, and the controls and fixed effects are the same as in previous regressions.

Table 10 displays the results for two separate samples: Panel A describes investors that did not own the company's stock in the quarter before the 10-K, and Panel B describes those that did. In Panel A, Bog Index and N_Negative are both positive predictors of 1(determined) and are significant at the 1% level. A one standard deviation increase in N_Negative (.373) raises the probability of an investor following up by 5.81 percentage points. A one standard deviation increase in Bog Index (6.56) leads to a 5.73 percentage points higher chance of following up. In contrast, Panel B shows that investors that already own the company's stock are not sensitive to these text attributes. The coefficients are still positive, but they are much smaller in magnitude and are not significant at the 10% level. This suggests that investors that already own the stock are content with their information set regardless of the textual style of disclosures. Both owners and non-owners of the stock are more likely to follow up when the 10-K uses fewer positive words. Perhaps owners do feel concern when the company is not using the expected optimistic platitudes.

I also analyze whether following up on a 10-K predicts trading in the stock. I use logit to estimate the model:

$$1(HoldingGain_{ift}) =$$

$$\beta_0 * Constant + \beta_1 * 1(FollowedUp)_{ft} + \beta_2 * TextAttributes_{ft} + \beta_3 * Controls_{ft} + \epsilon_{ft},$$

where $1(HoldingGain_{ift})$ is an indicator equal to one if investor i increased their stock holdings in firm *f* from quarters $[t-1, t+2]$, and the other variables are as in previous regressions. In the results displayed in Table 11, I find that following up on the 10-K increases the probability of owners increasing their holdings by 10.1 percentage points, which is significant at the 10% level. Bog Index, N_Positive, and N_ModalWeak predict a greater chance of increasing holdings, while N_Negative and N_Modal_Strong predict a lower chance. However, these coefficients have large standard errors and are not significant at the 10% level. This shows that while investors that own the firm's stock are less sensitive to readability and sentiment in the 10-K, they will follow up when considering increasing their holdings. Future research may be interested in this link between investor information acquisition and portfolio decisions.

7. CONCLUSION

In this paper, I show that the style of text in firm disclosures has a significant impact on investors' subsequent research patterns. When firms communicate in obscure ways, use more negative tone, and use more weak and strong modal words, investors tend to seek more information in the company's other SEC filings. This effect is stronger when the firm's information environment is weaker. Prior literature shows that low readability leads to confusion around the firm. I find no evidence that increased scrutiny from EDGAR users reduces these confusion effects, even when the extra research comes from sophisticated EDGAR users.

I also find that owners of the company's stock are less sensitive than non-owners to textual attributes. For example, readability of the 10-K has no significant effect on owner's propensity to conduct follow-up research, while non-owners are more likely to seek more information. This may be caused by owners being complacent with their information set and thinking they already know enough about the firm.

To my knowledge, this paper is the first to document how textual style affects how investors allocate their attention. This question has been impossible to answer before the public release of the EDGAR server log data set. My methodology of analyzing the sequence of individual activity on EDGAR introduces a novel way to study investor learning. It allows researchers to study why investors allocate their attention in certain ways. Future research should study the sequence of EDGAR usage to continue shedding light on the black box of investor information acquisition. This topic is crucial to understanding the efficiency of market prices.

APPENDICES

APPENDIX A: Figure and Tables

Figure 1
Timeline of Follow-up Activity

This figure shows the timeline of follow-up activity. First, an EDGAR user views a recently filed 10-K. Then, within 6 trading days of the 10-K's filing, they view another of the company's SEC filings on EDGAR.

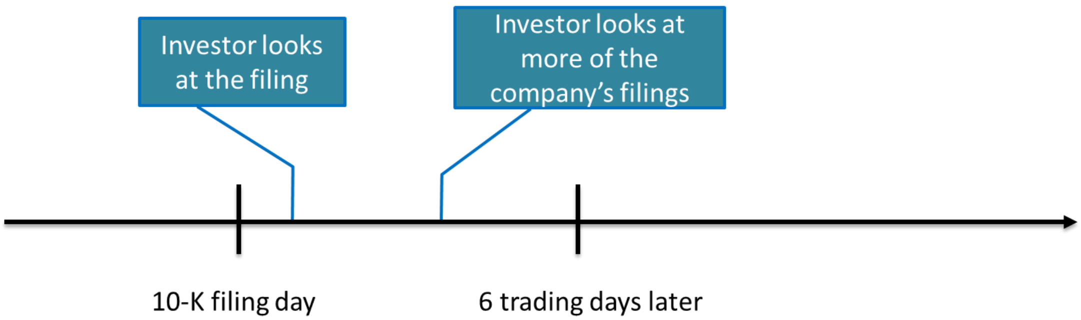


Table 1
Summary Statistics

This table displays summary statistics for the main sample from January 1st, 2003 to March 31st, 2017. *Alpha[-252,-6]*, *RMSE[-252,-6]*, and *Return[0,1]* are multiplied by 100 to represent percentage points. *File Size*, *N_Words*, and *Market Equity* are log-transformed when used in regressions. See the appendix for detailed descriptions of each variable.

	Mean	SD	Min	Max	N
Follow-up activity					
Determined Investors	13	39	0	6,121	33,793
PercentDetermined	17	9	0	69	33,793
Textual attributes					
Bog Index	86	7	54	131	33,590
File Size (bytes)	7,719,880	8,461,760	428,409	28,438,544	33,793
N_Words	50,279	24,678	20,072	113,306	33,793
N_Negative	1.70	0.40	0.46	8.84	33,793
N_Positive	0.72	0.17	0.13	1.94	33,793
N_ModalWeak	0.57	0.20	0.05	1.71	33,793
N_Modal_Strong	0.29	0.10	0.06	1.72	33,793
Control variables					
Alpha[-252,-6]	0.04	0.18	-1.18	4.81	33,793
RMSE[-252,-6]	2.51	1.51	0.31	56.97	33,793
Return[0,1]	-0.05	4.72	-60.91	124.64	33,793
Market Equity ($ thousands)	4,899.51	19,128.60	2.48	638,654.19	33,793
Analysts	9.44	8.51	0.00	65.00	33,793
Segments	2.44	1.74	1.00	11.00	33,793
Filings[-6,6]	7.74	8.37	1.00	228.00	33,793

Table 2
Readability and Follow-up Research

These OLS regression results show the effects of 10-K readability and firm controls on EDGAR follow-up research. In Panel A, the dependent variable is the logged number of "determined investors", measured by the number of IP addresses that view the 10-K and then another of the company's EDGAR filings within 6 trading days of the 10-K's release. In Panel B, the dependent variable is the percentage of "determined users" out of all "interested users". See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

Panel A: *Log(IPs determined)*	(1)	(2)	(3)	(4)	(5)
Bog Index		0.003			−0.000
		(0.002)			(0.002)
Log(Gross file size)			0.083***		0.049***
			(0.015)		(0.017)
Log(Words)				0.154***	0.129***
				(0.026)	(0.027)
Alpha[-252,-6]	−0.297***	−0.294***	−0.285***	−0.268***	−0.267***
	(0.066)	(0.066)	(0.066)	(0.067)	(0.067)
RMSE[-252,-6]	0.087***	0.085***	0.084***	0.078***	0.078***
	(0.015)	(0.014)	(0.015)	(0.014)	(0.014)
Abs(return[0,1])	0.022***	0.021***	0.021***	0.021***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Log(size)	0.212***	0.210***	0.198***	0.196***	0.191***
	(0.014)	(0.014)	(0.015)	(0.015)	(0.016)
Log(BM)	0.081***	0.080***	0.073***	0.072***	0.069***
	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)
Nasdaq	−0.074***	−0.073***	−0.070***	−0.069***	−0.065***
	(0.019)	(0.019)	(0.019)	(0.018)	(0.019)
Analysts	0.014***	0.014***	0.014***	0.014***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Segments	0.021***	0.020***	0.018***	0.017***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Log(Filings[-6,6])	0.093***	0.093***	0.089***	0.087***	0.087***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Constant	−0.290***	−0.502***	−0.187	−1.758***	−1.454***
	(0.112)	(0.170)	(0.121)	(0.251)	(0.320)
Observations	33,793	33,590	33,793	33,793	33,590

Table 2 (cont'd)

Panel B: *PercentDetermined*	(1)	(2)	(3)	(4)	(5)
Bog Index		0.082***			0.055***
		(0.015)			(0.014)
Log(Gross file size)			−0.057		−0.577**
			(0.310)		(0.272)
Log(Words)				1.645***	1.750***
				(0.185)	(0.158)
Alpha[-252,-6]	−2.051***	−1.913***	−2.059***	−1.742***	−1.705***
	(0.562)	(0.576)	(0.577)	(0.560)	(0.586)
RMSE[-252,-6]	0.538***	0.496***	0.539***	0.439***	0.423***
	(0.104)	(0.105)	(0.106)	(0.097)	(0.100)
Abs(return[0,1])	0.176***	0.172***	0.176***	0.169***	0.167***
	(0.024)	(0.025)	(0.025)	(0.024)	(0.024)
Log(size)	0.429***	0.373***	0.438***	0.260***	0.306***
	(0.089)	(0.094)	(0.096)	(0.093)	(0.099)
Log(BM)	0.056	0.014	0.062	−0.041	−0.018
	(0.096)	(0.096)	(0.097)	(0.101)	(0.099)
Nasdaq	−0.444**	−0.459**	−0.447**	−0.386**	−0.419**
	(0.189)	(0.196)	(0.184)	(0.186)	(0.189)
Analysts	−0.024	−0.022	−0.024	−0.026	−0.024
	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)
Segments	0.003	−0.027	0.005	−0.034	−0.035
	(0.043)	(0.040)	(0.042)	(0.042)	(0.040)
Log(Filings[-6,6])	1.068***	1.058***	1.070***	1.011***	1.022***
	(0.120)	(0.123)	(0.118)	(0.122)	(0.123)
Constant	7.208***	0.806	7.137***	−8.492***	−14.506***
	(0.831)	(1.225)	(0.906)	(1.868)	(1.967)
Observations	33,793	33,590	33,793	33,793	33,590

Table 3
Determinants of Readability

This table shows the determinants of readability using three measures. Each column shows the result of an OLS regression of readability on firm characteristics. See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

	(1) Bog Index	(2) Log(Gross file size)	(3) Log(Words)
Alpha[-252,-6]	-1.482^{***}	-0.142^{***}	-0.188^{***}
	(0.410)	(0.036)	(0.019)
RMSE[-252,-6]	0.514^{***}	0.030^{***}	0.060^{***}
	(0.086)	(0.010)	(0.007)
Abs(return[0,1])	0.051^{***}	0.002^{**}	0.005^{***}
	(0.010)	(0.001)	(0.001)
Log(size)	0.717^{***}	0.161^{***}	0.102^{***}
	(0.128)	(0.016)	(0.008)
Log(BM)	0.512^{***}	0.099^{***}	0.059^{***}
	(0.167)	(0.011)	(0.010)
Nasdaq	0.308	-0.050^{**}	-0.035^{**}
	(0.309)	(0.025)	(0.017)
Analysts	-0.025	0.000	0.001
	(0.017)	(0.002)	(0.002)
Segments	0.322^{***}	0.039^{***}	0.023^{***}
	(0.056)	(0.004)	(0.003)
Log(Filings[-6,6])	0.142^{**}	0.038^{***}	0.035^{***}
	(0.072)	(0.006)	(0.004)
Constant	76.770^{***}	-1.233^{***}	9.545^{***}
	(0.723)	(0.112)	(0.042)
Observations	33,590	33,793	33,793

Table 4
Other Text Attributes and Follow-up Research

This table shows how textual attributes of the 10-K affect follow-up research. In Panel A, the dependent variable is the logged number of "determined investors", measured by the number of IP addresses that view the 10-K and then another of the company's EDGAR filings within 6 trading days of the 10-K's release. In Panel B, the dependent variable is the percentage of "determined users" out of all "interested users". See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions include the same control variables as in Table 2.

Panel A: *Log(IPs determined)*

	(1)	(2)	(3)	(4)	(5)	(6)
Bog Index	0.003					0.003
	(0.002)					(0.002)
N_Negative		0.085***				0.107***
		(0.029)				(0.029)
N_Positive			−0.208***			−0.185***
			(0.061)			(0.062)
N_ModalWeak				−0.148**		−0.230***
				(0.060)		(0.060)
N_Modal_Strong					0.141**	0.188***
					(0.057)	(0.037)
Constant	−0.502***	−0.393***	−0.170	−0.229*	−0.327***	−0.471***
	(0.170)	(0.126)	(0.132)	(0.126)	(0.122)	(0.166)
Observations	33,590	33,793	33,793	33,793	33,793	33,590

Panel B: *PercentDetermined*

	(1)	(2)	(3)	(4)	(5)	(6)
Bog Index	0.082***					0.071***
	(0.015)					(0.015)
N_Negative		1.075***				0.671***
		(0.201)				(0.179)
N_Positive			−1.711***			−2.173***
			(0.613)			(0.733)
N_ModalWeak				2.068***		1.437***
				(0.545)		(0.550)
N_Modal_Strong					2.129***	1.518***
					(0.663)	(0.529)
Constant	0.806	5.905***	8.189***	6.363***	6.637***	1.117
	(1.225)	(0.820)	(1.044)	(0.805)	(0.831)	(1.357)
Observations	33,590	33,793	33,793	33,793	33,793	33,590

Table 5
XBRL Introduction and Sensitivity to Textual Style

This table shows how the introduction of XBRL formatting affects how sensitive investors are to textual attributes. The dependent variable is the percentage of "determined users" out of all "interested users". See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

	PercentDetermined					
	(1)	(2)	(3)	(4)	(5)	(6)
XBRL	−3.873***	−6.025***	−3.845***	−3.865***	−3.818***	−3.856***
	(0.572)	(1.728)	(0.570)	(0.563)	(0.560)	(0.573)
Bog Index		0.070***				
		(0.016)				
bogindex * XBRL		0.025				
		(0.017)				
N_Negative			1.438***			
			(0.232)			
N_Negative * XBRL			−0.043**			
			(0.019)			
N_Positive				−0.709		
				(0.780)		
N_Positive * XBRL				−0.112**		
				(0.054)		
N_ModalWeak					1.470**	
					(0.636)	
N_ModalWeak * XBRL					0.056	
					(0.045)	
N_Modal_Strong						1.282
						(0.921)
N_Modal_Strong * XBRL						0.085
						(0.061)
Constant	6.954***	1.575	5.134***	7.286***	6.327***	6.592***
	(0.807)	(1.126)	(0.811)	(1.129)	(0.794)	(0.725)
Observations	33,793	33,590	33,793	33,793	33,793	33,793

Table 6
Effects of Text Style in Small v.s. Large Firms

This table shows how EDGAR users' sensitivity to 10-K text varies between small and large firms. The dependent variable is the percentage of "determined users" out of all "interested users". See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

Panel A: Small firms

	PercentDetermined				
	(1)	(2)	(3)	(4)	(5)
Bog Index	0.052^{***}				
	(0.017)				
N_Negative		1.113^{***}			
		(0.255)			
N_Positive			-1.148		
			(0.741)		
N_ModalWeak				1.866^{**}	
				(0.751)	
N_Modal_Strong					2.163^{**}
					(0.943)
Constant	2.554^{**}	5.333^{***}	7.203^{***}	5.995^{***}	6.004^{***}
	(1.131)	(0.960)	(1.229)	(0.960)	(0.962)
Observations	18,183	18,295	18,295	18,295	18,295

Panel B: Large firms

	(1)	(2)	(3)	(4)	(5)
Bog Index	0.064^{***}				
	(0.024)				
N_Negative		-0.612^{*}			
		(0.366)			
N_Positive			-3.107^{***}		
			(0.785)		
N_ModalWeak				-0.222	
				(0.764)	
N_Modal_Strong					1.776
					(1.473)
Constant	6.250^{***}	12.531^{***}	13.663^{***}	11.832^{***}	11.211^{***}
	(1.961)	(1.643)	(1.685)	(1.528)	(1.604)
Observations	5,589	5,616	5,616	5,616	5,616

Table 7
Follow-up Research Does not Reduce Uncertainty

This table shows how follow-up research from determined EDGAR users affects confusion about the firm. The dependent variables in Panels A, B, and C are *RMSE[6,28]*, *abs(SUE)*, and *Analyst Dispersion*. See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

Panel A: RMSE[6,28]	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent determined		0.004^{***}		0.010		0.003^{***}		-0.038^{***}
		(0.001)		(0.008)		(0.001)		(0.012)
Bog Index			0.008^{***}	0.009^{***}				
			(0.002)	(0.003)				
bogindex * perDet				-0.000				
				(0.000)				
Log(Gross file size)					0.045^{***}	0.028		
					(0.016)	(0.018)		
logfilesize * perDet						0.001		
						(0.001)		
Log(words)							0.138^{***}	0.064^{**}
							(0.027)	(0.031)
logwords * perDet								0.004^{***}
								(0.001)
Observations	33,793	33,793	33,590	33,590	33,793	33,793	33,793	33,793

Table 7 (cont'd)

Panel B: abs(SUE)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent determined		0.028***		−0.072**		0.025***		−0.165
		(0.006)		(0.030)		(0.009)		(0.102)
Bog Index			0.048***	0.026**				
			(0.012)	(0.012)				
bogindex * perDet				0.001***				
				(0.000)				
Log(Gross file size)					0.431***	0.383**		
					(0.141)	(0.184)		
logfilesize * perDet						0.003		
						(0.004)		
Log(words)							0.994***	0.653***
							(0.180)	(0.244)
logwords * perDet								0.018*
								(0.009)
Observations	29,315	29,315	29,150	29,150	29,315	29,315	29,315	29,315

Table 7 (cont'd)

Panel C: Analyst Dispersion	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent determined		0.002***		−0.003		0.001***		−0.011***
		(0.000)		(0.002)		(0.000)		(0.003)
Bog Index			0.003***	0.002**				
			(0.001)	(0.001)				
bogindex * perDet				0.000**				
				(0.000)				
Log(Gross file size)					0.016***	0.012*		
					(0.005)	(0.006)		
logfilesize * perDet						0.000**		
						(0.000)		
Log(words)							0.055***	0.032*
							(0.013)	(0.016)
logwords * perDet								0.001***
								(0.000)
Observations	28,125	28,125	27,962	27,962	28,125	28,125	28,125	28,125

Table 8
Follow-up Activity by Professionals

This table shows how textual attributes of the 10-K affect follow-up research by "financial professionals". To identify "financial professionals", I merge in organization names of EDGAR users from ARIN based on their IP address in the log files. Then, I use a regular expression to identify the organization names that contain the terms "financial", "advisor", etc. and separately mark some large financial institutions, like Wells Fargo. The dependent variable is the percentage of "determined users" out of all "interested users" using the subset of "financial professionals". See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

	PercentDetermined					
	(1)	(2)	(3)	(4)	(5)	(6)
Textual Attributes:						
Bog Index	0.105^{***}					0.095^{***}
	(0.025)					(0.023)
N_Negative		0.802^{**}				0.425
		(0.319)				(0.284)
N_Positive			-2.160^{***}			-2.409^{***}
			(0.726)			(0.852)
N_ModalWeak				1.746^{**}		1.265
				(0.803)		(0.776)
N_Modal_Strong					2.876^{***}	2.549^{***}
					(1.047)	(0.961)
Observations	13,217	13,282	13,282	13,282	13,282	13,217

Table 9
Professionals Do Not Make Low Readability Worse

This table shows how follow-up research from determined "financial professionals" affects confusion about the firm. To identify "financial professionals", I merge in organization names of EDGAR users from ARIN based on their IP address in the log files. Then, I use a regular expression to identify the organization names that contain the terms "financial", "advisor", etc. and separately mark some large financial institutions, like Wells Fargo. The dependent variables in Panels A, B, and C are *RMSE[6,28]*, *abs(SUE)*, and *Analyst Dispersion*. See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

Panel A: RMSE[6,28]	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Percent determined	0.006***		−0.002		0.003*		0.024
	(0.002)		(0.010)		(0.002)		(0.025)
Bog Index		0.004	0.002				
		(0.003)	(0.004)				
bogindex_perDet			0.000				
			(0.000)				
Log(Gross file size)				0.172***	0.196***		
				(0.033)	(0.059)		
logfilesize_perDet					−0.002		
					(0.002)		
Log(words)						0.168***	0.192***
						(0.032)	(0.058)
logwords_perDet							−0.002
							(0.002)
Constant	2.544***	2.272***	2.399***	2.936***	2.888***	0.949***	0.619
	(0.217)	(0.215)	(0.257)	(0.243)	(0.251)	(0.319)	(0.575)
Observations	13,282	13,217	13,217	13,282	13,282	13,282	13,282

Table 9 (cont'd)

Panel B: abs(SUE)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Percent determined	0.022***		0.007		0.025***		−0.046
	(0.004)		(0.054)		(0.009)		(0.067)
Bog Index		0.028***	0.023				
		(0.008)	(0.015)				
bogindex_perDet			0.000				
			(0.001)				
Log(Gross file size)				0.511***	0.376*		
				(0.128)	(0.213)		
logfilesize_perDet					0.005		
					(0.007)		
Log(words)						0.479***	0.328
						(0.123)	(0.205)
logwords_perDet							0.006
							(0.006)
Constant	0.873	−1.160*	−0.993	2.091***	1.688**	−3.617***	−2.257
	(0.555)	(0.632)	(1.041)	(0.740)	(0.792)	(0.870)	(1.778)
Observations	12,376	12,317	12,317	12,376	12,376	12,376	12,376

Panel C: Analyst Dispersion	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Percent determined	0.002***		0.001		0.002***		−0.001
	(0.000)		(0.004)		(0.001)		(0.008)
Bog Index		0.003***	0.003*				
		(0.001)	(0.002)				
bogindex_perDet			0.000				
			(0.000)				
Log(Gross file size)				0.059***	0.050*		
				(0.014)	(0.027)		
logfilesize_perDet					0.000		
					(0.001)		
Log(words)						0.056***	0.048*
						(0.013)	(0.026)
logwords_perDet							0.000
							(0.001)
Constant	−0.153***	−0.396***	−0.391**	−0.021	−0.052	−0.687***	−0.622**
	(0.046)	(0.115)	(0.152)	(0.058)	(0.063)	(0.126)	(0.271)
Observations	12,354	12,292	12,292	12,354	12,354	12,354	12,354

Table 10
Follow-up Activity by Owners v.s. Non-Owners

This table shows how sensitivity to 10-K text varies between EDGAR users that own the stock versus those that do not. The dependent variable is an indicator equal to one if the "interested" EDGAR user proceeded to view another of the firm's EDGAR filings. Panel A shows the logit results for EDGAR users that owned the stock in the quarter before the 10-K's release, and Panel B shows the results for those that did not. See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

Panel A: Non-Owners

	1(Followed up)					
	(1)	(2)	(3)	(4)	(5)	(6)
Bog Index	0.009***					0.008***
	(0.003)					(0.003)
N_Negative		0.160***				0.132**
		(0.046)				(0.054)
N_Positive			−0.373***			−0.378***
			(0.132)			(0.143)
N_ModalWeak				0.147		0.067
				(0.096)		(0.115)
N_Modal_Strong					0.107	0.106
					(0.192)	(0.194)
Constant	−2.692***	−2.274***	−1.710***	−2.119***	−2.025***	−2.634***
	(0.462)	(0.363)	(0.348)	(0.363)	(0.351)	(0.504)
Observations	19,071	19,214	19,214	19,214	19,214	19,071

Panel B: Owners

	(1)	(2)	(3)	(4)	(5)	(6)
Bog Index	0.002					0.000
	(0.005)					(0.005)
N_Negative		0.051				0.055
		(0.082)				(0.082)
N_Positive			−0.379**			−0.376**
			(0.178)			(0.173)
N_ModalWeak				−0.064		−0.072
				(0.207)		(0.205)
N_Modal_Strong					0.112	0.142
					(0.261)	(0.259)
Constant	−1.796***	−1.778***	−1.402**	−1.629***	−1.723***	−1.504**
	(0.618)	(0.562)	(0.566)	(0.521)	(0.566)	(0.620)
Observations	7,791	7,803	7,803	7,803	7,803	7,791

Table 11
Owners Increase Holdings after Following Up

This table shows how following up on a 10-K affects an EDGAR user's propensity to increase their holdings. The dependent variable is an indicator equal to one if the EDGAR user increases their holdings in the firm's stock. See the appendix for detailed descriptions of all variables. Standard errors are in parentheses below each coefficient. All regressions control for industry and year fixed effects.

	1(HoldingGain)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1(Followed up)	0.101*	0.101*	0.102*	0.103*	0.102*	0.102*	0.104*
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
Bog Index		0.004					0.004
		(0.004)					(0.004)
N_Negative			−0.058				−0.104
			(0.079)				(0.084)
N_Positive				0.210			0.204
				(0.167)			(0.162)
N_ModalWeak					0.238		0.308
					(0.177)		(0.192)
N_Modal_Strong						−0.194	−0.259
						(0.220)	(0.219)
Constant	−0.816*	−1.121**	−0.711	−0.971**	−1.015**	−0.748*	−1.271**
	(0.430)	(0.535)	(0.447)	(0.457)	(0.424)	(0.420)	(0.534)
Observations	7,793	7,781	7,793	7,793	7,793	7,793	7,781

APPENDIX B: Variable Definitions

Readability Measures	
Bog Index	Readability rating given by a program called StyleWriter that identifies "plain English" violations, like passive voice and legal jargon; introduced by Bonsall et al (2017); data from Samuel Bonsall's website
Log(Gross file size)	Log of the size of the 10-K filing on EDGAR; data from Bill McDonald's website
Log(Words)	Log of the number of words in the 10-K filing; data from Bill McDonald's website
Attention Measures	
Log(IPs Determined)	Log of the number of IP addresses on EDGAR that view the 10-K and then subsequently view another of the same firm's SEC filings within 6 trading days of the 10-K's filing; recorded from the EDGAR log files (covering January 1, 2003 to May 31, 2017 with the exception of September 24, 2005 through May 10, 2006)
PercentDetermined	Number of IP addresses determined (as described above) divided by the number of IP addresses that viewed the 10-K filing
Sentiment Measures	
N_Negative	Proportion of negative words in the 10-K filing; all word lists are developed in Loughran and McDonald (2011) and the data by 10-K is from Bill McDonald's website; this list includes words such as "failure", "litigation", and "loss"; data from Bill McDonald's website
N_Negative	Proportion of negative words in the 10-K filing; this list includes words such as "failure", "litigation", and "loss"
N_Positive	Proportion of positive words in the 10-K filing; this list includes words such as "achive", "efficient", and "profitable"
N_ModalWeak	Proportion of weak modal words in the 10-K filing; wthis list includes words such as "might", "could", and "possibly"
N_Modal_Strong	Proportion of strong modal words in the 10-K filing; this list includes words such as "failure", "litigation", and "loss"

Variable Definitions (cont'd)

Confusion Measures	
RMSE[6,28]	Root Mean Squared Error from a FF 3-factor model measured from 6 trading days after filing to 28 trading days (40 days) after filing.
Abs(SUE)	$abs(\frac{\textit{mean(EPS estimates) - actual EPS}}{\textit{Stock price on the day before the EA}})$ using estimates made between the 10-K filing date and the next Earnings Announcement (EA). Data are from I/B/E/S. If an analyst makes multiple estimates, then I use the earliest estimate after the 10-K filing date.
Analyst Dispersion	*STDEV(EPS estimates included above)* if there are at least two estimates.

Control Variables	
Alpha[-252,-6]	FF 3-factor alpha from trading days [-252,-6] relative to 10-K filing date if there are at least 60 days of stock returns in CRSP.
RMSE[-252,-6]	FF 3-factor RMSE from trading days [-252,-6] relative to 10-K filing date if there are at least 60 days of stock returns in CRSP.
Return[0,1]	Two-day return from close of the day before the 10-K filing to close of the day after filing. In most regressions, I use the absolute value of this.
Market Equity ($ thousands)	Share price times the number of shares outstanding on the day before the 10-K filing. Data from CRSP.
Analysts	Number of analysts that make earnings estimates on the firm's previous quarter.
Segments	Number of Compustat Business Segments (BUSSEG) for the firm in the 10-K filing year.
Filings[-6,6]	Number of SEC filings that the firm discloses between [-6,6] trading days around the 10-K filing. Data from the EDGAR master index files.

BIBLIOGRAPHY

BIBLIOGRAPHY

Balachandran, B., H. Duong, R. Faff, and S. Shams. 2020. Does the Readability of Target Firms' Annual Reports Matter to Bidders?. *Working Paper*.

Baker, M. and J. Wurgler. Investor Sentiment and the Cross-Section of Stock Returns. 2006. *The Journal of Finance* 61:1645-1680.

Ben-Rephael, A., Z. Da, and R. Israelsen. 2017. It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *The Review of Financial Studies* 30:3009-3047.

Biddle, G., G. Hilary, and R. Verdi. 2009. How Does Financial Reporting Quality Relate to Investment Efficiency?. *Journal of Accounting and Economics* 48:112-131.

Bonsall, S., A. Leone, B. Miller, and K. Rennekamp. 2017. A Plain English Measure of Financial Reporting Readability. *Journal of Accounting and Economics* 63:329-357.

Bodnaruk, A., T. Loughran, and B. McDonald. 2015. Using 10-K Text to Gauge Financial Constraints. *The Journal of Financial and Quantitative Analysis* 50:623-646.

Bushee, B., I. Gow, and D. Taylor 2018. Linguistic Complexity in Firm Disclosures: Obfuscation or Information?. *Journal of Accounting Research* 56:85-121.

Cao, S., W. Jiang, B. Yang, and A. Zhang. 2020. How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI. *Working Paper*.

Chen, H., L. Cohen, U. Gurun, D. Lou, and C. Malloy. 2020. IQ from IP: Simplifying Search In Portfolio Choice. *Journal of Financial Economics* 138:118-137.

Cohen, L., C. Malloy, and Q. Nguyen. 2020. Lazy Prices. *The Journal of Finance* 75:1371-1415.

Crane, A., K. Crotty, and T. Umar. 2020. The Role of Public and Private Information in Fund Performance: Complements or Substitutes?. *Working Paper*.

Da, Z., J. Engelberg, and P. Gao. Using 10-K Text to Gauge Financial Constraints. *The Journal of Financial and Quantitative Analysis* 50:623-646.

Drake, M., P. Quinn, and J. Thornock. Who Uses Financial Statements? A Demographic Analysis of Financial Statement Downloads from EDGAR. *Accounting Horizons* 31:55-68.

DellaVigna, S. and J. Pollett. Investor Inattention and Friday Earnings Announcements. 2009. *The Journal of Finance* 64:709-49.

Drake M., D. Roulstone, and J. Thornock. 2014. The Determinants and Consequences of Information Acquisition via EDGAR. *Contemporary Accounting Research* 32:1128–61.

Ertugrul, M., J. Lei, J. Qiu, and C. Wan. 2020. Annual Report Readability, Tone Ambiguity, and the Cost of Borrowing. *Journal of Financial and Quantitative Analysis* 52:811-836.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal. 2010. Management's Tone Change, Post Earnings Announcement Drift and Accruals. *Review of Accounting Studies* 15:915-953.

Gibbons, B., P. Iliev, and J. Kalodimos. 2021. Analyst Information Acquisition via EDGAR. *Management Science* 2019

Grossman, S. and J. Stiglitz. 1976. Information and Competitive Price Systems. *The American Economic Review* 66:246-253.

Hirshleifer, D., S. S. Lim, and S. H. Teoh. 2009. Limited Investor Attention and Stock Market Misreactions to Accounting Information. *Review of Asset Pricing Studies* 1:35–73.

Kahneman, D.. Attention and Effort. Englewood Cliffs, *NJ: Prentice-Hall*.

Lee, C., P. Ma, and C. Wang. 2015. Search-Based Peer Firms: Aggregating Investor Perceptions through Internet Co-Searches. *Journal of Financial Economics* 116:410-431.

Lehavy, R., F. Li, and K. Merkley. 2011. The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review* 86:1087-1115.

Li, F. 2008. Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics* 45:221-247.

Li, R., X. Wang, Z. Yan, and Y, Zhao 2019.Sophisticated Investor Attention and Market Reaction to Earnings Announcements: Evidence From the SEC's EDGAR Log Files. *Journal of Behavioral Finance* 20:490-503.

Loughran, T., and B. McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *Journal of Finance* 66:35-65.

Loughran, T., and B. McDonald. 2014. Measuring Readability in Financial Disclosures. *Journal of Finance* 69:69.

Pely, D.J. and D.S. Schoch. 2020. Merger Rhetoric and the Credibility of Managerial Synergy Forecasts. *Working Paper*.

Tetlock, P.. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. 2007. *The Journal of Finance* 62:1139-1168.

Tucker, A., Y. Xia, and S. Smelcer. 2020. It Ain't Just What Funds Disclose (It's The Way That They Do It). *Working Paper*.

CHAPTER 2. Insider Trading Plans Not Going to Plan

1. INTRODUCTION

Stock selling plans under SEC Rule 10b5-1 provide corporate insiders with legal protection against allegations of trading on material non-public information (MNPI). The details of these plans are highly unrestricted, however, leaving insiders free to make strategically timed trades while maintaining this legal protection. There are no limits on when trading can start after a plan is made, plans can be amended or terminated without notice, and it is not even required to disclose the existence of plans. This dissertation chapter dials in on the most basic detail of 10b5-1 trading plans: do insiders even sell the number of shares that they plan to? What makes them follow through on their plan exactly as promised? To our knowledge, we are the first to analyze this key feature of 10b5-1 plans. The results help illuminate how insiders use these plans and may help inform lawmakers and courts in insider trading issues. We also use our large sample to update the literature's previous findings about the abuse of 10b5-1 trading plans.

Our first main result is that insiders sell the exact number of shares they promised in only 24% for which we have data. Only 35% of plans in our sample sell between 50% and 150% of the planned shares. 10b5-1 plans are meant to be set in stone, but it seems that insiders exercise some discretion in choosing when to sell their holdings. One possible explanation is that the stock price did not reach the limit price set forth in the plan's documentation. Without the data on these limit prices, we cannot control for them in our analysis. However, given that a substantial 76% of plans are not fully completed, we suspect that insiders arbitrarily decide not to follow through in some of these cases.

To better understand how plans with perfect follow through are different, we compare the variable means for this sample against those of other plans. Plans with exact follow through tend to have shorter cooling-off periods and fewer trades. These features of 10b5-1 plans have been linked to strategic trading by previous literature (recently, Larcker et al 2020). They find that shorter cooling-off periods, measured as the number of days between the plan adoption and the first trade, and fewer trades lead to lower subsequent stock returns. Lower returns after sales mean that insiders timed their trades well to avoid losses. We find that plans with perfect follow through do not have significantly different returns than other plans. However, when we split the sample into single-trade plans and those with longer cooling-off periods, we do see greater loss avoidance in plans with perfect follow through. Our hypothesis was that insiders who choose to complete their 10b5-1 plans exactly as promised would tend to make more strategic trades. They might use perfect follow through as extra cover to argue that they are good actors when they actually time trades based on MNPI. We do not find conclusive evidence of this, but it may be a fruitful area for future research.

What does drive the decision to follow through perfectly, then? We conduct a logit regression to test measures of litigation risk, corporate governance, and stock return volatility as determinants of perfect follow through. We find that insiders tend to follow through perfectly more often when their firm's litigation risk is higher and institutional ownership is lower. It could be that insiders feel the pressure of litigation risk want to show how honest they are by selling the number of shares that they promised. For these results and many others involving our novel *FollowThrough* variable, the interpretations are not totally clear. This exploratory research is valuable to those wishing to understand the unique SEC Rule 10b5-1.

Our next task is to use our large dataset on 10b5-1 plan announcements and insider trading to update some key results of the previous literature. Jagolinzer 2009 finds that 10b5-1 sales avoid 2.2 percentage points greater losses than other sales. While they test data from 2000 to 2005, we extend the analysis to the 2003 to 2020 time period and find similar results. We observe that 10b5-

1 sales precede stock returns 0.96 percentage points lower than other sales. It seems that insiders have become less bold over time and do not use 10b5-1 plans for strategic trades as often.

We also study Henderson et al 2015's finding that announced 10b5-1 sales avoid larger losses than unannounced ones. Their argument is that insiders announce plans publicly as a sort of "offensive disclosure", shielding them from allegations of illegal trading. However, we cannot replicate their results in either the full sample or the sample from 2003 to 2005. It may be that their result is specific to the time period from 2000 to 2003 for which we do not have insider trading data.

Lastly, we test whether 105b-1 plans with shorter cooling-off periods and those with only one trade avoid larger losses. Short cooling-off periods enable insiders to use more recent MNPI to time trades. Plans with only one trade clearly ignore Rule 10b5-1's intention of regular trading schedules. Expanding the starting year of Larcker et al 2020's sample from 2016 to 2003, we verify their result that single-trade plans precede abnormal stock declines in almost every cooling-off period category. While there isn't a consistent trend of loss avoidance across cooling-off periods, the largest loss avoidance is in plans with less than 30 days before the first trade. It is simple for insiders to have MNPI that might drop the stock price, start a 10b5-1 plan, and start selling their stock holdings to avoid the loss in value. In fact, there are several cases where the insider fulfills the entire 10b5-1 plan on the same day the plan is adopted!

There is clearly abuse going on in 10b5-1 plans, and this dissertation chapter helps paint the full picture of what is happening. The original rule enacted in 2000 allows far too much leeway in designing and modifying these plans. The SEC has recently discussed tightening the rule, but it has not yet done so.[9] This chapter's new findings and update to old results may help guide the legislation in this area.

2. LITERATURE REVIEW

Previous literature summarizes 10b5-1 trading plans, focusing on factors related to strategic trade timing. Jagolinzer 2009 finds that sales under 10b5-1 plans tend to precede greater price declines than preplanned trades do. This loss avoidance suggests that insiders take advantage of Rule 10b5-1's legal protection to abuse material non-public information (MNPI) and time their sales just before negative news events. Other papers studying 10b5-1 plans provide evidence for or against these findings, study the mechanisms underlying the results, and attempt to isolate subsamples where loss avoidance is greatest.

Sen 2008 argues that the peaking price patterns found by Jagolinzer 2009 are not evidence of intentional strategic timing but merely reflect the existence of price limits in 10b5-1 plans. They show that testing the phenomenon in specific ways - weighting the average post-trade returns equally for every firm - results in a negative bias even when returns are independent and identically distributed. In the final version of the paper, Jagolinzer 2009 counters this argument by weighting the observations equally instead of the firms, and their results remain unchanged. Fich et al 2018 picks up the debate by showing that the average cumulative abnormal return (CAR) after large 10b5-1 sales by CEOs is 0%. In contrast, non-105b-1 trades precede -8% CAR, suggesting that 10b5-1 plans *prevent* strategic loss avoidance in insider sales. This is a dramatic reversal of

[9] Available at https://www.sec.gov/spotlight/investor-advisory-committee-2012/20210916-10b5-1-recommendation.pdf

Jagolinzer 2009's results, but Fich et al 2018's results may be specific to large CEO sales. We update this loss avoidance result in Section 4.

Some papers seek to explain what drives stock prices around insider trades by studying firms' accounting numbers and news events. Fich et al 2018 show that CEOs may use accounting accruals and news events to manipulate stock prices around their trades, whether a 10b5-1 plan is in place or not. Lee 2020 finds that 10b5-1 sales take advantage of price movements around earnings announcements. Insiders clearly have MNPI during earnings season and may use 10b5-1 protection to take advantage of this, although many firms have "blackout periods" at these times when insiders are not allowed to trade. Shon and Veliotis 2013 show that CEOs and CFOs may even influence the firm's earnings before their 10b5-1 sales. They find that firms are 31% more likely to meet or beat earnings expectations in quarters preceding 10b5-1 sales, suggesting that executives use some discretionary accounting to nudge earnings just barely to or above expectations. On the other side, Lee 2020 argues that there is no evidence of earnings management, even when 10b5-1 trading plans are in place.

Mitts 2020 shows that 10b5-1 sales are much more likely on days when good news is released, suggesting that insiders either time their trades or time the news release. This evidence, however, is consistent with the presence of limit prices in 10b5-1 plans. When positive news is released, prices tend to increase and may reach the limit prices.[10]

This dissertation chapter adds new results and updates the findings of the literature that identifies subsets of 10b5-1 plans that are more strategic than others. Milian 2016 shows that plans

[10] Much of the literature points out that limit orders in 10b5-1 plans may drive some conclusions. It may be a focus of future research to gather this data for any plan announcements that describe it. Legislature should perhaps consider requiring its disclosure to ensure that trading in 10b5-1 plans is automatic and not up to the insiders' discretion.

with fewer trades and shorter duration avoid greater losses. These plans seem haphazard and perhaps a lazy attempt to gain 10b5-1 legal protection. Larcker et al 2020 finds similar results for a more recent sample (2016-2020) and breaks down 10b5-1 plan profitability by cooling-off period and single-trade plans. Plans with trades less than 60 days after plan adoption and those with only one trade avoid much larger losses. Jagolinzer et al 2011 finds that plans that did not require approval of a firm's general counsel tend to avoid greater losses.

The new results of this dissertation chapter add to those of Henderson et al 2015, who find that when 10b5-1 plans are announced, insiders make much more strategic trades. They reveal that loss avoidance is greatest for 10b5-1 plans with more specific announcements. This suggests that insiders make extra "offensive disclosure" to cover their tracks when they intend to make loss-avoiding strategic trades. We extend this analysis to insiders that follow through completely on their 10b5-1 plan's promises, hoping to find evidence of an "offensive obedience" – insiders cover their tracks for strategic 10b5-1 sales by obeying their plans perfectly. They work hard to play the part of an honest insider.

3. DATA AND SAMPLE CREATION

We obtain data on stock returns from CRSP and institutional ownership from Thomson Reuters 13-F database. Industry classification comes from matching the SIC codes attached to the EDGAR filings with the 12 categories described on Kenneth French's data repository.[11] We retrieve data on civil lawsuits matched with the SEC's company identifier, CIK, from the Federal Judicial Center's (FJC) Integrated Database, accessed through Wharton Research Data Services

[11] Available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

(WRDS). Data on 10b5-1 plan announcements and insider trades come from the SEC EDGAR website as described below.

3.1. 10b5-1 Plan Announcements

Our sample of 10b5-1 plan announcements starts with all 3,903 8-K filings from January 1st, 2003 to May 30th, 2020 that mention the term "10b5-1" and not "repurchase".[12] We run a Python script to parse these documents, identifying the insiders who adopted plans, the adoption date and end date of each plan, and the number of shares they promise to sell.[13]

While the automated code works well, we manually verify the data for the first 601 8-K filings (about a sixth of the sample), sorted by company CIK. From these documents, we identify 827 plans and find the number of shares for 595 of these. Figure 2 displays an example of an 8-K filing that details the start date, end date, and number of shares promised. The 595 plans that we find this data for is our sample to study *FollowThrough* on 10b5-1 plans. We use the full sample of 8-K announcements to study the literature's previous results on 10b5-1 plans.

3.2. Insider Common Stock Sales

We obtain data on all insider stock sales from January 1st, 2003 to May 30th, 2020 directly from Form 4 filings on the SEC EDGAR website. We use Python script to process these documents and retrieve the dates of the sales, the SEC's Central Identification Key (CIK) of the issuer company, the CIK of the insider, the number of shares sold, the price of the transaction, and the

---

[12] Rule 10b5-1 can also be used to protect firms' stock repurchase plans. Our study, however, focuses on insider trading, so we do not consider 8-K filings describing repurchase programs.

[13] When the 8-K does not explicitly state the adoption date of the plan, we use the filing date of the 8-K. When they do not describe the plan's end date, we fill this in as one year after the adoption date. One year is the most frequent plan duration, and much of the literature, including Jagolinzer 2009, uses this method to identify all plan end dates.

footnotes. If one of the footnotes of a filing mentions the term “10b5-1”, then we mark all of the filing’s sales as part of a plan. This method corrects some cases where the footnote that describes a 10b5-1 plan is only applied to one sale but states, “All of the sales in this form are made pursuant to a 10b5-1 plan.”

For many of these footnotes describing planned sales, we also extract the plan’s adoption date. Some of the Form 4 filings describe sales on multiple days and there is sometimes overlap with other filings, so we aggregate the number of shares sold and the total dollar value to the company-insider-day-10b5 level, referring to the observations as “sales”. We start with 1,031,963 of these observations. After requiring CRSP data, the sample is reduced to 962,297 sales. Finally, we reduce the sample to those companies that would be included in the manually checked 8-K sample, i.e., the company must have a CIK less than 350868. The final sample contains 101,026 insider sales.

3.3. Calculating Follow Through

To measure *FollowThrough*, we need to combine data on planned sales with the number of shares the insider ends up selling. We use two methods to match the 10b5-1 plans with Form 4 trades. First, we match the Form 4 filings with footnotes that explicitly state a 10b5-1 plan adoption date. Figure 3 shows an example of this type of document; footnote 2 references the 8-K plan announcement shown in Figure 2. Second, to account for documents that do not have footnotes listing the plan adoption date, we match every trade that occurs between the 105b-1 plans’ start and end dates. In some cases, insiders make additional sales within these time periods that are not part of trading plans, so we mistakenly measure extra sales. This perhaps represents an even more severe abuse of 10b5-1 plans because insiders may be arbitrarily deciding which trades need extra legal protection from Rule 10b5-1. That scenario is far from the SEC's vision of strict robotic trading plans.

To calculate *FollowThrough*, we simply divide the number of shares traded by the number of shares planned to be sold. We adjust for stock splits using the CFACPR variable in CRSP. If the insider sold fewer shares than promised, then *FollowThrough* will be between zero and one. If the insider sold the exact number of shares promised, then *FollowThrough* will be one. In this project, we mainly study *1[FollowPerfect]*, a variable equal to one if *FollowThrough* equals one for a plan and zero otherwise.

4. RESULTS

4.1. New Findings: Follow Through on 10b5-1 Plans

A first glance at our *FollowThrough* data makes it clear that 10b5-1 plans are not set in stone. Throughout all 595 trading plans in our sample, the average *FollowThrough* is 0.387, representing that 38.7% of promised shares were sold on average. Only 24% of announced plans in our sample sold exactly the number of shares that they promised, i.e., *FollowThrough* equal to one. 35% of plans sold between 50% and 150% of the promised shares. On the other hand, a whopping 53% of announced 105b-1 plans do not have any sales in their lifetime.

There are a few leading reasons why three quarters of 10b5-1 plans have *FollowThrough* not equal to one, listed here in order of most to least innocuous. First, we may have identified planned trades incorrectly. While this may be our mistake, it demonstrates that some 10b5-1 plan announcement are not described clearly. Second, the stock price may not have reached the limit prices throughout the plan, so *FollowThrough* is less than one. Most of the plan announcements in our sample mention that shares will be sold "subject to minimum price thresholds". Unfortunately, only a few of these 8-K filings list the exact limit prices, so we cannot control for that information in our analysis. Third, the insider may have amended or terminated the plan. The SEC allows 10b5-1 plans to be adjusted at will and does not require any public announcement. However, this activity seems to be far from the spirit of Rule 10b5-1's strict trading plans. The fourth and most suspicious reason is that insiders might simply decide to deviate from their plans, either selling fewer shares than planned or extra.

Figure 4 displays the trend of *1[FollowPerfect]* from 2004 to 2019. Early 10b5-1 plans averaged less than 20% with perfect *FollowThrough*, hitting a low in 2009. The 7.9% *FollowThrough* in the Great Recession can reasonably be explained by stocks not reaching the plans' limit prices. After this time, academic literature and mainstream media increased their scrutiny of 10b5-1 plans, and we see perfect *FollowThrough* stay above 25% from 2011 to 2019 with a peak of 50% in 2018.

Using the same variables that previous literature uses to summarize 105b-1 plans, we compare plans that follow through perfectly with those that do not. Table 12 shows the summary statistics for the two subsamples. It is important to note that the means are calculated by plan, so companies with more plans have greater weight. 105b-1 plans that have perfect *FollowThrough* tend to have shorter cooling-off periods (57 days versus 73 days), fewer trades (3.72 versus 5.67), and larger trade size ($1,827,726 versus $877,741). These variable differences suggest that perfect *FollowThrough* might correspond with abuse of 10b5-1 plans. Short cooling-off periods, measured as the time between plan adoption and the first trade, mean that insiders are more likely to have *material* non-public information when they start trading. Fewer trades allow insiders to pick the most opportune days to sell stock. Larger trade sizes mean that insiders can avoid greater losses. 10b5-1 plans with perfect *FollowThrough* also have lower institutional ownership, higher litigation risk (discussed below), and a greater chance of being sued in the following three years (also discussed below). However, we do not find a significant difference in stock returns after 10b5-1 sales. Figure 5 shows the average returns around 10b5-1 sales broken down by *1[FollowPerfect]*. Note that these averages are calculated by sale, not by plan. We see the same runup in the 90 days before sales that previous literature has documented, but we do not see subsequent declines and loss avoidance. In general, stock prices seem to continue rising in the 90 days after sales whether 10b5-1 plans were followed exactly or not.

Is there a subset of 10b5-1 plans for which perfect *FollowThrough* corresponds with greater loss avoidance? Previous literature, such as Larcker et al 2020, finds that plans with only a single trade and those with shorter cooling-off periods feature greater loss avoidance. Does breaking down the sample by these variables reveal anything more about *FollowThrough*? In Panel A of Figure 6, we see that returns after single trade plans are indeed much lower than those after multi-trade plan sales. Among plans with only one trade, those with perfect *FollowThrough* have lower subsequent returns than other plans by 0.5 percentage points. It makes sense that a 10b5-1 plan with only one trade that fulfills the entire plan is disingenuous and suspicious. Panel B shows that among plans with cooling-off periods greater than six months, those that follow through exactly precede lower stock returns. It may be that insiders that design plans with long cooling-off periods and end up following through exactly are masquerading as the most honest actors when they really seek to trade strategically.

Finally, we analyze why insiders decide to stick to their 10b5-1 plans or not. To answer this question, we fit the following regression using logit:

$$1[FollowPerfect]_i = \beta_0 * Intercept_i + \beta_1 * Determinants_i + \beta_2 * Year\ FEs + \epsilon_i,$$

where *1[FollowPerfect]* is equal to one if *FollowThrough* equals one for plan *i*. Our hypothesis is that perfect *FollowThrough* occurs more often when the firm is at higher risk of litigation, has weaker corporate governance, and has more opportunities to strategically time trades. The variables of interest are as in Henderson et al 2015: litigation risk, institutional ownership percentage, the proportion of directors that are also officers, and the prior year's stock price volatility. They estimate litigation risk as a combination of the prior year's minimum stock return, volatility of stock returns, turnover, firm size, beta, and industry. In column 1 of Table 13, we test those variables individually and find that only standard deviation, beta, and industry are significant predictors of *1[FollowPerfect]*. Lower total risk, higher systematic risk, and being in the retail industry correlate with a greater chance of perfect *FollowThrough*. Column 2 shows that higher litigation risk, calculated using the weights from Henderson et al 2015, predicts a greater likelihood of *1[FollowPerfect]*. It may be that the pressure of litigation risk compels insiders to adhere to their 10b5-1 plans exactly, but the relationship could be caused by industry which affects both litigation risk and *1[FollowPerfect]*.

Columns 3 and 4 of Table 13 test institutional ownership and the proportion of "insider directors", which are meant to be measures of corporate governance. The regression results show that a lower percentage of institutional ownership and higher proportion of insider directors correspond with a larger probability of perfect *FollowThrough*. The signs of these two variables suggest weaker corporate governance, according to previous literature, although the coefficient on insider directors is not statistically significant. In column 5, we test stock price volatility's effect on *1[FollowPerfect]*, and the coefficient is insignificantly different from zero. Large stock price swings create more opportunities for insiders to strategically time trades, so we expected this coefficient be positive; insiders that intend on making strategic trades will seek the protection of perfect *FollowThrough*. In column 6, we test litigation risk and institutional ownership percentage with corporate governance and volatility variables, and they maintain their sign and statistical significance. Overall, we find evidence that high litigation risk and less institutional ownership encourage insiders to complete their 10b5-1 plans exactly as promised.

The new facts presented in this section can be interpreted in several ways, but they may be useful in lawsuits involving illegal insider trading. We argue that perfect 10b5-1 plan *FollowThrough* does not mean that insiders are benevolent actors.

4.2. Updating Old Results: Strategic Timing of 10b5-1 Trades

Our dataset on 105b-1 plan announcements and insider trades covers the longest time period (2003-2020) of any in the previous literature. The next task of this dissertation chapter is to update some results of these papers to see if they maintain validity in the longer sample. We focus on findings related to loss avoidance because this is the main reason that insiders would abuse 10b5-1 plans.

First, we examine the holy grail result that insider sales under 10b5-1 plans avoid greater losses than other sales. If true, this result argues that insiders opt to use 10b5-1 plans when strategically timing stock sales based on MNPI. Jagolinzer 2009 finds that stock prices steadily decline by 2.2% after 10b5-1 sales but only by 0.7% after non-10b5-1 sales. This finding is already disputed by a few papers, such as Sen 2008.

In our larger sample period, we confirm the price pattern of Jagolinzer 2009. Figure 6 shows that the average 90-day abnormal return after 10b5-1 trades is 0.96 percentage points lower than that for non-10b5-1 trades. This difference is statistically significant at the 0.1% level with a t-value of 8.18. While unplanned trades precede slight positive returns in the next 3 months, planned trades precede consistently decreasing prices. This supports the conclusion that insiders may use the legal protection of 10b5-1 plans for more strategic (and illegal) trades, profiting on MNPI.

Next, we analyze Henderson et al 2015's result that announced 10b5-1 trades avoid greater losses than unannounced ones. The idea is that when insiders intend on making more strategic trades, they want to make their 10b5-1 plan as visible as possible. That way, if they are accused of trading on MNPI, they can point back to the announcement and more easily argue that the trade was preplanned. However, Figure 8 shows that, using our expanded sample, insider sales in announced plans actually have greater subsequent stock returns than those in unannounced plans. Even when we use the 2003 to 2005 subsample in Panel B, we cannot replicate the price patterns that they find for announced versus unannounced 105b-1 plans.[14]

Finally, we test the results of Larcker et al 2020 that plans with shorter cooling-off periods and those with only one trade avoid larger losses. Short cooling-off periods mean that insiders have access to more recent MNPI that might have a stronger effect on stock prices. Plans with only one trade go against the spirit of SEC Rule 10b5-1, which intends for insiders to make regular trades. This haphazard use of 10b5-1 plans seems to indicate strategic trading. While Larcker et al 2020 only use data from 2016 to 2020, we find very similar results by starting in 2003. Figure 9 shows that single-trade plans precede abnormal stock declines in almost every cooling-off period category. The greatest loss avoidance is in single-trade plans where the first trade occurs within 30 days of plan adoption. These plans are clearly not the pre-planned trades that the SEC intended to protect.

[14] Henderson et al 2015 use October 2000 through December 2005 as their sample period, but we lack insider trading data for 2000-2002. It may be that their result regarding announced 105b-1 plans is concentrated in that time period. In any case, the effect does not seem to extend any further.

With our longer sample period, we have a more complete view of 10b5-1 plans. The evidence that we present in this section supports the idea that 10b5-1 plans are being systematically abused.

5. CONCLUSION AND POTENTIAL FUTURE DIRECTIONS

While SEC Rule 10b5-1 was intended to provide legal protection for insiders to sell their stock holdings at regular pre-planned dates, it has been systematically abused. We document that insiders sell the exact number of shares promised in only 24% of 10b5-1 plans. Following through perfectly might not indicate an honest insider, however. We find that plans with perfect follow through tend to have shorter cooling-off periods and fewer trades, which have been linked by previous literature to strategic trading. We do not find a significant difference in the loss avoidance of 10b5-1 sales whose plans were followed perfect versus that in other sales.

We also take advantage of our longer sample size to update some important results in previous literature. Most significantly, we confirm Jagolinzer 2009's result that insider sales under 10b5-1 plans tend to avoid greater losses than other sales. We also verify that 105b-1 plans with shorter cooling-off periods and those with only one trade precede larger stock price declines.

Our *FollowThrough* variable is new to the literature and demonstrates how insiders see 10b5-1 plans as mere suggestions. They use them as flexible tools to protect strategic trades, rather than strict trading plans set in stone. In the future, we intend on expanding our sample of manually checked 10b5-1 plan announcements to improve the validity of our results. We also plan to test different variables as determinants of perfect *FollowThrough* and see if *1[FollowPerfect]* might have marginal predictive power over future firm outcomes.

APPENDIX

APPENDIX

Figure 2
Example of 10b5-1 Plan Announcement

This figure shows an example of a 10b5-1 plan announcement in Form 8-K. It lists the insider's name, the adoption date, the end date, and the number of shares promised to be sold.

Item 7.01 Regulation FD Disclosure.

On January 28, 2015, Joseph F. Puishys, Chief Executive Officer of Apogee Enterprises, Inc. ("Apogee"), entered into a trading plan in accordance with Rule 10b5-1 of the Securities Exchange Act, under which he intends to exercise 150,171 stock options and sell the underlying shares of Apogee common stock, subject to specified price limits, beginning March 23, 2015 and continuing from time to time through March 25, 2016.

Mr. Puishys continues to be one of Apogee's 25 largest shareholders. He holds 257,261 shares of Apogee common stock, well in excess of Apogee's stock holding guideline of five times his base salary. This process will facilitate orderly exercise of his stock options and sale of common stock to minimize any market impact and avoid any concerns about the timing of the transactions.

Rule 10b5-1 permits individuals who are not then in possession of material non-public information to establish pre-arranged plans to trade stock. The rule allows individuals to buy or sell shares of stock at a specific price in the future, regardless of any subsequent material non-public information.

Figure 3
Example of an Insider Trade in Form 4

This figure shows an example of a set of insider trades described in Form 4. The form describes two call option exercises and the subsequent sale of the resulting shares. Note that Footnote 2 states that the two common stock sales are pursuant to a 10b5-1 trading plan. *This footnote references the plan announcement shown in Figure 1*.

SEC Form 4

FORM 4

UNITED STATES SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

STATEMENT OF CHANGES IN BENEFICIAL OWNERSHIP

OMB APPROVAL	
OMB Number:	3235-0287
Estimated average burden hours per response:	0.5

☐ Check this box if no longer subject to Section 16. Form 4 or Form 5 obligations may continue. *See* Instruction 1(b).

Filed pursuant to Section 16(a) of the Securities Exchange Act of 1934 or Section 30(h) of the Investment Company Act of 1940

1. Name and Address of Reporting Person*
Puishys Joseph F.
(Last) (First) (Middle)
4400 WEST 78TH STREET
SUITE 520
(Street)
MINNEAPOLIS MN 55435
(City) (State) (Zip)

2. Issuer Name **and** Ticker or Trading Symbol
APOGEE ENTERPRISES, INC. [APOG]

3. Date of Earliest Transaction (Month/Day/Year)
03/23/2015

4. If Amendment, Date of Original Filed (Month/Day/Year)

5. Relationship of Reporting Person(s) to Issuer (Check all applicable)
X Director
10% Owner
X Officer (give title below)
Other (specify below)
Chief Executive Officer & Pres

6. Individual or Joint/Group Filing (Check Applicable Line)
X Form filed by One Reporting Person
Form filed by More than One Reporting Person

Table I - Non-Derivative Securities Acquired, Disposed of, or Beneficially Owned

1. Title of Security (Instr. 3)	2. Transaction Date (Month/Day/Year)	2A. Deemed Execution Date, if any (Month/Day/Year)	3. Transaction Code (Instr. 8)		4. Securities Acquired (A) or Disposed Of (D) (Instr. 3, 4 and 5)			5. Amount of Securities Beneficially Owned Following Reported Transaction(s) (Instr. 3 and 4)	6. Ownership Form: Direct (D) or Indirect (I) (Instr. 4)	7. Nature of Indirect Beneficial Ownership (Instr. 4)
			Code	V	Amount	(A) or (D)	Price			
Common Stock	03/23/2015		M		34,871	A	$8.34	282,132[1]	D	
Common Stock	03/23/2015		S[2]		34,871	D	$44.3082[3]	247,261	D	
Common Stock	03/24/2015		M		34,645	A	$8.34	281,906	D	
Common Stock	03/24/2015		S[2]		34,645	D	$44.107[4]	247,261	D	

Table II - Derivative Securities Acquired, Disposed of, or Beneficially Owned (e.g., puts, calls, warrants, options, convertible securities)

1. Title of Derivative Security (Instr. 3)	2. Conversion or Exercise Price of Derivative Security	3. Transaction Date (Month/Day/Year)	3A. Deemed Execution Date, if any (Month/Day/Year)	4. Transaction Code (Instr. 8)		5. Number of Derivative Securities Acquired (A) or Disposed of (D) (Instr. 3, 4 and 5)		6. Date Exercisable and Expiration Date (Month/Day/Year)		7. Title and Amount of Securities Underlying Derivative Security (Instr. 3 and 4)		8. Price of Derivative Security (Instr. 5)	9. Number of derivative Securities Beneficially Owned Following Reported Transaction(s) (Instr. 4)	10. Ownership Form: Direct (D) or Indirect (I) (Instr. 4)	11. Nature of Indirect Beneficial Ownership (Instr. 4)
				Code	V	(A)	(D)	Date Exercisable	Expiration Date	Title	Amount or Number of Shares				
Option (Right to Buy)	$8.34	03/23/2015		M			34,871	(5)	08/22/2021	Common Stock	34,871	$0.00	415,641	D	
Option (Right to Buy)	$8.34	03/24/2015		M			34,645	(5)	08/22/2021	Common Stock	34,645	$0.00	380,996	D	

Explanation of Responses:

1. Includes restricted stock awards under the 2009 Stock Incentive Plan.
2. The sale reported in this Form 4 was effected pursuant to a Rule 10b5-1 trading plan adopted by the Reporting Person on January 28, 2015 which was reported on a Form 8-K filed February 3, 2015.
3. The price reported is the weighted average sale price for the transactions reported. The prices received ranged from $44.13 to $45. The Reporting Person will provide to the issuer, a security holder of the issuer, or the SEC staff, upon request, full information regarding the number of shares sold at each price within the range.
4. The price reported is the weighted average sale price for the transactions reported. The prices received ranged from $43.835 to $44.62. The Reporting Person will provide to the issuer, a security holder of the issuer, or the SEC staff, upon request, full information regarding the number of shares sold at each price within the range.
5. Currently 100% exercisable.

Remarks:

/s/ Joseph F. Puishys — 03/25/2015
** Signature of Reporting Person — Date

Reminder: Report on a separate line for each class of securities beneficially owned directly or indirectly.

* If the form is filed by more than one reporting person, ***see*** Instruction 4 (b)(v).

** Intentional misstatements or omissions of facts constitute Federal Criminal Violations ***See*** 18 U.S.C. 1001 and 15 U.S.C. 78ff(a).

Note: File three copies of this Form, one of which must be manually signed. If space is insufficient, ***see*** Instruction 6 for procedure.

Persons who respond to the collection of information contained in this form are not required to respond unless the form displays a currently valid OMB Number.

Figure 4
Perfect Follow Through over Time

This figure shows the trend of *1[FollowPerfect]* over time from 2004 to 2019.

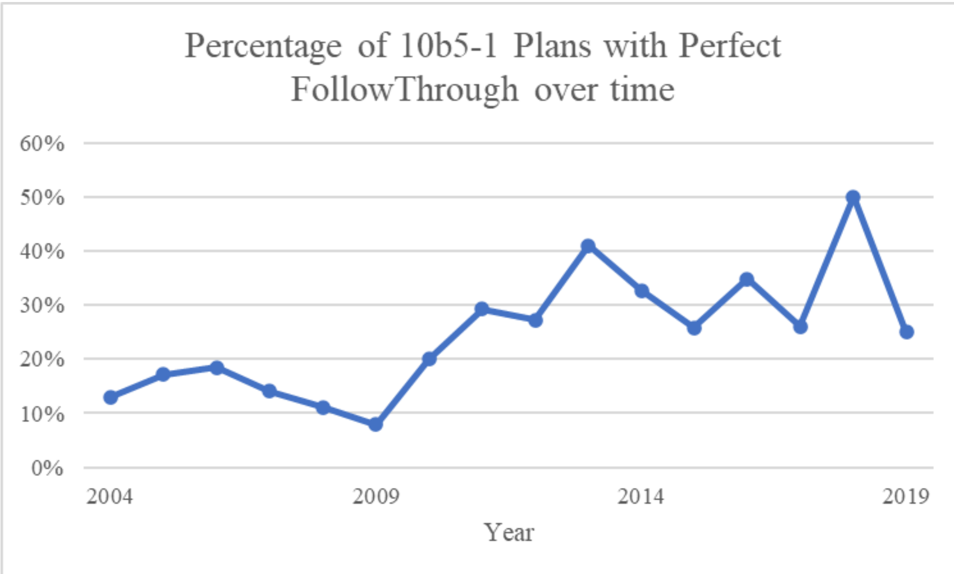


Figure 5
Returns around 10b5-1 Sales Sorted by Perfect Follow Through

This figure shows the average Cumulative Abnormal Returns (CARs) in the 180 days around 10b5-1 sales, broken down by perfect *FollowThrough* and imperfect *FollowThrough*. Averages are taken by trade, not by plan.

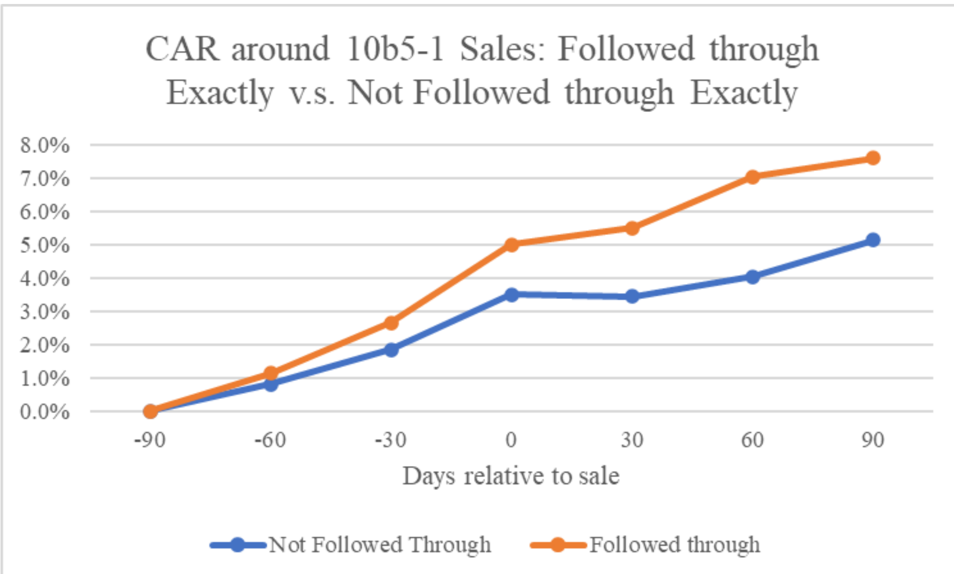
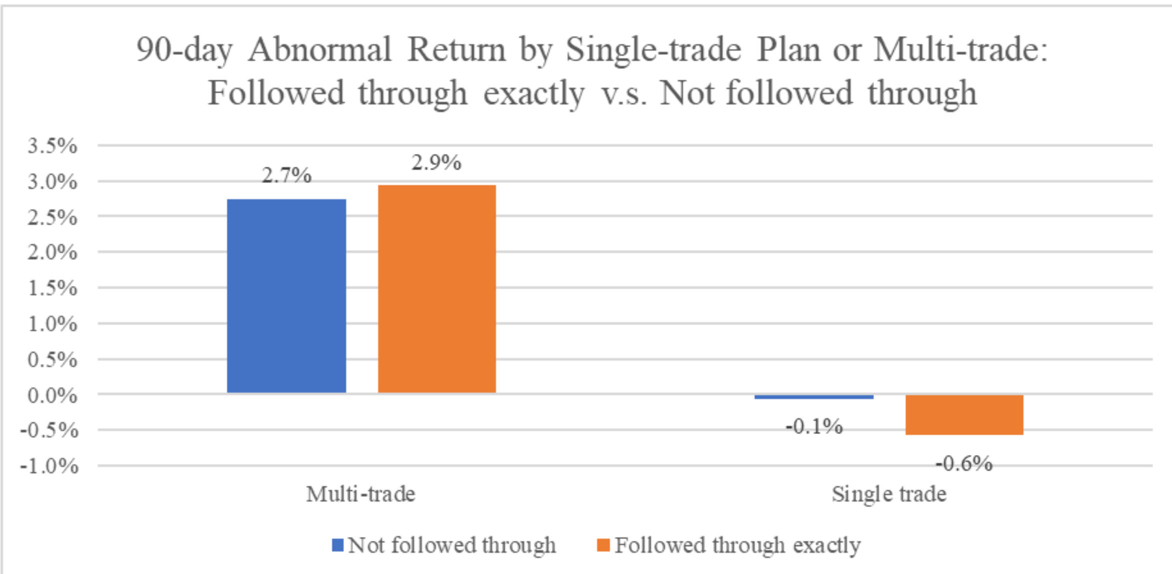


Figure 6
Returns after 10b5-1 Sales Sorted by Perfect Follow Through and other Variables

This figure shows the average 90-day abnormal return after 10b5-1 sales. In Panel A, we sort by whether the plan was single-trade or not and whether it was followed through or not. In Panel A, we sort by cooling-off period and whether the plan was followed through or not. Cooling-off period is defined as the time between plan adoption and the first trade. The averages are calculated by plan.

Panel A



Panel B

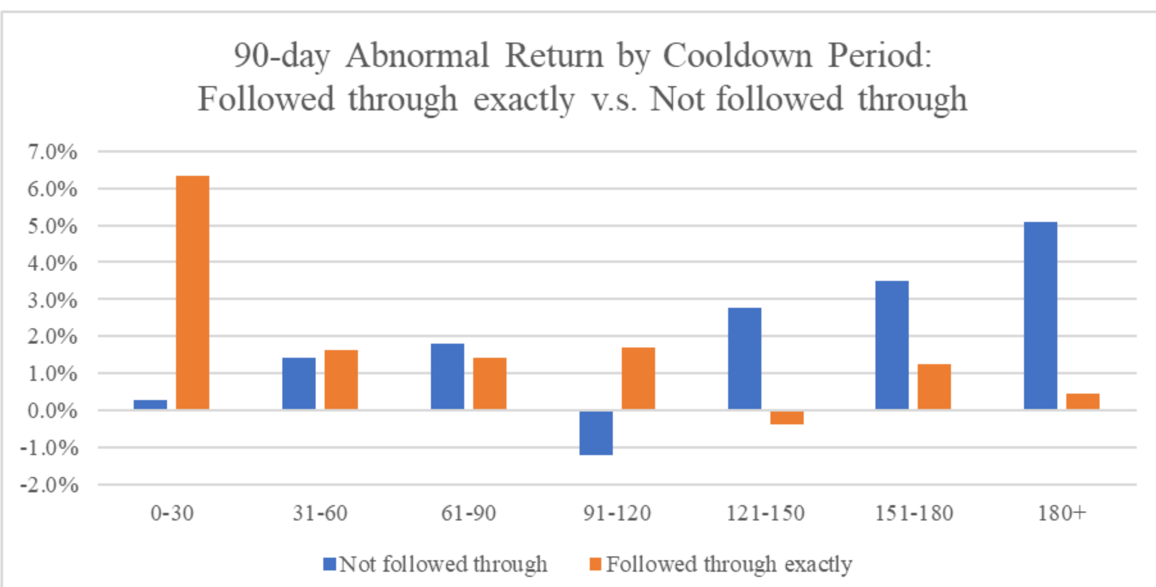
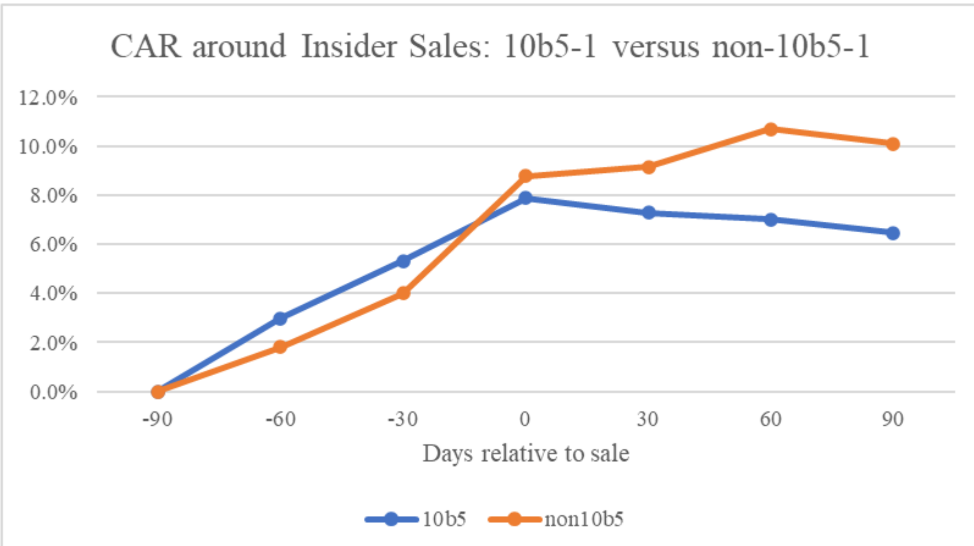


Figure 7
Loss Avoidance for 10b5-1 Sales versus Other Sales

This figure shows the Cumulative Abnormal Returns (CARs) in the 180 days around 10b5-1 sales (Panel A) and, specifically, the 90 days after (Panel B), broken down by 10b5-1 and non-10b5-1. Averages are taken by trade, not by plan.

Panel A



Panel B

This figure shows the Cumulative Abnormal Returns (CARs) in the 90 days after 10b5-1 sales, broken down by 10b5-1 and non-10b5-1. Averages are taken by trade, not by plan.

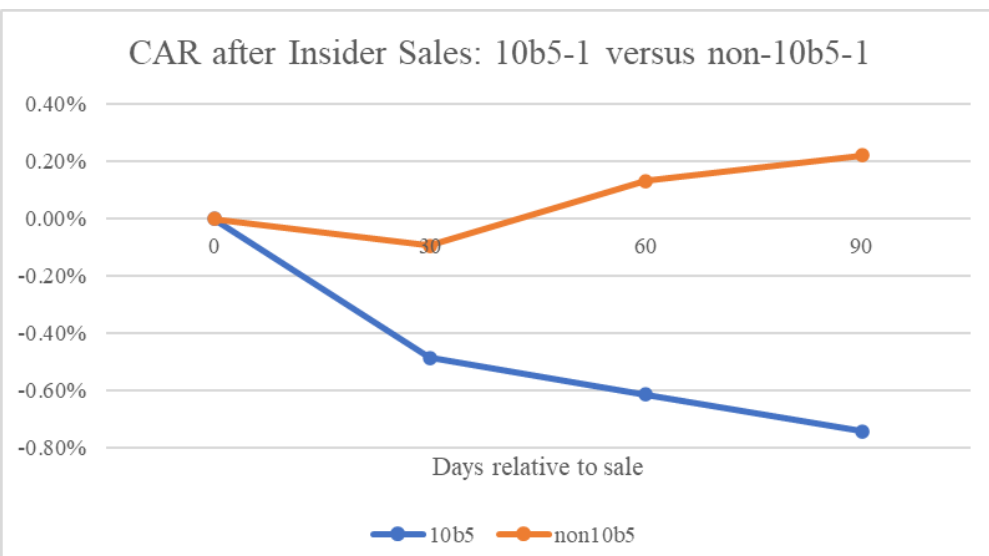


Figure 7 (cont'd)

Panel C

This figure shows the Cumulative Abnormal Returns (CARs) in the 30 days after 10b5-1 sales, broken down by 10b5-1 and non-10b5-1. Averages are taken by trade, not by plan.

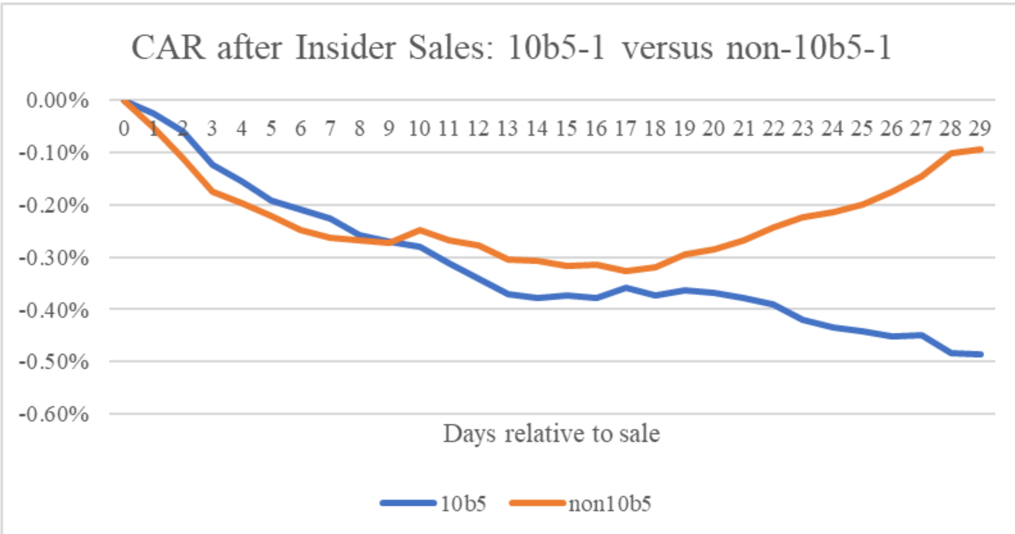
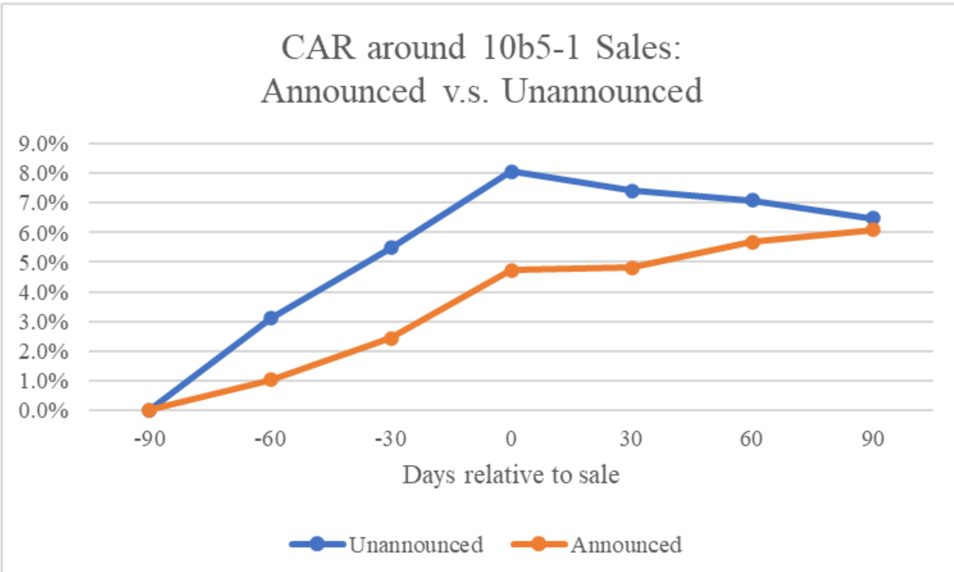


Figure 8
Returns around 10b5-1 Sales: Announced versus Unannounced Plans

This figure shows the Cumulative Abnormal Returns (CARs) in the 180 days around 10b5-1 sales, broken down by announced versus unannounced plans. Panel B shortens the sample period to 2003-2005 to match Henderson et al 2015 more closely. Averages are taken by trade, not by plan.

Panel A



Panel B

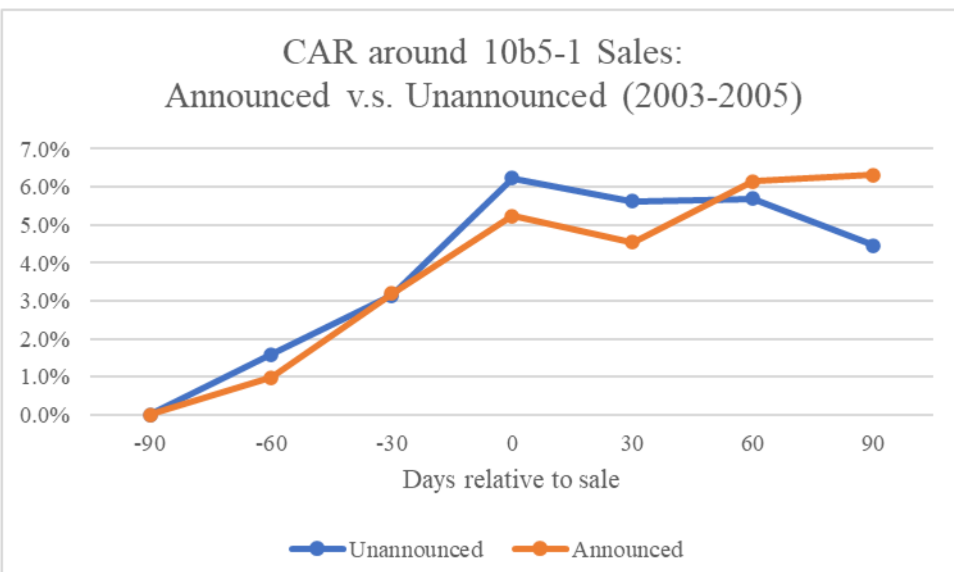


Figure 9
Loss Avoidance by Cooling-off Period and Single Trade

This figure shows the average 90-day abnormal return after 10b5-1 sales, broken down by cooling-off period and by single-trade. The averages are calculated by plan.

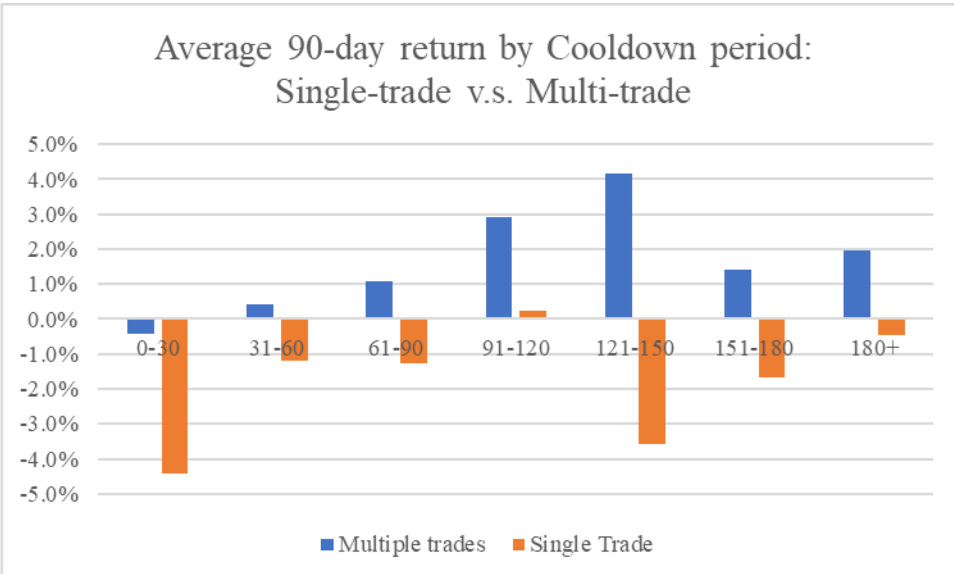


Table 12
Summary Statistics Sorted by Perfect Follow Through

This table compares summary statistics for 10b5-1 plans with *FollowThrough* not equal to one versus those with perfect *FollowThrough*. Variable definitions are found in Appendix B. Stars in column 3 represent statistical significance at the 10%, 5%, and 1% levels for the differences in means.

	(1)	(2)	(3)
	FollowThrough != 1	*FollowThrough* == 1	(1) - (2)
Plan Variables			
Cooling-off Period	73.24	57.08	16.16
Trades per Plan	5.67	3.72	1.95*
1[Single Trade]	0.29	0.36	-0.07
Trade Size	877,741	1,827,726	-949,985**
Number of Shares Planned	168,361	62,704	105,657***
Followthrough	0.23	1.00	-0.77***
1[FollowPerfect]	0.00	1.00	-1.00
Company Traits			
Abn Return[0,90]	0.00	0.01	-0.01
Inside Directors	0.20	0.19	0.01
Insitutional Ownership	0.80	0.73	0.06***
Market Value	12,454,308	1,424,0127	-1,785,819
Min Return[-252,0]	-0.08	-0.07	-0.01*
Std Returns[-252,0]	0.02	0.02	0.00***
Beta[-252,0]	0.23	0.26	-0.03
Average Volume[-252,0]	0.01	0.01	0.00
Litigation Risk	0.61	0.64	-0.03***
1[Gets sued]	0.41	0.53	-0.12*
Observations	473	122	595

Table 13
Determinants of Perfect Follow Through

This table shows how some firm-level variables predict perfect *FollowThrough*. The regression is estimated using logit. Standard errors are in parentheses below each coefficient. Stars represent statistical significance at the 10%, 5%, and 1% levels.

	(1)	(2)	(3)	(4)	(5)	(6)
Min Return[-252,0]	0.954					
	(4.076)					
Std Returns[-252,0]	−42.218*					
	(23.126)					
Average Volume[-252,0]	−6.849					
	(25.327)					
Beta[-252,0]	0.666**					
	(0.334)					
ind_tech	0.413					
	(0.284)					
ind_retail	1.483***					
	(0.284)					
Litigation Risk		3.712***				4.207***
		(1.049)				(1.155)
Insitutional Ownership			−2.272***			−2.646**
			(0.573)			(1.259)
Inside Directors				0.245		2.338
				(0.539)		(1.911)
Volatility					−15.525	−5.251
					(12.762)	(47.290)
volat_instOwn						2.230
						(56.045)
volat_insideDirs						−110.447
						(116.021)
Constant	−13.601***	−16.964***	−13.572***	−1.159*	−13.783***	−2.827**
	(0.717)	(0.732)	(0.607)	(0.666)	(0.524)	(1.357)
Observations	588	588	553	573	588	540

BIBLIOGRAPHY

BIBLIOGRAPHY

Fich, E., R. Parrino, and A. Tran. 2018. *Working Paper*.

Jagolinzer, A. 2009. SEC Rule 10b5-1 and Insiders' Strategic Trade. *Management Science* 55: 224-239.

Jagolinzer, A., D. Larcker, and D. Taylor. 2011. Corporate Governance and the Information Content of Insider Trades. *Journal of Accounting Research* 49:1249-1274.

Henderson, T., A. Jagolinzer, and K. Muller. 2015. Offensive Disclosure: How Voluntary Disclosure Can Increase Returns from Insider Trading. *The Georgetown Law Journal* 103:1275-1306.

Larcker, D., B. Lynch, P. Quinn, B. Tayan, and D. Taylor. Gaming the System: Three “Red Flags” of Potential 10b5-1 Abuse. *Stanford Closer Look Series*.

Lee, Y. 2020. Insiders’ Foreknowledge of Earnings Results and Rule 10b5-1 Sales Trades. *Working Paper*.

Millian, J. 2016. Insider Sales Based on Short-Term Earnings Information. *Review of Quantitative Financial Accounting* 47:109–128.

Mitts, J. 2020. Insider Trading and Strategic Disclosure. *Working Paper*.

Sen, R. 2008. Are Pre-planned Insider Sales Strategically Timed?. *Working Paper*.

Shon, J., and S. Veliotis. 2013. Insiders’ Sales Under Rule 10b5-1 Plans and Meeting or Beating Earnings Expectations. *Management Science* 59:1988-2002.