

Protecting Forward Looking Statements

Daniela De la Parra
Rice University
dd37@rice.edu

K. Ramesh
Rice University
rameshk@rice.edu

Macleon Gaulin
University of Utah
mac.gaulin@utah.edu

February, 2020

Abstract

We provide the first examination of the determinants of firms' decision to use a list of keywords in SEC filings to identify forward-looking statements and obtain 'safe harbor' protection under the Private Securities Litigation Reform Act. We show that proxies for ex ante litigation risk, disclosure supply, economic uncertainty, and disclosure herding are strongly associated with the decision to include the keyword list. In addition, when we examine the determinants of the number of keywords, we find that both structural variables and proxies for transient forces are statistically significant, with the latter being consistent with lower disclosure costs. Finally, using exploratory factor analysis, we identify five disclosure attributes that capture the most frequent keywords that managers choose. We find that managers use specific keywords that evolve over time, potentially to tailor the language of their forward-looking statements to reflect the economic circumstances they face. Together this evidence provides an important first look at the determinants of firms' decisions regarding a central feature of 'safe harbor' protection.

JEL Classification: G14, G32, D83

Keywords: Disclosure, Forward Looking Statements, Litigation Risk

Data Availability: Data are available from the public sources cited in the manuscript.

Preliminary working draft. Please do not circulate or quote without permission. Prepared for presentation at the 2020 Lone Star Conference. All remaining errors are our own. We thank Allen Huang for providing us with the *Liberal Court* variable. The internet appendix is available upon request.

I. INTRODUCTION

The Private Securities Litigation Reform Act (PSLRA) of 1995 affords firms protection for disclosures they make about uncertain future events. These legal protections are contingent upon identifying which statements are forward-looking, and including clear cautionary language that the forward-looking statements (FLS) made by the firm are subject to uncertainty.¹ Johnson, Kasznik, and Nelson (2001) find that the passage of PSLRA led to increases in voluntary and credible disclosures of forward-looking information by high technology firms, especially so when firms faced higher ex-ante risk prior to PSLRA. In addition, extant literature shows that firms disclose FLS to convey private information to markets and that the capital markets find such information to be value relevant (see Section II). However, the question of how firms identify FLS in order to be eligible for ‘safe harbor’ protection has not been explored, and without a clear identification of FLS, courts have held that protection for FLS is not guaranteed.

We study an increasingly prevalent managerial disclosure practice of listing specific keywords in SEC filings that firms assert will be used in constructing forward-looking statements, thereby clearly identifying FLS to support their claim for ‘safe harbor’ protection. Firms decide *whether* to have a list of keywords, and if so, *which* specific keywords to include. Moreover, both decisions can potentially change from year to year depending on the cost-benefit calculus. The purpose of our study is to explore the factors underlying these disclosure decisions, and thereby, provide evidence on an important precursor to receiving legal protection to safely convey valuable forward-looking information to the marketplace.

We do not expect the demand for forward-looking information to directly shift in response to PSLRA, so we focus on supply-side factors to explain managerial choice regarding keyword inclusion. Moreover, given that firms may face significant disclosure costs from abandoning a keyword list (Einhorn and Ziv 2008), the decision to include the list is likely determined by structural factors rather than transitory issues. Specifically, we predict that ex ante litigation

¹Pub. L. 104-67, 109 Stat. 737

risk ([Johnson et al. 2001](#)), disclosure propensity, economic uncertainty, and disclosure herding or bandwagon effects ([Hirshleifer and Teoh 2003](#)) are associated with the decision to include a keyword list. However, firms may change keywords from year to year to respond to changing economic circumstances affecting the forward-looking information disclosed. As a result, we predict that both structural and time-series effects are likely to explain the variations in the number of keywords. Finally, using exploratory factor analysis, we identify different combinations of words that managers chose to represent different disclosure attributes that they would like to include when conveying FLS.

We parse the text in the 10-Ks of every SEC registrant from 1996 to 2017 and extract the list of FLS keywords when present. Our filtering procedure results in a sample of 132,303 annual reports, of which 66% contain FLS keywords. We find considerable variation in the cautionary narrative, both across firms, across economic conditions, and across time. The adoption of these FLS keyword lists increases drastically from 13% to 91% over the two decades since the passage of the PSLRA.

As suggested by the litigation risk argument, we find that firms located in federal circuits with more liberal judges ([Huang, Hui, and Li 2019](#)) have a higher probability of including FLS keywords in their 10-K reports. In terms of disclosure supply, firms that issue guidance and provide more disclosures in their 10-K reports are more likely to have a keyword list. Consistent with the predicted influence of uncertainty, we find that younger firms, firms with R&D investments, firms facing losses, and those engaged in significant M&As are more likely to include keyword lists and thereby gain ‘safe harbor’ for their forward-looking disclosures intended to mitigate investor concerns. The inter-temporal increase in the adoption of a keyword list is at least partially driven by firms learning from the disclosure practices of industry peers, consistent with herding or bandwagon effects.

We obtain additional insight from examining sub-samples based on firms’ disclosure practices. In firms that are persistent adopters (always or never disclose a keyword list), our regressions have much higher explanatory power, and firms subject 10b-5 lawsuits and those

with legal expertise in senior management tend to include keywords more often. For those firms that made a structural decision to switch (only once during our sample), none of the litigation-related variables are significant, consistent with other structural shifts post-PSLRA effecting the switch. To shed light on the inter-temporal trend in adoption, we re-run our regressions by dividing the sample into three sequential time periods. A key finding is that all of the disclosure herding behavior occurs in the first seven years (of our 22-year sample), which provides some preliminary evidence on the speed with which firms in the economy achieved equilibrium in response to a new disclosure regulation.

When we examine the determinants of the number of keywords in the list, we find both structural variables and proxies for transient forces are statistically significant, with the latter being consistent with our expectation that the disclosure costs of changing specific keywords is not prohibitive. Specifically, firms with more left skewed returns, those that recently entered the capital market, or those experiencing an executive turnover include more keywords.

We next conduct an exploratory factor analysis to identify the different disclosure attributes that firms may employ to prepare forward-looking statements. By analyzing the top-20 most used keywords, we identify five disclosure attributes that substantially capture the variations in the overall choice of the keyword list in our sample. We call these attributes *future state*, *modal verbs*, *likelihood*, *quantitative*, and *future action*. We find that firms consistently use words describing the likelihood of events or actions since the passage of the PSLRA, while the use of words capturing other disclosure attributes contribute to the inter-temporally increasing use of keywords.

We obtain several insights from separating the bag of keywords into specific disclosure attributes. Consistent with higher litigation risk, firms in liberal courts include keywords that can be used to describe future state or provide quantitative information, thereby limiting any shareholder lawsuits from unmet expectations. Managers that frequently guide may be expected to also explain their plans, so they include keywords that convey future action and

quantitative information. Younger firms use all disclosure attributes, except quantitative information, to mitigate investor concerns regarding uncertainty. In addition, firms with high stock return volatility avoid keywords that convey potentially uncertain quantitative information, but rely more on words that can help qualitatively convey information regarding future action or future state. Taken together, we find that managers selectively choose words that capture different disclosure attributes to obtain legal protection under varying economic circumstances.

Our study contributes to the literature on forward-looking statements by highlighting an important aspect of forward-looking statements, namely the requirement under the PSLRA that they be clearly identified. We also demonstrate that managers choose their cautionary language in a non-random or boilerplate fashion. Prior research studying forward-looking information in corporate disclosures focuses on the determinants of FLS, the quantity and quality of these voluntary disclosures, and the channels through which firms convey their FLS. To our knowledge, we are the first to empirically explore the aspect of identifying forward-looking information as required under the PSLRA.

Our study also contributes to the broader accounting literature focusing on textual disclosures. Our findings that managers selectively choose the language with which they convey information to markets suggests that a careful approach to linguistic analysis in general may be warranted. Traditional bag of word dictionary approaches, while providing important information about broader disclosure trends, may not always be suited to the study of idiosyncratic firm behavior. Our results suggest that a ‘one size fits all’ methodology may overlook important firm and managerial decisions, which could apply broadly to how the research community evaluates textual disclosures.

II. LITERATURE AND RESEARCH DEVELOPMENT

Background and Literature

To meet the information demand from various stakeholders, firms convey their forward-looking information through voluntary disclosures. One potential cost of these disclosures is that they are inimically tied to legal liability from 10b-5 securities litigation. On the one hand, managers may believe that by voluntarily disclosing adverse information in a timely fashion, they can lessen any negative impact on stock prices—which could otherwise trigger a securities litigation (Skinner 1994, 1997; Field, Lowry, and Shu 2005).² Consistent with this theory, Naughton, Rusticus, Wang, and Yeung (2019) find that after a plausibly exogenous reduction in litigation risk, firms reduce their overall level of voluntary forecast disclosures. On the other hand, a 10b-5 litigation can be triggered by a misleading or fraudulent disclosure.³ Managers may thus see voluntary disclosure as a potential catalyst for litigation risk (Rogers and Van Buskirk 2009).

The question of how forward-looking disclosures affect firm’s risk is relatively new, because for the first 40 years of its existence, the SEC prohibited firms from voluntarily disclosing: forward-looking information about financial projections, management plans and future operations, or expected future economic performance. In 1973, the SEC provided an interpretative release that lifted this prohibition, allowing reporting companies to disclose projected sales and earnings information. The new rules required that firms disclose updates to any material changes in forecasts, and file with the SEC any projections disclosed through other channels.

Despite the SEC’s intent to promote disclosure of FLS, their proposed safe-harbor protections were rarely pleaded in securities litigation. Instead, courts developed the ‘bespeaks caution doctrine,’ which held that forward-looking statements would not be considered misleading if accompanied by adequate risk disclosure cautioning readers about specific risks

²We assume incentive alignment between the firm and the manager, and use the terms firms and managers interchangeably throughout the paper.

³Specifically, U.S.C §240.10b-5 states that, “It shall be unlawful for any person [...] To make any untrue statement of a material fact”

that could materially impact forecasts and predictions. Over the years, numerous cases were litigated over misleading FLS without adequate cautionary language (e.g. *Inst. Investors Group v. Avaya, Inc.* 564 F.3d 242, 256; 3d Cir. 2009), and numerous were dismissed as the cautionary language was found to be sufficient (e.g. *Convergent Technologies Security Litigation*, 948 F.2d 507, 515; 9th Cir. 1991). The disparate application of the bespeaks caution doctrine created significant uncertainty regarding liability stemming from FLS.

In an effort to ameliorate this uncertainty, the Private Securities Litigation Reform Act was passed in 1995. The PSLRA provides an avenue for managers to disclose valuable, forward-looking information without fearing the financial and managerial costs of unmeritorious lawsuits. However, there are three requirements to receive protection under PSLRA: first, firms must clearly identify statements that are forward-looking, second, they must ensure that FLS are not made with the knowledge that they are false or misleading, and third, FLS must be accompanied by cautionary language. [Nelson and Pritchard \(2016\)](#) examine the impact of the changes in cautionary language after the passage of the PSLRA, and find that firms' equilibrium level of voluntary disclosure is increasing in the protections afforded. In addition, extant research focuses on the incentives of firms to disclose FLS to convey their private information to markets, and the market reaction to these disclosures.⁴ These studies presume protection of the forward-looking disclosures, but they do not explicitly address the issue empirically.

Our study focuses specifically on the PSLRA requirement that firms must identify forward-looking statements to gain the 'safe harbor' protection. We identify the increasingly prevalent managerial disclosure practice of listing specific keywords that are meant to signal forward-looking statements, to obtain 'safe harbor' protection (see [Appendix C](#) for examples).

Keyword-lists are suggested in contemporary practitioner literature ([Chivers and Quinn 2009](#)), as well as in complaints and court cases. However, managers might believe that a

⁴See, e.g., [Johnson et al. \(2001\)](#); [Kimbrough and Louis \(2011\)](#); [Muslu, Radhakrishnan, Subramanyam, and Lim \(2015\)](#); [Bochkay, Chychyla, and Nanda \(2016\)](#); [Bozanic, Dietrich, and Johnson \(2017\)](#); [Bozanic, Roulstone, and Van Buskirk \(2018\)](#); [Hutton, Miller, and Skinner \(2003\)](#); [Li \(2010\)](#); [Hassanein and Hussainey \(2015\)](#); [Bozanic and Kraft \(2018\)](#); [Bozanic et al. \(2018\)](#)

boilerplate list of keywords would be sufficient to gain protection. For example, [McClane \(2019\)](#) suggests that “the skill [lawyers entering corporate securities practice] employ most is the ability to cut and paste.” Consistent with this, [Dyer, Lang, and Stice-Lawrence \(2017\)](#) demonstrate that over the past two decades, the use of boilerplate language has been increasing. To derive keywords lists, lawyers could easily utilize generic word-lists such as those used in [Bozanic et al. \(2018\)](#).

Given the focus and significance that courts have placed on identifying forward-looking statements, the argument that boilerplate disclosure garners sufficient protection appears unlikely. In *Southland Securities v. INSpire Insurance Solutions Inc.* (365 F.3d 353, 2004), the Fifth Circuit US Court of Appeals emphasized the importance of identifying FLS: “the defendants have not shown that these statements were identified as forward-looking statements. Accordingly, the plaintiffs may properly allege a claim based on these statements...” Similarly, In re *Griffin v. GK Intelligent Systems, Inc.* (87 F. Supp. 2d 684, 1999): “the challenged statements do not fall within the ‘safe-harbor’ provision because they were also not clearly identified as ‘forward looking.’” Most notably, in *Slayton v. American Express Company* (604 F.3d 758, 2010), the Second Circuit US Court of Appeals specifically mentioned the inclusion of a list of words is sufficient to identify FLS: “The Company wrote, “[t]he words ‘believe’, ‘expect’, [...] are intended to identify such forward-looking statements.” [...] The May 15 statement is plainly forward-looking—it projects results in the future. It is also accompanied by a statement of the common-sense proposition that words such as ‘expect’ identify forward-looking statements.” This evidence suggests that FLS keyword lists are an important part of gaining PSLRA protection. We seek to add to the literature by providing novel evidence on the determinants of FLS keyword disclosures.

Key Determinants of Protecting Forward-Looking Statements

Given that the legal protection standards were altered by a change in securities laws, we focus largely on supply-side considerations in explaining managerial choice to include a list of keywords. We do not expect the demand for forward-looking information to directly shift in

response to PSLRA.⁵ In addition, we view the decision to include a keyword list as more of a structural decision rather than being transitory in nature. Once included, firms may face significant disclosure costs from abandoning a keyword list (Einhorn and Ziv 2008) although structural shifts in economic conditions could more easily justify the creation of a list.

We view litigation risk as a key determinant of including a list of keywords. Previous research has largely focused on the relationship between voluntary disclosures and litigation risk, with mixed evidence on whether disclosures deter or trigger shareholder litigation. In our setting, we focus on the specific decision to garner legal protection through FLS keyword disclosure that is designed to meet the legal standards. Thus our prediction regarding litigation risk is more unambiguous than those for disclosures in general. We expect firms to be more likely to seek ‘safe harbor’ when their ex-ante litigation risk is increasing (Johnson et al. 2001). If disclosure mitigates litigation risk, then firms’ use of disclosure will increase in litigation risk, and therefore, the value and use of a keyword list should increase. If, on the other hand, disclosure exacerbates litigation risk, then protections of that disclosure are even more valuable. Thus, we expect the relationship between ex-ante litigation risk and FLS keyword disclosure to be unconditionally positive.

To the extent PSLRA reduced disclosure costs, firms are bound to disclose more quantitative and qualitative FLS at the margin. If managers did not have private and credible forward-looking information in the first place, a new statute’s ability to offer protection for such statements is unlikely to lead to more disclosures. As a result, we predict a positive association between the supply of disclosures and the propensity to include a keyword list given the expected reduction in disclosure costs.

Firms with uncertain future outlook may be hesitant to provide FLS for fear of shareholder litigation. The uncertainty may arise from risky growth opportunities, financial distress,

⁵We do not rule out indirect demand-side effects. For instance, analysts who abandoned firms that could not provide forward-looking information due to litigation uncertainty could gravitate towards such firms post PSLRA and demand more forward-looking information. In our empirical specifications we include controls for analyst following and institutional ownership, but we are agnostic about the nature of these economic effects.

or strategic long-term investments that could cause information fractions and limit the ability of the firm to convey value-relevant information. Alternatively, firms could respond to uncertainty by reducing or maintaining their disclosure levels, for example if the uncertainty implies positive skewness in expected cash flows (Heinle, Smith, and Verrecchia 2018). Ceteris paribus, to the extent PSLRA reduces the cost of sharing information on potentially risky outcomes, managers facing uncertain environments are, at the margin, more likely to include a keyword list to facilitate their disclosure strategy.

A potential alternative rationale for adoption of FLS keyword disclosure is herding or bandwagon effects. Informational cascades require some coarseness to the signal, such that there is not a continuous correlation between private information and outcome (Hirshleifer and Teoh 2003). Given our signals are quite coarse (a binary disclosure decision and keyword counts), it is plausible that herding might occur, whereby firms adopt a legal disclosure strategy based on the public observation of other firm disclosure (Bikhchandani, Hirshleifer, and Welch 1998). We are agnostic as to whether this herding may arise due to rational or irrational agents, as it is difficult to disentangle the empirical predictions (Hirshleifer and Teoh 2003). Given the ability of firms to observe both disclosure decisions and litigation outcomes of their peer firms, it is plausibly beneficial for firms to adopt the legal protection strategies that appear to be successful for other firms. This is consistent with prior findings on herding in other managerial disclosures (e.g. Tse and Tucker 2010; Jung 2013; Armstrong, Glaeser, and Kepler 2019). As discussed in Section IV, we find that while less than 20 percent of the firms included a keyword immediately after the passage of the PSLRA, more than 90 percent do so 20 years later. To the extent herding behavior can explain this trend, it could provide new insights regarding the long-term evolution of corporate disclosure strategy.⁶

We distinguish between the decision to disclose FLS keywords and the decision of how many or which words to include. Unlike the inclusion of a dictionary, the potential disclosure

⁶Such inter-temporal trends are not uncommon in the disclosure area (e.g., the sharp decline in the number of firms announcing earnings during regular trading hours over the last three decades). However, only limited evidence is available on the economic forces at work that contributed to such trends.

costs of adding or excluding some of the keywords are not expected to be as prohibitive. In fact, the investors may view such changes as an indication of managerial efforts to convey contextual forward-looking information.⁷ As a result, both structural and time-series effects are likely to explain the variations in the number of keywords. Consistent with this expectation, empirically there is no inter-temporal convergence in the number of keywords across firms (the standard deviation in number of keywords increases from 2 to 4.6 over the course of our sample period). In other words, there are economically meaningful cross-sectional and time-series variations in the number of keywords used to identify FLS. We, therefore, predict that the decision as to which keywords to include or exclude is likely influenced by economic events firms are likely to discuss in their FLS when they enter the capital market, when they experience executive turnover, when the current market perceptions of risk are high, etc.

III. DATA AND METHODOLOGY

Sample Selection

We begin with the intersection of firm-years from Compustat and EDGAR. As in [Dyer et al. \(2017\)](#), we focus on 10-K reports, especially given the substantial variations in the type and extent of disclosures between quarterly and annual reports, with more extensive disclosures in the 10-K reports. We use the CIK from CRSP's Comphist file to match these firm-years to their appropriate 10-K filing from EDGAR. This match results in 158,744 firm-years with EDGAR data. Given that we parse the text to extract any FLS keyword lists, we then exclude filings that are unavailable for scraping or whose HTML format is either missing or broken. For a small sample of firm-years (564), we fill in missing keywords information using the FLS keyword lists from the two adjoining years (only when they are identical). To maintain a consistent legal environment for the disclosures we study, we also exclude 10-Ks filed before the PSLRA became effective on December 22, 1995. We drop firms with one or more gaps in their EDGAR data (that we can not interpolate in the previous step) and then

⁷Our discussions with the Chief Accounting Officer of a large publicly-traded bank suggest that the bank made a conscious decision to include or exclude certain keywords when engaged in an acquisition and following implementation of the Dodd-Frank Act.

we drop firms with fewer than two years of EDGAR data. Lastly, we drop firm-years with fiscal year end before 1996, when the EDGAR system started gaining widespread adoption. Our resulting sample includes 132,303 firm-years from 14,722 firms, spanning the fiscal years from 1995 to 2017. The sample selection is detailed in Table I.

[Table I about here.]

Computation of Disclosure Measures

We use regular expressions to parse the text from 10-K filings and extract the list of keywords used by firms to identify forward-looking statements. The FLS keyword list is often found in one of three places in the 10-K filing: at the beginning or end of the Management’s Discussion and Analysis section, immediately preceding the Risk Factor section, or at the beginning of the 10-K under a subsection called “Forward-Looking Statements” (or similar). We search for quoted lists of words near forward-looking phrases (e.g. “forward-looking,” “cautionary statements,” and “safe harbor”) and extract the FLS keyword list when present. We use a similar procedure to extract the FLS keyword list from the fourth-quarter earnings press releases submitted as Item 2.02 in 8-K filings. Appendix B provides a detailed description of the text extraction procedure.

From the extracted data, we measure *FLS Keywords Inclusion* as an indicator variable, which is equal to one if the 10-K report for a given fiscal year includes an FLS keyword list. We then measure $\# \text{ Keywords}$ as the number of keywords included in the list. To verify the quality and accuracy of our extraction technique, we hand-verified several random samples. In the rare case our methodology resulted in errors, it was typically due to technical issues such as filing encoding or significant English or grammatical errors.

Research Design

We estimate the following probit regression (baseline regression) to test our predictions regarding the firm, industry, and macro characteristics that are associated with the decision

to include a list of FLS keywords:

$$\begin{aligned}
FLS\ Keywords\ Inclusion_{it} = & \alpha + \beta_1 Liberal\ Court_{it} + \beta_2 10b-5\ Firm_{it} \\
& + \beta_3 FPS_{it} + \beta_4 Legal_{it} + \beta_5 Guidance_{it} + \beta_6 Log(10-K\ Size)_{it-1} \\
& + \beta_7 Log(Age + 1)_{it} + \beta_8 R\&D_{it} + \beta_9 Loss\ Year_{it} + \beta_{10} Acquisition\ Sales\ Ratio_{it} \\
& + \beta_{11} Industry\ Adoption_{it-1} + \Sigma \delta_j Controls_{it} + \epsilon_{it}
\end{aligned}$$

See Appendix A for variable definitions. We consider three proxies for litigation risk (*Liberal Court*, *10b-5 Firm*, *FPS*), and one measure of legal expertise at the top management level (*Legal*). *Liberal Court*, from Huang et al. (2019), measures the probability that the three-judge panel assigned to a case in a circuit court is composed by at least two judges appointed by Democratic presidents. The higher the measure, the higher the probability that the case will be decided in favor of shareholders; thus, the higher the plausibly exogenous ex-ante litigation risk. Using data from the Stanford Law School’s Securities Class Action Clearinghouse, we construct the indicator variable *10b-5 Firm* as a firm-level measure of litigation propensity, which is set equal to one if the firm is sued at least once during our sample period. Following Francis, Philbrick, and Schipper (1994), *FPS* is an indicator for high litigation industries.⁸ We expect positive coefficients on the litigation risk proxies, consistent with firms with higher litigation risk including a list of FLS keywords to garner protection under PSLRA. Lastly, *Legal* is an indicator variable that takes a value of one when one of the officers identified under Item 401 of Regulation S-K is a lawyer or a corporate secretary (using data from the BoardEx database). We expect that this legal expertise will increase the propensity for disclosure preferences which are more legally risk averse, thus a higher likelihood of FLS keyword list inclusion.⁹

⁸These industries are defined using SIC codes, and they include biotechnology (2833-2836, 8731-8734), computers (3570-3577, 7370-7374), electronics (3600-3674), and retailing (5200-5961).

⁹The tenor of our results holds when we limit *Legal* to only lawyers in senior management.

We consider two measures of disclosure which we predict are positively associated with the likelihood of FLS keywords inclusion. The first, *Guidance*, is an indicator variable equal to one if management issues any guidance during the fiscal year. As the primary intent of cautionary language is to protect forward-looking disclosures such as financial projections, we expect guiding managers to derive more value from the PSLRA, consistent with [Johnson et al. \(2001\)](#). The second, *Log (10-K Size)*, is the length (logged) of the previous year’s 10-K, which is a proxy for the overall amount of disclosure. We expect that as the amount of disclosure increases, the exposure to litigation risk increases commensurately, and so to the value of “safe harbor.”

We consider four measures related to idiosyncratic firm uncertainty. Following [Li \(2010\)](#), we use firm age ($\text{Log}(\text{Age} + 1)$) as an inverse proxy for uncertainty. If younger firms are more hesitant to discuss future outlook, they are more likely to rely on a keyword list to obtain litigation immunity. We use an *R&D* indicator consistent with [Kothari, Laguerre, and Leone \(2002\)](#), who argue that the benefits from R&D investments are much more uncertain than those from PP&E. Similarly, mergers and acquisitions (*Acquisition Sales Ratio*) potentially induce both uncertainty and litigation risk, thus managers would have heightened incentives to discuss future prospects. [Bloomfield \(2008\)](#) suggests that loss firms (*Loss*) may require greater amount of disclosures to better “describe” the economic circumstances, and as a result, we expect a greater need for forward-looking statements. Our proxy for disclosure herding is $\text{Industry Adoption}_{t-1}$, which is the percentage of firms within a GICS industry group that adopted a FLS keyword list in their 10-K during the fiscal year $t - 1$.

The focus of our study is on the change in disclosure costs and benefits due to a change in securities laws. We, therefore, focus largely on supply side factors to explain the disclosure choice. However, some of our additional control variables could be viewed, at the margin, as capturing the demand-side response. Specifically, we include an indicator for whether the firm has any analyst following, and the percentage of stock owned by institutional investors (*Inst_Own*) as potential demand-side controls. Although the marginal effect of a Big-N

auditor on the inclusion of a keyword list is unclear, we include the *Big N Auditor* indicator variable as a control as well.

We then expand our base regression specification to include transient institutional ownership and the change in the 10-K size. Because our proxies for uncertainty reflect a parsimonious set of firm economic conditions, we include additional indicators for firm life-cycle stages (Dickinson (2011)), with the indicator for mature stage being the omitted base-stage.

We view the decision to include of a list of FLS keywords as a structural decision, which is unlikely to change periodically. However, firms may decide to include or eliminate keywords in response to structural changes in economic circumstances, despite the fact that theory suggests the cost of abandoning a keyword list may be higher (Einhorn and Ziv 2008). Regardless, we expect stickiness in the decision to include a list. As a result, our probit regression results are more likely to reflect the structural decisions rather than any time-varying changes. To address this, we exploit the structural choice argument in our empirical analysis by examining the disclosure behavior of different sub-groups of firms.

We acknowledge that our models are not designed to necessarily identify a causal relationship. Variables such as *Liberal Court* and lagged industry adoption are plausibly exogenous (or at least pre-determined), so identification issues such as reverse causality are unlikely. Similarly, without an economic incentive to supply forward-looking information in the first place, management is unlikely to provide guidance solely due to the inclusion of a list of FLS keywords. We do not directly test for time-series variations in the consequences of firms' decision regarding the inclusion of a keyword list. For instance, it is possible that analyst following and institutional ownership change in predictable ways following the inclusion or elimination of a keyword list. We hope to explore such possibilities and other plausible causal effects in future revisions.

Conditional on including a list of keywords, we estimate the following Poisson regression (baseline regression) to test our predictions regarding the firm, industry, and macro characteristics that are associated with the number of keywords to identify forward-looking

statements:¹⁰

$$\begin{aligned}
\# \text{ Keywords}_{it} = & \alpha + \beta_1 \text{ Liberal Court}_{it} + \beta_2 \text{ Legal}_{it} + \beta_3 \Delta \text{ Log}(10\text{-K Size})_{it} \\
& + \beta_4 \text{ Log}(10\text{-K Size})_{it-1} + \beta_5 \text{ Log}(\text{Age} + 1)_{it} + \beta_6 \text{ R\&D}_{it} + \beta_7 \text{ Loss Year}_{it} \\
& + \beta_8 \text{ Acquisition Sales Ratio}_{it} + \beta_9 \text{ ROA Deciles}_{it} + \beta_{10} \text{ IPOed within 3yrs}_{it} \\
& + \beta_{11} \text{ Industry Avg. \# KW}_{it-1} + \beta_{12} \text{ Big N Auditor}_{it} + \beta_{13} \text{ CEO Change}_{it} \\
& + \beta_{14} \text{ CFO Change}_{it} + \sum \delta_j \text{ Controls}_{it} + \epsilon_{it}
\end{aligned}$$

We retain many of the variables from the probit specification. We substitute the average number of keywords within the firm’s industry (*Industry Avg. # KW_{t-1}*) as the disclosure herding proxy. Unlike the decision regarding the inclusion of an FLS keyword list, it is rational to expect that managers could update the list of FLS keywords depending on informational demand and supply consideration or other changes in their FLS disclosures. As a result, we include a series of additional explanatory variables to reflect such time-varying considerations. To capture any differential demand for forward-looking information of firms that are new to the equity market, we include an indicator variable for firms that had an IPO within the last three years (*IPOed within 3yrs*). We include two executive turnover variables (*CEO Change* and *CFO Change*) and two returns-based metrics (*Return Volatility* and *Return Skewness*) to capture short-term demand effects of shocks to business conditions. We consider regression specifications with (market) and without (baseline) the market-based metrics as the inclusion of the two variables significantly restricts our sample size. Managers’ reliance on the number of keywords to convey forward-looking information could vary with firm performance, so we include *ROA Deciles* as a final control.

¹⁰We do not believe this is a case of self-selection, which requires well-defined counterfactual outcomes in the two states. Without a list of keywords, the choice of the number of keywords is moot. Given that the decision to have a list of keywords and the number of words to include are made by the same economic agent, we could estimate a single Tobit regression by including zeroes for the number of keywords for firm-years without a list. The results of this Tobit regression are available upon request. While the Tobit results are comparable to our probit/Poisson regression results, they cannot provide differential insights regarding the two dimensions of the decision: inclusion of the list and the choice of the number of keywords.

IV. RESULTS

Descriptive Statistics

Table II presents descriptive statistics on our sample. Panel A shows the rate of FLS keyword list adoption over time, distinguishing between the three sub-periods that we examine separately in our analyses below. For each fiscal year, we report the number and row percentage of firms that include an FLS keyword list in their 10-K filing. The portion of firms with an FLS keyword list has been steadily increasing, starting with only 13% of the firms one year after the passage of the PSLRA, and reaching 91% by 2017.

Panel B presents the portion of firms with FLS keywords by industry (based on the Global Industry Classification Standard codes, or GICS).¹¹ The heading “All Years” reports the number (*Obs*) and percentage of observations (*Obs %*) for each of the twenty four industry groups in our sample period. Each of the remaining headings shows the rate of FLS keyword adoption during a subperiod, with *Obs %* representing the number of firm-years within each industry group during a given subsample as a percentage of the total number of firm-years for that subperiod. Column *KW %* shows the percentage of firm-years within each industry group that has an FLS keyword list during the subperiod. The distribution of industries is generally stable across the sub-periods, although there are some inter-temporal variations. Pharma and Energy sectors represent a larger proportion of the sample in the last sub-period, with consumer durables and technology equipment declining by more than two percentage points.

The dispersion in FLS keyword adoption has decreased dramatically from the first sub-period (29.9 to 52.6) to the last (82.7 to 92.8). Consistent with this result, although all industry sectors experience a large increase in the rate of inclusion of a keyword list, there are notable differences. While the inclusion rate increased by more than 60 percentage points

¹¹We first collect GICS codes from CRSP’s Compustat dataset. We substitute the missing codes with back and forwardfilled missing GICS codes from Compustat’s Compustat’s Co hmic dataset. We make sure that the fiscal year end of the firm is between the effective dates of the dataset GICS codes (*indfrom* and *indthru*). Lastly, we substitute the remaining missing codes with back and forwardfilled GICS codes from Compustat’s Company dataset.

in the Materials sector, it was less than 40 percentage points in semiconductors. Although there is an increase in $KW \%$ across all three sub-periods, the inter-industry variation has declined substantially in the second sub-period, with a coefficient of variation of 0.053 that is much closer to that of the third sub-period than the first.

Panel C provides the average number of keywords ($\# KW$) by industry group for each of the subperiods, for the subsample of firms that include FLS keywords. Overall, the number of keywords has increased from an average of 7 words to almost 12 words by the end of our sample period, with comparable increases across all industry sectors. This is further substantiated by a comparable coefficient of variation in inter-industry means of number of keywords across the three sub-periods. In other words, the inter-temporal increase in the number of keywords has not altered its cross-industry variation.

[Table II about here.]

Table III Panel A provides univariate statistics for our main explanatory variables by FLS keyword list inclusion. The t-statistics on the differences are calculated with standard errors clustered by firm. The differences between firms with and without FLS keywords are largely in the expected directions (with the exception of $\Delta \text{Log}(10 - K \text{ Size})$ and *Acquisition Sales Ratio*). Firms with FLS keyword lists are more likely to have analyst following and larger institutional ownership. Mature firms are less likely to have a keyword list. Panel B reports the Pearson correlation matrix of the explanatory variables.

[Table III about here.]

FLS Keywords Inclusion Determinants

Table IV provides results of probit regressions for the baseline and expanded models. Given that the inferences are largely identical, we focus on the expanded regression and highlight circumstances when the results differ. Among the litigation variables, only *Liberal Court* is statistically significant, consistent with firms in liberal circuits facing increased risk of

class-action lawsuits, and therefore seeking safe harbor protection by using a keyword list to explicitly identify FLS. Variables proxying for disclosure supply, firm uncertainty, and disclosure herding are all statistically significant in the predicted directions. Firms providing forward-looking guidance and with large volume of disclosure choose to have a keyword list, likely to facilitate their equilibrium disclosure strategy. The behavior of R&D intensive firms, firms engaged in larger acquisitions, and those experiencing losses is consistent with the need for FLS to help investors resolve uncertainty. A firm’s propensity to adopt a list is strongly associated with that of its industry peers, providing support for the posited herding behavior.

We find a significant positive relation between analyst following and the inclusion of FLS keywords, but an insignificant association with the overall level of institutional ownership. However, when we decompose the institutional ownership by types, we find that firms with larger transient ownership specifically are more likely to seek increased legal protection for their forward-looking information.¹² To the extent that FLS improve price efficiency and mitigate trading costs, transient institutions are more likely to gravitate towards such stocks (Bushee and Noe 2000). Firms at the end of the life cycle spectrum show a need for protecting FLS, consistent with the investors’ concerns about residual value.

As we discussed above, the choice of including a list of keywords is a structural decision. To better understand its determinants, we divide the sample into firms that have one keyword disclosure policy throughout their inclusion in our sample period (persistent adopters) and those that switched their policy at least once (switchers). We further subdivide the latter group into those who switch only once in our sample period (structural switchers) and those who switch more than once (transient switchers). The sub-sample regression results are provided in columns (3) through (6) of Table IV.¹³

¹²For sake of parsimony, we include only transient ownership in our regression, although none of the other categories of institutional ownership (dedicated, quasi indexers, and other) significantly loads in the expanded regression.

¹³The propensity to switch to include a keyword list likely increases over time, and by definition, firms age over time. To avoid any spurious effects, we exclude $\text{Log}(\text{Age} + 1)$ from the switchers regressions (columns (4) through (6) of Table IV). Given that the focus is on inter-temporal switching, we also exclude $10b-5$ *Firm* and *FPS* from these regressions as they are time invariant. None of the three variables is significant for the structural or the transient switchers. When we include the three time invariant indicators, $R\&D$

The amount and growth of disclosure is significantly associated with the inclusion of the keyword list in all sub-samples. Whether the reasons for the disclosure policy are structural or transient, firms that disclose more are prone to obtain explicit legal protection for FLS. Among the sub-samples, the Pseudo R-squared increases by roughly 80% when we focus only on the persistent adopters (19.6% to 35.2%). Unlike the results for the overall sample, persistent adopters with litigation exposure (*10b-5 Firm*) or with legal expertise in the senior management (*Legal*) are keen to protect their FLS from frivolous shareholder litigation. The latter is consistent with the finding in Kwak, Ro, and Suk (2012) that legal expertise in top management increases the likelihood of management forecasts, thereby supporting the need for their legal protection. However, the existence of the transitory measures: losses, M&A intensity, and disclosure herding do not explain the disclosure policy choice of persistent adopters, which is expected. For these firms, litigation risk, disclosure supply, firm age, and innovation uncertainty are factors that influence their decision to include a keyword list initially.

In contrast, structural switchers are likely to be swayed by the propensity of their industry peers to obtain legal protection. In addition, occurrence of losses and increased M&A intensity further contributes to their desire to adopt a keyword list. Among the control variables, the significance of *Analyst Following* in the overall sample is largely due to its effect among the sample of structural switchers. Taken together, by subdividing our sample, we show that while fundamental firm characteristics and litigation risk influence the behavior of unitary adopters, disclosure herding and changing business conditions influence the structural switchers.¹⁴

[Table IV about here.]

Given the steep inter-temporal trend in the adoption behavior, we re-estimate our expanded regressions for each of the three sub-periods; the results are reported in Table V. The purpose

becomes significant at the 0.10 level with a negative coefficient for transient switchers, i.e., R&D firms are less likely to be stopgap adopters.

¹⁴We have not fully explored the disclosure behavior of transient adopters, but we find that these firms are more likely to be in the declining stage of their life cycle.

of this analysis is to test for varying effects of the disclosure policy determinants over time to better understand the steep increase in the adoption behavior over our sample period. Disclosure supply and firm age are significant in all sub-periods.¹⁵ More importantly, disclosure herding documented in Table IV is confined to the first sub-period. In other words, any equilibrium adjustments in response to peer behavior occurred in the early years after the passage of PSLRA, as reflected in the structural switchers sample in Table IV. In Table II Panel B we show that inter-industry variations in the adoption behavior reached “steady state” by the second sub-period, consistent with the disclosure herding behavior being limited to the first sub-period. Overall, the sub-period analysis sheds additional light on the determinants of the inter-temporal adoption behavior.

[Table V about here.]

Determinants of Number of Keywords

After investigating the determinants of the decision to include an FLS keyword list, we then turn to the determinants of *how many* keywords to include. The distribution of keywords varies widely across firms, as demonstrated in Figure I.

[Figure I about here.]

Table VI reports results of the determinants model for the number of keywords. The Poisson regressions are limited to those firm-years with an FLS keyword list. Specifications (1) and (2) include the full sample of firm-years, and specifications (3) and (4) re-estimate the same models on the sub-sample of firms that consistently have keywords every year they appear in our sample (i.e. those firms from the Persistent sample from Table IV that *always* disclose an FLS list). Given litigation risk can plausibly influence both the inclusion of a keyword list and the number of keywords, we include two of the litigation variables in

¹⁵The one exception is the lack of significance for management guidance in the first sub-period, which could be due to the higher likelihood of measurement error in the guidance database in early years (Chuk, Matsumoto, and Miller 2013).

Table VI regressions that were significant in the keywords inclusion model. We exclude a number of variables that are more likely to be associated with the structural decision to include a keyword list.¹⁶

In the full-sample, measures of court ideology, disclosure supply, uncertainty, and disclosure herding are significantly associated with the number of keywords, with court ideology losing significance in the sub-set of firms that always have keywords. In other words, while the decision to always have a list of keywords was driven by court liberalism, the choice of the number of keywords is not. However, while the keywords inclusion decision is unrelated to auditor choice, clients of Big-N firms include a greater number of keywords. We include several proxies for business circumstances that could impact the need for altering the number of keywords. Firms that are new entrants to the capital markets face heightened uncertainty (Lang (1991)), and therefore, have increased incentives to convey forward-looking information to mitigate information frictions. Executive turnover likely occurs when firms experience business volatility, and may require management to provide a clearer picture of the business prospects. Consistent with these expectations, firms that had a recent IPO or a change in CFO have a larger keyword list, although we observe no effect for a CEO turnover in specifications (2) and (4).

We also consider several accounting and stock return measures as proxies for the level and volatility of current performance. Specifically, the number of keywords increases with the negative skewness of stock returns and with accounting rate of return. Moreover, return volatility explains the variations in the number of keywords of firms that persistently have a list. Taken together, in addition to some of the structural factors that explain the inclusion of a keyword list, several circumstantial factors significantly influence the managerial choice of how many keywords will optimally convey forward-looking information to market participants. Overall, neither the inclusion of the list nor the choice of the number of keywords is boilerplate nor ad hoc. Our analysis identifies a set of firm, industry, and macro characteristics that help

¹⁶Specifically, we do not include *10b-5 Firm*, *FPS*, *Guidance*, *Analyst Following*, and *Trans_Own*. If we do include all these variables, none have significant coefficients.

explain one key aspect of managerial disclosure strategy for communicating value relevant information to help investors assess the firm’s future prospects.

[Table VI about here.]

Factor Analysis of FLS Keywords

So far our analysis has focused on the number of keywords used by firms with the intent of obtaining legal protection for forward-looking information. We next turn to understanding the choice of keywords that firms include in their list. Table VII Panel A reports the top 20 stemmed words used by firms in their keyword lists. We limit it to 20 words as that captures more than 90% of the total words in our population, and no other words are included in more than five percent of the 10-K reports in our sample. While our research design is meant to isolate the determinants of the most popular keywords, it does not help us understand unique words that firms may use in unique circumstances. We plan to study the properties of sporadically used words in future versions of the manuscript.

To study the determinants of keyword choice, we create 20 indicator variables equal to one if the firm includes the given word in its FLS keyword list. We then conduct an exploratory factor analysis to potentially identify underlying disclosure attributes these different words convey. Because we model the words as indicator variables, we use tetrachoric correlations as input to the factor analysis. We report the top-ten Eigenvalues in Panel B and, limiting to values greater than one, consider five factors that together account for roughly 77% of the total variance.

Panel C reports the factor loadings based on a varimax rotation. We find the factor loadings to be quite intuitive, and label the five factors as follows. Using an ad hoc rule, we focus on keywords with factor loadings of at least 0.60 to label the underlying disclosure attributes. The disclosure attributes that we identify are “future state” (factor 1), “likelihood” (factor 3), “quantitative” (factor 4), and “future action” (factor 5). In Panel C we provide the appropriate definitions from the Merriam-Webster online dictionary as a basis for the

chosen labels. Factor 2 represents modal verbs, which are usually used with another verb to express different ideas such as possibility (might or may), probability or necessity (must), promise or willingness (will), preferences or desires (would), ability (can), possibility (could), and uncertain prediction (should).

Before we study the determinants of the identified disclosure attributes, in Panel D we provide the inter-temporal incidence of the use of the top-20 keywords. Given that the average number of keywords in 10-K reports has increased from less than six to over 12 in our sample (see Figure IV), Table VII Panel D shows which keywords and disclosure attributes contribute to this increase. The three words that capture the disclosure attribute of likelihood (“expect,” “anticipate,” and “believe”) are the most highly used words in the keyword list throughout the sample period. Managers seem most concerned with legal risk when expressing the likelihood of economic activities or events, and therefore, have taken efforts to protect such FLS since the inception of the PSLRA.

To discuss future state, firms largely rely on “estimate” throughout the sample period, with increasing tendency to use other words representing the disclosure attribute (“predict,” “potential,” and “continue”). In terms of words that facilitate description of future actions, firms have dramatically increased the use of the words “intend” (37.4% to 87.5%) and “plan” (29.6% to 82.4%) with limited use of “seek.” Apart from a meaningful increase in the use of the word “project,” firms are using other words that could convey quantitative information much less frequently. The use of modal verbs has substantially increased over time, indicating a desire to identify potential uncertain FLS for legal protection. Overall, while firms have always used keywords to convey the likelihood of economic activities or events, they have increased reliance on words that can convey future state, future actions, and the uncertain nature of FLS.¹⁷

¹⁷Our analysis is based on keywords from 10-K reports. Given earnings releases are more likely to include quantitative FLS, we might see an increased incidence of those keywords. We provide some preliminary evidence in Table IX on the commonality of keywords in 10-K reports and Q4 earnings releases (filed in a Form 8-K). We intend to conduct additional analysis on the two sets of keywords and the reasons why they may differ.

[Table VII about here.]

We next examine the association between the disclosure attributes and the determinants of the overall number of keywords from Table VI. We expect that the segregation of keywords by disclosure attribute can shed light on opposite forces that may be in effect at times. For instance, circumstances that may require quantitative FLS may be orthogonal to cases when firms may discuss more qualitative plans. We measure each of the disclosure attributes as the number of keywords with factor loading of at least 0.6 in the list of keywords. For instance, we measure the disclosure attribute “Future Action” as the sum of the indicator variables *intend*, *plan*, and *seek*.

The results of Poisson regressions of each of the disclosure attributes on the determinants from Table VI are presented in Table VIII Panel A, with column (1) reporting the results for the total number of keywords.¹⁸ Consistent with a lack of inter-temporal variation in the words that describe the likelihood of events or activities, the pseudo R-squared for specification (4) is 0.1 %. Only when managers issue guidance, or experience loss, do they more often include such words.

For the remaining attributes, we focus on key empirical regularities that shed light on the incentives to include the related words. When firms face higher ex-ante litigation risk due to court ideology, they include a list of keywords that can protect them from FLS used to describe future state or provide quantitative information. Obviously, such statements are more likely to trigger a class-action lawsuit if the expectation is not realized. Managers that frequently provide quantitative guidance may be compelled to explain future actions that could help achieve the guidance. Consistent with this intuition, we find *Guidance Intensity* to be significantly associated with the keywords that convey quantitative information or future action.

¹⁸We do include *Guidance* in these regressions, which was omitted from Table VI as we did not expect it to explain the number of keywords. Our column (1) result confirms that belief. In addition, we also include *Guidance Intensity* to capture the number of different guidance forecasts provided by managers during a fiscal year.

Older firms with longer histories should have better forecasting ability, which is consistent with firm age having a positive loading only when explaining the choice of quantitative words. Firms facing higher stock return volatility are less likely to include quantitative words in their keyword list due to the uncertain environment that may deter the disclosure of quantitative FLS. Consistent with a desire to fall back on qualitative disclosures, they instead include keywords that can describe future state or future action, thereby conveying relevant forward-looking information. Interestingly, several of the determinants that are not significantly related to the total number of keywords (e.g., *Guidance*, *Guidance Intensity*, and *Return Volatility*) are associated with some of the disclosure attributes in meaningful ways. Taken together, our analysis provides compelling evidence on the selective choice of different disclosure attribute keywords that managers use to obtain legal protection under different economic circumstances.¹⁹

[Table VIII about here.]

Comparison of 10-K vs. 8-K FLS Keyword Lists

Although firms can use multiple disclosure channels to convey forward-looking information, our analysis so far focused on protection of FLS in 10-K reports. Prior research finds that managers use earnings press releases to convey value-relevant forward-looking statements (Bozanic et al. 2018). Beginning March 28, 2003, SEC registrants must file all publicly-issued earnings press releases in a form 8-K.²⁰ Given the importance of earnings press releases, we

¹⁹Our measures are derived from the managerial perspective of choosing what specific disclosure attributes to include in the list of keywords. The outcome of the managerial decision is a set of binary variables that captures the decision to include or exclude a list of keywords. In contrast, the purpose of an exploratory factor analysis is to reduce the dimensionality by identifying several factors that are linear combinations of a large set of variables, but with the objective of explaining as much of the variance in the system as possible. Although our measures are designed to reflect managerial decision making, we also rerun our regressions using the factor scores as the dependent variables and report the results in Table VIII Panel B. The results in the two panels lead to similar inferences in most cases. The exception is *Return Volatility* which has the correct signs in the Quantitative and Future Action regressions, but is insignificant in both.

²⁰See SEC Release Nos. 33-8176; 34-47226; FR-65.

next examine firms' disclosure of an FLS keyword list in earnings press releases and report the results of our preliminary analysis.²¹

Using the same scrapping strategy as that for 10-K reports, we identify and download all 8-K reports that include an earnings press release, and extract any keyword list included therein. Given that earnings releases are generally filed with the SEC only after March 27, 2003, this results in a much smaller sample size for our analysis. Untabulated data shows that we have 59,241 firm-years with 10-K reports and the corresponding earnings release. When earnings release includes a list of keywords (n=22,497), roughly 95% of the corresponding 10-Ks include a list as well. However, when the 10-K includes a keyword list (n=48,776), only 44% of the earnings releases contain a list.

Our analysis is based on 21,346 firm-year observations that contain FLS keyword lists in both the 10-K and Q4 8-K earnings release. Figure II provides the distribution of the number of keywords contained in each report. Although the distributions are reasonably similar, the 10-K pdf is slightly shifted to the right. Figure III plots the cdf of the number of 10-K keywords minus the number of 8-K keywords. About 40% of the observations have the same number of keywords, 40% have more keywords in the 10-K, and the remaining 20% have fewer. The distributions of the FLS list adoption and the mean number of FLS keywords in the lists of 10-Ks vs. 8-Ks are shown in Figure IV. While the mean number of keywords in 10-Ks vs. 8-Ks follow a similar trend over the years, the 8-K mean is consistently lower and does not appear to converge to the 10-K mean over time. This suggests that there are persistent systematic differences in the disclosures. The graph also shows that the adoption rate of 8-Ks has a more linear pattern than for 10-Ks. Figure V plots the mean of the number of 10-K and 8-K keywords by year, as well as the number of keywords that overlap in the two sets. All three means increase almost monotonically over time with parallel trends.

[Figures II–V about here.]

²¹Because 8-K earnings releases are usually “furnished,” not “filed” with the SEC, they are not subject to liability under Section 18 of the Exchange Act, though they still face liability under Rule 10b-5.

We next provide preliminary evidence on the determinants of the similarity and differences between the two lists of keywords. We consider three measures: (1) Jaccard similarity index; (2) absolute value of the difference between the number of keywords; and (3) number of 10-K minus number 8-K keywords. In Figure VI we provide inter-temporal descriptive statistics for the Jaccard similarity index. At least a quarter of the firms have identical word lists, with mean (median) of the similarity index consistently above 0.7 (0.8). In addition, we find little time-series variations in the distribution of the Jaccard index.

[Figure VI about here.]

In our regressions, we consider a limited set of variables from our list of determinants that we expect would explain the variations in these three measures. In addition, we include the ratio of the length in characters of earnings release to 10-K release as an additional explanatory variable.

[Table IX about here.]

Results in Table IX indicate that firms with greater ex-ante litigation risk tend to use the same set of words in the two disclosure channels. Consistent use of FLS keywords could avoid the potential for litigation risk increasing due to ambiguity. Older firms, firms with a Big-N auditor, or firms that issue longer earnings press releases tend to use more similar words and have similar number of words in the two lists. In contrast, when firms face an uncertain environment (*R&D*, *Loss Year*, and *Acquisition Sales Ratio*), there is more divergence in both length and word choice between the two keyword lists. Overall, our preliminary analysis provides some insight into the managerial decision when choosing the list of keywords in two important disclosure channels. It also provides further evidence to the claim that these keyword lists, both the inclusion decision and composition, are not mere boilerplate disclosures.

V. CONCLUSION

This study provides evidence on the determinants of one aspect of voluntary disclosures, namely the identification of forward-looking statements. Firms have increasingly adopted the practice of listing keywords which they use to signal forward-looking statements, and thereby afford them protection under the PSLRA's 'safe harbor' provision. We study the determinants of including an FLS keyword list, as well as the linguistic contents of this disclosure. Our findings generally support the claim that these lists and their content are chosen based on economic factors, firm characteristics, and macro forces, providing little credence to the claim that this is a boilerplate disclosure.

We find that the inclusion of an FLS keyword list has increased from 13% to 91% over the past 22 years. Most of that adoption occurred in the earlier years, and has since seemed to reach steady state. This adoption is associated with ex-ante litigation risk, the extent of mandatory and voluntary disclosure, firm-specific uncertainty, and bandwagon effects. Further, litigation risk and bandwagon effects are concentrated in the early years of adoption, suggesting that the forces underlying the disclosure decision evolved over time.

Investigation into the content of the list reveals that similar but distinct forces are associated with the two decisions of *whether* to include a list, and *what* to include in the list. We find that ex-ante litigation risk is strongly associated with the adoption of more keywords, and the length of the keyword list is increasing in firm-specific uncertainty and performance. Again we find evidence of potential bandwagon effects at play. A content analysis of *what* words managers choose reveals variation in the language used. We find that in general, the choice of language firms employ varies systematically with firms' economic circumstances.

We currently plan on extending the project to include more in-depth linguistic contextual analysis of forward-looking statements, as defined by the inclusion of one or more forward-looking words. Prior literature has studied, among other things, the tone and sentiment of these sentences. However, it is still unclear the extent to which specific word usage choices reflect informational differences in the signal being conveyed by managers. We also intend

to extend our study to include other disclosures which are afforded protection under the PSLRA, specifically earnings announcements and other press releases. By comparing these two sources of disclosure, one significantly less discretionary than the other, we hope to gain insight into the specific linguistic choices managers make as a function of the regulatory burden.

References

- Armstrong, C. S., S. Glaeser, and J. D. Kepler. 2019. Strategic Reactions in Corporate Tax Planning. *Journal of Accounting and Economics* 68 (1): 101232.
- Bikhchandani, S., D. Hirshleifer, and I. Welch. 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *The Journal of Economic Perspectives* 12 (3): 151–170.
- Bloomfield, R. 2008. Discussion of “Annual Report Readability, Current Earnings, and Earnings Persistence”. *Journal of Accounting and Economics* 45 (2–3): 248–252.
- Bochkay, K., R. Chychyla, and D. Nanda. 2016. Dynamics of CEO Communication Style. *Working Paper* 57.
- Bozanic, Z., J. R. Dietrich, and B. A. Johnson. 2017. SEC Comment Letters and Firm Disclosure. *Journal of Accounting and Public Policy* 36 (5): 337–357.
- Bozanic, Z. and P. Kraft. 2018. Qualitative Corporate Disclosure and Credit Analysts’ Soft Rating Adjustments. *Working Paper* .
- Bozanic, Z., D. T. Roulstone, and A. Van Buskirk. 2018. Management Earnings Forecasts and Other Forward-Looking Statements. *Journal of Accounting and Economics* 65 (1): 1–20.
- Bushee, B. J. 1998. The Influence of Institutional Investors on Myopic R&D Investment Behavior. *The Accounting Review* 73 (3): 305–333.
- Bushee, B. J. and C. F. Noe. 2000. Corporate Disclosure Practices, Institutional Investors, and Stock Return Volatility. *Journal of Accounting Research* 38: 171–202.
- Chivers, C. and G. Quinn. 2009. Recent Court Case Confirms the Importance of Well Drafted Forward-Looking Statement Disclosures. *Finance Digest* 1–9.
- Chuk, E., D. Matsumoto, and G. S. Miller. 2013. Assessing methods of identifying management forecasts: CIG vs. researcher collected. *Journal of Accounting and Economics* 55 (1): 23 – 42.
- Dickinson, V. 2011. Cash Flow Patterns as a Proxy for Firm Life Cycle. *The Accounting Review* 86 (6): 1969–1994.
- Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2): 221–245.
- Einhorn, E. and A. Ziv. 2008. Intertemporal Dynamics of Corporate Voluntary Disclosures. *Journal of Accounting Research* 46 (3): 567–589.
- Field, L., M. Lowry, and S. Shu. 2005. Does Disclosure Deter or Trigger Litigation? *Journal of Accounting and Economics* 39 (3): 487–507.

- Francis, J., D. Philbrick, and K. Schipper. 1994. Shareholder Litigation and Corporate Disclosures. *Journal of Accounting Research* 32 (2): 137–164.
- Hassanein, A. and K. Hussainey. 2015. Is Forward-Looking Financial Disclosure Really Informative? Evidence from UK Narrative Statements. *International Review of Financial Analysis* 41: 52–61.
- Heinle, M. S., K. C. Smith, and R. E. Verrecchia. 2018. Risk-Factor Disclosure and Asset Prices. *The Accounting Review* 93 (2): 191–208.
- Hirshleifer, D. and S. H. Teoh. 2003. Herd Behaviour and Cascading in Capital Markets: A Review and Synthesis. *European Financial Management* 9 (1): 25–66.
- Huang, A., K. W. Hui, and R. Z. Li. 2019. Federal Judge Ideology: A New Measure of Ex Ante Litigation Risk. *Journal of Accounting Research* 57 (2): 431–489.
- Hutton, A. P., G. S. Miller, and D. J. Skinner. 2003. The Role of Supplementary Statements with Management Earnings Forecasts. *Journal of Accounting Research* 41 (5): 867–890.
- Johnson, M. F., R. Kasznik, and K. K. Nelson. 2001. The Impact of Securities Litigation Reform on the Disclosure of Forward-Looking Information By High Technology Firms. *Journal of Accounting Research* 39 (2): 297–327.
- Jung, M. J. 2013. Investor Overlap and Diffusion of Disclosure Practices. *Review of Accounting Studies* 18 (1): 167–206.
- Kimbrough, M. D. and H. Louis. 2011. Voluntary Disclosure to Influence Investor Reactions to Merger Announcements: An Examination of Conference Calls. *The Accounting Review* 86 (2): 637–667.
- Kothari, S. P., T. E. Laguerre, and A. J. Leone. 2002. Capitalization versus Expensing: Evidence on the Uncertainty of Future Earnings from Capital Expenditures versus R&D Outlays. *Review of Accounting Studies* 7 (4): 355–382.
- Kwak, B., B. T. Ro, and I. Suk. 2012. The Composition of Top Management with General Counsel and Voluntary Information Disclosure. *Journal of Accounting and Economics* 54 (1): 19–41.
- Lang, M. 1991. Time-Varying Stock Price Response to Earnings Induced by Uncertainty about the Time-Series Process of Earnings. *Journal of Accounting Research* 29 (2): 229–257.
- Li, F. 2010. The Information Content of Forward-Looking Statements in Corporate Filings—A Naïve Bayesian Machine Learning Approach. *Journal of Accounting Research* 48 (5): 1049–1102.
- McClane, J. 2019. Boilerplate and the Impact of Disclosure in Securities Dealmaking. *Vanderbilt Law Review* 72 (1): 191–295.

- Muslu, V., S. Radhakrishnan, K. R. Subramanyam, and D. Lim. 2015. Forward-Looking MD&A Disclosures and the Information Environment. *Management Science* 61 (5): 931–948.
- Naughton, J. P., T. O. Rusticus, C. Wang, and I. Yeung. 2019. Private Litigation Costs and Voluntary Disclosure: Evidence from the Morrison Ruling. *The Accounting Review* 94 (3): 303–327.
- Nelson, K. and A. C. Pritchard. 2016. Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors. *Journal of Empirical Legal Studies* 13 (2): 266–297.
- Rogers, J. L. and A. Van Buskirk. 2009. Shareholder Litigation and Changes in Disclosure Behavior. *Journal of Accounting and Economics* 47 (1–2): 136–156.
- Skinner, D. J. 1994. Why Firms Voluntarily Disclose Bad News. *Journal of Accounting Research* 32 (1): 38–60.
- Skinner, D. J. 1997. Earnings Disclosures and Stockholder Lawsuits. *Journal of Accounting and Economics* 23 (3): 249–282.
- Tse, S. and J. W. Tucker. 2010. Within-Industry Timing of Earnings Warnings: Do Managers Herd? *Review of Accounting Studies* 15 (4): 879–914.

Appendix A Variable Descriptions

The following table defines the variables used in this paper. Variable names and calculations provided in brackets correspond to source database. For the regressions presented in the tables, the continuous variables are winsorized at the 1st and 99th percentiles.

Variable	Description
<i>Textual Disclosure Variables</i>	
FLS Keywords Inclusion	A dummy variable equal to one if a firm has a FLS keyword list included in the 10-K during the fiscal year
# Keywords	The number of FLS terms included in the firm's FLS keyword list in the 10-K for that fiscal year. Some FLS terms may include more than one word (e.g. "will likely result," "may affect," "is likely"). For these cases, we do not separately count the words in the term, instead we treat the whole expression as one FLS term/keyword
#10K-#8K	The difference between the number of keywords in the 10-K FLS list and the number of keywords in the 8-K list
8-K/10-K Length Ratio	The ratio of the length of the 8-K to the length of the 10-K in characters
Industry Adoption	Percentage of firms within a GICS industry group that include a FLS keywords list in their 10-K during a fiscal year. We measure this variable for each firm-year, excluding the focal firm from the calculation
Industry Avg. # KW	Average number of keywords in the 10-K FLS list of firms within a GICS industry group during a fiscal year. We measure this variable for each firm-year, excluding the focal firm from the calculation
Jaccard Index	Intersection of the sets of stemmed words in the 10-K and 8-K FLS lists divided by the union of the sets
Log(10-K Size)	Natural logarithm of the length of the 10-K filing text, in characters (bytes)
<i>Accounting Variables</i>	
Acquisition Sales Ratio	The sales contribution of an acquisition divided by total sales { $AQS/SALE$ }
Analyst Following	An indicator variable equal to one if the firm is covered by at least one analyst during the fiscal year
Big N Auditor	An indicator equal to one if the firm engages the services of a Big-N auditor { $AU \in (1, 4, 5, 6, 7)$ }
CEO Change	An indicator variable equal to one if there was a change of CEO during the current fiscal year
CFO Change	An indicator variable equal to one if there was a change of CFO during the current fiscal year
Decline	A dummy variable equal to one if the firm's life cycle is in the <i>Decline</i> stage, as defined by Dickinson (2011)
Age	The age of the firm in years, calculated as the number of years since the firm first appeared in Compustat { $(datadate - MIN(datadate))/365.25$ }
Growth	A dummy variable equal to one if the firm's life cycle is in the <i>Growth</i> stage, as defined by Dickinson (2011)
Guidance	An indicator variable equal to one if the firm issued at least one quarterly or annual management forecast during the fiscal year

Continued on next page...

...continued from previous page.

Variable	Description
Guidance Intensity	Number of quarterly or annual managerial forecasts issued during the fiscal year
II Total	Total percentage of firms' stock owned by institutional investors, as defined by 13-F filings
II Transient	Percentage of firms' stock owned by institutional investors classified as <i>Transient</i> by Bushee (1998)
Introduction	A dummy variable equal to one if the firm's life cycle is in the <i>Introduction</i> stage, as defined by Dickinson (2011)
IPOed within 3yrs	An indicator variable equal to one if the firm's IPO occurred within the last three years, using Jay Ritter's IPO data
Legal	An indicator variable equal to one if the firm employs an executive director, supervisory director, or senior manager whose role title includes any of the following words: "legal," "counsel," "secretary."
Loss Year	A dummy variable equal to one if income before extraordinary items is negative $\{IB < 0\}$
Mature	A dummy variable equal to one if the firm's life cycle is in the <i>Mature</i> stage, as defined by Dickinson (2011)
Out	A dummy variable equal to one if the firm's life cycle is in the <i>Out</i> stage, as defined by Dickinson (2011)
R&D	An indicator variable equal to one if research and development expenses divided by total assets is positive $\{COALESCE(XRD, 0)/AT > 0\}$
Return Skewness	Skewness of daily raw stock returns from the previous to the current fiscal year end
Return Volatility	Standard deviation of daily raw stock returns from the previous to the current fiscal year end
ROA Deciles	Deciles of operating income after depreciation to total assets $\{OIADP/AT\}$
<i>Other Variables</i>	
10b-5 Firm	A dummy variable equal to one if the firm is the subject of a securities litigation at any point in its inclusion in the sample
FPS	A dummy variable equal to one if the firm's industry is included in the Francis et al. (1994) high litigation industries (biotechnology, computers, electronics, and retailing)
Liberal Court	The federal circuit level judge ideology measure developed by Huang et al. (2019) , calculated at the monthly level, and matched to a firm's fiscal-year-end month and headquarters

Appendix B Textual Data Collection

Our sample of forward-looking statement keyword lists was gathered from firms' 10-K filings, which we downloaded from the SEC's EDGAR website. Both the primary 10-K exhibit, as well as any exhibit 13s were searched for a match to the following regular expression (here, in Python syntax):

```
# all possible quotes, equivalent to \p{Pi} and \p{Pf} in PERL
quotes = ("'"'\u00ab\u2018\u201b\u201c\u201f\u2039"
          "\u2e02\u2e04\u2e09\u2e0c\u2e1c\u00bb\u2019"
          "\u201d\u203a\u2e03\u2e05\u2e0a\u2e0d\u2e1d")

# Python regular expression for finding quoted word lists:
quote_regex = (
r"( # begin word capture group, to capture each word in the list
 r"((?<!\w)[\"+quotes+r\"{1,3})" # opening quote, not following a word
 r"(?P<word>[\w ()\[\] <> . , : ; ~ / \ \ | ! ? @ # $ % ^ & * _ = + \ - ] {1,100})" # the word
 r"\s*(["+quotes+r"]{1,3}(?!\w))" # closing quote mark
 r"\s*,?\s*(and/or|and|or)?\s*,?\s*" # maybe comma/and/or delimited
r"){2,}" # close word capture, requiring at least two words
)
# compiled with re.I flag
```

We then require that this quoted list includes one of the following words: “expect,” “expects,” “anticipate,” “anticipates,” “intend,” “intends,” “estimate,” “estimates,” “may,” or “should.” Further, if such a quoted list is not found, we remove the quotation requirement from the regular expression, and re-search for unquoted lists which occur within 1000 bytes before or 2000 after a forward-looking statement phrase. Forward-looking statements were defined as the following phrases: “forward-looking,” “cautionary statements,” “safe harbor,” “private securities litigation reform act,” “PSLRA,” “similar expressions,” or “not historical in nature.”

Appendix C FLS Keyword List Examples

This appendix shows examples of FLS keyword disclosures from 10-K and 8-K filings.

Teleflex Inc. 10-K. CIK: 96943, Filing Date: 2016-02-25.

All statements made in this Annual Report on Form 10-K, other than statements of historical fact, are forward-looking statements. The words “anticipate,” “believe,” “estimate,” “expect,” “intend,” “may,” “plan,” “will,” “would,” “should,” “guidance,” “potential,” “continue,” “project,” “forecast,” “confident,” “prospects” and similar expressions typically are used to identify forward-looking statements. Forward-looking statements are based on the then-current expectations, beliefs, assumptions, estimates and forecasts about our business and the industry and markets in which we operate. These statements are not guarantees of future performance and are subject to risks, uncertainties and assumptions which are difficult to predict. Therefore, actual outcomes and results may differ materially from what is expressed or implied by these forward-looking statements due to a number of factors...

Seachange International Inc 10-K. CIK: 1019671, Filing Date: 2003-05-01.

Any statements contained in this Form 10-K that do not describe historical facts may constitute forward-looking statements as that term is defined in the Private Securities Litigation Reform Act of 1995. These statements relate to future events or our future financial performance and are identified by words such as “may,” “will,” “could,” “should,” “expect,” “plan,” “intend,” “seek,” “anticipate,” “believe,” “estimate,” “potential,” or “continue” or other comparable terms or the negative of those terms. Forward-looking statements in this Form 10-K include certain statements regarding the effect of certain accounting standards on our financial position and results of operations, the effect of certain legal claims against us, projected changes in our revenues, earnings and expenses, exchange rate sensitivity, interest rate sensitivity, liquidity, product introductions and general market conditions. Our actual future results may differ significantly from those stated in any forward-looking statements. Any such forward-looking statements contained herein are based on current expectations, but are subject to a number of risks and uncertainties that may cause actual results to differ materially from expectations. Factors that may cause such differences include, but are not limited to, the factors discussed below. Each of these factors, and others, are discussed from time to time in our filings with the SEC.

Regions Financial Corp 8-K. CIK: 1281761, Filing Date: 2014-07-22.

This release may include forward-looking statements as defined in the Private Securities Litigation Reform Act of 1995, which reflect Regions current views with respect to future events and financial performance. Forward-looking statements are not based on historical information, but rather are related to future operations, strategies, financial results or other developments. Forward-looking statements are

based on managements expectations as well as certain assumptions and estimates made by, and information available to, management at the time the statements are made. Those statements are based on general assumptions and are subject to various risks, uncertainties and other factors that may cause actual results to differ materially from the views, beliefs and projections expressed in such statements. These risks, uncertainties and other factors include, but are not limited to, those described below:

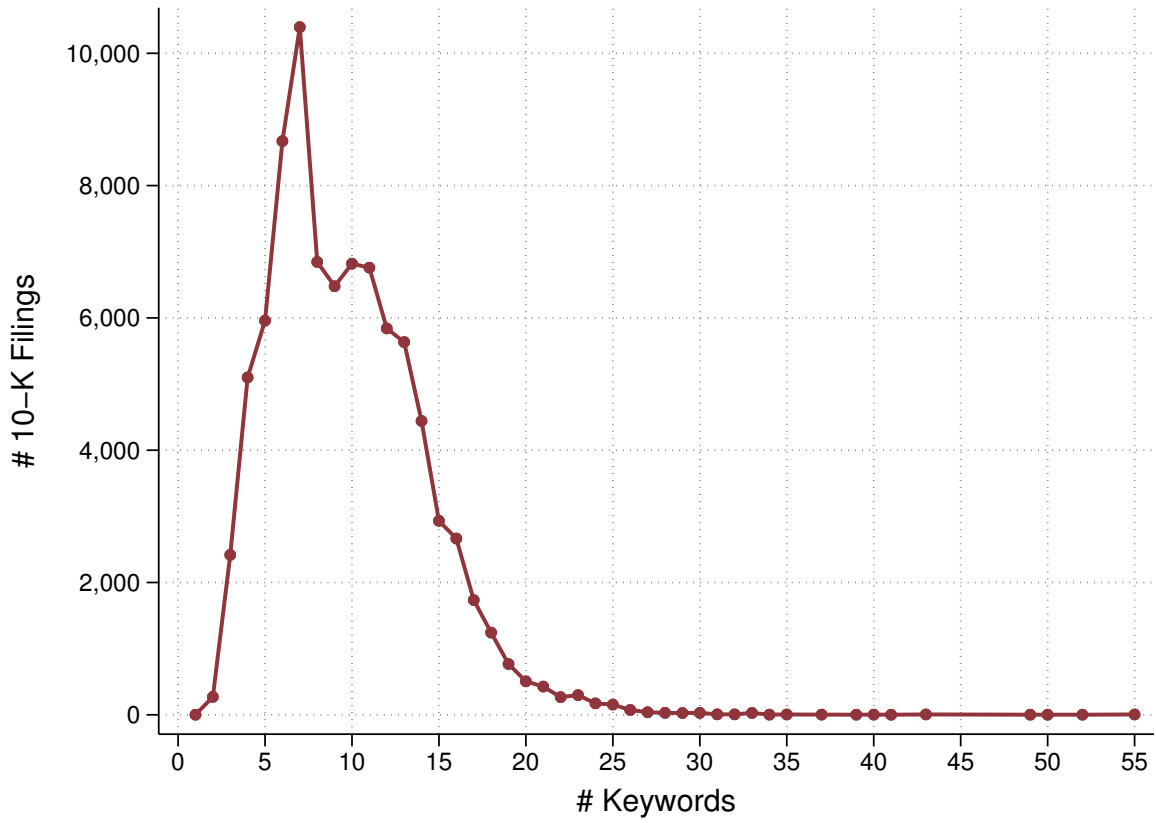
...

The words “anticipates,” “intends,” “plans,” “seeks,” “believes,” “estimates,” “expects,” “targets,” “projects,” “outlook,” “forecast,” “will,” “may,” “could,” “should,” “can,” and similar expressions often signify forward-looking statements. You should not place undue reliance on any forward-looking statements, which speak only as of the date made. We assume no obligation to update or revise any forward-looking statements that are made from time to time.

Lydall Inc. 8-K. CIK: 60977, Filing Date: 2017-10-31.

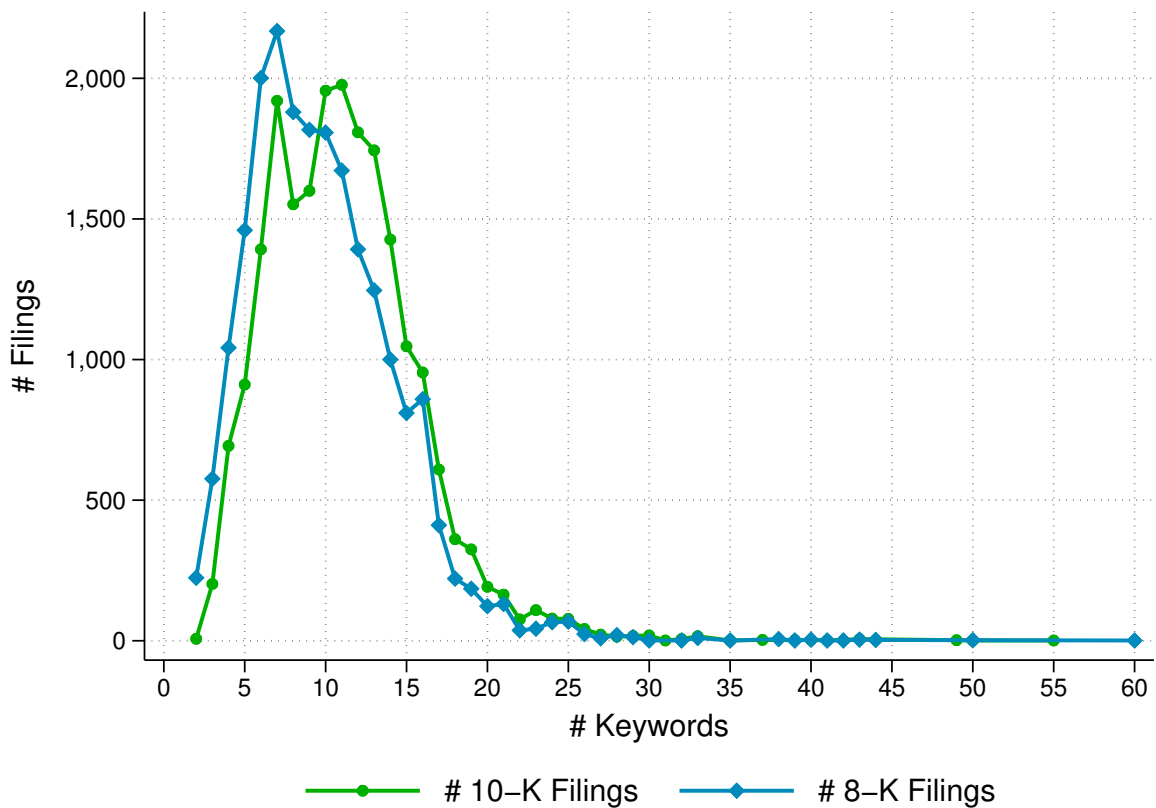
This press release contains “forward-looking statements” within the Private Securities Litigation Reform Act of 1995. Any statements contained in this press release that are not statements of historical fact, including statements about the outlook for the remainder of 2017 and into 2018, may be deemed to be forward-looking statements. All such forward-looking statements are intended to provide managements current expectations for the future operating and financial performance of the Company based on current expectations and assumptions relating to the Companys business, the economy and other future conditions. Forward-looking statements generally can be identified through the use of words such as “believes,” “anticipates,” “may,” “should,” “will,” “plans,” “projects,” “expects,” “expectations,” “estimates,” “forecasts,” “predicts,” “targets,” “prospects,” “strategy,” “signs,” and other words of similar meaning in connection with the discussion of future operating or financial performance. Because forward-looking statements relate to the future, they are subject to inherent risks, uncertainties and changes in circumstances that are difficult to predict. Such risks and uncertainties which include, among others, worldwide economic or political changes...

Figure I: Number of 10-K Keywords Distribution



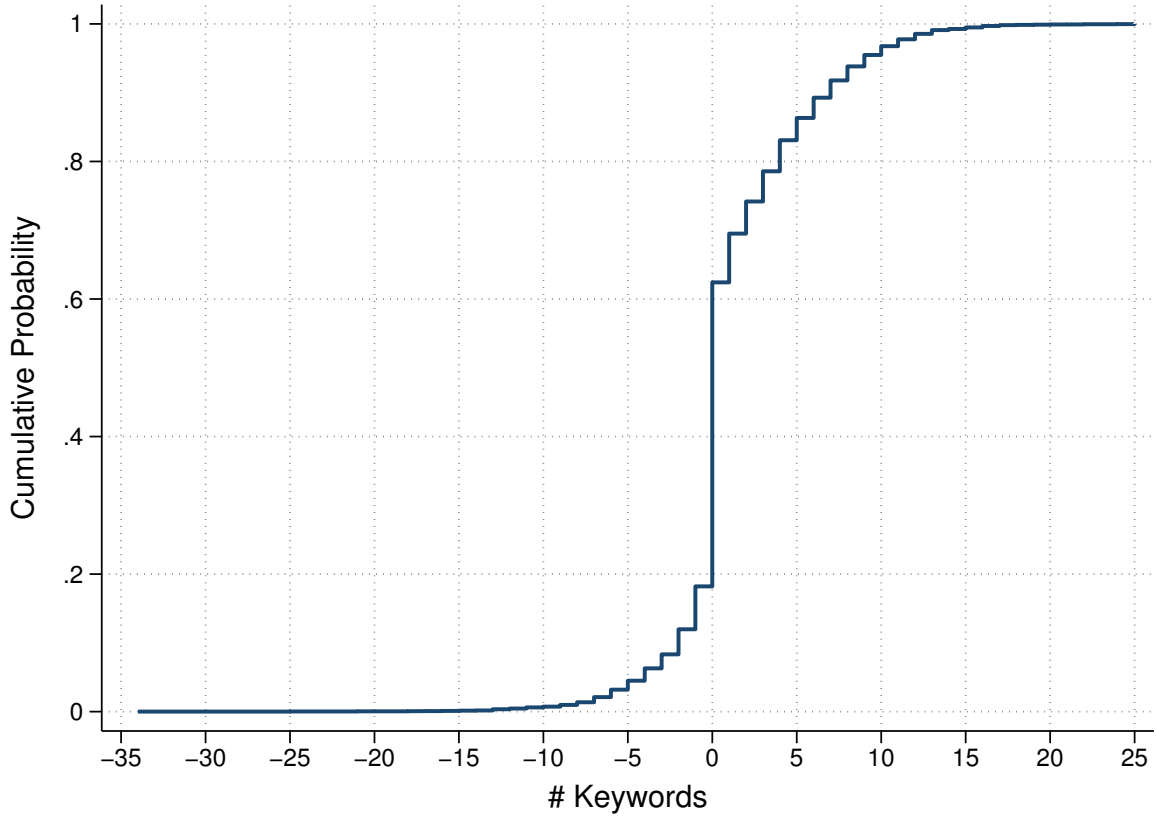
This figure shows the distribution of number of FLS keywords in the 87,079 10-Ks filings with a nonempty FLS keyword list throughout our sample period.

Figure II: Number of Keywords Distribution in 10-Ks and 8-Ks



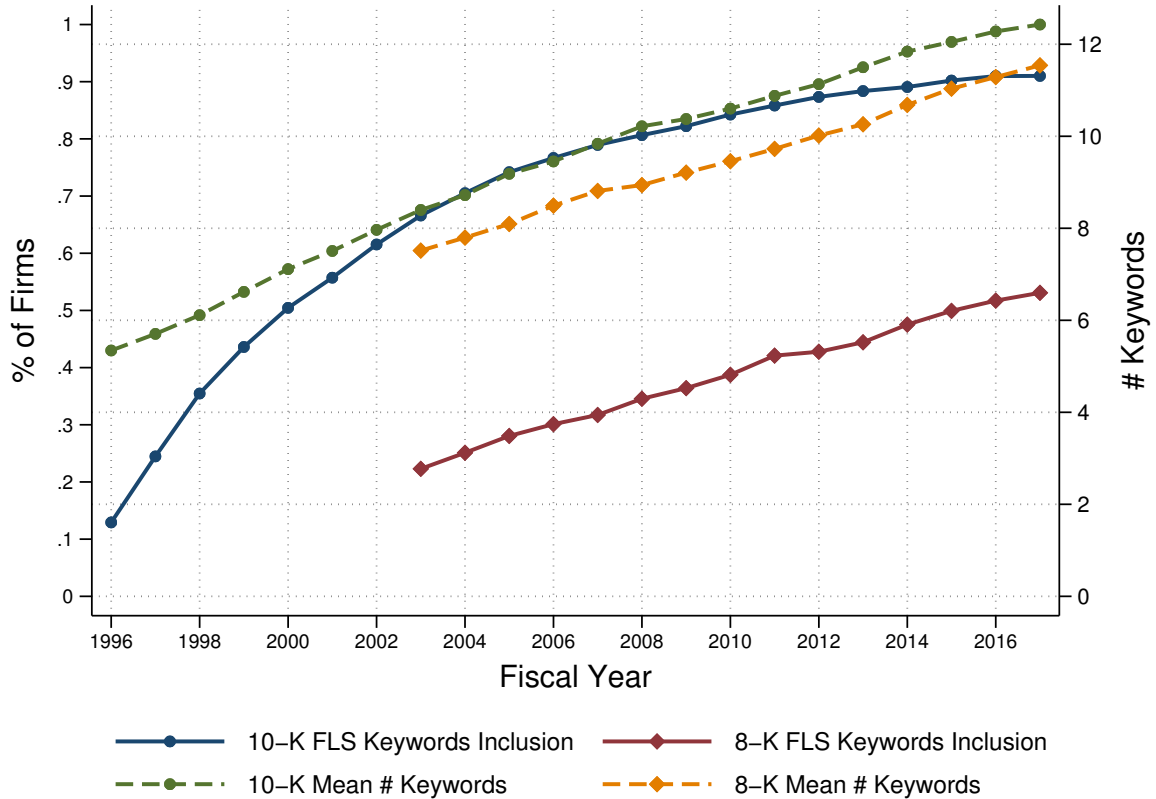
This figure shows the distribution of number of FLS keywords in 10-Ks and 8-Ks for the observations in our sample where filing 10-K was successfully matched to its corresponding 8-K filing. The graph shows only the 21,341 observations where both 10-K and 8-K had an FLS keyword list, and it covers the fiscal years from 2003 to 2017. We exclude fiscal year 2002 because there are only five matched 10-K/8-K pairs with nonempty FLS lists during this year.

Figure III: Cumulative Distribution Function of the Difference in Number of Keywords



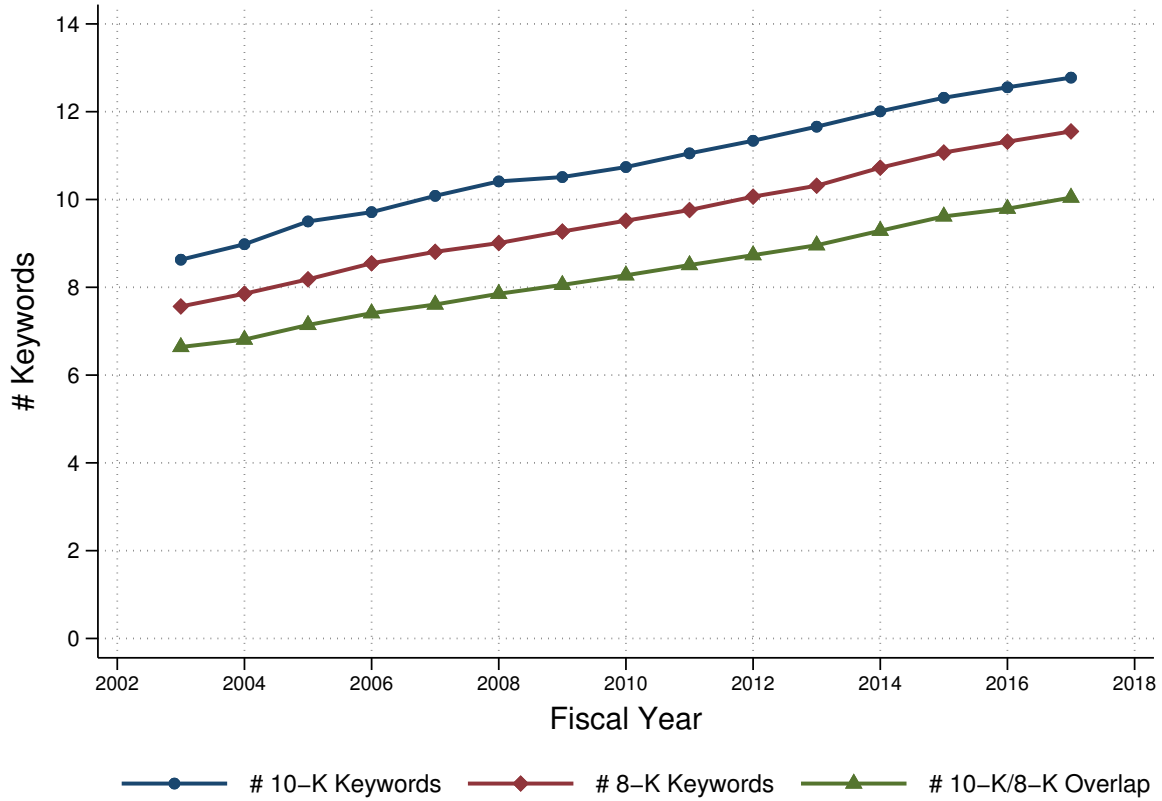
This figure shows the cumulative distribution function of number of FLS keywords in the 10-K filing minus the number of keywords in the 8-K filing for 21,341 of the matched 10-K/8-K pairs where both filings include a FLS keyword list. Since we only have five observations that meet this condition in 2002, we exclude this year. The graph thus covers the sample period from 2003 to 2017.

Figure IV: Distribution of FLS Keywords Inclusion and Number of Keywords by Year



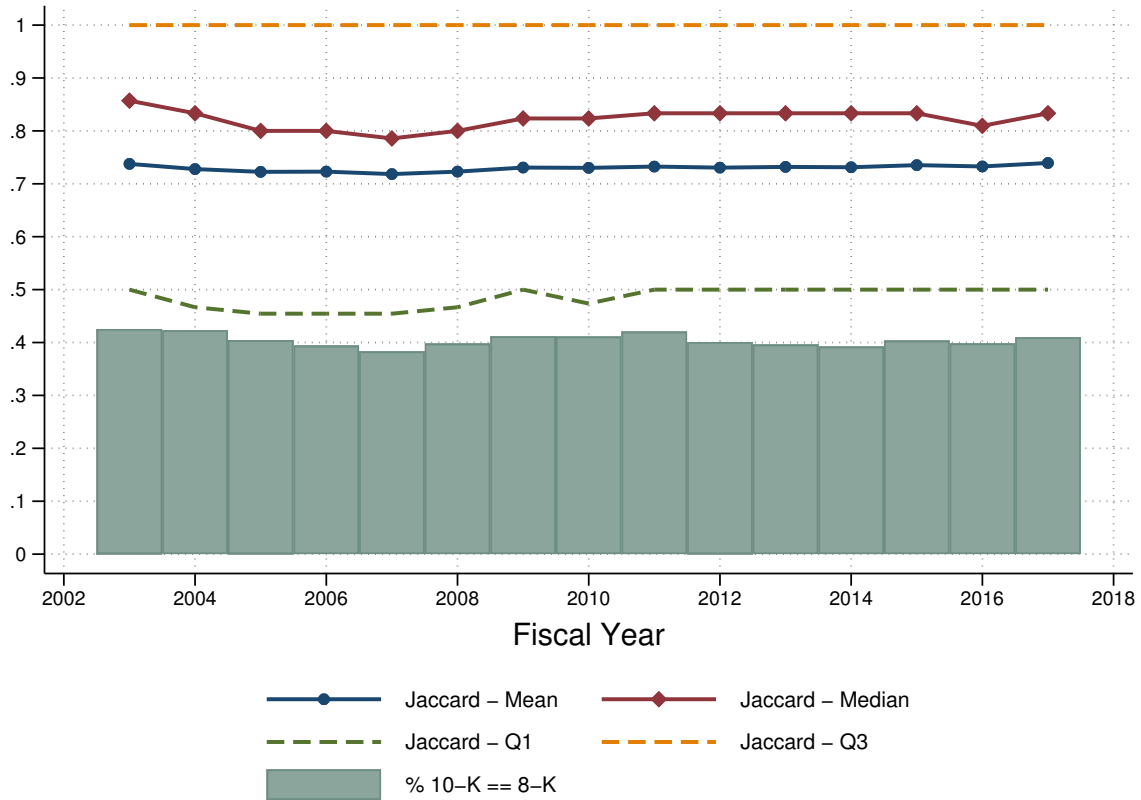
This figure shows the distribution of FLS keywords inclusion and number of keywords in 10-K and 8-K filings throughout our sample period. The percentages of 10-K and 8-K FLS Keywords Inclusion are calculated over our sample of 132,303 10-Ks and 59,241 matched 8-Ks, respectively. The 10-K Mean # Keywords is calculated over the 87,079 10-K filings with a FLS keyword list, while the 8-K Mean # Keywords is calculated over 22,491 8-Ks with a nonempty list. We do not show 8-K statistics for fiscal year 2002 because there are only twenty six 8-K filings matched to 10-K filings during this year.

Figure V: Distribution of Number Keywords And Overlap in 10-Ks and 8-Ks



This figure shows the distribution of mean number of keywords in 10-Ks vs. 8-Ks, as well as the mean overlap, during 2003-2007. Similar to Figures II and III, we use the 21,341 matched 10-K/8-K pairs with an FLS list.

Figure VI: Comparison of 10-K vs. 8-K FLS Keyword Lists



This figure compares the content of the FLS keyword list in 10-K vs. 8-K filings during 2003 to 2017 for the sample of 21,341 annual reports that have a matched 8-K filing, and where both filings have an FLS list. The lines show the time series variation in the mean, median, and inter-quartiles of the Jaccard similarity index of the 10-K/8-K pairs, calculated as the number of FLS keywords that appear in both 10-K's and 8-K's (stemmed) lists divided by the number of keywords in the union of the lists. The teal bar chart indicates the percentage of 10-K/8-K pairs where the stemmed FLS keywords in the 10-K are the same as those in the 8-K. We use NLTK's PorterStemmer to find the stem of each FLS keyword.

Table I: Sample Selection

Table I reports the sample selection statistics. Our final sample includes 132,303 observations from 14,722 firms, and extends from 1996 to 2017.

Initial sample of firm-years from Compustat merged with EDGAR CIKs	158,744
Less: firm-years with 10-K unavailable for scraping	(10,521)
Plus: firm-years with 10-K FLS keywords list filled	564
Less: firm-years with 10-K filed before December 22, 1995	(549)
Less: firms with years of missing data	(12,541)
Less: firms with less than 2 years of data	(1,295)
Less: firm-years with fiscal year before 1996	(2,099)
Final Sample Size	132,303

Table II: FLS Keywords Inclusion Tabulations

This table presents descriptive statistics on FLS Keywords. Panel A presents the FLS keywords inclusion by fiscal year. The horizontal lines delineate the sub-periods used in subsequent analyses. Panel B presents a tabulation of FLS keywords inclusion by industry (using the GICS industry definitions). The first two columns under the “All Years” heading report the total number of observations for each industry (Obs), and the column percentage (Obs %). The remaining columns show FLS keywords inclusion by industry during three sub-periods. Each sub-period reports the number of observations in each industry as a percentage of the total number of observations during that sub-period (Obs %), and the percentage of observations within each industry that included an FLS keyword list in their 10-K during that sub-period (KW %). Panel C presents descriptive statistics on the number of words included in the FLS keyword list by industry. The first two columns under the “All Years” heading report the number of 10-K filings with an FLS keyword list (Obs), and the average number of words in the list (# KW) by industry. Each sub-period column includes the percentage of observations with an FLS keyword list within each industry (Obs %), and the average number of words in the lists of that industry (KW %).

Panel A: FLS Keywords Inclusion by Fiscal Year				
Fiscal Year	Has Keywords		No Keywords	
	Obs	Row %	Obs	Row %
1996	868	13	5,849	87
1997	1,858	24	5,735	76
1998	2,674	35	4,865	65
1999	3,336	44	4,314	56
2000	3,769	50	3,702	50
2001	4,028	56	3,200	44
2002	4,217	62	2,636	38
2003	4,333	67	2,173	33
2004	4,458	71	1,865	29
2005	4,554	74	1,585	26
2006	4,592	77	1,398	23
2007	4,608	79	1,229	21
2008	4,438	81	1,064	19
2009	4,386	82	948	18
2010	4,420	84	826	16
2011	4,508	86	744	14
2012	4,567	87	662	13
2013	4,640	88	611	12
2014	4,637	89	569	11
2015	4,494	90	488	10
2016	4,300	91	426	9
2017	3,394	91	335	9
Total	87,079	66	45,224	34

Panel B: FLS Keywords Inclusion by Industry

Industry	All Years N = 132,303		1996-2002 N = 51,051		2003-2010 N = 46,877		2011-2017 N = 34,375	
	Obs	Obs %	Obs %	KW %	Obs %	KW %	Obs %	KW %
	Automobiles & Components	1,583	1.2	1.3	39.2	1.1	75.9	1.1
Banks	12,848	9.7	9.1	42.5	10.5	75.1	9.5	88.8
Capital Goods	9,549	7.2	7.2	35.1	7.2	70.9	7.3	85.0
Commercial & Professional Services	5,275	4.0	4.9	38.5	3.6	75.7	3.1	85.8
Consumer Durables & Apparel	5,224	3.9	5.0	35.5	3.7	71.7	2.8	86.7
Consumer Services	4,770	3.6	3.8	38.7	3.5	75.5	3.4	90.0
Diversified Financials	4,774	3.6	3.2	37.7	3.4	74.3	4.5	88.0
Energy	8,034	6.1	4.5	41.4	6.3	77.5	8.0	89.5
Food & Staples Retailing	955	0.7	0.8	37.7	0.7	68.8	0.6	89.8
Food, Beverage & Tobacco	2,854	2.2	2.2	29.9	2.0	68.8	2.3	88.1
Health Care Equipment & Services	8,836	6.7	7.0	43.5	6.7	77.6	6.2	89.6
Household & Personal Products	1,359	1.0	0.9	38.9	1.0	72.8	1.2	89.2
Insurance	3,254	2.5	2.4	37.1	2.6	75.1	2.4	88.9
Materials	6,878	5.2	5.1	31.6	5.1	76.5	5.4	92.3
Media	3,726	2.8	3.1	42.3	2.8	73.8	2.4	83.0
Pharmaceuticals, Biotechnology & Life Sciences	9,272	7.0	5.1	43.8	7.2	80.2	9.5	92.8
Real Estate	5,446	4.1	3.5	40.5	3.9	77.7	5.3	90.7
Retailing	5,165	3.9	4.5	41.9	3.7	76.5	3.3	83.5
Semiconductors & Semiconductor Equipment	2,692	2.0	1.5	52.6	2.5	85.8	2.2	89.7
Software & Services	11,285	8.5	9.3	51.0	8.3	80.9	7.8	92.2
Technology Hardware & Equipment	9,256	7.0	8.4	41.0	6.8	78.1	5.1	86.9
Telecommunication Services	2,189	1.7	2.0	49.8	1.7	81.2	1.1	89.3
Transportation	2,170	1.6	1.6	36.8	1.7	79.5	1.5	91.2
Utilities	4,909	3.7	3.4	35.2	3.9	71.4	4.0	85.1
		Mean		40.1		75.9		88.3
		Std		5.5		4.0		2.9
		Coefficient of Variation		0.1371		0.0533		0.0326

Panel C: Number of Keywords by Industry

Industry	All Years N = 87,079		1996-2002 N = 20,750		2003-2010 N = 35,789		2011-2017 N = 30,540	
	Obs	# KW	Obs %	# KW	Obs %	# KW	Obs %	# KW
Automobiles & Components	979	9.0	1.3	6.6	1.1	8.9	1.0	11.1
Banks	8,575	9.2	9.5	6.8	10.4	9.0	9.5	11.1
Capital Goods	5,802	9.7	6.2	6.8	6.7	9.6	7.0	11.5
Commercial & Professional Services	3,163	9.0	4.7	6.4	3.6	9.3	3.0	11.3
Consumer Durables & Apparel	2,963	8.9	4.4	6.6	3.5	9.1	2.7	11.0
Consumer Services	3,062	9.9	3.6	6.6	3.5	9.8	3.5	12.3
Diversified Financials	3,163	10.7	3.0	6.9	3.3	10.6	4.5	12.4
Energy	5,718	10.9	4.6	7.3	6.4	10.4	8.1	12.7
Food & Staples Retailing	568	9.0	0.8	6.2	0.6	8.2	0.6	12.2
Food, Beverage & Tobacco	1,690	9.6	1.6	6.4	1.8	9.2	2.3	11.3
Health Care Equipment & Services	5,904	9.8	7.4	7.0	6.8	9.7	6.3	12.1
Household & Personal Products	896	9.7	0.9	6.5	1.0	9.4	1.2	11.5
Insurance	2,102	9.4	2.1	7.0	2.5	9.4	2.4	10.9
Materials	4,376	9.5	4.0	6.8	5.1	9.2	5.6	11.2
Media	2,322	9.1	3.2	6.8	2.7	9.3	2.2	11.2
Pharmaceuticals, Biotechnology & Life Sciences	6,892	10.8	5.5	7.3	7.6	10.3	9.9	12.7
Real Estate	3,797	9.5	3.5	6.4	3.9	9.3	5.4	11.0
Retailing	3,241	9.4	4.6	7.0	3.7	9.6	3.1	11.6
Semiconductors & Semiconductor Equipment	2,092	9.9	1.9	6.9	2.8	9.6	2.2	12.2
Software & Services	8,011	9.6	11.6	7.3	8.8	9.8	8.0	11.7
Technology Hardware & Equipment	5,786	9.3	8.5	7.2	7.0	9.6	5.0	11.2
Telecommunication Services	1,483	8.5	2.5	7.3	1.8	8.7	1.1	10.2
Transportation	1,429	9.2	1.5	7.0	1.8	9.1	1.6	10.7
Utilities	3,065	10.3	2.9	7.6	3.6	10.2	3.8	11.9
		Mean		6.9		9.5		11.5
		Std		0.4		0.5		0.6
		Coefficient of Variation		0.0521		0.057		0.0558

Table III: Descriptive Statistics

This table reports descriptive statistics for the variables used in our analyses. Panel A presents univariate statistics for the sample of 10-K filings with vs. without FLS keywords. The difference is calculated as (*Has Keywords* - *No Keywords*), and the t-statistics are calculated using clustered standard errors. Panel B presents Pearson correlations. Variables are defined in Appendix A. For Panel A, ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively. In Panel B, for brevity, significance is limited to bold font, which indicates significance at the 1% level.

Panel A: Univariate Statistics by FLS Keywords Inclusion							
	Has Keywords			No Keywords			Difference
	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	
Liberal Court	83,189	0.410	0.190	44,238	0.387	0.157	0.02***
10b-5 Firm	87,079	0.167	0.373	45,224	0.112	0.316	0.05***
FPS	87,079	0.266	0.442	45,224	0.236	0.425	0.03***
Legal	87,079	0.429	0.495	45,224	0.205	0.403	0.22***
Guidance	87,079	0.333	0.471	45,224	0.185	0.389	0.15***
$\Delta \text{Log}(10\text{-K Size})$	80,817	0.049	0.223	38,863	0.053	0.316	-0.00*
$\text{Log}(10\text{-K Size})_{t-1}$	80,817	12.436	0.546	38,863	12.036	0.616	0.40***
Age	87,079	16.716	14.265	45,224	17.354	14.237	-0.64**
R&D	87,079	0.382	0.486	45,224	0.336	0.472	0.05***
Loss Year	85,605	0.410	0.492	44,631	0.362	0.481	0.05***
Acquisition Sales Ratio	85,604	0.022	0.110	44,631	0.024	0.114	-0.00**
Industry Adoption $_{t-1}$	80,817	0.716	0.197	38,862	0.475	0.262	0.24***
Analyst Following	87,079	0.511	0.500	45,224	0.420	0.494	0.09***
Inst_Own	87,079	0.286	0.352	45,224	0.194	0.286	0.09***
Trans_Own	87,079	0.068	0.101	45,224	0.048	0.087	0.02***
Big N Auditor	87,079	0.614	0.487	45,224	0.626	0.484	-0.01
Introduction	87,079	0.176	0.381	45,224	0.167	0.373	0.01*
Growth	87,079	0.278	0.448	45,224	0.257	0.437	0.02***
Mature	87,079	0.291	0.454	45,224	0.300	0.458	-0.01
Out	87,079	0.099	0.299	45,224	0.096	0.294	0.00
Decline	87,079	0.109	0.312	45,224	0.098	0.297	0.01***

Panel B: Pearson Correlations										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Liberal Court	1.00									
(2) 10b-5 Firm	0.05	1.00								
(3) FPS	0.11	0.14	1.00							
(4) Legal	-0.02	0.19	-0.02	1.00						
(5) Guidance	-0.02	0.25	0.08	0.41	1.00					
(6) Log(10-K Size)	0.02	0.18	-0.05	0.43	0.29	1.00				
(7) Age	-0.11	0.09	-0.11	0.23	0.22	0.15	1.00			
(8) R&D	0.09	0.13	0.37	-0.00	0.13	-0.03	0.00	1.00		
(9) Loss Year	0.12	-0.03	0.19	-0.17	-0.19	-0.09	-0.24	0.21	1.00	
(10) Acquisition Sales Ratio	0.01	0.01	0.01	-0.01	0.01	0.01	-0.05	0.02	0.04	1.00
(11) Industry Adoption _{t-1}	0.07	0.11	0.06	0.43	0.28	0.49	0.11	0.05	0.06	-0.04
(12) Analyst Following	-0.01	0.21	0.05	0.34	0.44	0.28	0.13	0.09	-0.25	0.00
(13) Inst_Own	-0.00	0.21	-0.01	0.40	0.47	0.32	0.19	0.10	-0.25	-0.00
(14) Trans_Own	0.03	0.19	0.04	0.27	0.36	0.22	0.07	0.12	-0.17	0.02
(15) Big N Auditor	-0.03	0.17	0.06	0.22	0.32	0.23	0.20	0.09	-0.17	0.03
(16) Introduction	0.09	-0.03	0.13	-0.15	-0.16	-0.09	-0.19	0.17	0.41	0.05
(17) Growth	-0.03	0.02	-0.06	0.09	0.08	0.12	0.00	-0.09	-0.21	0.10
(18) Mature	-0.09	0.05	-0.05	0.15	0.20	0.03	0.26	-0.03	-0.28	-0.09
(19) Out	0.00	-0.01	-0.02	0.02	-0.02	0.01	0.01	-0.02	0.02	-0.03
(20) Decline	0.09	-0.02	0.11	-0.12	-0.14	-0.11	-0.11	0.10	0.33	-0.01

	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
(11) Industry Adoption _{t-1}	1.00									
(12) Analyst Following	0.11	1.00								
(13) Inst_Own	0.22	0.68	1.00							
(14) Trans_Own	0.12	0.57	0.80	1.00						
(15) Big N Auditor	-0.09	0.31	0.32	0.28	1.00					
(16) Introduction	-0.01	-0.17	-0.18	-0.13	-0.16	1.00				
(17) Growth	0.01	0.12	0.13	0.12	0.11	-0.27	1.00			
(18) Mature	0.03	0.13	0.16	0.09	0.22	-0.30	-0.42	1.00		
(19) Out	0.04	-0.03	-0.01	-0.01	-0.01	-0.15	-0.21	-0.23	1.00	
(20) Decline	0.04	-0.15	-0.15	-0.11	-0.13	-0.15	-0.21	-0.23	-0.12	1.00

Table IV: FLS Keywords Inclusion Determinants Model

This table reports results from cross-sectional probit models at the firm-year level. The dependent variable in all specifications is *FLS Keywords Inclusion*. Our base and expanded models (specifications (1) and (2) respectively) include the full sample of firms. Specification (3) includes only “Persistent” firms—those who consistently either do, or do not provide FLS keywords during our sample. Specifications (4)–(6) include only “Switchers”—those firms who both do, and do not provide FLS keywords during our sample. Specification (4) includes all firms that switch at least one time during our sample period, while specification (5) includes firms that switch *only once*, and specification (6) includes those firms that switch *multiple* times. All models include federal circuit and year fixed effects. We report average marginal effects with the coefficient z-statistics in parenthesis. Standard errors are clustered at the firm level. Variables are defined in Appendix A. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Dependent Variable = FLS Keywords Inclusion					
	(1) Base	(2) Expanded	(3) Persistent	Switchers		
				(4) All	(5) Structural	(6) Transient
Liberal Court	0.096*** (3.03)	0.076** (2.38)	0.091** (2.30)	0.061 (1.37)	0.051 (1.11)	0.116 (0.96)
10b-5 Firm	0.009 (0.77)	0.004 (0.32)	0.052*** (2.69)			
FPS	0.006 (0.73)	0.006 (0.73)	0.001 (0.05)			
Legal	0.007 (0.93)	0.006 (0.77)	0.023** (1.99)	-0.010 (-1.04)	-0.009 (-0.93)	-0.029 (-1.21)
Guidance	0.015** (2.06)	0.015** (2.11)	0.028*** (2.59)	0.009 (0.99)	0.005 (0.58)	0.029 (1.27)
$\Delta \text{Log}(10\text{-K Size})$		0.074*** (13.39)	0.061*** (7.86)	0.065*** (9.16)	0.055*** (7.62)	0.108*** (5.44)
$\text{Log}(10\text{-K Size})_{t-1}$	0.087*** (13.04)	0.100*** (12.88)	0.127*** (10.51)	0.061*** (6.74)	0.060*** (6.16)	0.075*** (3.51)
$\text{Log}(\text{Age} + 1)$	-0.067*** (-15.07)	-0.066*** (-14.37)	-0.195*** (-24.42)			
R&D	0.027*** (3.43)	0.025*** (3.16)	0.041*** (3.41)	0.003 (0.34)	0.006 (0.66)	-0.034 (-1.41)
Loss Year	0.017*** (3.29)	0.010* (1.90)	0.003 (0.37)	0.015** (2.39)	0.018*** (2.74)	0.017 (0.93)
Acquisition Sales Ratio	0.054*** (4.08)	0.033** (2.54)	-0.013 (-0.74)	0.053*** (3.04)	0.054*** (2.99)	0.017 (0.36)
Industry Adoption $_t - 1$	0.155** (2.26)	0.153** (2.24)	0.132 (1.28)	0.181** (2.23)	0.199** (2.34)	0.061 (0.29)
Analyst Following	0.017** (2.23)	0.015** (2.00)	-0.006 (-0.52)	0.037*** (4.21)	0.036*** (3.83)	0.030 (1.30)
Inst_Own	0.008 (0.53)					
Trans_Own		0.065* (1.95)	0.125*** (2.58)	-0.003 (-0.07)	0.026 (0.60)	-0.137 (-1.27)
Big N Auditor	0.006 (0.89)	-0.001 (-0.10)	0.004 (0.42)	-0.008 (-0.87)	0.001 (0.12)	-0.024 (-0.94)
Introduction		0.019*** (2.79)	0.017* (1.72)	0.009 (1.04)	0.014* (1.66)	0.021 (0.87)
Growth		0.012** (2.31)	0.020*** (2.64)	-0.003 (-0.45)	-0.003 (-0.39)	0.005 (0.27)
Out		-0.000 (-0.02)	0.010 (1.11)	-0.007 (-0.85)	0.001 (0.11)	-0.018 (-0.81)
Decline		0.013* (1.72)	0.023** (2.12)	0.000 (0.01)	-0.007 (-0.62)	0.055** (2.03)
Federal Court F.E.	Y	Y	Y	Y	Y	Y
Fiscal Year F.E.	Y	Y	Y	Y	Y	Y
Pseudo R ²	0.193	0.196	0.352	0.167	0.219	0.047
Observations	114,607	109,777	47,339	62,438	52,308	10,130

Table V: FLS Keywords Inclusion by Sub-periods

This table presents regression results by sub-periods based on the expanded probit model from Table IV. The dependent variable in all specifications is *FLS Keywords Inclusion*. We report average marginal effects with the coefficient z-statistics in parenthesis. Standard errors are clustered at the firm level. Variables are defined in Appendix A. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Dependent Variable = FLS Keywords Inclusion			
	(1) All Years	(2) 1996-2002	(3) 2003-2010	(4) 2011-2017
Liberal Court	0.076** (2.38)	0.011 (0.27)	0.019 (0.30)	0.006 (0.15)
10b-5 Firm	0.004 (0.32)	0.015 (0.99)	0.006 (0.45)	-0.008 (-0.67)
FPS	0.006 (0.73)	-0.011 (-1.00)	0.019 (1.53)	0.012 (1.09)
Legal	0.006 (0.77)	0.016 (1.28)	0.008 (0.82)	0.006 (0.63)
Guidance	0.015** (2.11)	0.008 (0.88)	0.026*** (2.63)	0.020** (2.15)
$\Delta \text{Log}(10\text{-K Size})$	0.074*** (13.39)	0.093*** (13.87)	0.076*** (5.96)	0.041** (2.55)
$\text{Log}(10\text{-K Size})_{t-1}$	0.100*** (12.88)	0.133*** (13.76)	0.091*** (8.30)	0.072*** (7.08)
$\text{Log}(\text{Age} + 1)$	-0.066*** (-14.37)	-0.066*** (-10.94)	-0.070*** (-10.46)	-0.057*** (-10.22)
R&D	0.025*** (3.16)	0.039*** (3.94)	0.021** (1.98)	0.010 (1.09)
Loss Year	0.010* (1.90)	0.006 (0.82)	0.011 (1.42)	0.013* (1.66)
Acquisition Sales Ratio	0.033** (2.54)	0.055*** (2.68)	0.022 (0.93)	0.009 (0.41)
$\text{Industry Adoption}_{t-1}$	0.153** (2.24)	0.283*** (3.31)	0.110 (1.04)	0.034 (0.25)
Analyst Following	0.015** (2.00)	0.029*** (2.85)	0.007 (0.64)	0.007 (0.69)
Trans_Own	0.065* (1.95)	0.046 (1.03)	0.055 (1.08)	0.063 (1.38)
Big N Auditor	-0.001 (-0.10)	0.030*** (2.76)	-0.002 (-0.14)	-0.025** (-2.50)
Introduction	0.019*** (2.79)	0.029*** (3.02)	0.027** (2.53)	-0.007 (-0.64)
Growth	0.012** (2.31)	0.013* (1.71)	0.023*** (3.00)	-0.006 (-0.81)
Out	-0.000 (-0.02)	-0.011 (-1.10)	0.003 (0.35)	0.008 (0.95)
Decline	0.013* (1.72)	0.020* (1.77)	0.019 (1.56)	-0.006 (-0.54)
Federal Court F.E.	Y	Y	Y	Y
Fiscal Year F.E.	Y	Y	Y	Y
Pseudo R ²	0.196	0.105	0.049	0.058
Observations	109,777	37,735	41,730	30,312

Table VI: Number of Keywords Determinants Model

This table reports results from Poisson regressions of $\#$ *Keywords*. Specification (1) is our baseline model. Specification (2) adds *Return Volatility* and *Return Skewness*. Specifications (3) and (4) re-estimate these regressions for the sub-sample of firms that always include an FLS keyword list in their 10-K filing. All models include federal circuit and year fixed effects. We report average marginal effects with the coefficient z-statistics in parenthesis. Standard errors are clustered at the firm level. Variables are defined in Appendix A. ***, **, and * indicate significance at the 1%, 5%, and 10% level, respectively.

	Dependent Variable = # Keywords			
	(1)	(2)	Always	
			(3)	(4)
	Baseline	Market	Baseline	Market
Liberal Court	1.191*** (2.95)	1.238*** (2.70)	0.929 (1.50)	1.102 (1.55)
Legal	-0.003 (-0.04)	-0.009 (-0.10)	0.174 (1.48)	0.044 (0.32)
Δ Log(10-K Size)	0.967*** (12.02)	0.878*** (8.83)	0.913*** (7.08)	0.755*** (4.85)
Log(10-K Size) $_{t-1}$	1.206*** (13.72)	1.341*** (10.92)	1.275*** (8.95)	1.373*** (6.48)
Log(Age + 1)	-0.335*** (-5.29)	-0.444*** (-5.56)	-0.595*** (-5.36)	-0.793*** (-5.46)
R&D	0.164* (1.84)	0.155 (1.45)	0.347*** (2.65)	0.325** (1.99)
Loss Year	0.238*** (3.17)	0.227*** (2.69)	0.361*** (3.10)	0.294** (2.19)
Acquisition Sales Ratio	0.382*** (2.78)	0.361** (2.27)	0.454** (2.23)	0.311 (1.31)
ROA Deciles	0.047*** (2.66)	0.052** (2.52)	0.058** (2.16)	0.065** (2.06)
Return Volatility		1.686 (1.34)		5.193*** (3.05)
Return Skewness		-0.043*** (-3.53)		-0.067*** (-3.76)
IPOed within 3yrs	0.600*** (8.31)	0.560*** (6.55)	0.478*** (5.33)	0.345*** (3.14)
Industry Avg. # KW $_{t-1}$	0.255*** (4.25)	0.239*** (3.37)	0.393*** (4.82)	0.362*** (3.67)
Big N Auditor	0.292*** (3.33)	0.267*** (2.67)	0.511*** (3.99)	0.452*** (3.12)
CEO Change	0.093* (1.75)	0.061 (1.07)	0.158* (1.95)	0.132 (1.49)
CFO Change	0.177*** (3.54)	0.171*** (3.32)	0.224*** (2.93)	0.230*** (2.87)
Introduction	0.376*** (4.60)	0.330*** (3.51)	0.455*** (3.52)	0.423*** (2.81)
Growth	0.038 (0.70)	-0.019 (-0.33)	-0.099 (-1.14)	-0.147 (-1.50)
Out	-0.035 (-0.47)	-0.045 (-0.53)	-0.089 (-0.71)	-0.095 (-0.65)
Decline	0.113 (1.18)	0.048 (0.42)	0.059 (0.39)	-0.016 (-0.09)
Federal Court F.E.	Y	Y	Y	Y
Fiscal Year F.E.	Y	Y	Y	Y
Pseudo R ²	0.073	0.073	0.084	0.085
Observations	74,198	58,114	31,428	23,797

Table VII: Factor Analysis of the Most Frequent Words

This table reports the results from a factor analysis of the top words in the FLS keyword lists from 10-K filings. Panel A presents the frequencies and cumulative frequencies in percentage of the 20 most frequent keywords. The last column shows the percentage of 10-K filings in which each word appears. Panel B reports the eigenvalues and the variance explained from the first 10 factors. Based on these results, we further examine factors 1 through 5 and present the loadings and their uniqueness from a varimax rotation in Panel C. Lastly, Panel D presents the inter-temporal means of the most frequent 20 words in the 10-K FLS list. These means capture the percentage of 10-Ks where each word has appeared throughout our sample period. The keywords used in these analyses have been stemmed using the Porter Stemming algorithm from the NLTK Python package. The factor analysis was performed on tetrachoric correlations, due to the binary nature of FLS keyword inclusion.

Panel A: Keyword Frequency in 10-K Filings			
Keyword	Frequency	Cumulative %	% of 10-Ks
expect	84,095	10.1	96.6
anticip	82,598	20.0	94.8
believ	81,280	29.8	93.3
estim	71,617	38.4	82.2
intend	65,209	46.2	74.9
plan	58,929	53.3	67.7
may	50,499	59.3	58.0
will	44,797	64.7	51.4
should	39,149	69.4	44.9
project	36,228	73.7	41.6
could	27,405	77.0	31.5
continu	24,971	80.0	28.7
potenti	19,196	82.3	22.0
predict	18,288	84.5	21.0
would	15,080	86.3	17.3
seek	13,174	87.9	15.1
forecast	9,187	89.0	10.5
target	8,026	90.0	9.2
goal	5,421	90.6	6.2
might	4,421	91.2	5.1
...
Total	833,152	100.0	100.0

Panel B: Eigenvalues and Explained Variance				
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	7.639	5.240	0.399	0.399
Factor 2	2.399	0.193	0.125	0.524
Factor 3	2.205	0.833	0.115	0.639
Factor 4	1.372	0.323	0.072	0.711
Factor 5	1.049	0.140	0.055	0.766
Factor 6	0.908	0.104	0.047	0.813
Factor 7	0.804	0.114	0.042	0.855
Factor 8	0.690	0.134	0.036	0.891
Factor 9	0.556	0.129	0.029	0.920
Factor 10	0.427	0.086	0.022	0.942
...				

Panel C: Varimax Rotated Factor Loadings

Keyword	Factor 1 (Future State ^a)	Factor 2 (Modal ^b)	Factor 3 (Likelihood ^c)	Factor 4 (Quantitative ^d)	Factor 5 (Future Action ^e)	Uniqueness
expect	0.20	0.18	0.90	0.01	0.16	0.09
anticip	0.25	0.08	0.77	0.02	0.18	0.30
believ	0.04	0.15	0.69	0.07	0.30	0.41
estim	0.67	0.13	-0.26	0.26	0.50	0.15
intend	-0.06	0.33	0.35	0.16	0.63	0.34
plan	0.38	0.19	0.30	0.24	0.61	0.29
may	0.58	0.75	0.16	0.02	-0.03	0.08
will	0.45	0.60	0.26	-0.01	-0.11	0.36
should	0.48	0.63	0.24	0.16	0.07	0.29
project	0.10	0.21	-0.34	0.78	0.17	0.19
could	0.28	0.83	0.12	0.22	0.16	0.14
continu	0.84	0.24	0.17	-0.04	-0.09	0.21
potenti	0.85	0.29	0.20	0.11	0.06	0.14
predict	0.85	0.19	0.16	0.09	0.11	0.19
would	0.07	0.87	0.07	0.09	0.24	0.17
seek	0.02	0.07	0.28	0.05	0.83	0.23
forecast	0.12	0.06	0.18	0.81	-0.06	0.29
target	0.06	0.18	0.21	0.69	0.23	0.40
goal	0.05	0.07	0.11	0.60	0.21	0.58
might	0.27	0.64	-0.04	0.15	0.06	0.49

^aFrom the Merriam-Webster online dictionary: **Estimate**: to judge tentatively or approximately the value, worth, or significance of. **Continue**: to maintain without interruption a condition, course, or action. **Potential**: existing in possibility: capable of development into actuality. **Predict**: to declare or indicate in advance

^bA modal verb is usually used with another verb to express different ideas such as possibility (might or may), probability or necessity (must), promise or willingness (will), preferences or desires (would), ability (can), possibility (could), and uncertain prediction (should).

^cFrom the Merriam-Webster online dictionary: **Anticipate**: to look forward to as certain. **Expect**: to consider probable or certain. **Believe**: to consider to be true or honest

^dFrom the Merriam-Webster online dictionary: **Project**: to plan, figure, or estimate for the future. **Forecast**: to calculate or predict (some future event or condition) usually as a result of study and analysis of available pertinent data. **Target**: a goal to be achieved. **Goal**: the end toward which effort is directed

^eFrom the Merriam-Webster online dictionary: **Intend**: to design for a specified use or future. **Plan**: to devise or project the realization or achievement of. **Seek**: to try to acquire or gain

Panel D: Inter-temporal Means of Top 20 Words in 10-K FLS Lists

Fiscal Year	expect	anticip	believ	estim	intend	plan	may	will	should	project
1996	0.882	0.844	0.781	0.621	0.374	0.296	0.156	0.147	0.085	0.238
1997	0.904	0.875	0.826	0.629	0.461	0.343	0.197	0.173	0.113	0.224
1998	0.924	0.889	0.847	0.666	0.505	0.392	0.246	0.229	0.155	0.234
1999	0.932	0.900	0.868	0.689	0.558	0.448	0.316	0.286	0.195	0.237
2000	0.934	0.907	0.880	0.720	0.596	0.496	0.369	0.345	0.243	0.253
2001	0.945	0.920	0.898	0.736	0.617	0.529	0.406	0.392	0.274	0.277
2002	0.949	0.925	0.907	0.761	0.658	0.563	0.439	0.431	0.311	0.313
2003	0.956	0.937	0.922	0.783	0.699	0.603	0.482	0.454	0.350	0.338
2004	0.962	0.944	0.926	0.794	0.723	0.626	0.514	0.472	0.372	0.358
2005	0.968	0.950	0.932	0.815	0.751	0.665	0.552	0.496	0.413	0.384
2006	0.970	0.954	0.940	0.826	0.758	0.689	0.577	0.508	0.437	0.404
2007	0.973	0.956	0.944	0.841	0.781	0.713	0.610	0.526	0.462	0.418
2008	0.977	0.960	0.951	0.850	0.797	0.729	0.636	0.550	0.493	0.444
2009	0.976	0.961	0.951	0.860	0.800	0.738	0.649	0.560	0.510	0.458
2010	0.978	0.965	0.956	0.865	0.810	0.755	0.666	0.573	0.531	0.472
2011	0.980	0.969	0.957	0.872	0.814	0.768	0.685	0.587	0.556	0.482
2012	0.980	0.970	0.960	0.880	0.828	0.780	0.703	0.602	0.572	0.498
2013	0.981	0.971	0.963	0.886	0.845	0.793	0.722	0.628	0.589	0.519
2014	0.983	0.974	0.966	0.895	0.856	0.805	0.746	0.648	0.613	0.534
2015	0.985	0.975	0.971	0.903	0.862	0.815	0.758	0.664	0.629	0.548
2016	0.986	0.975	0.973	0.907	0.870	0.821	0.770	0.679	0.642	0.562
2017	0.986	0.974	0.973	0.907	0.875	0.824	0.771	0.690	0.652	0.568
Total	0.966	0.948	0.933	0.822	0.749	0.677	0.580	0.514	0.450	0.416

Fiscal Year	could	continu	potenti	predict	would	seek	forecast	target	goal	might
1996	0.018	0.082	0.018	0.035	0.005	0.069	0.020	0.008	0.020	0.000
1997	0.033	0.096	0.020	0.036	0.011	0.079	0.032	0.016	0.028	0.003
1998	0.055	0.111	0.033	0.044	0.021	0.075	0.033	0.016	0.025	0.004
1999	0.083	0.148	0.063	0.071	0.032	0.077	0.035	0.019	0.027	0.006
2000	0.111	0.175	0.094	0.097	0.048	0.084	0.042	0.021	0.029	0.011
2001	0.142	0.193	0.112	0.112	0.063	0.090	0.053	0.024	0.030	0.016
2002	0.173	0.212	0.129	0.131	0.079	0.096	0.065	0.033	0.038	0.022
2003	0.205	0.225	0.144	0.150	0.098	0.108	0.074	0.043	0.046	0.026
2004	0.237	0.243	0.163	0.165	0.115	0.117	0.082	0.052	0.047	0.029
2005	0.268	0.266	0.189	0.183	0.132	0.129	0.087	0.065	0.051	0.037
2006	0.294	0.281	0.209	0.206	0.149	0.133	0.092	0.072	0.053	0.043
2007	0.328	0.306	0.228	0.223	0.170	0.144	0.096	0.082	0.056	0.054
2008	0.351	0.313	0.242	0.229	0.189	0.155	0.110	0.094	0.061	0.059
2009	0.367	0.314	0.247	0.234	0.192	0.160	0.117	0.102	0.066	0.058
2010	0.380	0.321	0.260	0.240	0.203	0.170	0.123	0.113	0.070	0.057
2011	0.404	0.338	0.280	0.255	0.223	0.178	0.132	0.125	0.075	0.063
2012	0.419	0.353	0.289	0.272	0.238	0.187	0.139	0.129	0.078	0.067
2013	0.446	0.370	0.315	0.291	0.261	0.203	0.152	0.142	0.083	0.079
2014	0.477	0.382	0.339	0.314	0.287	0.215	0.157	0.159	0.093	0.088
2015	0.498	0.390	0.351	0.322	0.305	0.226	0.166	0.170	0.098	0.094
2016	0.520	0.397	0.357	0.328	0.321	0.232	0.171	0.185	0.103	0.100
2017	0.536	0.396	0.363	0.329	0.341	0.244	0.184	0.197	0.116	0.099
Total	0.315	0.287	0.220	0.210	0.173	0.151	0.106	0.092	0.062	0.051

Future State: estim, continu, potenti, predict
Modal: may, will, should, could, would, might
Likelihood: expect, anticip, believ
Quantitative: project, forecast, target, goal
Future Action: intend, plan, seek

Panel B: Factor Scores						
	(1)	(2)	(3)	(4)	(5)	(6)
	# Keywords	Future State	Modal	Likelihood	Quantitative	Future Action
Liberal Court	0.998** (2.30)	0.325*** (3.77)	-0.030 (-0.32)	-0.049 (-0.52)	-0.070 (-0.69)	0.271*** (2.73)
Legal	-0.010 (-0.11)	-0.022 (-1.07)	-0.018 (-0.84)	0.012 (0.62)	0.039* (1.79)	0.005 (0.21)
Guidance	-0.050 (-0.56)	0.018 (0.90)	-0.079*** (-3.81)	0.044** (2.40)	-0.001 (-0.03)	0.023 (1.08)
Guidance Intensity	0.006 (0.74)	0.001 (0.73)	-0.003 (-1.31)	-0.001 (-0.54)	0.007*** (3.26)	0.002 (1.06)
$\Delta \text{Log}(10\text{-K Size})$	0.826*** (9.44)	0.065*** (3.13)	0.125*** (6.06)	-0.050** (-2.26)	0.077*** (3.56)	0.057*** (2.65)
$\text{Log}(10\text{-K Size})_{t-1}$	1.298*** (10.86)	0.107*** (3.85)	0.178*** (5.92)	-0.045 (-1.58)	0.130*** (4.07)	0.092*** (3.05)
$\text{Log}(\text{Age} + 1)$	-0.446*** (-5.67)	-0.114*** (-7.04)	-0.098*** (-5.25)	-0.019 (-1.11)	0.137*** (6.52)	-0.039** (-2.02)
R&D	0.136 (1.29)	0.042* (1.80)	0.076*** (3.12)	0.056** (2.35)	-0.077*** (-2.88)	-0.039 (-1.55)
Loss Year	0.235*** (2.75)	0.087*** (4.77)	-0.042** (-2.00)	0.008 (0.43)	-0.016 (-0.80)	0.023 (1.19)
Acquisition Sales Ratio	0.316** (2.01)	0.085** (2.26)	-0.017 (-0.44)	0.001 (0.03)	0.023 (0.56)	-0.010 (-0.26)
ROA Deciles	0.052** (2.50)	0.012*** (2.63)	-0.007 (-1.46)	-0.000 (-0.00)	0.011** (2.12)	-0.000 (-0.10)
Return Volatility	1.277 (1.03)	0.368 (1.21)	0.011 (0.04)	-0.056 (-0.18)	-0.479 (-1.60)	0.492 (1.49)
Return Skewness	-0.043*** (-3.29)	-0.002 (-0.59)	-0.008*** (-2.65)	0.002 (0.73)	0.001 (0.47)	-0.006* (-1.93)
IPOed within 3yrs	0.593*** (6.95)	0.103*** (5.63)	0.042** (2.14)	-0.009 (-0.44)	0.074*** (3.66)	0.049** (2.33)
Industry Avg. # KW _{t-1}	0.281*** (3.71)	0.084*** (5.21)	-0.002 (-0.11)	0.053*** (3.61)	0.004 (0.21)	0.015 (0.84)
Big N Auditor	0.275*** (2.69)	0.040* (1.70)	0.018 (0.71)	0.017 (0.79)	0.012 (0.48)	0.038 (1.52)
CEO Change	0.063 (1.05)	0.016 (1.17)	-0.004 (-0.30)	-0.000 (-0.03)	0.022 (1.47)	-0.004 (-0.25)
CFO Change	0.187*** (3.33)	0.003 (0.28)	0.053*** (3.68)	-0.003 (-0.32)	0.013 (0.94)	0.021 (1.58)
Introduction	0.333*** (3.61)	0.043** (2.04)	0.066*** (2.86)	-0.030 (-1.50)	-0.003 (-0.12)	0.009 (0.39)
Growth	-0.022 (-0.37)	-0.012 (-0.97)	0.027* (1.94)	0.026** (2.16)	-0.021 (-1.34)	-0.003 (-0.20)
Out	-0.045 (-0.54)	-0.014 (-0.82)	0.050*** (2.60)	0.005 (0.31)	-0.060*** (-3.05)	-0.034* (-1.79)
Decline	0.044 (0.39)	0.003 (0.11)	0.056** (2.03)	-0.028 (-1.17)	-0.057** (-2.24)	0.000 (0.01)
Federal Court F.E.	Y	Y	Y	Y	Y	Y
Fiscal Year F.E.	Y	Y	Y	Y	Y	Y
R ²	0.216	0.067	0.099	0.010	0.079	0.035
Observations	58,114	58,114	58,114	58,114	58,114	58,114

Table IX: Comparison of 10-K vs. 8-K FLS Keyword Lists

This table reports results from a comparison of the FLS keyword lists in 10-K filings vs. in 8-K filings. Specification (1) presents results from a fractional regression of the Jaccard similarity index computed as the intersection of the stemmed words in the 10-K and 8-K keyword lists divided by the union of the words in the lists. Specification (2) shows results from OLS regressions of the absolute difference between the number of keywords in the 10-K FLS list and the number of keywords in the 8-K list (Abs[# 10-K KWs — # 8-K KWs]). Lastly, specification (3) reports results from OLS regressions of the signed value of this difference.

	(1) Jaccard Index	(2) #10K-#8K	(3) #10K-#8K
Liberal Court	0.289* (1.68)	-0.492 (-0.78)	-0.013 (-0.02)
Legal	-0.002 (-0.06)	0.044 (0.39)	0.143 (1.09)
8-K/10-K Length Ratio	1.902*** (4.22)	-5.415*** (-4.30)	-5.289*** (-3.59)
Log(Age + 1)	0.155*** (6.46)	-0.492*** (-6.04)	-0.182* (-1.90)
R&D	-0.140*** (-4.31)	0.601*** (4.96)	0.553*** (3.77)
Loss Year	-0.105*** (-3.82)	0.338*** (3.26)	0.194 (1.50)
Acquisition Sales Ratio	-0.113* (-1.87)	0.711*** (2.58)	0.182 (0.53)
ROA Deciles	-0.011* (-1.67)	0.030 (1.16)	0.036 (1.18)
IPOed within 3yrs	0.009 (0.31)	0.308*** (2.63)	0.039 (0.27)
Big N Auditor	0.212*** (6.75)	-0.530*** (-4.15)	-0.280* (-1.81)
CEO Change	-0.005 (-0.24)	-0.005 (-0.06)	0.056 (0.60)
CFO Change	-0.035* (-1.81)	0.135* (1.72)	0.141 (1.49)
Introduction	-0.031 (-0.94)	0.154 (1.10)	0.418** (2.41)
Growth	-0.004 (-0.19)	0.041 (0.52)	0.138 (1.44)
Out	-0.007 (-0.27)	0.066 (0.63)	0.125 (1.01)
Decline	-0.067* (-1.68)	0.177 (1.08)	0.352* (1.76)
Federal Court F.E.	Y	Y	Y
Fiscal Year F.E.	Y	Y	Y
Pseudo R ² /R ²	0.032	0.071	0.025
Observations	20,137	20,137	20,137