Rev Account Stud (2014) 19:396–455
DOI 10.1007/s11142-013-9258-3

The information content of mandatory risk factor disclosures in corporate filings

John L. Campbell · Hsinchun Chen · Dan S. Dhaliwal · Hsin-min Lu · Logan B. Steele

Published online: 8 October 2013

Abstract Beginning in 2005, the Securities and Exchange Commission (SEC) mandated firms to include a "risk factor" section in their Form 10-K to discuss "the most significant factors that make the company speculative or risky." In this study, we examine the information content of this newly created section and offer two main results. First, we find that firms facing greater risk disclose more risk factors, and that the type of risk the firm faces determines whether it devotes a greater portion of its disclosures towards describing that risk type. That is, managers provide risk factor disclosures that meaningfully reflect the risks they face. Second, we find that the information conveyed by risk factor disclosures is reflected in systematic risk, idiosyncratic risk, information asymmetry, and firm value. Overall, our evidence supports the SEC's decision to mandate risk factor disclosures, as the disclosures appear to be firm-specific and useful to investors.

J. L. Campbell (✉)
J.M. Tull School of Accounting, University of Georgia, 228 Brooks Hall, 310 Herty Drive, Athens, GA 30602, USA
e-mail: johnc@uga.edu

H. Chen
Management Information Systems Department, University of Arizona, McClelland Hall 430Z, 1130 E. Helen St., Tucson, AZ 85721, USA

D. S. Dhaliwal
Department of Accounting, University of Arizona and Korea University Business School, McClelland Hall 301D, 1130 E. Helen St., Tucson, AZ 85721, USA

H. Lu
Department of Information Management, National Taiwan University, Room 509, Management Building II, No. 1, Sec. 4, Roosevelt Rd., Taipei City 106, Taiwan

L. B. Steele
Department of Accounting, University of Connecticut, 2100 Hillside Road, Unit 1041, Storrs, CT 06269, USA

Keywords Textual analysis · Risk disclosures · Disclosure tone · Regulation · Information content

JEL Classification D8 · G24 · G12 · M4

1 Introduction

> Risk factors are looked upon as boilerplate, 'Have we considered all our bases just in case the worst case does happen?'… The irony of it is that risk factors are almost meant not to be read, or relied upon.–Tom Taulli, IPO analyst (Reuters 2005)

Beginning in 2005, the Securities and Exchange Commission (SEC) required all firms to include a new section in their annual filings (Secttion 1A of the Annual Report on Form 10-K) to discuss "the most significant factors that make the company speculative or risky" (Regulation S–K, Item 305(c), SEC 2005). Prior to this change, companies were only required to provide this information in registration statements for equity and debt offerings.[1] By mandating this disclosure for all firms, the SEC is suggesting that risk factor disclosures are informative and that investors benefit from this information.

Critics of the new disclosure requirements argue that risk factor disclosures are unlikely to be informative for at least two reasons. First, firms do not have to estimate the likelihood that a disclosed risk will ultimately be realized. Second, firms do not have to quantify the impact that a disclosed risk might have on their current and future financial statements. Thus, managers simply disclose all possible risks and uncertainties, regardless of the likelihood that they will ultimately affect the firm, and the disclosure surrounding each of these risks and uncertainties is likely to be vague and boilerplate in nature (Reuters 2005). The SEC warned firms in 2010 to "avoid generic risk factor disclosure that could apply to any company" (SEC 2010), and has repeatedly called for increased focus and specificity in risk factor disclosures through the comment letter process (CFO 2010). Seven years after the adoption of the SEC's risk factor disclosure requirement (Regulation S–K, Item 305c), we have little evidence on whether these newly created disclosures are informative. Our study addresses this gap in the literature.

Understanding the information conveyed by risk factor disclosures is important to regulators, investors, and academic researchers, as on average these disclosures represent 11.0 % of the words in firms' Form 10-K. Regulators are often criticized for failing to mandate sufficient disclosure regarding firms' risk and uncertainties

[1] Some firms voluntarily provide risk disclosures in MD&A if they also provide forward-looking statements about future performance. That is not our focus. We focus on the newly created risk factor disclosure section because it is mandatory for all firms, and throughout all of our tests, we explicitly control for MD&A risk disclosures (and their risk related keywords). We acknowledge that some firms may have moved their voluntary risk disclosures from MD&A to the risk factors disclosure section after it was mandated. Thus, for these firms, the newly created section may not provide completely new information. Nevertheless, our tests are designed to incorporate investors' expectations of risk disclosure and suggest that the newly created disclosures are informative.

and historically react to periods of market volatility by requiring enhanced risk disclosures. In light of the recent financial crisis, the SEC has made firms' risk factor disclosures a focus of its comment letter process, asking firms to provide more details about the specific risks they face. In fact, the SEC recently added a project to its agenda that revisits the risk factor disclosure rules (CFO 2010). Therefore, it is important to document whether risk factor disclosures convey information that is useful to investors.

Academic research is also interested in the effect of disclosure on investors' risk perceptions. Although theoretical work suggests that increased disclosure reduces the cost of capital, empirical studies have found mixed evidence. Kothari et al. (2009a) argue that the reason prior empirical findings are mixed is because disclosure *tone* affects the relationship between disclosure and the cost of capital. While they predict that negative/pessimistic disclosures increase the cost of capital, they are unable to find evidence of this association when the source of the disclosure is the firm itself. The newly created risk factor disclosure section provides a powerful test of the effect of firm-provided disclosure *tone* on cost of capital for two reasons. First, these disclosures are likely to convey information about firms' systematic risk, which is a major component of the cost of capital. Second, by definition, the tone of this section is negative/pessimistic.[2]

Our research offers two important results. First, we find a positive association between risk factor disclosures and *pre-disclosure* measures that proxy for firm risk. To provide further assurance that this association is not spurious, we decompose the risk factor disclosures into five subcategories: (1) financial, (2) tax, (3) legal, (4) other-systematic, and (5) other-idiosyncratic. We find that when firms' pre-disclosure risk proxies are more strongly associated with a given subcategory of risk, firms devote more of their disclosures towards describing that risk.[3] Overall, our evidence suggests that—contrary to critics' assertions—risk factor disclosures are *not* boilerplate but instead meaningfully reflect the risks a firm faces.

Second, we examine whether investors incorporate the unexpected portion of these disclosures into their assessments of firm risk and value. We find a positive association between the unexpected portion of risk factor disclosures and the post-disclosure level of market beta and stock return volatility, suggesting that the disclosures are positively associated with investors' assessment of firms' fundamental risk. That is, investors revise their estimate of the *level* of the risk parameters in the distribution function of future cash flows.

After controlling for the effect of risk disclosure on post-disclosure beta and stock return volatility, we find a negative association between the unexpected

[2] We acknowledge that it is possible for firms to provide risk factor disclosure that describes a negative/pessimistic event that is actually *less* negative/pessimistic than the market expects it to be. Thus, for our market tests, we explicitly control for investors' expectations of risk factor disclosure. See Sects. 3.1 and 3.3.

[3] Appendices 1 and 2 describe our text analysis procedures and our methods for classifying key words into the five risk subcategories. We first classify key words into financial, tax, and legal risk subcategories. With the remaining words, we classify them as "other-systematic" if they relate to economy-wide risk and "other-idiosyncratic" if they relate to firm-specific risk. As shown in Table 2, 69 % of keywords in the average risk factor disclosure are comprised of words that fall into the "other" categories.

portion of risk factor disclosures and the post-disclosure level of information asymmetry among investors (i.e., bid-ask spread). This suggests that, after controlling for the fact that risk factor disclosure *increases* the market's assessment of a firm's risk, the public availability of risk factor disclosure *decreases* information asymmetry among that same firm's shareholders.

To further ensure that our results are capturing investor reaction to the risk factor disclosures (rather than events subsequent to the 10-K release), we document a negative association between the unexpected portion of risk factor disclosures and short-window abnormal returns around the Form 10-K release date. These results remain after controlling for market risk, size, growth, changes in analysts' estimates of future cash flows, and other information in the 10-K. This suggests that investors incorporate the information from risk factor disclosures into stock price.

Our research contributes to the emerging literature on the effect of disclosure *tone* on firms' cost of capital. Theoretical and empirical research that examines the effect of disclosure on market-based assessments of firm risk generally assumes that the tone of disclosure is neutral (Barry and Brown 1985; Lang and Lundholm 1996; Botosan 1997). When disclosure tone is neutral, the expectation is unidirectional—disclosure decreases the cost of capital. Kothari et al. (2009a) argue that disclosure tone affects the association between disclosure and firms' cost of capital in a directional manner. Although Kothari et al. find that negative/pessimistic disclosure increases cost of capital when a combined set of information sources is analyzed (i.e., corporations, analysts, and business press), they are unable to document this relation when the source of the disclosure is the corporation itself (i.e., MD&A disclosure). They suggest two possible explanations for this result: (1) the software they use to analyze tone is inaccurate, or (2) incentives bias managers against providing informative negative/pessimistic disclosures (Kothari et al. 2009a, p. 1661). By using the newly created risk factor section as a powerful setting where the tone is clear, we provide evidence that negative/pessimistic disclosure is positively related to market-based assessments of firm risk, *even when the source of the news is the corporation itself*. This is an interesting finding since negative/pessimistic disclosures increase investors' assessment of systematic risk, and thus the incentive might be strong for managers to resist disclosure. Despite this incentive, we find that the information is, on average, informative.

Second, we contribute to the risk disclosure literature in two key ways. We document that the type of risk a firm faces (i.e., financial, tax, legal, other-systematic, and other-idiosyncratic) affects the type of risks that are disclosed. That is, we provide direct evidence that managers' qualitative risk factor disclosures about risks and uncertainties are not boilerplate on average but instead meaningfully reflect the specific risks the firm faces across a broad spectrum of risk types. Second, we show that *qualitative* risk disclosures are incorporated into firms' market-based measures of risk and stock price. Although prior research has examined mandatory *quantitative* risk disclosures (Item 7A of Form 10-K) (Schrand 1997; Tufano 1998; Roulstone 1999; Wong 2001; Linsmeier et al. 2002), it is an empirical question whether and how investors would use qualitative disclosures that do not provide investors with (1) the likelihood a risk occurs or (2) the quantified impact a risk would have on the firm's financial statements.

Third, we contribute to the emerging literature on textual analysis. In a contemporaneous paper, Li (2008a) counts risk words in the 10-K during time periods prior to the newly created risk factor disclosures and finds that, when risk sentiment increases, future earnings decrease and investors underreact (or fail to react altogether) to the disclosures. We complement Li (2008a) by providing evidence that investors do at least partially react to risk disclosures, at least during time periods after the disclosures became more prominent and concentrated in one section of the Form 10-K. Kravet and Muslu (2013) count risk sentences in the entire 10-K across years before and after the newly created risk factor section. They find that changes in the number of risk sentences are associated with investors' assessments of firms' idiosyncratic risk (i.e., stock return volatility), as well as analyst earnings forecast properties. We complement Kravet and Muslu (2013) in several ways. By focusing strictly on the risk factor section during time periods when it was mandated, we are uniquely positioned to assess the merits of the new risk factor disclosure section, and we confirm their results related to investors revisions of idiosyncratic risk. We also show that investors incorporate the disclosures into nondiversifiable, systematic risk measures, as well as into firm value. Finally, rather than relying on investors' reaction to the disclosures, we use a word counting and classification scheme to provide direct evidence that managers' risk disclosures are not boilerplate but instead reflect the actual risks the firm faces. Overall, our study complements these contemporaneous papers to show a more complete picture of the effects of qualitative risk factor disclosures. Further, since our study examines the determinants of the newly created risk factor section when all firms were mandated to provide this disclosure, we are uniquely able to address the debate on the usefulness of a mandatory risk factor disclosure and the criticism currently facing the SEC.

Moving beyond risk disclosures to textual analysis more broadly, Lehavy et al. (2011) find that the readability of 10-K filings affect analyst forecast dispersion, accuracy, and effort. Similarly, You and Zhang (2009) show that investors underreact to disclosures provided in longer and more complex 10-K filings. We find that risk factor disclosures have an incremental effect on market-based assessments of risk, after controlling for these readability and complexity measures.

Finally, the SEC has recently incorporated a project into its agenda to re-examine the disclosure rules related to risk factors. Its concern is that companies are not accurately disclosing their *firm-specific* risk but instead are providing "generic" and "mind-numbing risk factor discourse" (CFO 2010). By providing evidence that risk factor disclosures are positively related to pre-disclosure measures that proxy for firm risk, while simultaneously documenting that this disclosure is incorporated into post-disclosure market measures, we provide the SEC with empirical evidence suggesting that firms are disclosing at least some of their specific risks and uncertainties in the newly created risk factor section, and this disclosure is informative to investors. However, our results do not necessarily imply that critics who call for improvements to risk factor disclosure are wrong. Requiring firms to provide even more firm-specific information (and less generic, boilerplate language) could further benefit users, as our results suggest that even in their potentially "flawed" state, the disclosures are informative.

2 Hypothesis development

2.1 Risk factor disclosures and pre-disclosure proxies for firm risk

Prior academic literature on corporate disclosure suggests that managers have a self-serving bias to disclose favorable information about the firm. Accounting information is used by current and potential creditors in debt contract negotiation, by boards of directors in assessing executive performance and compensation, by labor unions in contract renegotiations, and by capital markets in setting firm value and in determining the cost of capital (Watts and Zimmerman 1986; Fields et al. 2001; Kothari et al. 2009a). Similarly, this information will also impact manager's career concerns (Kothari et al. 2009b), and could also reveal proprietary information to a firm's competitors or investors (Scott 1994; Ke et al. 2003). As a result, managers are likely biased against providing unfavorable disclosures (Kothari et al. 2009a). Since risk factor disclosure is designed to convey information regarding unfavorable risks and uncertainties of the firm, it seems reasonable that managers would avoid providing meaningful risk factor disclosure and instead have a bias towards vague and boilerplate risk factor disclosures that "attempt to catalog every conceivable problem that might surface: valued employees quitting, tie-ups with suppliers or clients disintegrating, even acts of God, terrorism, and war" (Spindler 2006).

A review of recent 10-K filings presents exactly this sort of vague and boilerplate risk factor disclosure. For example, the 2009 risk factor disclosure for Remington Arms Company states:

> Our business is subject to economic, market and other factors beyond our control or ability to predict. … Any significant or prolonged economic downturn could have a material adverse impact on our financial condition and results of operations. … Moreover, additional risks and uncertainties not currently known to us or that we currently deem to be immaterial may also materially and adversely affect our business and operations. These risks and uncertainties include, but are not limited to, weather and other Acts of God that could result in the loss or destruction of warehoused inventory and stoppages in our ability to manufacture our key products for a sustained period of time. (SEC 2009)

It seems likely that this type of risk factor disclosure is not informative about the firm's specific risks and could be why the SEC states that the disclosures should "avoid generic risk factor disclosures that could relate to any company" (SEC 2010). This example also illustrates that risk factor disclosures may be used by managers to catalog all possible risks and uncertainties, regardless of the likelihood that they will ultimately affect the firm. Consistent with this view, a corporate securities lawyer recently wrote that the disclosures seem to be a way for managers to effectively tell investors: "Seriously–anything can happen. … By investing in our business, you are agreeing that we owe you no duty of care other than not being crooks. We can promise you nothing else" (Corporate Counsel 2006).

Although managers face business and career incentives not to disclose negative information (Watts and Zimmerman 1986; Scott 1994; Fields et al. 2001; Kothari et al. 2009a, b), they also face legal incentives to at least partially provide meaningful, firm-specific risk factor disclosures. Managers may be exposed to high litigation costs if they fail to disclose bad news in a timely manner (Skinner 1994). If a material risk comes to fruition and managers failed to disclose it, they could be sued, face legal liability, and owe damages. Furthermore, lawyers could use anecdotal evidence and the SEC's criticisms regarding risk factor disclosures against the manager to argue that the firm's risk factor disclosures were a laundry list of generic risks that did not adequately warn investors of the material risk. For instance, in a recent class action securities lawsuit, plaintiffs argued that Credit Suisse concealed the degree of its risk exposure to mortgage-backed securities in its SEC filings. A lead plaintiff's complaint heavily referenced an SEC comment letter discussing various deficiencies with Credit Suisse's 2007 annual report specifically mentioning issues with the risk factor section. Credit Suisse settled the lawsuit in 2011, paying $70 million to the plaintiffs.[4]

If there is an insufficient amount of variation in risk factor disclosures across firms or if variation in the disclosures is not related to variation in risks faced, there should be no association between risk factor disclosures and pre-disclosure measures that proxy for firm risk. On the other hand, if the disclosures are firm-specific, we should find a positive relation between disclosure levels and pre-disclosure, firm-specific risk proxies. That is, managers should provide risk factor disclosures when firms' quantitative measures suggest that they face more risk. Ultimately, it is an empirical question whether risk factor disclosures are meaningful. Accordingly, we test the following hypothesis:

Hypothesis 1 The length and content of risk factor disclosures are associated with pre-disclosure proxies for firm risk.

2.2 Risk factor disclosures and post-disclosure market-based measures of firm risk

In H1, we investigate whether managers' risk factor disclosure meaningfully reflects the pre-disclosure risks a firm faces. To the extent that it does, this suggests that managers do not simply disclose a laundry list of generic or boilerplate risks which do not vary across firms. However, H1 does not assess whether managers use risk factor disclosures to only discuss risks that the market already knows about. If so, investors would not find the disclosures informative, as there would be little information conveyed to cause investors to revise their ex ante beliefs about the firm's risk. Theoretical research suggests that incremental disclosure reduces the cost of equity in one of three ways. First, higher levels of disclosure reduce information asymmetry among current and prospective shareholders, and thus disclosure improves the liquidity of a security (Amihud and Mendelson 1986; Easley and O'Hara 2004). Second, investors may have difficulty estimating firms'

[4] Cornwell et al. v. Credit Suisse Group et al., No. 08 Civ. 3758, 2010 U.S. Dist.

true expected rates of return on securities, and these errors in estimation are nondiversifiable and thus positively related to firms' cost of capital (Klein and Bawa 1976; Barry and Brown 1985). Third, more precise disclosures decrease the assessed covariance between a firm's cash flows and the market's cash flows (i.e., market beta), and thus increased precision of disclosure reduces the cost of equity (Lambert et al. 2007). In sum, the theoretical literature suggests that the effect of disclosure on firm risk is unidirectional. That is, increases in disclosure reduce the cost of capital.

In contrast, Kothari et al. (2009a) note that prior empirical tests of this hypothesis have found some evidence (Lang and Lundholm 1996; Botosan 1997; Botosan and Plumlee 2002, among others), but the results across the literature are mixed. They argue that one explanation for this lack of broad-based empirical support could be that the relation between disclosure and firms' cost of capital (as well as stock return volatility) is directional. That is, "… on average, unfavorable information would be associated with an increase in cash flow risk and, conversely, favorable information would be associated with a decline in the risk of expected future cash flows" (p. 1646). Their argument is, in part, based on the fact that prior research suggests that, on average, unfavorable news is associated with higher volatility of future earnings, and thus unfavorable disclosure is likely to decrease the certainty with which investors can forecast a firm's future cash flows (Chan 1988; Ball and Kothari 1989).

The newly created risk factor section of the 10-K is a unique form of disclosure where all the news relates to "unfavorable" information. More specifically, by definition, the information in this section relates to *uncertainty* about future firm performance. Post-disclosure measures of market beta (for systematic risk) and stock return volatility (for idiosyncratic risk) should proxy for investors' ex ante, expected risk associated with the firms' future cash flows, after incorporating the information conveyed by risk factor disclosures. Therefore, if the disclosures are informative and investors incorporate this information into their risk assessments, we should find a positive association between risk factor disclosures and post-disclosure market-based measures of firm risk (i.e., market beta and stock return volatility). Thus, our second hypothesis:

Hypothesis 2 The unexpected portion of risk factor disclosure is positively associated with post-disclosure market-based measures of firm risk (i.e., beta and stock return volatility).

2.3 Risk factor disclosures and post-disclosure information asymmetry

Theoretical research argues that voluntary disclosure reduces information asymmetries among informed and uninformed investors (Diamond and Verrecchia 1991; Kim and Verrecchia 1994). Consequently, these studies suggest that "for firms with high levels of disclosure, investors can be relatively confident that any stock transactions occur at a 'fair price,' increasing liquidity in the firm's stock (Healy and Palepu 2001, p. 429). Consistent with these arguments, prior empirical studies document a negative relation between the level of disclosure and firms' bid-ask spread. For instance, Leuz and Verrecchia (2000) find that firms listed on the Neuer

Market have lower bid-ask spreads than firms listed on the Frankfurt Exchange and attribute this finding to differences in disclosure requirements between the two.

A maintained assumption of this prior literature is that incremental disclosure is credible and informative and not boilerplate in nature or otherwise not meaningful. For reasons discussed in H1 and H2, we expect that managers will provide risk factor disclosures that meaningfully reflect the risks their firm faces and that the disclosures will be informative to market participants. Thus, we also predict that higher levels of risk factor disclosure will reduce information asymmetries among informed and uninformed investors, result in greater liquidity, and therefore will be negatively associated with information asymmetry (as measured by firms' bid-ask spread). It is important to note, however, that in H2 we predict that risk factor disclosures will increase investors' perception of a firm's risk, and prior literature suggests that, when investors perceive a firm to face more risk, the bid-ask spread increases because informed investors may have an even greater information advantage (Kyle 1985; Demsetz 1986; Jayaraman 2008). This presents an empirical challenge when assessing the effects of risk factor disclosure on information asymmetry because it implies that we already know that the phenomenon we have tested with H2 (i.e., investors' perceptions of risk) also has an effect on the relation between risk disclosures and information asymmetry. Our interest is not in testing two hypotheses surrounding the same phenomenon, but rather more generally on whether risk factor disclosures are informative and thus whether the disclosures reduce the information gap between informed and uninformed investors *conditional on investor risk perceptions*. Therefore, we control for both pre-disclosure and post-disclosure measures of firm risk. Our third hypothesis follows:

Hypothesis 3 After controlling for investor risk perceptions, the unexpected portion of risk factor disclosure is negatively associated with post-disclosure market-based measures of information asymmetry (i.e., bid-ask spread).

2.4 Investor reaction to risk factor disclosures

So far we predict that managers provide informative risk factor disclosures, and investors incorporate this information into their post-disclosure assessments of firm risk and information asymmetry. However, we have not explicitly examined the timeliness with which investors incorporate this information. This is an important point, as it is possible that investors do not use the information provided in risk factor disclosures. Instead, the information could be correlated with subsequent information released by the firm, analysts, or the business press over the course of the following year for which we do not adequately control, and this subsequent release of information could manifest itself in post-disclosure market-based measures of firm risk at that time. By examining returns around the 10-K release date, we attempt to overcome this potential correlated omitted variable problem.[5]

[5] We acknowledge that short-window returns at the 10-K release date could be a function of changes in either (1) firms' expected future cash flows, or (2) the assessment of firm risk. We interpret our results as being related to firm risk. For more assurance regarding this interpretation, in Table 8 we control for both firms' earnings surprise and changes in analysts' estimates of future earnings, as well as other variables in

Beginning with Ball and Brown (1968) and Beaver (1968), prior event study literature uses short-window tests around accounting report release dates to determine whether and how investors react to accounting disclosures. We expect that investors react to the information conveyed by risk factor disclosures at the time the disclosure is made available. We expect that risk factor disclosure will be negatively related to short-window abnormal stock returns around the 10-K release date based on the predicted increase in firms' risk (H2). Although we predict that information asymmetry will improve (H3), we expect that risk factor disclosures have an overall negative association with short-window stock returns for at least two reasons. First, our tests of H3 control for the extent to which increases in firm risk result in an increase in information asymmetry. Thus, it is not clear ex ante whether the overall effect of risk factor disclosures on information asymmetry would be an increase or decrease. Second, the relation between information asymmetry and a firm's cost of capital (and thus its market value over short windows around the 10-K release) is still debated in the literature. Lambert et al. (2012) argue that, in models of perfect competition, information asymmetry *per se* does not affect the cost of capital. Therefore, we predict that risk factor disclosures should be negatively related to market returns. Thus, our final hypothesis:

Hypothesis 4 The unexpected portion of risk factor disclosures are negatively associated with abnormal returns surrounding the Form 10-K release date.

3 Research design and empirical results

3.1 Textual analysis data collection process

Appendix 1 describes the data collection process we use to extract qualitative risk disclosures from firms' Form 10-K and describes how we then calculate the measures used in our study. Additionally, Fig. 1 presents a summary illustration of the specific sequence of steps used in our textual analysis. As can be seen in Fig. 1, annual Form 10-K filings are downloaded from the SEC's Electronic Data Gathering and Retrieval (EDGAR) database and are then processed to generate appropriate counting measures that objectively quantify firms' risk disclosures.[6] We wrote our text analysis program to identify the various Form 10-K subsections based on the assumption that most EDGAR filings contain visual clues for human readers to recognize section boundaries (i.e., subsection titles, boldface fonts, extra spacing, etc.). Our program searches for various combinations of these clues in the HTML code for each filing and partitions the subsections accordingly. Full details of our

Footnote 5 continued
prior literature that could indicate a change in firms' cash flows. Our tests suggest that the inferences with respect to abnormal returns reflect a change in investors' perception of firm risk.

[6] In our main tests, we only use qualitative disclosure information from Item 1A "Risk Factors" and Item 7 "Management's Discussion & Analysis of Financial Condition and Results of Operation" (MD&A). In sects. 4.1 and 4.3, we also control for disclosures in "Quantitative and Qualitative Disclosures about Market Risk" (Item 7A) and our results are unaffected.

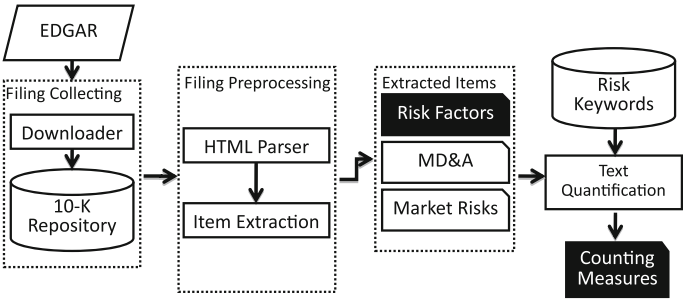

Fig. 1 Analysis steps in constructing qualitative measures of risk factor disclosure for each company

programming procedures are outlined in Appendix 1. To verify that our program is accurate, we randomly selected 300 firms in the EDGAR population and manually examined the words from Item 1A Risk Factors, Item 7 MD&A, and Item 7A Market Risk, sections respectively. For cases where text could be extracted, we found that, in over 98 % of the cases across all three subsections, the software extracted the correct subsection, and only the correct subsection, of text.

After having extracted the appropriate subsection of text, we then compute three measures for each subsection: (1) total word count, (2) total key word count, and (3) key word count by risk subcategory. We use the natural logarithm of each of these three measures. We also formally calculate "unexpected disclosure" measures, which control for the risk factor disclosure in the prior year as well as the quantitative determinants of the disclosure. We do not, however, use a simple change in the disclosure measures because Nelson and Pritchard (2007) show that once risk disclosures are made, they very rarely are removed in a subsequent year. Thus, a simple change approach could be misspecified if investors have more reason to be alarmed both when disclosures are removed as well as when they are significantly increased.[7]

Our list of key words is presented in Appendix 3. We develop our key words in two steps. First, we begin with key risk words used by prior literature (Nelson and Pritchard 2007). Then, our software identified a list of additional words that repeatedly appeared in firms' risk factor sections using a document clustering approach known as the Latent Dirichlet Allocation (Blei et al. 2003). To our knowledge, we are the only study to apply this methodology to risk disclosures, and

[7] Alternative methods to measure changes in textual risk disclosures include the rate of change in the frequency of specific words used within text or the frequency of word groups within a sentence (Brown and Tucker 2011; Nelson and Pritchard 2007). We choose an expectations model as our proxy for changes in risk factor disclosure due to its relative empirical simplicity and since it explains approximately 80 % of the variation in risk factor disclosures (see Table 9).

consequently our key word list of over 300 words is larger than most others in the literature.[8]

To classify each identified key word into subcategories, we first identify those key words associated with (1) financial risk, (2) litigation risk, and (3) tax risk. In general, words were classified as financial risk if they related to liquidity, debt, covenants, or capital structure. Similarly, as in Nelson and Pritchard 2007), words were classified as litigation risk if they related to legal matters, lawsuits, intellectual property, or environmental issues. Finally, words were classified as tax risk if they related to the accounting for income taxes or tax avoidance in general. This process covered approximately 30 % of the key word list. With the remaining 70 %, we classified the terms as either "other-systematic" or "other-idiosyncratic" based on whether the key word most closely related to an economy-wide risk or a firm-specific risk, respectively.

3.2 Sample selection and descriptive statistics

As we are interested in the effects of the newly created risk factor section, we begin our sample selection with the first year in which the SEC mandated this disclosure. Accordingly, our sample of firms begins in 2005. We begin with the overall universe of Compustat firms and then remove firm-year observations that do not have an industry classification or that are missing necessary control variables for our multivariate regression models. As previously mentioned, our textual analysis software requires firms to have a machine-readable (i.e., HTML) filing on the SEC's EDGAR database that includes the newly created "Risk Factors" disclosure. Accordingly, we remove firm-year observations that do not have filings with risk factor disclosures that allow for textual analysis, though in Panel B we consider whether this leads to a bias in our sample by comparing it to the Compustat universe of firms. Finally, we remove all observations that lack necessary equity market data from CRSP and truncate all necessary data at the top and bottom 1 % to mitigate the influence of outliers. The final sample consists of 9,076 firm-year observations. Details concerning the sample selection for our main empirical tests are described in Panel A of Table 1.

Panel B of Table 1 presents an industry breakdown of our sample into 48 industry groupings (see Fama and French 1993). Panel B also reports the industry groupings for the overall universe of Compustat firms. The most striking observation is that the trading industry is under-represented in our sample relative to the overall Compustat universe by 9.0 %. However, as discussed, our sample requires that firms have a filing in the SEC's EDGAR database that includes the newly created "Risk Factors" disclosure. General Instruction J to Form 10-K states that asset-backed issuers are not required to disclose the risk factors item. Thus, the

[8] We acknowledge that there is no consensus in the literature as to which key word list is most appropriate. There are two other relevant studies that report their list of key risk words. Nelson and Pritchard identify 75 risk factor terms, and Kravet and Muslu (2013) identify 20 risk-related keywords based on their reading of 100 randomly selected annual reports. Our key word list includes the vast majority of these identified key words but is expanded considerably using the Latent Dirichlet Allocation method described in the paper.

under-representation of firms in the trading industry is not surprising.[9] The remaining industries in our sample are relatively close to the overall composition of the Compustat universe, except for the banking industry, which is over-represented relative to the overall Compustat universe by 4.7 %. To reduce concerns regarding the effects of industry concentration, we control for differences in industry in all our regression models using industry fixed effects.

Table 2 presents descriptive statistics for the 9,076 firm-year observations. Panel A presents descriptive statistics for our primary risk factor disclosure variables and shows that our sample firms, on average, write 4,902 words in the risk factor section (Item 1A) of their Form 10-K. This is a strikingly large portion of the 10-K (i.e., 4,902/45,911, or 10.7 % of the entire document). For comparison purposes, the risk factor section contains approximately one-half the number of words as the MD&A section (Item 7). In addition to the number of overall words, we examine how many of the key words from Appendix 3 appear in the text. We find that, on average, 293 of the key words from Appendix 3 appear in the risk factor section, and most of these key words relate to "other-systematic" (36 %) and "other-idiosyncratic" risks (33 %). We also find that 15, 13, and 2 % of these words specifically relate to litigation, financial, and tax risks, respectively.

Panel B of Table 2 provides the descriptive statistics for our other variables. The average firm in our sample has a log market value of equity of 6.463 (which is equivalent to $640 million), and size appears to vary considerably across the sample. Additionally, the average firm has a book to market ratio of 0.608 and an annual stock return of -0.003 %. This suggests that the severe market downturn in 2008 overwhelms the generally positive returns from the years 2005–2007. The firms in our sample appear to be profitable, as average net income deflated by the beginning of period market value of equity is 0.006. Finally, the firms in our sample appear to have similar market-based risk measures as in prior research. For example, the mean standard deviation of abnormal daily returns is 0.031 for the 250 trading days following the release of the 10-K and 0.025 for the 250 trading days prior to the release of the 10-K. This is comparable to Kothari et al. (2009a), where the mean of daily return volatility is 0.027. Similarly, the mean value of market beta in the 250 trading days after the 10-K release date is 0.915, which is less than 1, indicating that firms in our sample are slightly less risky than average. The mean bid-ask spread for the 250 trading days following the release of the 10-K is 0.68 %. Overall, the evidence from Table 2 suggests that the descriptive statistics of our sample firms are representative of prior studies and are consistent with the time period during which our sample spans. Consequently, our results should be generalizable.

3.3 Risk factor disclosures and pre-disclosure proxies for firm risk

In our first hypothesis, we predict that risk factor disclosures will be positively associated with pre-disclosure proxies for firm risk. To examine H1, we model risk factor disclosures as a function of these measures identified by prior literature that

[9] Specifically, the trading industry includes the following firm descriptions: "security and commodity brokers," "closed-end management investments," "trusts," and "unit investment trusts."

Table 1 Sample selection and industry concentration

Panel A: Sample selection	
All Compustat firms (Fiscal Years = 2005, 2006, 2007, 2008)	43,081
Less	(7,335)
Missing *SIC*	35,746
Firms with missing control variables	
Missing *CIK*	(2,228)
Missing *AT*	(2,869)
Missing *MVE*	(2,637)
Missing *BVD*	(62)
Missing *NI*	(67)
Missing *BVE*	(2,641)
Missing *Lagged MVE or AT*	(241)
	25,001
Merge with text analysis	
Missing text analysis	(9,777)
	15,224
Merge with CRSP data	
Missing *RET*	(2,542)
Missing current or next year *BETA* or *STDERET*	(485)
Key word count < 10	(227)
Stock Price < $2	(1,022)
Truncation at the top and bottom 1 %	(1,872)
Observations in final sample	9,076

Panel B: Industry concentration

Industry	Sample		Compustat	
	Number	Percent	Number	Percent
Banking	1,260	13.88	3,286	9.19
Business services	930	10.25	3,424	9.58
Trading	678	7.47	5,886	16.47
Electronic equipment	484	5.33	1,798	5.03
Pharmaceutical products	407	4.48	2,122	5.94
Insurance	403	4.44	933	2.61
Retail	399	4.40	1,118	3.13
Petroleum and natural gas	353	3.89	1,356	3.79
Wholesale	292	3.22	840	2.35
Computers	291	3.21	1,065	2.98
Medical equipment	271	2.99	918	2.57
Machinery	267	2.94	749	2.10
Transportation	237	2.61	828	2.32
Utilities	223	2.46	1,200	3.36
Measuring and control equipment	205	2.26	545	1.52

Table 1 continued

Panel B: Industry concentration

Industry	Sample		Compustat	
	Number	Percent	Number	Percent
Communication	174	1.92	1,114	3.12
Chemicals	169	1.86	566	1.58
Healthcare	150	1.65	419	1.17
Construction materials	137	1.51	361	1.01
Food products	132	1.45	386	1.08
Apparel	116	1.28	310	0.87
Electrical equipment	115	1.27	416	1.16
Other	113	1.25	927	2.59
Restaurants, Hotels, Motels	112	1.23	420	1.17
Automobiles and trucks	102	1.12	367	1.03
Personal services	102	1.12	290	0.81
Construction	100	1.10	261	0.73
Business supplies	96	1.06	262	0.73
Remaining industries (Less < 1 %)	758	8.35	3,579	10.01
	9,076	100.00	35,746	100.00

relate to systematic, idiosyncratic, financial, legal, and tax risk (Fama and French 1993; Jones and Weingram 1996; Nelson and Pritchard 2007). Specifically, to test H1, we estimate the following regression equation:

$$\begin{aligned} \text{RF\_DISC}_{i,t} = {} & \beta_0 + \beta_1 \text{SIZE}_{i,t} + \beta_2 \text{BTM}_{i,t} + \beta_3 \text{RET}_{i,t} + \beta_4 \text{LEV}_{i,t} + \beta_5 \text{STDERET}_{i,t} \\ & + \beta_6 \text{BETA}_{i,t} + \beta_7 \text{SKEW}_{i,t} + \beta_8 \text{TURN}_{i,t} + \beta_9 \text{BIGN}_{i,t} + \beta_{10} \text{ETR}_{i,t} \\ & + \beta_{11} \text{DNI}_{i,t} + \beta_{12} \text{NUMEST}_{i,t} + \beta_{13} \text{INSTOWN}_{i,t} + \text{year effects} \\ & + \text{industry effects} + \varepsilon_{i,t} \end{aligned} \quad (1)$$

where the independent variables are defined in Table 2, and the dependent variables are allowed to vary as follows:

RF_DISC = variable capturing risk factor disclosures, which we vary as follows:

(1) ALL_WORDS = the log of the total number of words that appear in the firm's risk factor disclosure section,
(2) KEY_WORDS = the log of the total number of key words identified in Sect. 3.1 that appear in the firm's risk factor disclosure section,
(3) RF_SYS = the percentage of the total number of key words identified in Sect. 3.1 that relate to systematic risks and appear in the firm's risk factor disclosure section,
(4) RF_IDIO = the percentage of the total number of key words identified in Sect. 3.1 that relate to idiosyncratic risks and appear in the firm's risk factor disclosure section,

Table 2 Descriptive statistics

Panel A Descriptive statistics for risk factor disclosures (N = 9,076)							
	Mean	Std	P10	Q1	Median	Q3	P90
Word count							
RF_DISC	4,902	3,301	1,503	2,451	4,130	6,589	9,376
MDA_DISC (N = 8,099)	10,174	5,690	3,982	6,871	9,636	12,994	17,082
ALL_DISC	45,911	17,634	25,664	34,851	44,288	55,124	67,732
Key word count							
RF_DISC	293	190	95	151	247	396	562
MDA_DISC (N = 8,099)	425	259	146	268	391	544	748
ALL_DISC	1,110	534	562	769	1,016	1,337	1,765
Key word intensity by category							
RF_SYS	36 %	12 %	22 %	28 %	34 %	43 %	53 %
RF_IDIO	33 %	16 %	13 %	19 %	33 %	46 %	54 %
RF_FIN	13 %	10 %	3 %	5 %	10 %	20 %	29 %
RF_TAX	2 %	4 %	0 %	0 %	1 %	3 %	8 %
RF_LR	15 %	8 %	6 %	10 %	15 %	20 %	25 %
Key word count by category							
RF_SYS	101	74	34	53	82	126	183
RF_IDIO	103	95	18	36	71	142	236
RF_FIN	36	37	5	12	25	48	80
RF_TAX	9	18	–	–	2	9	25
RF_LR	45	37	10	20	35	60	93
FOG index							
RF_FOG	21	2	19	20	21	22	23
MDA_FOG (N = 8,099)	18	5	17	18	19	20	21
ALL_FOG	19	1	18	19	19	20	21

	Mean	Std	P10	Q1	Median	Q3	P90
Panel B: Descriptive statistics for other variables (N = 9,076)							
$STDERET_{it+1}$	0.031	0.017	0.014	0.019	0.027	0.038	0.053
$STDERET_{it}$	0.025	0.012	0.012	0.017	0.023	0.031	0.041
$BETA_{it+1}$	0.915	0.502	0.194	0.586	0.947	1.250	1.532
$BETA_{it}$	0.902	0.493	0.182	0.579	0.942	1.239	1.512
$SPREAD_{it+1}$	0.678	1.331	0.087	0.125	0.207	0.517	1.720
$SPREAD_{it}$	0.495	0.881	0.085	0.123	0.198	0.447	1.184
$SIZE_{it}$	6.463	1.604	4.431	5.280	6.381	7.534	8.600
Descriptive statistics							
BTM_{it}	0.608	0.437	0.205	0.325	0.514	0.766	1.091
RET_{it}	−0.003	0.385	−0.469	−0.255	−0.025	0.202	0.461
LEV_{it}	0.178	0.181	0.000	0.012	0.129	0.286	0.452
$SKEW_{it}$	0.246	0.949	−0.689	−0.137	0.229	0.647	1.236
$TURN_{it}$	8.307	6.854	1.275	3.179	6.696	11.360	17.482

Table 2 continued

	Mean	Std	P10	Q1	Median	Q3	P90
$BIGN_{it}$	0.763	0.425	0	1	1	1	1
DNI_{it}	0.006	0.029	−0.017	0.002	0.009	0.019	0.031
ETR_{it}	0.264	0.197	0	0.064	0.313	0.373	0.414
NUM_EST_{it}	5.631	5.620	0	1	4	8	14
$INST\_OWN_{it}$	0.202	0.157	0	0.048	0.198	0.324	0.411

Panel A presents descriptive statistics for various word counts across different risk categories. See Appendix 3 for a list of risk category key words. $\boldsymbol{RF\_DISC}_{it}$ is the word count or key word count in the risk factor disclosure (Item 1A). $\boldsymbol{MDA\_DISC_{it}}$ is the word count or key word count in the management discussion and analysis (Item 7). $\boldsymbol{ALL\_DISC_{it}}$ is the total word count or the total key word count in the 10-K filing. $\boldsymbol{RF\_SYS_{it}}$ is the key word percentage or key word count in the risk factor section referring to "other-systematic" risk exposure. $\boldsymbol{RF\_IDIO_{it}}$ is the key word percentage or key word count in the risk factor section referring to "other-idiosyncratic" risk exposure. $\boldsymbol{RF\_FIN_{it}}$ is the key word percentage or key word count in the risk factor section referring to financial risk exposure. $\boldsymbol{RF\_TAX_{it}}$ is the key word percentage or key word count in the risk factor section referring to tax risk exposure. $\boldsymbol{RF\_LR_{it}}$ is the key word percentage or key word count in the risk factor section referring to legal and regulatory risk exposure. $\boldsymbol{RF\_FOG_{it}}$ is the fog index for the risk factor section. $\boldsymbol{MDA\_FOG_{it}}$ is the fog index for the management discussion and analysis section. $\boldsymbol{ALL\_FOG_{it}}$ is the fog index for the 10-K filing

Panel B presents descriptive statistics for variables used in regressions presented on Tables 3 through 7. $\boldsymbol{STDERET_{it+1}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period beginning two trading days after the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns, $\boldsymbol{STDERET_{it}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period ending two trading days before the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{BETA_{it+1}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period beginning two trading days after the 10-K release. $\boldsymbol{BETA_{it}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{SPREAD_{it+1}}$ is the firm's average ending bid-ask spread (divided by ending stock price) for the 250 trading days beginning two trading days after the 10-K release date. $\boldsymbol{SPREAD_{it}}$ is the firm's average ending bid-ask spread (divided by ending stock price) for the 250 trading days ending two trading days before the 10-K release date. $\boldsymbol{SIZE_{it}}$ is the log of the market value of equity (PRCC_F*CSHO). $\boldsymbol{BTM_{it}}$ is the book value of equity (SEQ) divided by the market value of equity. $\boldsymbol{RET_{it}}$ is the 12-month stock return beginning in the fourth month of the fiscal year and ending in the third month after fiscal year-end. A continuous stream of 12 monthly returns is required. $\boldsymbol{LEV_{it}}$ is the firm's book value of debt (DLTT + DD1) divided by total assets (AT). Missing values of debt due in one year (DD1) are replaced with zero. $\boldsymbol{SKEW_{it}}$ is the skewness of daily stock returns for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{TURN_{it}}$ is the average daily share turnover (expressed as a percentage) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{BIGN_{it}}$ is a dummy variable taking on a value of one for firms with a Big N auditor. $\boldsymbol{DNI_{it}}$ is net income before extraordinary items (NI) divided by the lagged market value of equity. $\boldsymbol{ETR_{it}}$ is total tax expense (TXT) divided by pre-tax income. $\boldsymbol{NUM\_EST_{it}}$ is the number of analysis following the firm as reported by I/B/E/S. $\boldsymbol{INST\_OWN_{it}}$ is sum of shares owned by intuitional investors as reported by Thompson Financial divided by total shares outstanding (CSHO)

(5) RF_FIN = the percentage of the total number of key words identified in Sect. 3.1 that relate to financial risks and appear in the firm's risk factor disclosure section,

(6) RF_TAX = the percentage of the total number of key words identified in Sect. 3.1 that relate to tax risks and appear in the firm's risk factor disclosure section,

(7) RF_LR = the percentage of the total number of key words identified in Sect. 3.1 that relate to litigation risks and appear in the firm's risk factor disclosure section.

We expect the coefficient on SIZE (β_1) to be negative if larger firms are more stable and thus have lower risk (Fama and French 1993). However, if larger firms face a greater risk of litigation and thus would be more likely to disclose risk factors (i.e., the political cost hypothesis from Watts and Zimmerman 1986), then we would find a positive coefficient on SIZE. Therefore, we do not sign the coefficient on SIZE. The coefficient on BTM (β_2) will be negative if growth opportunities are positively related to firm risk (Fama and French 1993). However, if market participants have a negative outlook for the future prospects of firms with a high BTM, the firm could be riskier, and we would find a positive coefficient on BTM. Therefore, we do not sign the coefficient on BTM. If realized returns are a proxy for investors' expected rate of return (i.e., cost of capital) beyond those controls already incorporated in the model, then the coefficient on RET (β_3) would be positive (Vuolteenaho 2002). On the other hand, the coefficient on RET (β_3) should be negative if recent negative stock returns indicates that firms have become more fundamentally risky (Nelson and Pritchard 2007). Thus, we make no prediction for RET (β_3). LEV (β_4) should be positively related to risk factor disclosures if higher levels of debt increase a firm's default risk. STDERET (β_5) should be positively related to risk factor disclosures since stock return volatility is often used as a proxy for firm risk (Kothari et al. 2009a, among others). Similarly, BETA (β_6) should be positively related to risk factor disclosures since market beta is often used as a proxy for firm risk (Fama and French 1993, among others).

Consistent with Nelson and Pritchard (2007), we expect a negative coefficient on SKEW (β_7) since higher skewness in stock returns is likely to be associated with recent large and positive stock price movements and thus recently released information that reduces the firm's uncertainty. TURN (β_8) should be positively related to risk factor disclosures if turnover is a proxy for information uncertainty (Jiang et al. 2005). The coefficient on BIGN (β_9) is expected to be negative if firms that hire the services of a big N auditor are associated with increased audit quality and therefore less risk (Mansi, Maxwell, and Miller 2004). However, the coefficient on BIGN could be positive if big N auditors face more litigation risk and therefore require their clients to disclose more risks and uncertainties in their risk factor section (Lennox 1999). Therefore, we do not make a prediction for the coefficient on BIGN. We expect a negative coefficient on ETR (β_{10}) if firms that engage in greater tax avoidance are subject to greater audit risk from the Internal Revenue Service (Guedhami and Pittman 2008). We expect a negative coefficient on DNI (β_{11}) if firms with lower profitability are riskier. The coefficients on NUMEST (β_{12}) and INSTOWN (β_{13}) could be positive if a sophisticated investor base demands that firms provide greater disclosures and transparency (Lang and Lundholm 1996). On the other hand, these coefficients could be negative if a sophisticated investor base faces less information asymmetry with the firm and thus there is less demand for the risk factor disclosure (Healy and Palepu 2001). Thus, we make no prediction on the coefficients for NUMEST and INSTOWN. Finally, to control for varying macro-

economic factors across time and across industry, and since boilerplate and generic disclosures are likely to be sticky across time and across industry, we include both industry and year fixed effects.

3.3.1 Association between risk measures and subcategories of risk

Appendix 3 provides the list of key risk words that we use in our analysis and separates these key words into the following five subcategories discussed in Sect. 3.1 and Appendix 1: (1) financial risk (RF_FIN), (2) tax risk (RF_TAX), (3) litigation risk (RF_LR), (4) other-systematic risk (RF_SYS), and (5) other-idiosyncratic risk (RF_IDIO). Furthermore, prior literature suggests that risk from financial leverage increases both systematic risk and idiosyncratic risk (Hamada 1972; Modigliani and Miller 1958; among others). Thus, in our analyses on the subcategories of risk, we also examine our systematic (RF_SYS) and idiosyncratic (RF_IDIO) risk measures after adding financial risk (RF_FIN) to each of them. We do not include the tax and litigation risk categories in either systematic or idiosyncratic measures because it is not clear from prior literature whether these categories are primarily driven by firm-specific characteristics.

If risk factor disclosures are informative about the specific risks a firm faces, we should find that firms that face greater exposure to a subcategory of risk are more likely to devote a greater portion of the risk disclosures towards describing that risk. Specifically, we expect market risk (BETA) to be more strongly associated with our "other-systematic" risk subcategory (RF_SYS). Similarly, we expect abnormal stock return volatility (STDERET) to be more strongly associated with our "other – idiosyncratic" risk subcategory (RF_IDIO). Finally, we expect financial leverage (LEV) to be more strongly associated with the financial risk subcategory (RF_FIN), effective tax rate (ETR) to be more strongly associated with the tax risk subcategory (RF_TAX), and, consistent with the political cost hypothesis (Watts 1977, Watts and Zimmerman 1986), firm size (SIZE) to be more strongly associated with the litigation risk subcategory (RF_LR).

3.4 Results on risk factor disclosures and pre-disclosure proxies for firm risk

Table 3 presents multivariate evidence regarding H1.[10] The first column of Table 3 uses the key word count of the risk factor section as the dependent variable. The evidence suggests that nine out of the thirteen risk proxies are statistically associated with risk factor disclosure. Specifically, we find that firms increase their risk factor disclosure when they have higher expected returns (RET), have more leverage (LEV), have higher stock return volatility (STDERET) and turnover (TURN), engage a Big N auditor (BIGN), and have greater analyst following (NUM_EST). Additionally, we find that firms increase their risk factor disclosures

[10] Throughout the analysis, we evaluate the effects of multicollinearity with variance inflation factors (VIFs). In their textbook, Kutner et al. (2004) indicate that multicollinearity is not a problem when VIFs are less than 10. The results indicate that multicollinearity is not a serious concern in any of our multivariate regressions. Thus, for expositional purposes, we do not tabulate or discuss these results for each model.

Table 3 Multivariate analysis of risk factor disclosure determinants using word count

$$RF\_DISC_{it} = \beta_0 + \beta_1 SIZE_{it} + \beta_2 BTM_{it} + \beta_3 RET_{it} + \beta_4 LEV_{it} + \beta_5 STDERET_{it} + \beta_6 BETA_{it} + \beta_7 SKEW_{it} + \beta_8 TURN_{it} + \beta_9 BIGN_{it} + \beta_{10} ETR_{it} + \beta_{11} DNI_{it} + \beta_{12} NUM\_EST_{it} + \beta_{13} INST\_OWN_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: RF_DISC	
			KEW WORDS	ALL WORDS
Intercept	ß$_0$	?	5.1217***	8.1123***
			(0.1057)	(0.1211)
$SIZE_{it}$	ß$_1$	?	−0.0074	−0.0182*
			(0.0102)	(0.0107)
BTM_{it}	ß$_2$	?	−0.0179	−0.0109
			(0.0241)	(0.0241)
RET_{it}	ß$_3$	−	0.0406**	0.0506***
			(0.0183)	(0.0187)
LEV_{it}	ß$_4$	+	0.3941***	0.3814***
			(0.0588)	(0.0610)
$STDERET_{it}$	ß$_5$	+	9.6940***	10.0205***
			(1.0216)	(1.0486)
$BETA_{it}$	ß$_6$	+	−0.0056	−0.0244
			(0.0196)	(0.0202)
$SKEW_{it}$	ß$_7$	−	−0.0209***	−0.0258***
			(0.0064)	(0.0066)
$TURN_{it}$	ß$_8$	+	0.0120***	0.0130***
			(0.0016)	(0.0016)
$BIGN_{it}$	ß$_9$	?	0.1094***	0.1094***
			(0.0277)	(0.0278)
ETR_{it}	ß$_{10}$	−	−0.3756***	−0.3944***
			(0.0413)	(0.0426)
DNI_{it}	ß$_{11}$	−	−2.3562***	−2.4668***
			(0.3283)	(0.3211)
NUM_EST_{it}	ß$_{12}$	?	0.0919v	0.0963***
			(0.0143)	(0.0144)
$INST\_OWN_{it}$	ß$_{13}$	?	−0.0558	−0.0976
			(0.0727)	(0.0745)
YEAR FIXED EFFECTS			INCLUDED	INCLUDED
INDUSTRY FIXED EFFECTS			INCLUDED	INCLUDED
Number of observations			9,076	9,076

Table 3 continued

$$RF\_DISC_{it} = \beta_0 + \beta_1 SIZE_{it} + \beta_2 BTM_{it} + \beta_3 RET_{it} + \beta_4 LEV_{it} + \beta_5 STDERET_{it} + \beta_6 BETA_{it} + \beta_7 SKEW_{it} + \beta_8 TURN_{it} + \beta_9 BIGN_{it} + \beta_{10} ETR_{it} + \beta_{11} DNI_{it} + \beta_{12} NUM\_EST_{it} + \beta_{13} INST\_OWN_{it} + \varepsilon_{it}$$

	Prediction	Dependent variable: RF_DISC	
		KEW WORDS	ALL WORDS
Adjusted R-square		0.333	0.322

*** Significant at the 1 % level, ** Significant at the 5 % level, * Significant at the 10 % level

This is a regression of risk factor disclosures on pre-disclosure measures that proxy for firm risk, estimated using Eq. (1). Coefficient standard errors are shown in parentheses below the coefficient loading. Standard errors are Huber-White heteroskedastic robust and are clustered by firm. Fixed effects are included (FF48 and Year). $\boldsymbol{RF\_DISC}_{\text{-it}}$ is the word count or key word count in the risk factor disclosure (Item 1A). $\boldsymbol{SIZE_{it}}$ is the log of the market value of equity (PRCC_F*CSHO). $\boldsymbol{BTM_{it}}$ is the book value of equity (SEQ) divided by the market value of equity. $\boldsymbol{RET_{it}}$ is the 12-month stock return beginning in the fourth month of the fiscal year and ending in the third month after fiscal year-end. A continuous stream of 12 monthly returns is required. $\boldsymbol{LEV_{it}}$ is the firm's book value of debt (DLTT + DD1) divided by total assets (AT). Missing values of debt due in one year (DD1) are replaced with zero if missing. $\boldsymbol{STDERET_{it}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period ending two trading days before the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{BETA_{it}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{SKEW_{it}}$ is the skewness of daily stock returns for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{TURN_{it}}$ is the average daily share turnover (expressed as a percentage) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{BIGN_{it}}$ is a dummy variable taking on a value of one for firms with a Big N auditor. $\boldsymbol{ETR_{it}}$ is total tax expense (TXT) divided by pre-tax income. Pre-tax income is calculated as net income before extraordinary items (NI) plus total tax expense. $\boldsymbol{DNI_{it}}$ is net income before extraordinary items (NI) divided by the lagged market value of equity. $\boldsymbol{NUM\_EST_{it}}$ is the number of analysis following the firm as reported by I/B/E/S. $\boldsymbol{INST\_OWN_{it}}$ is sum of shares owned by intuitional investors as reported by Thompson Financial divided by total shares outstanding (CSHO)

when they have lower profitability (NI), lower effective tax rates (ETR), and lower stock return skewness (SKEW). The results are similar in the second column of Table 3 when we use total word count as the dependent variable rather than key word count. Overall, these results are consistent with H1 and suggest that firms with more risks disclose more risk factors.

Table 4 presents multivariate evidence regarding H1 across our five identified risk subcategories. In all five cases, we find that the magnitude and statistical significance of the coefficient on the risk proxy that most closely captures a given subcategory of risk is largest when the dependent variable in our regression analysis is the portion of the disclosure that is allocated to that subcategory.[11] Specifically, for the "other—systematic" risk subcategory (RF_SYS), we find that market beta (BETA) is positively associated with the portion of disclosure allocated to

[11] Since each left-hand side variable is specified as a percentage of the total key words in the risk factor section, and since the sample (and dependent variables) are the same across all regressions, the coefficients for a particular risk proxy are comparable across all of the regressions and indicate the percentage increase in total key words resulting from that risk proxy.

Table 4 Multivariate analysis of risk factor disclosure determinants using intensity

$$RF\_DISC_{it} = \beta_0 + \beta_1 SIZE_{it} + \beta_2 BTM_{it} + \beta_3 RET_{it} + \beta_4 LEV_{it} + \beta_5 STDERET_{it} + \beta_6 BETA_{it} + \beta_7 SKEW_{it} + \beta_8 TURN_{it} + \beta_9 BIGN_{it} + \beta_{10} ETR_{it} + \beta_{11} DNI_{it} + \beta_{12} NUM\_EST_{it} + \beta_{13} INST\_OWN_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: intenstity						
			RF_SYS	*RF_IDIO*	*RF_FIN*	*RF_TAX*	*RF_LR*	*RF_(SYS + FIN)*	*RF_(IDIO + FIN)*
Intercept	ß$_0$	?	0.3216***	0.3999***	0.1095***	0.0228***	0.1461***	0.4311***	0.5094***
			(0.0200)	(0.0180)	(0.0137)	(0.0081)	(0.0153)	(0.0196)	(0.0166)
$SIZE_{it}$	ß$_1$	?	0.0021	−0.0119***	−0.0006	0.0033***	**0.0072****	0.0014	−0.0125***
			(0.0016)	(0.0016)	(0.0012)	(0.0007)	**(0.0011)**	(0.0019)	(0.0015)
BTM_{it}	ß$_2$	?	0.0180***	−0.0405***	0.0248***	0.0034**	−0.0056**	0.0428***	−0.0158***
			(0.0039)	(0.0039)	(0.0030)	(0.0015)	(0.0026)	(0.0045)	(0.0037)
RET_{it}	ß$_3$	−	0.0014	−0.0031	0.0095***	−0.0049***	−0.0030	0.0110***	0.0064**
			(0.0031)	(0.0031)	(0.0020)	(0.0011)	(0.0020)	(0.0034)	(0.0029)
LEV_{it}	ß$_4$	+	−0.0376***	−0.1447***	**0.1454****	0.0337***	0.0031	0.1078***	0.0008
			(0.0091)	(0.0103)	**(0.0076)**	(0.0044)	(0.0066)	(0.0109)	(0.0093)
$STDERET_{it}$	ß$_5$	+	−0.2671*	**0.5984****	−0.2609**	−0.2101***	0.1397	−0.5280***	**0.3375****
			(0.1625)	**(0.1664)**	(0.1134)	(0.0618)	(0.1107)	(0.1851)	**(0.1588)**
$BETA_{it}$	ß$_6$	+	**0.0100****	−0.0057*	0.0050**	−0.0042***	−0.0051**	**0.0150****	−0.0007
			(0.0031)	(0.0032)	(0.0024)	(0.0011)	(0.0023)	**(0.0036)**	(0.0031)
$SKEW_{it}$	ß$_7$	−	0.0013	−0.0001	−0.0006	0.0000	−0.0006	0.0007	−0.0007
			(0.0011)	(0.0011)	(0.0008)	(0.0004)	(0.0008)	(0.0012)	(0.0010)
$TURN_{it}$	ß$_8$	+	−0.0004	0.0015***	−0.0004***	−0.0003***	−0.0004**	−0.0008***	0.0011***
			(0.0003)	(0.0003)	(0.0002)	(0.0001)	(0.0002)	(0.0003)	(0.0003)
$BIGN_{it}$	ß$_9$	?	0.0024	0.0059	−0.0073**	0.0036**	−0.0047	−0.0048	−0.0014
			(0.0041)	(0.0044)	(0.0033)	(0.0015)	(0.0032)	(0.0047)	(0.0042)
ETR_{it}	ß$_{10}$	−	0.0165**	0.0108	−0.0088*	**−0.0316*****	0.0131***	0.0077	0.0020
			(0.0064)	(0.0070)	(0.0048)	**(0.0032)**	(0.0043)	(0.0073)	(0.0066)

Table 4 continued

$$RF\_DISC_{it} = \beta_0 + \beta_1 SIZE_{it} + \beta_2 BTM_{it} + \beta_3 RET_{it} + \beta_4 LEV_{it} + \beta_5 STDERET_{it} + \beta_6 BETA_{it} + \beta_7 SKEW_{it} + \beta_8 TURN_{it} + \beta_9 BIGN_{it} + \beta_{10} ETR_{it} + \beta_{11} DNI_{it} + \beta_{12} NUM\_EST_{it} + \beta_{13} INST\_OWN_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: intenstity						
			RF_SYS	*RF_IDIO*	*RF_FIN*	*RF_TAX*	*RF_LR*	*RF_(SYS + FIN)*	*RF_(IDIO + FIN)*
DNI_{it}	$ß_{11}$	–	0.3511***	−0.3305***	−0.0726**	0.1104***	−0.0584*	0.2784***	−0.4031***
			(0.0538)	(0.0512)	(0.0296)	(0.0171)	(0.0352)	(0.0594)	(0.0471)
NUM_EST_{it}	$ß_{12}$	?	−0.0082***	0.0184***	−0.0066***	−0.0020**	−0.0016	−0.0148***	0.0118***
			(0.0022)	(0.0023)	(0.0016)	(0.0009)	(0.0016)	(0.0026)	(0.0021)
$INST\_OWN_{it}$	$ß_{13}$	?	0.0018	−0.0042	−0.0043	0.0007	0.0060	−0.0026	−0.0085
			(0.0121)	(0.0123)	(0.0076)	(0.0049)	(0.0080)	(0.0133)	(0.0117)
YEAR FIXED EFFECTS			INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED
YEAR FIXED EFFECTS			INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED
INDUSTRY FIXED EFFECTS			INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED	INCLUDED
Number of observations			9,076	9,076	9,076	9,076	9,076	9,076	9,076
Adjusted R-square			0.429	0.654	0.599	0.240	0.276	0.576	0.422

*** Significant at the 1 % level, ** Significant at the 5 % level, * Significant at the 10 % level

This is a regression of risk factor disclosures on pre-disclosure measures that proxy for firm risk, estimated using Eq. (1). Coefficient standard errors are shown in parentheses below the coefficient loading. Standard errors are Huber-White heteroskedastic robust and are clustered by firm. Fixed effects are included (FF48 and Year). $\boldsymbol{RF\_SYS_{it}}$ is the key word percentage in the risk factor section referring to "other-systematic" risk exposure. $\boldsymbol{RF\_IDIO_{it}}$ is the key word percentage in the risk factor section referring to "other-idiosyncratic" risk exposure. $\boldsymbol{RF\_FIN_{it}}$ is the key word percentage in the risk factor section referring to financial risk exposure. $\boldsymbol{RF\_TAX_{it}}$ is the key word percentage in the risk factor section referring to tax risk exposure. $\boldsymbol{RF\_LR_{it}}$ is the key word percentage in the risk factor section referring to legal and regulatory risk exposure. $\boldsymbol{SIZE_{it}}$ is the log of the market value of equity (PRCC_F*CSHO). $\boldsymbol{BTM_{it}}$ is the book value of equity (SEQ) divided by the market value of equity. $\boldsymbol{RET_{it}}$ is the 12-month stock return beginning in the fourth month of the fiscal year and ending in the third month after fiscal year-end. A continuous stream of 12 monthly returns is required. $\boldsymbol{LEV_{it}}$ is the firm's book value of debt (DLTT + DD1) divided by total assets (AT). Missing values of debt due in one year (DD1) are replaced with zero if missing. $\boldsymbol{STDERET_{it}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period ending two trading days before the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{BETA_{it}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{SKEW_{it}}$ is the skewness of daily stock returns for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{TURN_{it}}$ is the average daily share turnover (expressed as a percentage) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{BIGN_{it}}$ is a dummy variable taking on a value of one for firms with a Big N auditor. $\boldsymbol{ETR_{it}}$ is total tax expense (TXT) divided by pre-tax income. Pre-tax income is calculated as net income before extraordinary items (NI) plus total tax expense. $\boldsymbol{DNI_{it}}$ is net income before extraordinary items (NI) divided by the lagged market value of equity. $\boldsymbol{NUM\_EST_{it}}$ is the natural log of the number of analysis, plus one, following the firm as reported by I/B/E/S. $\boldsymbol{INST\_OWN_{it}}$ is sum of shares owned by intuitional investors as reported by Thompson Financial divided by total shares outstanding (CSHO)

systematic risk ($\beta_6 = 0.0100$), and the coefficient and statistical significance of BETA is larger than for any of the other four risk subcategories. Similarly, for the "other—idiosyncratic" risk subcategory (RF_IDIO), we find that abnormal stock return volatility (STDERET) is positively associated with (and the coefficient and statistical significance is largest for) the portion of disclosure allocated to idiosyncratic risk. For the financial risk subcategory (RF_FIN), we find that financial leverage (LEV) is positively associated with (and the coefficient and statistical significance is largest for) the portion of disclosure allocated to financial risk. For the tax risk subcategory (RF_TAX), we find the effective tax rate (ETR) is negatively associated with (and the coefficient and statistical significance is largest for) the portion of disclosure allocated to tax risk. Finally, for the litigation risk subcategory (RF_LR), we find that firm size (SIZE) is positively associated with (and the coefficient and statistical significance is largest for) the portion of disclosure allocated to litigation risk.

In the last two columns of Table 4, we examine systematic and idiosyncratic risk after adding financial risk to each of them. We continue to find that BETA is more strongly associated with systematic risk, while STDERET is more strongly associated with idiosyncratic risk. Overall, the evidence from Table 4 suggests that firms with greater subcategories of risk are more likely to devote a greater portion of their risk disclosures towards describing that risk and, consistent with H1, provides further assurance that managers provide informative risk factor disclosures.

3.5 Risk factor disclosures and post-disclosure market-based measures of firm risk

In our second hypothesis, we predict that risk factor disclosures are positively associated with post-disclosure market-based measures of firm risk (i.e., beta and stock return volatility). In order to test H2, we follow the research design of Kothari et al. (2009a) and modify it to include the pre-disclosure risk controls from Eq. (1). Specifically, we estimate the following regression:

$$\begin{aligned}\text{BETA}_{i,t+1}\text{or STDERET}_{i,t+1} &= \beta_0 + \beta_1\text{RF\_DISC}_{i,t} + \beta_2\text{MDA\_DISC}_{i,t} \\ &+ \beta_3\text{ALL\_DISC}_{i,t} + \beta_4\text{SIZE}_{i,t} + \beta_5\text{BTM}_{i,t} + \beta_6\text{RET}_{i,t} + \beta_7\text{LEV}_{i,t} \\ &+ \beta_8\text{STDERET}_{i,t} + \beta_9\text{BETA}_{i,t} + \beta_{10}\text{SKEW}_{i,t} + \beta_{11}\text{TURN}_{i,t} + \beta_{12}\text{BIGN}_{i,t} \\ &+ \beta_{13}\text{ETR}_{i,t} + \beta_{14}\text{DNI}_{i,t} + \beta_{15}\text{NUMEST}_{i,t} + \beta_{16}\text{INSTOWN}_{i,t} + \beta_{17}\text{RF\_FOG}_{i,t} \\ &+ \beta_{18}\text{MDA\_FOG}_{i,t} + \beta_{19}\text{ALL\_FOG}_{i,t} + \text{year effects} + \text{industry effects} + \varepsilon_{i,t}\end{aligned} \tag{2}$$

where all variables are as defined in Eq. (1), and RF_DISC = log of the word (key word) count in the risk factor disclosure section, and MDA_DISC = log of the word (key word) count in the MD&A section. If risk factor disclosures are positively associated with post-disclosure market-based measures of firm risk (H2), we expect that $\beta_1 > 0$. Since our interest is in risk factor disclosures where the tone is negative, and not in disclosures regarding the firm's current and future operating performance made in MD&A (Bryan 1997; Clarkson et al. 1999; Li 2010; Brown

Table 5 Multivariate analysis of risk factor disclosure informativeness using word count–predicting future riskiness

$$BETA_{it+1}\ or\ STDERET_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SIZE_{it} + \beta_5 BTM_{it} + \beta_6 RET_{it} + \beta_7 LEV_{it} + \beta_8 STDERET_{it} + \beta_9 BETA_{it} + \beta_{10} SKEW_{it} + \beta_{11} TURN_{it} + \beta_{12} BIGN_{it} + \beta_{13} ETR_{it} + \beta_{14} DNI_{it} + \beta_{15} NUM\_EST_{it} + \beta_{16} INST\_OWN_{it} + \beta_{17} RF\_FOG_{it} + \beta_{18} MDA\_FOG_{it} + \beta_{19} ALL\_FOG_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: $BETA_{it+1}$		Prediction	Dependent variable: $STDERET_{it+1}$	
			KEY WORDS	*ALL WORDS*		*KEY WORDS*	*ALL WORDS*
Intercept	ß$_0$	?	−0.1592**	−0.3273	?	0.0158***	0.0096***
			(0.0762)	***		(0.0024)	(0.0035)
				(0.1176)			
RF_DISC$_{it}$	**ß$_1$**	**+**	**0.0227*****	**0.0182****	**+**	**0.0006*****	**0.0005*****
			(0.0072)	**(0.0071)**		**(0.0002)**	**(0.0002)**
MDA_DISC$_{it}$	ß$_2$	+	0.0109***	0.0077***	+	0.0002**	0.0001**
			(0.0027)	(0.0019)		(0.0001)	(0.0001)
ALL_DISC$_{it}$	ß$_3$	**?**	0.0051	0.0168	**?**	0.0004	0.0007**
			(0.0095)	(0.0109)		(0.0003)	(0.0003)
SIZE$_{it}$	ß$_4$	?	0.0409***	0.0398***	–	−0.0017***	−0.0018***
			(0.0044)	(0.0045)		(0.0001)	(0.0001)
BTM$_{it}$	ß$_5$	?	−0.0357**	−0.0365**	+	0.0023***	0.0022***
			(0.0151)	(0.0151)		(0.0005)	(0.0005)
RET$_{it}$	ß$_6$	?	0.0930***	0.0933***	–	−0.0013***	−0.0013***
			(0.0148)	(0.0148)		(0.0004)	(0.0004)
LEV$_{it}$	ß$_7$	+	−0.0184	−0.0200	+	0.0051***	0.0050***
			(0.0272)	(0.0273)		(0.0008)	(0.0008)
STDERET$_{it}$	ß$_8$	?	2.6501***	2.6693***	+	0.5594***	0.5589***
			(0.6608)	(0.6599)		(0.0200)	(0.0200)
BETA$_{it}$	ß$_9$	+	0.4872***	0.4879***	?	0.0006**	0.0006**
			(0.0113)	(0.0113)		(0.0003)	(0.0003)

Table 5 continued

$$BETA_{it+1}\ or\ STDERET_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SIZE_{it} + \beta_5 BTM_{it} + \beta_6 RET_{it} + \beta_7 LEV_{it} + \beta_8 STDERET_{it} + \beta_9 BETA_{it} + \beta_{10} SKEW_{it} + \beta_{11} TURN_{it} + \beta_{12} BIGN_{it} + \beta_{13} ETR_{it} + \beta_{14} DNI_{it} + \beta_{15} NUM\_EST_{it} + \beta_{16} INST\_OWN_{it} + \beta_{17} RF\_FOG_{it} + \beta_{18} MDA\_FOG_{it} + \beta_{19} ALL\_FOG_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: $BETA_{it+1}$		Prediction	Dependent variable: $STDERET_{it+1}$	
			KEY WORDS	*ALL WORDS*		*KEY WORDS*	*ALL WORDS*
$SKEW_{it}$	β_{10}	?	−0.0101**	−0.0100**	?	−0.0004***	−0.0004***
			(0.0044)	(0.0044)		(0.0001)	(0.0001)
$TURN_{it}$	β_{11}	?	0.0017**	0.0017**	?	0.0001***	0.0001***
			(0.0008)	(0.0008)		(0.0000)	(0.0000)
$BIGN_{it}$	β_{12}	?	0.0634***	0.0641***	–	−0.0009**	−0.0009**
			(0.0126)	(0.0126)		(0.0004)	(0.0004)
ETR_{it}	β_{13}	?	−0.0140	−0.0142	–	−0.0013*	−0.0013*
			(0.0229)	(0.0229)		(0.0008)	(0.0008)
DNI_{it}	β_{14}	?	−0.4204**	−0.4101**	–	−0.0644***	−0.0638***
			(0.1917)	(0.1923)		(0.0055)	(0.0055)
NUM_EST_{it}	β_{15}	?	0.0051	0.0057	?	−0.0003	−0.0003
			(0.0064)	(0.0064)		(0.0002)	(0.0002)
$INST\_OWN_{it}$	β_{16}	?	0.1352***	0.1358***	?	−0.0045***	−0.0045***
			(0.0293)	(0.0293)		(0.0009)	(0.0009)
RF_FOG_{it}	β_{17}	?	−0.0001	−0.0001	?	0.0000	0.0000
			(0.0001)	(0.0001)		(0.0000)	(0.0000)
MDA_FOG_{it}	β_{18}	?	−0.0003**	−0.0003**	?	0.0000	0.0000
			(0.0001)	(0.0001)		(0.0000)	(0.0000)
ALL_FOG_{it}	β_{19}	?	0.0002*	0.0002*	?	0.0000	0.0000
			(0.0001)	(0.0001)		(0.0000)	(0.0000)
YEAR FIXED EFFECTS			INCLUDED	INCLUDED		INCLUDED	INCLUDED
INDUSTRY FIXED EFFECTS			INCLUDED	INCLUDED		INCLUDED	INCLUDED

Table 5 continued

$$BETA_{it+1}\ or\ STDERET_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SIZE_{it} + \beta_5 BTM_{it} + \beta_6 RET_{it} + \beta_7 LEV_{it} + \beta_8 STDERET_{it} + \beta_9 BETA_{it} + \beta_{10} SKEW_{it} + \beta_{11} TURN_{it} + \beta_{12} BIGN_{it} + \beta_{13} ETR_{it} + \beta_{14} DNI_{it} + \beta_{15} NUM\_EST_{it} + \beta_{16} INST\_OWN_{it} + \beta_{17} RF\_FOG_{it} + \beta_{18} MDA\_FOG_{it} + \beta_{19} ALL\_FOG_{it} + \varepsilon_{it}$$

	Prediction	Dependent variable: $BETA_{it+1}$		Prediction	Dependent variable: $STDERET_{it+1}$	
		KEY WORDS	*ALL WORDS*		*KEY WORDS*	*ALL WORDS*
Number of observations		9,076	9,076		9,076	9,076
Adjusted R-square		0.449	0.449		0.561	0.561

*** Significant at the 1 % level, ** Significant at the 5 % level, * Significant at the 10 % level

This is a regression of post-disclosure market-based measures of risk on risk factor disclosures estimated using Eq. (2). Coefficient standard errors are shown in parentheses below the coefficient loading. Standard errors are Huber-White heteroskedastic robust and are clustered by firm. Fixed effects are included (FF48 and Year). $\boldsymbol{BETA_{it+1}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period beginning two trading days after the 10-K release. $\boldsymbol{STDERET_{it+1}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period beginning two trading days after the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{RF\_DISC_{it}}$ is the log of the word count or the log of the key word count in the risk factor disclosure (Item 1A). $\boldsymbol{MDA\_DISC_{it}}$ is the log of the word count or the log of the key word count in the management discussion and analysis (Item 7). $\boldsymbol{ALL\_DISC_{it}}$ is the log of the total word count or the log of the key word count in the 10-K filing. $\boldsymbol{SIZE_{it}}$ is the log of the market value of equity (PRCC_F*CSHO). $\boldsymbol{BTM_{it}}$ is the book value of equity (SEQ) divided by the market value of equity. $\boldsymbol{RET_{it}}$ is the 12-month stock return beginning in the fourth month of the fiscal year and ending in the third month after fiscal year-end. A continuous stream of 12 monthly returns is required. $\boldsymbol{LEV_{it}}$ is the firm's book value of debt (DLTT + DD1) divided by total assets (AT). Missing values of debt due in one year (DD1) are replaced with zero if missing. $\boldsymbol{STDERET_{it}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period ending two trading days before the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{BETA_{it}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{SKEW_{it}}$ is the skewness of daily stock returns for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{TURN_{it}}$ is the average daily share turnover (expressed as a percentage) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{BIGN_{it}}$ is a dummy variable taking on a value of one for firms with a Big N auditor. $\boldsymbol{ETR_{it}}$ is total tax expense (TXT) divided by pre-tax income. Pre-tax income is calculated as net income before extraordinary items (NI) plus total tax expense. $\boldsymbol{DNI_{it}}$ is net income before extraordinary items (NI) divided by the lagged market value of equity. $\boldsymbol{NUM\_EST_{it}}$ is the natural log of the number of analysis, plus one, following the firm as reported by I/B/E/S. $\boldsymbol{INST\_OWN_{it}}$ is sum of shares owned by institutional investors as reported by Thompson Financial divided by total shares outstanding (CSHO). $\boldsymbol{RF\_FOG_{it}}$ is the fog index for the risk factor section. $\boldsymbol{MDA\_FOG_{it}}$ is the fog index for the management discussion and analysis section. $\boldsymbol{ALL\_FOG_{it}}$ is the fog index for the 10-K filing

and Tucker 2011), we control for the number of words (and the number of key words relating to risk) in MD&A disclosures at time t.[12] In addition, since prior research suggests that the complexity and length of 10-K disclosure is associated with investors' risk perceptions (You and Zhang 2009; Lehavy et al. 2011), we include controls for the length of disclosure in the entire 10-K (ALL_DISC) as well as the fog index of the disclosures in the risk factor section (RF_FOG), in the MD&A section (MDA_FOG), and in the rest of the 10-K (ALL_FOG).[13] Finally, we use industry and year fixed effects since boilerplate or generic disclosures are likely to be sticky across time for the same firm or are likely to be homogenous across firms in the same industry, as well as standard errors that control for both serial correlation and heteroscedasticity.

Finally, there are two important research design choices in all of our market-based tests (i.e., tests of H2 using Eq. (2), as well as tests of H3 and H4 in Eqs. (3) and (4), which follow). First, even though the tone of risk factor disclosure is negative by definition, it is possible that firms convey less negative information than investors expected and therefore the overall news from the disclosure is positive. That is, it is important to control for investors' expectations of risk factor disclosures. To do so, we control for pre-disclosure measures that proxy for firm risk (i.e., all of the determinants of risk disclosure from Eq. (1) and tests of H1). By doing so, our measure for risk factor disclosures (RF_DISC) captures the portion of risk factor disclosures that are "unexpected." See Sect. 4.2 for sensitivity tests where we formally estimate a two-stage regression model. Second, we model our dependent variable as the level of post-disclosure risk at time $t + 1$. By including a control for the level of risk at time t, this ensures that our model is capturing the change in the risk measure subsequent to the release of the risk factor disclosure. That is, our model is equivalent to a pure changes model, except that in a pure changes model the coefficient on the lagged value of the dependent variable is constrained to equal one rather than being allowed to vary. Not surprisingly, all of our results are quantitatively and qualitatively unchanged if our dependent variable is the change in risk from time t to time $t + 1$. For additional discussion of the pure changes model and tabulated regression results, see Sect. 4.3.

3.5.1 Association between market-based risk measures and subcategories of risk

As previously mentioned, we decompose the risk factor disclosures into subcategories based on different types of risk, including systematic and idiosyncratic risks. Our two market-based measures of risk are market beta and stock return volatility, and prior literature suggests that market beta mostly captures systematic risk while stock return volatility mostly captures firm-specific idiosyncratic risk. Thus, we

[12] We acknowledge that we do not explicitly control for the MD&A tone. As noted by Kothari et al. (2009), this is not easy to do as it requires software-reading technologies that are not particularly accurate. However, we follow prior literature that assesses the tone of MD&A (Tetlock 2007; Kravet and Muslu 2013) and count the number of words that relate to risk, assuming that the context of these words is negative/pessimistic (i.e., our variables MDA_DISC, MDA_SYS, MDA_IDIO, MDA_FIN, MDA_LIT, and MDA_TAX).

[13] The fog index is defined by Li (2008b) as (words per sentence + percent of complex words) * 0.4.

re-estimate Eq. (2) and decompose the key word count into our measures of "other-systematic" (RF_SYS) and "other-idiosyncratic" (RF_IDIO) risk.[14] If investors incorporate risk factor disclosures into post-disclosure measures of firm risk, the association between risk factor disclosures and BETA (STDERET) should mostly be driven by the RF_SYS (RF_IDIO) component of risk disclosures.

3.6 Results on risk factor disclosures and post-disclosure market-based measures of firm risk

Table 5 presents multivariate evidence regarding H2. The first column of Table 5 uses market beta (BETA) as the dependent variable and the key word count of the risk factor section as the independent variable of interest. We find a positive and significant association between RF_DISC and BETA ($\beta_1 = 0.0227$, std. error $= 0.0072$, t-statistic $= 3.15$). The third column of Table 5 uses abnormal stock return volatility (STDERET) as the dependent variable rather than BETA. As with BETA, we find a positive and significant association between RF_DISC and STDERET ($\beta_1 = 0.0006$, std. error $= 0.0002$, t-statistic $= 3.00$). This suggests that, if a firm moves from the 25th percentile to the 75th percentile of risk disclosure, investors incorporate an additional 2.5 % (1.9 %) into their assessment of the following period's BETA (STDERET).[15] The results are unchanged in the second and fourth columns when we substitute total word count as the independent variable rather than key word count. Overall, these results are consistent with H2 and suggest that investors incorporate the information conveyed by risk factor disclosures into post-disclosure market-based measures of risk.

As previously discussed, we expect that the positive association between risk factor disclosures and post-disclosure market beta is primarily driven by systematic risk factor disclosures, while the positive association between risk factor disclosures and abnormal stock return volatility is primarily driven by firm-specific idiosyncratic risk factor disclosures. Table 6 presents multivariate evidence regarding H2 across these two risk subcategories. The third column in Table 6 presents results for market beta when both systematic and idiosyncratic risk disclosure measures are included in the regression model. As expected, the coefficient on systematic risk disclosure is positive and significant ($\beta_1 = 0.0382$, std. error $= 0.0103$, t-statistic $= 3.71$), while the coefficient on idiosyncratic risk disclosure is not statistically significant. This suggests that, when managers disclose more systematic risk factors, investors respond by incorporating this risk into the systematic market-based measure BETA. On the other hand, the sixth column in Table 6 presents results for

[14] As before, we also add financial risk to each of these measures since prior literature shows that financial risk affects both systematic and idiosyncratic risk. Similarly, we do not include legal or tax risk in either of these categories since it is difficult to determine whether these risks are firm-specific, and prior literature does not provide much guidance in this respect.

[15] For comparison purposes, Kothari et al. (2009) examine the effect of negative/pessimistic disclosure across three sources of disclosure (corporations, analysts, business press), and their results suggest that moving from the 25th percentile to the 75th percentile increases firms' cost of capital by 2.0%. However, as previously mentioned, they find no such relation when the source of the disclosure is the corporation itself.

Table 6 Multivariate analysis of risk factor disclosure informativeness by category–predicting future riskiness

$$BETA_{it+1} \text{ or } STDRET_{it+1} = \beta_0 + \beta_1 RF\_SYS_{it} + \beta_2 RF\_IDIO_{it} + \beta_3 MDA\_SYS_{it} + \beta_4 MDA\_IDIO_{it} + \beta_5 ALL\_SYS_{it} + \beta_6 ALL\_IDIO_{it} + \beta_7 SIZE_{it} + \beta_8 BTM_{it} + \beta_9 RET_{it} + \beta_{10} LEV_{it} + \beta_{11} STDRET_{it} + \beta_{12} BETA_{it} + \beta_{13} SKEW_{it} + \beta ß_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} + \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: $BETA_{it+1}$			Prediction	Dependent variable: $STDERET_{it+1}$		
			1	2	3		1	2	3
Intercept	$ß_0$	?	−0.1745***	−0.1225*	−0.2125***	?	0.0190***	0.0166***	0.0166***
			(0.0631)	(0.0657)	(0.0682)		(0.0020)	(0.0021)	(0.0022)
***RF_SYS**$_{it}$*	**ß$_1$**	**+**	**0.0303****		**0.0382****	**?**	0.0005**		0.0001
			(0.0073)		**(0.0103)**		(0.0002)		(0.0003)
***RF_IDIO**$_{it}$*	**ß$_2$**	**?**		0.0085	−0.0092	**+**		**0.0005*****	**0.0005***
				(0.0060)	(0.0085)			**(0.0002)**	**(0.0003)**
MDA_SYS$_{it}$	$ß_3$	+	0.0129***		0.0218**	?	0.0002**		0.0002
			(0.0031)		(0.0089)		(0.0001)		(0.0002)
MDA_IDIO$_{it}$	$ß_4$	?		0.0113***	(0.0094)	+		0.0002**	(0.0000)
				(0.0032)	(0.0089)			(0.0001)	(0.0002)
ALL_SYS$_{it}$	$ß_5$	?	0.0112		−0.0064	–	0.0001		−0.0004
			(0.0085)		(0.0127)		(0.0002)		(0.0003)
ALL_IDIO$_{it}$	$ß_6$	?		0.0158*	0.0256*	?		0.0005*	0.0009**
				(0.0094)	(0.0139)			(0.0003)	(0.0004)
SIZE$_{it}$	$ß_7$	?	0.0394***	0.0402***	0.0386***	–	−0.0017***	−0.0017***	−0.0017***
			(0.0044)	(0.0045)	(0.0045)		(0.0001)	(0.0001)	(0.0001)
BTM$_{it}$	$ß_8$	?	−0.0396***	−0.0341**	−0.0409***	+	0.0022***	0.0024***	0.0024***
			(0.0151)	(0.0151)	(0.0152)		(0.0005)	(0.0005)	(0.0005)

Table 6 continued

$$BETA_{it+1} or\ STDRET_{it+1} = \beta_0 + \beta_1 RF\_SYS_{it} + \beta_2 RF\_IDIO_{it} + \beta_3 MDA\_SYS_{it} + \beta_4 MDA\_IDIO_{it} + \beta_5 ALL\_SYS_{it} + \beta_6 ALL\_IDIO_{it} + \beta_7 SIZE_{it} + \beta_8 BTM_{it} + \beta_9 RET_{it} + \beta_{10} LEV_{it} + \beta_{11} STDRET_{it} + \beta_{12} BETA_{it} + \beta_{13} SKEW_{it} + \beta ß_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} + \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: $BETA_{it+1}$			Prediction	Dependent variable: $STDERET_{it+1}$		
			1	2	3		1	2	3
RET_{it}	$ß_9$	?	0.0929***	0.0946***	0.0936***	–	−0.0013***	−0.0013***	−0.0012***
			(0.0148)	(0.0148)	(0.0147)		(0.0004)	(0.0004)	(0.0004)
LEV_{it}	$ß_{10}$	+	−0.0233	−0.0092	−0.0253	+	0.0053***	0.0054***	0.0055***
			(0.0270)	(0.0270)	(0.0271)		(0.0008)	(0.0008)	(0.0008)
$STDRET_{it}$	$ß_{11}$	?	2.6079***	2.6595***	2.5899***	+	0.5616***	0.5568***	0.5566***
			(0.6573)	(0.6627)	(0.6610)		(0.0199)	(0.0201)	(0.0200)
$BETA_{it}$	$ß_{12}$	+	0.4858***	0.4876***	0.4849***	?	0.0006**	0.0006**	0.0006**
			(0.0114)	(0.0113)	(0.0114)		(0.0003)	(0.0003)	(0.0003)
$SKEW_{it}$	$ß_{13}$	?	−0.0101**	−0.0103**	−0.0100**	?	−0.0004***	−0.0004***	−0.0004***
			(0.0044)	(0.0044)	(0.0044)		(0.0001)	(0.0001)	(0.0001)
$TURN_{it}$	$ß_{14}$	?	0.0016**	0.0017**	0.0016**	?	0.0001***	0.0001***	0.0001***
			(0.0008)	(0.0008)	(0.0008)		(0.0000)	(0.0000)	(0.0000)
$BIGN_{it}$	$ß_{15}$	?	0.0623***	0.0633***	0.0620***	–	−0.0009**	−0.0009***	−0.0009***
			(0.0126)	(0.0126)	(0.0126)		(0.0004)	(0.0004)	(0.0004)
ETR_{it}	$ß_{16}$	?	−0.0114	−0.0174	−0.0117	–	−0.0013*	−0.0013*	−0.0014*
			(0.0229)	(0.0228)	(0.0230)		(0.0008)	(0.0008)	(0.0008)
DNI_{it}	$ß_{17}$	?	−0.4316**	−0.4029**	−0.4403**	–	−0.0655***	−0.0637***	−0.0640***
			(0.1910)	(0.1920)	(0.1915)		(0.0055)	(0.0055)	(0.0055)

Table 6 continued

$$BETA_{it+1} or\ STDRET_{it+1} = \beta_0 + \beta_1 RF\_SYS_{it} + \beta_2 RF\_IDIO_{it} + \beta_3 MDA\_SYS_{it} + \beta_4 MDA\_IDIO_{it} + \beta_5 ALL\_SYS_{it} + \beta_6 ALL\_IDIO_{it} + \beta_7 SIZE_{it} + \beta_8 BTM_{it} + \beta_9 RET_{it} + \beta_{10} LEV_{it} + \beta_{11} STDRET_{it} + \beta_{12} BETA_{it} + \beta_{13} SKEW_{it} + \beta ß_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} + \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \varepsilon_{it}$$

		Prediction	Dependent variable: $BETA_{it+1}$			Prediction	Dependent variable: $STDERET_{it+1}$		
			1	2	3		1	2	3
NUM_EST_{it}	$ß_{18}$	?	0.0055	0.0050	0.0061	?	−0.0003	−0.0003*	−0.0003*
			(0.0064)	(0.0064)	(0.0064)		(0.0002)	(0.0002)	(0.0002)
$INST\_OWN_{it}$	$ß_{19}$	?	0.1368***	0.1335***	0.1357***	?	−0.0045***	−0.0045***	−0.0045***
			(0.0293)	(0.0293)	(0.0293)		(0.0009)	(0.0009)	(0.0009)
RF_FOG_{it}	$ß_{20}$	?	−0.0001	−0.0001	−0.0001	?	0.0000	0.0000	0.0000
			(0.0001)	(0.0001)	(0.0001)		(0.0000)	(0.0000)	(0.0000)
MDA_FOG_{it}	$ß_{21}$	?	−0.0003**	−0.0002*	−0.0002*	?	0.0000	0.0000	0.0000
			(0.0001)	(0.0001)	(0.0001)		(0.0000)	(0.0000)	(0.0000)
ALL_FOG_{it}	$ß_{22}$	?	0.0002*	0.0002*	0.0002	?	0.0000	0.0000	0.0000
			(0.0001)	(0.0001)	(0.0001)		(0.0000)	(0.0000)	(0.0000)
YEAR FIXED EFFECTS			INCLUDED	INCLUDED	INCLUDED		INCLUDED	INCLUDED	INCLUDED
INDUSTRY FIXED EFFECTS			INCLUDED	INCLUDED	INCLUDED		INCLUDED	INCLUDED	INCLUDED
Number of Observations			9,076	9,076	9,076		9,076	9,076	9,076

Table 6 continued

$$BETA_{it+1} or\ STDRET_{it+1} = \beta_0 + \beta_1 RF\_SYS_{it} + \beta_2 RF\_IDIO_{it} + \beta_3 MDA\_SYS_{it} + \beta_4 MDA\_IDIO_{it} + \beta_5 ALL\_SYS_{it} + \beta_6 ALL\_IDIO_{it} + \beta_7 SIZE_{it} + \beta_8 BTM_{it} + \beta_9 RET_{it} + \beta_{10} LEV_{it} + \beta_{11} STDRET_{it} + \beta_{12} BETA_{it} + \beta_{13} SKEW_{it} + \beta\beta_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} + \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \varepsilon_{it}$$

	Prediction	Dependent variable: $BETA_{it+1}$			Prediction	Dependent variable: $STDERET_{it+1}$		
		1	2	3		1	2	3
Adjusted R-Square		0.450	0.449	0.450		0.561	0.561	0.561

*** Significant at the 1 % level, ** Significant at the 5 % level, * Significant at the 10 % levelThis is a regression of post-disclosure market-based measures of risk on risk factor disclosures estimated using Eq. (2). Coefficient standard errors are shown in parentheses below the coefficient loading. Standard errors are Huber-White heteroskedastic robust and are clustered by firm. Fixed effects are included (FF48 and Year). $\boldsymbol{BETA_{it+1}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period beginning two trading days after the 10-K release. $\boldsymbol{STDERET_{it+1}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period beginning two trading days after the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{RF\_SYS_{it}}$ is the log of the count of key words in the risk factor section referring to "other-systematic" risk exposure (Item 1A). $\boldsymbol{RF\_IDIO_{it}}$ is the log of the count of key words in the risk factor section referring to "other-idiosyncratic" risk exposure (Item 1A). $\boldsymbol{MDA\_SYS_{it}}$ is the log of the count of key words in the management discussion and analysis referring to "other-systematic" risk exposure (Item 7). $\boldsymbol{MDA\_IDIO_{it}}$ is the log of the count of key words in the management discussion and analysis referring to "other-idiosyncratic" risk exposure (Item 7). $\boldsymbol{ALL\_SYS_{it}}$ is the log of the count of key words in the 10-K referring to "other-systematic" risk exposure. $\boldsymbol{ALL\_IDIO_{it}}$ is the log of the count of key words in the 10-K referring to "other-idiosyncratic" risk exposure. $\boldsymbol{SIZE_{it}}$ is the log of the market value of equity (PRCC_F*CSHO). $\boldsymbol{BTM_{it}}$ is the book value of equity (SEQ) divided by the market value of equity. $\boldsymbol{RET_{it}}$ is the 12-month stock return beginning in the fourth month of the fiscal year and ending in the third month after fiscal year-end. A continuous stream of 12 monthly returns is required. $\boldsymbol{LEV_{it}}$ is the firm's book value of debt (DLTT + DD1) divided by total assets (AT). Missing values of debt due in one year (DD1) are replaced with zero if missing. $\boldsymbol{STDERET_{it}}$ is the standard deviation of daily abnormal stock returns for the 250 trading day period ending two trading days before the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. $\boldsymbol{BETA_{it}}$ is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{SKEW_{it}}$ is the skewness of daily stock returns for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{TURN_{it}}$ is the average daily share turnover (expressed as a percentage) for the 250 trading day period ending two trading days before the 10-K release. $\boldsymbol{BIGN_{it}}$ is a dummy variable taking on a value of one for firms with a Big N auditor. $\boldsymbol{ETR_{it}}$ is total tax expense (TXT) divided by pre-tax income. Pre-tax income is calculated as net income before extraordinary items (NI) plus total tax expense. $\boldsymbol{DNI_{it}}$ is net income before extraordinary items (NI) divided by the lagged market value of equity. $\boldsymbol{NUM\_EST_{it}}$ is the natural log of the number of analysis, plus one, following the firm as reported by I/B/E/S. $\boldsymbol{INST\_OWN_{it}}$ is sum of shares owned by institutional investors as reported by Thompson Financial divided by total shares outstanding (CSHO). $\boldsymbol{RF\_FOG_{it}}$ is the fog index for the risk factor section. $\boldsymbol{MDA\_FOG_{it}}$ is the fog index for the management discussion and analysis section. $\boldsymbol{ALL\_FOG_{it}}$ is the fog index for the 10-K filing

abnormal stock return volatility when both systematic and idiosyncratic risk disclosure measures are included in the regression model. As expected, the coefficient on idiosyncratic risk disclosure is positive and significant ($\beta_1 = 0.0005$, std. error = 0.00025, t-statistic = 2.00), while the coefficient on systematic risk disclosure is not statistically significant. This suggests that, when managers disclose more idiosyncratic risk factors, investors respond by incorporating this risk into the idiosyncratic market-based measure STDERET. Taken together, the evidence from Tables 5 and 6 suggest that investors incorporate the information conveyed by risk factor disclosures into post-disclosure market-based measures of firm risk.

3.7 Risk factor disclosures and post-disclosure information asymmetry

In our third hypothesis, we predict that—conditional on their impact on investors' risk perceptions as tested above in H2—risk factor disclosures will be negatively associated with information asymmetry. To examine H3, we follow the prior literature and use bid-ask spread as a proxy for information asymmetry (i.e., Khan and Watts 2009, Lang et al. 2012) and add additional control variables designed to capture a firm's information environment (Stoll 2000). In order to ensure we capture the incremental association between risk disclosures and information asymmetry, we also include the control variables for pre-disclosure risk and performance from Eq. (1), as well as post-disclosure risk. Specifically, we estimate the following regression:

$$\begin{aligned}\text{SPREAD}_{i,t+1} = {} & \beta_0 + \beta_1 \text{RF\_DISC}_{i,t} + \beta_2 \text{MDA\_DISC}_{i,t} + \beta_3 \text{ALL\_DISC}_{i,t} \\ & + \beta_4 \text{SPREAD}_{i,t} + \beta_5 \text{SIZE}_{i,t} + \beta_6 \text{BTM}_{i,t} + \beta_7 \text{RET}_{i,t} + \beta_8 \text{LEV}_{i,t} \\ & + \beta_9 \text{STDERET}_{i,t} + \beta_{10} \text{STDERET}_{i,t+1} + \beta_{11} \text{BETA}_{i,t} + \beta_{12} \text{BETA}_{i,t+1} \\ & + \beta_{13} \text{SKEW}_{i,t} + \beta_{14} \text{TURN}_{i,t} + \beta_{15} \text{BIGN}_{i,t} + \beta_{16} \text{ETR}_{i,t} + \beta_{17} \text{DNI}_{i,t} \\ & + \beta_{18} \text{NUMEST}_{i,t} + \beta_{19} \text{INSTOWN}_{i,t} + \beta_{20} \text{RF\_FOG}_{i,t} \\ & + \beta_{21} \text{MDA\_FOG}_{i,t} + \beta_{22} \text{ALL\_FOG}_{i,t} + \beta_{23} \text{INV\_PRC}_{i,t} \\ & + \text{year effects} + \text{industry effects} + \varepsilon_{i,t} \end{aligned} \quad (3)$$

where all variables are as defined in Eq. (2), SPREAD = the average daily difference between a firm's bid and ask price measured over 12-months, and INV_PRC = the inverse of fiscal year ending stock price. We include post-disclosure market-based measures of risk to control for the effect of disclosures on firm risk, which can in turn affect bid-ask spread. If risk factor disclosures are negatively associated with post-disclosure market-based measures of information asymmetry (H3), we expect that $\beta_1 < 0$. As before, we control for the disclosures that are made in MD&A at time t, use industry and year fixed effects, and use standard errors that control for both serial correlation and heteroscedasticity.

3.8 Results on risk factor disclosures and post-disclosure information asymmetry

Table 7 presents evidence regarding H3. The first column of Table 7 uses the key word count of the risk factor section as the independent variable of interest. We find

a negative and significant association between RF_DISC and SPREAD ($\beta_1 = -0.0440$, std. error = 0.0069, *t*-statistic = 6.38). This suggests that, if a firm moves from the 25th percentile to the 75th percentile of risk disclosure, the following period's SPREAD is lower by 2.7 %. The results are unchanged in the second column when we substitute total word count as the independent variable rather than key word count. Overall, the results are consistent with H3 and suggest that, after controlling for the fact that risk factor disclosure *increases* the market's assessment of a firm's risk, risk factor disclosure also *decreases* information asymmetry among that same firm's shareholders.

3.9 Investor reaction to risk factor disclosures

In our final hypotheses, we predict that risk factor disclosures will be negatively associated with abnormal returns at the 10-K filing date. We follow prior literature that examines information content around the 10-K release (Bernard and Stober 1989; Han et al. 1992; Wahlen 1994, among others) and estimate the following regression equation:

$$\begin{aligned} \mathrm{CAR}_{i,t} = {} & \beta_0 + \beta_1 \mathrm{RF\_DISC}_{i,t} + \beta_2 \mathrm{MDA\_DISC}_{i,t} + \beta_3 \mathrm{ALL\_DISC}_{i,t} + \beta_4 \Delta \mathrm{NI}_{i,t} \\ & + \beta_5 \mathrm{AVG\_ACC}_{i,t} + \beta_6 \mathrm{LOSSES}_{i,t} + \beta_7 \Delta \mathrm{EST}_{i,t} + \beta_8 \mathrm{BTM}_{i,t} + \beta_9 \mathrm{SIZE}_{i,t} \\ & + \beta_{10} \mathrm{NUMEST}_{i,t} + \beta_{11} \mathrm{INSTOWN}_{i,t} + \beta_{12} \mathrm{RF\_FOG}_{i,t} + \beta_{13} \mathrm{MDA\_FOG}_{i,t} \\ & + \beta_{14} \mathrm{ALL\_FOG}_{i,t} + \text{year effects} + \text{industry effects} + \varepsilon_{i,t} \end{aligned} \quad (4)$$

where RF_DISC and MDA_DISC are defined as in Eq. (3), CAR = abnormal return calculated over the three days around the Form 10-K filing date and adjusted for the effects of market returns, size, and growth (Fama and French 1993). We obtain the Form 10-K filing date with the text analysis described in Appendix 1. As before, we control for disclosures made in MD&A and use industry and year fixed effects, as well as standard errors that control for serial dependence and heteroscedasticity. If risk factor disclosures are negatively associated with abnormal returns around the Form 10-K release date (H4), we expect that $\beta_1 < 0$.

Short-window returns at the 10-K release date could be a function of changes in either (1) firms' expected future cash flows or (2) investors' assessment of firm risk (Kothari 2001). Thus, we include controls for the information in the 10-K that is likely to lead to changes in firms' expected future cash flows. First, we control for firms' earnings surprise (ΔNI), defined as change in net income before extraordinary items divided by lagged market value of equity. Earnings surprises are likely to be correlated with changes in expected future cash flows. Second, we control for the change in analysts' consensus earnings estimate for the next year (ΔEST), where the change is measured as the difference between the consensus estimate prior to and after the Form 10-K release date. If our short-window abnormal returns results are attributable to changes in markets' perceptions of firm risk (and not cash flows), we should continue to find a negative association between risk factor disclosures and abnormal returns after controlling for ΔNI and ΔEST. Finally, we include controls for the effect of firms' information environment on returns around the 10-K release date. Specifically, we include LOSSES, a dummy variable that takes on a value of

Table 7 Multivariate analysis of risk factor disclosure informativeness–effect on information asymmetry

$$SPREAD_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SPREAD_{it} + \beta_5 SIZE_{it} + \beta_6 BTM_{it} + \beta_7 RET_{it} + \beta_8 LEV_{it} + \beta_9 STDERET_{it} + \beta_{10} STDERET_{it+1} + \beta_{11} BETA_{it} + \beta_{12} BETA_{it+1} + \beta_{13} SKEW_{it} + \beta_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \beta_{23} INV\_PRC_{it} + \varepsilon_{it}$$

		Prediction	Independent variable: RF_DISC	
			KEW WORDS	*ALL WORDS*
Intercept	ß$_0$	?	0.0100	0.2890***
			(0.0776)	(0.1119)
RF_DISC$_{it}$	**ß$_1$**	**–**	**−0.0440****	**−0.0472****
			(0.0069)	**(0.0067)**
MDA_DISC$_{it}$	ß$_2$	–	−0.0058**	−0.0044***
			(0.0023)	(0.0017)
ALL_DISC$_{it}$	ß$_3$	?	0.0011	−0.0117
			(0.0087)	(0.0101)
SPREAD$_{it}$	ß$_4$	+	0.6782***	0.6764***
			(0.0124)	(0.0125)
SIZE$_{it}$	ß$_5$	–	−0.0595***	−0.0593***
			(0.0049)	(0.0050)
BTM$_{it}$	ß$_6$	?	−0.0023	−0.0014
			(0.0125)	(0.0125)
RET$_{it}$	ß$_7$	–	−0.1084***	−0.1075***
			(0.0125)	(0.0125)
LEV$_{it}$	ß$_8$	+	−0.0679***	−0.0622***
			(0.0237)	(0.0239)

Table 7 continued

$$SPREAD_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SPREAD_{it} + \beta_5 SIZE_{it} + \beta_6 BTM_{it} + \beta_7 RET_{it} + \beta_8 LEV_{it} + \beta_9 STDERET_{it} + \beta_{10} STDERET_{it+1} + \beta_{11} BETA_{it} + \beta_{12} BETA_{it+1} + \beta_{13} SKEW_{it} + \beta_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \beta_{23} INV\_PRC_{it} + \varepsilon_{it}$$

		Prediction	Independent variable: RF_DISC	
			KEW WORDS	*ALL WORDS*
$STDERET_{it}$	ß$_9$	?	−9.6710***	−9.5824***
			(0.6985)	(0.6985)
$STDERET_{it+1}$	ß$_{10}$	?	21.3781***	21.3908***
			(0.4536)	(0.4534)
$BETA_{it}$	ß$_{11}$	?	−0.0353***	−0.0368***
			(0.0117)	(0.0117)
$BETA_{it+1}$	ß$_{12}$	?	−0.2282***	−0.2284***
			(0.0116)	(0.0116)
$SKEW_{it}$	ß$_{13}$	–	−0.0337***	−0.0340***
			(0.0039)	(0.0039)
$TURN_{it}$	ß$_{14}$	–	−0.0062***	−0.0061***
			(0.0008)	(0.0008)
$BIGN_{it}$	ß$_{15}$	–	−0.0616***	−0.0614***
			(0.0116)	(0.0116)
ETR_{it}	ß$_{16}$	–	−0.0429**	−0.0451**
			(0.0213)	(0.0213)
DNI_{it}	ß$_{17}$	–	0.2003	0.1788
			(0.1712)	(0.1702)

Table 7 continued

$$SPREAD_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SPREAD_{it} + \beta_5 SIZE_{it} + \beta_6 BTM_{it} + \beta_7 RET_{it} + \beta_8 LEV_{it} + \beta_9 STDERET_{it} + \beta_{10} STDERET_{it+1} + \beta_{11} BETA_{it} + \beta_{12} BETA_{it+1} + \beta_{13} SKEW_{it} + \beta_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \beta_{23} INV\_PRC_{it} + \varepsilon_{it}$$

		Prediction	Independent variable: RF_DISC	
			KEW WORDS	*ALL WORDS*
NUM_EST_{it}	ß$_{18}$	–	−0.0391***	−0.0392***
			(0.0057)	(0.0057)
$INST\_OWN_{it}$	ß$_{19}$	–	−0.1251***	−0.1282***
			(0.0281)	(0.0281)
RF_FOG_{it}	ß$_{20}$	?	0.0000	0.0000
			(0.0001)	(0.0001)
MDA_FOG_{it}	ß$_{21}$	?	0.0001	0.0001
			(0.0001)	(0.0001)
ALL_FOG_{it}	ß$_{22}$	?	−0.0002**	−0.0002*
			(0.0001)	(0.0001)
INV_PRC_{it}	ß$_{23}$	+	0.5722***	0.5801***
			(0.0759)	(0.0760)
YEAR FIXED EFFECTS			INCLUDED	INCLUDED
INDUSTRY FIXED EFFECTS			INCLUDED	INCLUDED
Number of observations			9,076	9,076

Table 7 continued

$$SPREAD_{it+1} = \beta_0 + \beta_1 RF\_DISC_{it} + \beta_2 MDA\_DISC_{it} + \beta_3 ALL\_DISC_{it} + \beta_4 SPREAD_{it} + \beta_5 SIZE_{it} + \beta_6 BTM_{it} + \beta_7 RET_{it} + \beta_8 LEV_{it} + \beta_9 STDERET_{it} + \beta_{10} STDERET_{it+1} + \beta_{11} BETA_{it} + \beta_{12} BETA_{it+1} + \beta_{13} SKEW_{it} + \beta_{14} TURN_{it} + \beta_{15} BIGN_{it} + \beta_{16} ETR_t + \beta_{17} DNI_{it} \beta_{18} NUM\_EST_{it} + \beta_{19} INST\_OWN_{it} + \beta_{20} RF\_FOG_{it} + \beta_{21} MDA\_FOG_{it} + \beta_{22} ALL\_FOG_{it} + \beta_{23} INV\_PRC_{it} + \varepsilon_{it}$$

	Prediction	Independent variable: RF_DISC	
		KEW WORDS	*ALL WORDS*
Adjusted R-square		0.911	0.911

*** Significant at the 1 % level, ** Significant at the 5 % level, * Significant at the 10 % level

This is a regression of post-disclosure information asymmetry on risk factor disclosures estimated using Eq. (3). Coefficient standard errors are shown in parentheses below the coefficient loading. Standard errors are Huber-White heteroskedastic robust and are clustered by firm. Fixed effects are included (FF48 and Year). ***SPREAD*$_{it+1}$** is the firm's average ending bid-ask spread (divided by ending stock price) for the 250 trading days beginning two trading days after the 10-K release date. ***RF_DISC*$_{it}$** is the log of the word count or the log of the key word count in the risk factor disclosure (Item 1A). ***MDA_DISC*$_{it}$** is the log of the word count or the log of the key word count in the management discussion and analysis (Item 7). ***ALL_DISC*$_{it}$** is the log of the total word count, or the log of the key word count in the 10-K filing. ***SPREAD*$_{it}$** is the firm's average ending bid-ask spread (divided by ending stock price) for the 250 trading days ending two trading days before the 10-K release date. ***SIZE*$_{it}$** is the log of the market value of equity (PRCC_F*CSHO). ***BTM*$_{it}$** is the book value of equity (SEQ) divided by the market value of equity. ***RET*$_{it}$** is the 12-month stock return beginning in the fourth month of the fiscal year and ending in the third month after fiscal year-end. A continuous stream of 12 monthly returns is required. ***LEV*$_{it}$** is the firm's book value of debt (DLTT + DD1) divided by total assets (AT). Missing values of debt due in one year (DD1) are replaced with zero if missing. ***STDERET*$_{it}$** is the standard deviation of daily abnormal stock returns for the 250 trading day period ending two trading days before the 10-K release. Abnormal stock returns are calculated using the error term from the market model, with a firm specific coefficient on market returns. ***BETA*$_{it}$** is the firm's coefficient loading on the market excess return (Fama & French) for the 250 trading day period ending two trading days before the 10-K release. ***SKEW*$_{it}$** is the skewness of daily stock returns for the 250 trading day period ending two trading days before the 10-K release. ***TURN*$_{it}$** is the average daily share turnover (expressed as a percentage) for the 250 trading day period ending two trading days before the 10-K release. ***BIGN*$_{it}$** is a dummy variable taking on a value of one for firms with a Big N auditor. ***ETR*$_{it}$** is total tax expense (TXT) divided by pre-tax income. Pre-tax income is calculated as net income before extraordinary items (NI) plus total tax expense. ***DNI*$_{it}$** is net income before extraordinary items (NI) divided by the lagged market value of equity. ***NUM_EST*$_{it}$** is the log of the number of equity analysts per I/B/E/S as of the last month before fiscal year-end. ***INV_PRC*$_{it}$** is the inverse of fiscal year ending stock price (PRCC_F). ***NUM_EST*$_{it}$** is the natural log of the number of analysis, plus one, following the firm as reported by I/B/E/S. ***INST_OWN*$_{it}$** is sum of shares owned by institutional investors as reported by Thompson Financial divided by total shares outstanding (CSHO). ***RF_FOG*$_{it}$** is the fog index for the risk factor section. ***MDA_FOG*$_{it}$** is the fog index for the management discussion and analysis section. ***ALL_FOG*$_{it}$** is the fog index for the 10-K filing

one if there is one or more loss years (NI < 0) over the previous 5 years, and AVG_ACC, defined as the average absolute value of accruals calculated using the statement of cash flows deflated by average total assets.

3.10 Results on investor reaction to risk factor disclosures

Table 8 presents evidence regarding H4. The first column shows a negative and significant association between the key word count from risk factor disclosures (RF_DISC) and CAR ($\beta_1 = -0.0073$, std. error = 0.0025, *t*-statistic = 2.92). This suggests that, if a firm moves from the 25th percentile to the 75th percentile of risk disclosure, the abnormal return around the 10-K release date is lower by 68 basis points. The results are similar in Column 2 when we use total word count. Overall, the results are consistent with H4 and suggest that investors incorporate the information conveyed by risk factor disclosures into firm stock price.

4 Sensitivity analysis and robustness

4.1 Controls for risk disclosures in Item 7A of Form 10-K

As previously mentioned, risk factor disclosure in Item 1A of Form 10-K conveys information about anything that makes the company "speculative or risky." Thus, the tone of the disclosure is, by definition, negative, and so this disclosure provides a powerful setting in which to examine whether disclosure tone affects the cost of capital and firm value. However, another section of the Form 10-K conveys qualitative and quantitative disclosure about risk—Item 7A "Quantitative and Qualitative Disclosures about Market Risk" (hereafter referred to as "market risk" disclosures).

In our primary regression analysis, we do not control for market risk disclosures since, just as with the risk factor disclosures in Item 1A, they provide *qualitative* disclosure relating to items that make the company risky. Thus, including a control for these disclosures could "over-control" for the information we are trying to capture with the Item 1A Risk Factor disclosure. However, these market risk disclosures are different from risk factor disclosures in that, in addition to providing qualitative risk disclosures about a narrow set of risks, they also *quantify* the impact that these certain risks have on firms' financial statements. Since our measures are designed only to capture qualitative disclosures, this additional quantitative information contained by the disclosures in Item 7A should not affect our results. Nevertheless, we control for the information in this section in order to provide further assurance that our results strictly hold for qualitative risk factor disclosure rather than a mixture of qualitative *and* quantitative disclosures.

In untabulated results, we re-test all of our hypotheses but include controls for the Item 7A market risk disclosures. Specifically, when our variable of interest is the total word count for the risk factor section, we control for the total word count in the market risk section, and when our variable of interest is key word count in the risk factor section, we control for the total number of key words in the market risk

Table 8 Multivariate analysis of risk factor disclosure informativeness–short window returns

$$CAR_{it} = ß_0 + ß_1 RF\_DISC_{it-1} + ß_2 MDA\_DISC_{it} + ß_3 ALL\_DISC_{it} + ß_4 \Delta NI_{it} + ß_5 AVG\_ACC_{it} + ß_6 LOSSES_{it} + ß_7 \Delta EST_{it} + ß_8 BTM_{it} + ß_9 SIZE_{it} + ß_{10} NUM\_EST_{it} + ß_{11} INST\_OWN_{it} + ß_{12} RF\_FOG_{it} + ß_{13} MDA\_FOG_{it} + ß_{14} ALL\_FOG_{it} + \varepsilon_{it}$$

		Prediction	Independent variable: RF_DISC	
			KEW WORDS	*ALL WORDS*
Intercept	$ß_0$	?	−0.0340	−0.0333
			(0.0261)	(0.0388)
***RF_DISC*$_{it}$**	**$ß_1$**	**–**	**−0.0073*****	**−0.0059****
			(0.0025)	**(0.0025)**
MDA_DISC_{it}	$ß_2$	?	−0.0012	−0.0007
			(0.0009)	(0.0006)
ALL_DISC_{it}	$ß_3$	?	(0.0074)**	(0.0039)
			(0.0031)	(0.0036)
ΔNI_{it}	$ß_4$	?	−0.0440	−0.0435
			(0.0280)	(0.0280)
AVG_ACC_{it}	$ß_5$	?	−0.0102	−0.0106
			(0.0799)	(0.0800)
$LOSSES_{it}$	$ß_6$	?	−0.0047	−0.0052
			(0.0035)	(0.0035)
ΔEST_{it}	$ß_7$	+	0.0607*	0.0592
			(0.0368)	(0.0369)
BTM_{it}	$ß_8$	?	0.0513***	0.0507***
			(0.0055)	(0.0055)
$SIZE_{it}$	$ß_9$	?	0.0059***	0.0056***
			(0.0013)	(0.0013)
$NUMEST_{it}$	$ß_{10}$	?	0.0080***	0.0079***
			(0.0021)	(0.0021)
$INST\_OWN_{it}$	$ß_{11}$	?	0.0013	0.0006
			(0.0105)	(0.0105)
RF_FOG_{it}	$ß_{12}$	?	0.0000	0.0000
			(0.0000)	(0.0000)
MDA_FOG_{it}	$ß_{13}$	?	0.0001*	0.0001*
			(0.0000)	(0.0000)
ALL_FOG_{it}	$ß_{14}$	?	0.0000	0.0000
			(0.0000)	(0.0000)
YEAR FIXED EFFECTS			INCLUDED	INCLUDED
INDUSTRY FIXED EFFECTS			INCLUDED	INCLUDED
Number of observations			8,193	8,193

Table 8 continued

$$CAR_{it} = ß_0 + ß_1 RF\_DISC_{it-1} + ß_2 MDA\_DISC_{it} + ß_3 ALL\_DISC_{it} + ß_4 \Delta NI_{it} + ß_5 AVG\_ACC_{it} + ß_6 LOSSES_{it} + ß_7 \Delta EST_{it} + ß_8 BTM_{it} + ß_9 SIZE_{it} + ß_{10} NUM\_EST_{it} + ß_{11} INST\_OWN_{it} + ß_{12} RF\_FOG_{it} + ß_{13} MDA\_FOG_{it} + ß_{14} ALL\_FOG_{it} + \varepsilon_{it}$$

	Prediction	Independent variable: RF_DISC	
		KEW WORDS	*ALL WORDS*
Adjusted R-square		0.179	0.178

*** Significant at the 1 % level, ** Significant at the 5 % level, * Significant at the 10 % level

This is a regression of short-window abnormal returns on risk factor disclosures estimated using Eq. (4). Coefficient standard errors are shown in parentheses below the coefficient loading. Standard errors are Huber-White heteroskedastic robust and are clustered by firm. Fixed effects are included (FF48 and Year). $\boldsymbol{CAR_{it}}$ is the three-day buy-and-hold return starting one trading day before the 10-K release and ending one trading day after less the expected return. The expected return is calculated using the firm's loading on the market return as well as two hedge portfolio returns (HML and SMB from the Fama & French database). The firm's loadings are measured over the previous 250 trading days. Each factor loading is then multiplied by the portfolio return to the factor over the three-day buy-and-hold return window. $\boldsymbol{RF\_DISC_{it}}$ is the log of the key word count in the risk factor disclosure (Item 1A). $\boldsymbol{MDA\_DISC_{it}}$ is the log of the key word count in the management discussion and analysis (Item 7). $\boldsymbol{ALL\_DISC_{it}}$ is the log of the total word count, or the log of the key word count in the 10-K filing. $\boldsymbol{\Delta NI_{it}}$ is the change in net income before extraordinary items (NI) divided by the lagged market value of equity. Set to zero when missing, $\boldsymbol{\Delta EST_{it}}$ is the change in one-year-ahead mean EPS estimate based on the I/B/E/S summary file, deflated by stock price. The change is calculated as the difference between the estimate for the first full month after the release of the 10-K less that in the first full month before release. Set to zero when missing, $\boldsymbol{LOSSES_{it}}$ is a dummy variable that takes on a value of one if the firm has one or more loss years ($NI < 0$) over the previous five years. $\boldsymbol{AVG\_ACC_{it}}$ is the average absolute value of accruals calculated using the cash flows statement (NI-OANCF), deflated by average total assets (AT). $\boldsymbol{SIZE_{it}}$ is the log of the market value of equity (PRCC_F*CSHO). $\boldsymbol{BTM_{it}}$ is the book value of equity (SEQ) divided by the market value of equity. $\boldsymbol{NUM\_EST_{it}}$ is the natural log of the number of analysis, plus one, following the firm as reported by I/B/E/S. $\boldsymbol{INST\_OWN_{it}}$ is sum of shares owned by institutional investors as reported by Thompson Financial divided by total shares outstanding (CSHO). $\boldsymbol{RF\_FOG_{it}}$ is the fog index for the risk factor section. $\boldsymbol{MDA\_FOG_{it}}$ is the fog index for the management discussion and analysis section. $\boldsymbol{ALL\_FOG_{it}}$ is the fog index for the 10-K filing

section. All of the results from our hypothesis testing are quantitatively and qualitatively unchanged by these additional controls for market risk disclosures.

4.2 Expectations model for risk factor disclosures

In our market tests of H2 through H4, we control for the market's pre-disclosure expectations of disclosure by controlling for pre-disclosure measures of firm risk (i.e., all of the variables from the risk factor determinants model used to test H1). Therefore, the coefficient loading on our variable of interest is purged of the effect of the market's expectations of risk factor disclosures and instead captures the effect of the abnormal amount of risk factor disclosure. To provide further assurance, we re-examine our hypotheses with a two-stage regression model where we formally calculate abnormal risk factor disclosure. Specifically, we estimate our determinants model in Eq. (1) after also including lagged risk factor disclosure as a determinant. We take the residual from this regression and label it "unexpected risk factor

disclosure." Finally, we use this unexpected risk factor disclosure variable to re-test all of our hypotheses related to H2 through H4. All of the results from our hypothesis testing are unaffected, providing further assurance that our results are not sensitive to explicitly controlling for the market's expectations regarding risk factor disclosure.

4.3 Alternate regression specifications

In our primary tests, we examine the effect of disclosure tone on the future level of market beta, stock return volatility, and information asymmetry. To ensure this is capturing a response to risk disclosures, we include controls for the lagged level of market beta, stock return volatility, and information asymmetry. This allows the coefficient on the lagged level to vary, depending on the persistence of the variable in question. In contrast, a pure changes model tests a similar construct but instead constrains the coefficient on the lagged value of the dependent variable to equal one.

As a final robustness test, we re-estimate our hypotheses by substituting the change in the dependent variable for the level of the dependent variable. For brevity, and to be comprehensive with all of our robustness checks in this section, we also include controls for market risk disclosures (Sect. 4.1) as well as the two-stage expectations model for the risk factor disclosures (Sect. 4.2), although it should be noted that the results from Tables 5 through 8 are unchanged if we simply change the dependent variable from a level to a change. In untabulated results, all of our results for H2 through H4 continue to hold under a pure changes model, providing further assurance that our results are not due to correlated omitted variables.

5 Conclusion

This study examines the information content of risk factor disclosures for market-based risk, information asymmetry, and stock returns and offers four results. First, we find that risk factor disclosures are positively associated with *pre-disclosure* measures that proxy for firm risk, and this association varies predictably across subcategories of firm risk (i.e., systematic, idiosyncratic, financial, tax, and litigation risks). This suggests that—contrary to critics' assertions—risk factor disclosures are *not* boilerplate. Instead, managers appear to provide informative disclosures. Second, we find that risk factor disclosures are positively associated with *post-disclosure* market-based measures of firm risk. This suggests that market participants incorporate the information conveyed by risk factor disclosures into their assessments of firm risk. Third, after controlling for the investors' assessment of fundamental risk, we find that risk factor disclosures are negatively associated with post-disclosure information asymmetry. This suggests that, while risk factor disclosures *increase* the market's assessment of a firm's fundamental risk, the public availability of the disclosures simultaneously *decreases* information differences across that firm's shareholders (after controlling for the effect of fundamental risk on information asymmetry). Finally, we find that risk factor disclosures are negatively associated with abnormal returns at the Form 10-K release date. Overall,

our results suggest that managers provide useful risk factor disclosures and investors incorporate this information into market values.

Future research might examine whether firms that provide incremental risk disclosures, and thus increase the market's assessment of their overall risk, are more likely to experience negative future outcomes such as decreased investment or increased probability of bankruptcy and credit default. Additionally, our study focuses strictly on the effect of risk factor disclosures on the equity market's assessment of firm risk. Future studies could examine whether risk factor disclosures are informative to debt markets either through investors' pricing of debt or through future changes in credit ratings.

Acknowledgments We are grateful for the helpful comments and suggestions of an anonymous reviewer, Kris Allee, Terry Baker, Liz Chuk, James Cotter, Jon Duchac, Fabio Gaertner, Ronen Gal-Or, Andrea Kelton, Feng Li, Russell Lundholm, Bill Marcum, Dale Martin, Rick Mergenthaler, Karen Nelson, Deon Strickland, Jake Thornock, the doctoral students at the University of Arizona, workshop participants at Wake Forest University, and conference participants at the 2011 AAA Annual Meeting and the 22nd Annual Conference on Financial Economics and Accounting (FEA) at Indiana University. Finally, we thank both the Securities and Exchange Commission (SEC) and the Institute of Chartered Accountants in England and Wales (ICAEW) for their interest in this paper. Portions of this paper were included in the ICAEW's October 2011 report titled "Reporting Business Risks: Meeting Expectations." Professor Lu gratefully acknowledges assistance from the National Science Council of Taiwan (grant number NSC100-2410-H-002-025-MY3).

Appendix 1: Summary of textual analysis procedures to generate risk factor disclosure measures

Data collection technique

In this appendix, we discuss the textual analysis procedures employed in the collection of risk disclosures from annual filings in the SEC's Electronic Data Gathering and Retrieval (EDGAR) database (for a thorough treatment on who uses the EDGAR database to obtain information, and the most commonly retrieved forms, see Drake et al. 2011). Figure 1 illustrates the specific sequence of steps used in our textual analysis. As presented in Fig. 1, annual filings are downloaded and processed to generate appropriate counting measures that can objectively quantify firms' risk disclosure. The rest of this section presents our system design in detail.

File collection

Our system starts by collecting all relevant 10-K filings and storing them in a relational database. Since subsequent analysis and processing may require several runs of prototyping and testing, we create our own repository in order to increase the efficiency of subsequent activities. EDGAR provides convenient filing download mechanisms based on the File Transfer Protocol (FTP). The downloader first retrieves all index files beginning in 2005. All 10-K filings are downloaded according to the index files. The procedure ensures that all annual filings that were uploaded to the EDGAR database between January 1, 2005 and December 31, 2009

are collected. In total, 44,998 10-K filings were downloaded and stored in our repository.

File preprocessing

The goal of file preprocessing is to extract important items, including risk factors, from individual filings for subsequent text analysis. The EDGAR system requires firms to upload their Form 10-K reports in either (1) plain text format or (2) HTML format. Our software is unable to reliably scan plain text filings, so we delete these filings from our sample. These filings account for 21 % of the population.[16] On the other hand, our software is able to reliably scan HTML filings, and we use an automatic procedure to extract text from the following subsections of the Form 10-K: risk factors (Form 10-K, Item 1A), MD&A (Item 7), and market risk disclosures (Item 7A).

Our item extraction procedure is based on the assumption that HTML filings contain visual clues for human readers to recognize item boundaries easily. These visual clues include the use of subsection titles, boldface fonts, extra spacing, and so on. Most filings also use standard item headings that start with "Item," followed by an item number and a description (e.g., "ITEM 1A. RISK FACTORS"). The HTML format, however, allows visually the same display to be achieved using different tags. One example is that < b > and < strong > have the same visual effect in most browsers. The flexibility in composing HTML filings creates a technical challenge when designing the item extraction procedure. Our design overcomes this challenge by first converting an HTML file into an intermediate representation that combines HTML tags that have similar effects for human readers when deciding item boundaries. A list of candidate item heading locations than can be identified using the intermediate representation. Finally, items are extracted based on the locations of candidate heading locations. We describe the three steps in details below.

In the first step, the input HTML filing is parsed into a tree structure, where the leaf nodes are text segments of the content and the internal nodes are HTML tags such as "b," "title," or "li." The tree structure allows us to associate the characteristics of a text segment by traveling upward and inspecting the parent nodes.

To facilitate subsequent extraction, we further convert the parse tree into a flat structure by traversing the tree and calculating two scores for each text segment: (1) an emphasizing score and (2) a segmentation score. Both of these scores are important to ensure that we have fully extracted all of the text within a Form 10-K "Item" or subsection. The emphasizing score is designed to determine the prominence of the text within the disclosure section. To compute the emphasizing score, we examine the set of HTML tags surrounding the text. Examples include the "strong" tag and the "underline" style within a "div" tag. A complete list of

[16] To ensure that the removal of plain text filings does not bias our sample, in Sect. 3, we compare our sample to the overall universe of Compustat firms. We find that our sample is generalizable across industries and years. In addition, we include industry and year fixed effects in all of our multivariate analyses.

emphasizing tags can be found in Appendix 2. Emphasizing scores are computed for each text segment according to the number of emphasizing tags in its parent nodes. A positive emphasizing score indicates that the text segment is visually more prominent and may be the heading of an item.

The segmentation score is designed to indicate whether and how the text is visually broken down into sections. HTML tags such as "
" and "<p>" are two common tags used to separate text. By tracking the number of "separation tags" in the parent nodes when traversing the parse tree, we can detect the locations where text segments are visually separated. The segmentation score is the number of increased or decreased "separation tags."

The second step is to construct candidate item heading locations using the flat structure. We assume that the item headings are visually more prominent (emphasizing score > 0) with a text segment that "looks like" item headings (using regular expression "(Item\s +\d\D?\.?)," case ignored). The output of this step is a list of locations that may be the beginning of an item.

The last step is to extract items of interest using the list of candidate item heading locations. To extract the risk factor subsection (Item 1A), we scan the text around the candidate item heading locations for the keyword "risk factors." The text from a matching item heading location until the beginning of the following candidate headings locations are assigned to the risk factor subsection. Other items of interests are processed in the same manner.

Extracted items go through consistency checks before recording back to the database. The assumption is that Item 1A (Risk Factors) should precede Item 7 and Item 7A. Item heading locations that do not match the ordering are rejected. We conduct performance analysis to ensure the quality of the item extraction procedure. Two aspects are of particular interest. The first aspect is the proportion of HTML filings that can be extracted. The proportion is referred to as the coverage of selected items. The following table summarizes the coverage of our procedure:

Fiscal year	10 K Filings	HTML filing	Extracted Item 1A	Extracted Item 1A and Item 7	Extracted Item 1A, Item 7, and Item 7A
2005	8,996	6,326	4,255	3,593	3,558
2006	8,941	6,842	5,833	4,833	4,764
2007	7,318	7,005	5,936	5,021	4,863
2008	9,236	8,624	7,081	5819	5,480
Totals	34,491	28,797	23,105	19,266	18,665
Coverage			80 %	67 %	65 %

Among all the filings collected, 34,491 filings are from fiscal years 2005 to 2008. Eighty-three percent (28,797) of them are HTML filings. Our procedure is able to extract 80 % of the risk factor subsections from the HTML filings. If we restrict to the subset that have successfully extracted both Item 1A and Item 7, then the number drops to 67 %. Our procedure can extract Item 1A, Item 7, and Item 7A from 65 % of HTML filings

The second performance aspect of importance is the precision of the extracted items. We visually inspect 300 random filings from the subset that Item 1A was extracted using the procedure. We were unable to extract nine of these filings.[17] Among the remaining 291, all of them correctly contain only the appropriate subsection. Thus, the precision of Item 1A extraction is 100 %. Similarly, our software was unable to extract Item 7 from 47 firms in this subset. Among the remaining 253, 249 of them were extracted correctly. Thus, the precision for Item 7 is 98 % (249/253). Finally, our software extracted Item 7A from 281 of the 300 firms in this subset and five extracted items contain other subsections, for a precision of 98 % (281/286)

Overall, the results show that our software extracted the risk factors from 80 % of HTML filings. Moreover, for over 98 % of the extracted items, the right subsection, and only the right subsection was extracted. Among the small number of incorrectly extracted subsections, most of them contain the text from the target item, with small chunks of text from other subsections

Text quantification using risk keywords

The extracted items need to be quantified in order to be included in our empirical models. We first count the total words within a subsection and label it as word count (ALL_WORDS). Then, we identify key words using the predefined dictionary from Appendix 3. We developed our list of key words in three steps. First, we begin with key risk words used by prior literature (Nelson and Pritchard 2007). Second, we add additional key words to the list that, based on our review of risk factor disclosures, appear to be common across firms. Third, we classify the list of key words as relating to financial, litigation, tax, other-systematic, or other-idiosyncratic risk subcategories as described in Sect. 3.1 of the text. To enhance the coverage of our key words, we reviewed important terms identified by a document clustering approach, known as the Latent Dirichlet Allocation (Blei et al. 2003). By inspecting important terms in the same cluster, keywords that were previously missing were included.

Finally, the text quantification module computes the frequency of each term and aggregates the number of key words according to the risk subcategory (KEY_WORDS). To increase the precision of the key word matching process, the text quantification module allows term matching to be case sensitive. This constraint is especially useful for acronyms such as IRS (i.e., Internal Revenue Service) and EU (i.e., European Union). If there are variations of terms that need to be matched, these variations are explicitly specified in the keyword list. For example, the criterion "(lease|leases|leasing)" is used to match the variation of "lease." By performing these techniques, we compute three measures for each extracted subsection of the Form 10-K: (1) total word count (ALLWORDS), (2)

[17] We consider whether our final sample is biased as a result of this sample size reduction by comparing our sample to the Compustat universe of firms in Sect. 3 and in Table 1.

total key word count (KEYWORDS), and (3) key word count (KEYWORDS) by risk subcategory.

Appendix 2

HTML Tags Used to Identify the Text from Applicable Subsections of Form 10-K.

This appendix provides the specific HTML tags that were used to determine the "emphasizing score" and the "separation score." These tags are used to help us fully extract on those subsections of the Form 10-K that are used in our analysis. For example, the emphasizing tags help to identify when text is bolded or underlined, and the separation tags help to identify when text is segmented by paragraph breaks or section headings. This procedure is fully explained in Appendix 1, Textual Analysis Data Collection Process.

Emphasizing tags		Separation tags	
Tag	Attribute	Tag	Attribute
b		br	
strong		dd	
h1		dl	
h2		dt	
h3		form	
h4		hr	
h5		li	
u		p	
font	style = (bold|underline)	pre	
div	style = (bold|underline)	table	
p	style = (bold|underline)	tbody	
		td	
		th	
		title	
		tr	
		ul	
		div	.

Appendix 3

See Table 9.

Table 9 Key words list by risk category

Risk category	Keyword	Risk category	Keyword
Financial	Anti-takeover (provisions|provision)	Financial	Reserves
Financial	Bank debt	Financial	Revolver
Financial	Capital (expenditure|expenditures)	Financial	Sale of productive assets
Financial	Capital (lease|leases)	Financial	Stock market listing
Financial	Chapter 11	Financial	Stock price drop
Financial	Chapter 7	Financial	Stock price volatility
Financial	Chapter 9	Financial	Underfunded pensions
Financial	Collateral	Financial	Underwriting
Financial	Concentrated ownership	Financial	Volatility of operating results
Financial	(Covenant|covenants)	Financial	Volatility of revenues
Financial	Credit (facility|facilities)	Financial	Volatility of sales
Financial	Credit rating	Financial	Working capital
Financial	Credit risk	Other-Idiosyncratic	Acquisition
Financial	Debt burden	Other-Idiosyncratic	Adequate staffing
Financial	Decline in stock price	Other-Idiosyncratic	Advertising
Financial	Default	Other-Idiosyncratic	Asset (impairment|impairments)
Financial	Defined Benefit	Other-Idiosyncratic	Asset (securitization|securitizations)
Financial	Dilution	Other-Idiosyncratic	Assimilation
Financial	Dividends	Other-Idiosyncratic	Backlog
Financial	Downgrade	Other-Idiosyncratic	Brand
Financial	Family	Other-Idiosyncratic	Brand recognition
Financial	Financial condition	Other-Idiosyncratic	California power crisis
Financial	Financing costs	Other-Idiosyncratic	Certification
Financial	Funded status	Other-Idiosyncratic	Clinical (trial|trials)
Financial	Illiquid market	Other-Idiosyncratic	Commercialize
Financial	Improvements	Other-Idiosyncratic	Concentration
Financial	Indebtedness	Other-Idiosyncratic	Consolidation
Financial	Insider sales	Other-Idiosyncratic	Construction

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Financial	Investment in equipment	Other-Idiosyncratic	(Contract|contracts)
Financial	Investment in plant	Other-Idiosyncratic	(Copyright|copyrights)
Financial	(lease|leases|leasing)	Other-Idiosyncratic	Corporate culture
Financial	Lease (commitment|commitments)	Other-Idiosyncratic	Cost control
Financial	Leverage	Other-Idiosyncratic	Customer concentration
Financial	Leveraged (lease|leases)	Other-Idiosyncratic	Customer service
Financial	Limited trading	Other-Idiosyncratic	Delivery
Financial	Liquidity	Other-Idiosyncratic	Distribution
Financial	Loan	Other-Idiosyncratic	(Distributor|distributors)
Financial	Locked-in (lease|leases)	Other-Idiosyncratic	Downsizing
Financial	Mandatory contribution	Other-Idiosyncratic	Economies of scale
Financial	Maturity	Other-Idiosyncratic	Embargo
Financial	Negative operating cash flow	Other-Idiosyncratic	Enron
Financial	New financing	Other-Idiosyncratic	Expand
Financial	Obligations	Other-Idiosyncratic	Expanding
Financial	OPEB	Other-Idiosyncratic	Expansion
Financial	O.P.E.B.	Other-Idiosyncratic	(Export|exports)
Financial	Operating losses	Other-Idiosyncratic	Facilities
Financial	Penny stock	Other-Idiosyncratic	Franchise
Financial	Postretirement	Other-Idiosyncratic	Franchisee
Financial	Rating	Other-Idiosyncratic	Goodwill
Financial	Refinance	Other-Idiosyncratic	Goodwill (impairment|impairments)
Financial	Refinancing	Other-Idiosyncratic	Impairment

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Financial	Reinsurance	Other-Idiosyncratic	Information technology
Financial	Renegotiation	Other-Idiosyncratic	Innovation
Financial	Reorganization	Other-Idiosyncratic	Insurance coverage
Other-Idiosyncratic	Intangible	Other-Idiosyncratic	(Secret|secrets)
Other-Idiosyncratic	(Integrate|integrating|integration)	Other-Idiosyncratic	Security
Other-Idiosyncratic	Intellectual	Other-Idiosyncratic	Shortages
Other-Idiosyncratic	Internal (control|controls)	Other-Idiosyncratic	Single customer
Other-Idiosyncratic	Internet	Other-Idiosyncratic	Single supplier
Other-Idiosyncratic	Investment in (subsidiary|subsidiaries)	Other-Idiosyncratic	Software
Other-Idiosyncratic	IT	Other-Idiosyncratic	Sole (supplier|suppliers)
Other-Idiosyncratic	I.T.	Other-Idiosyncratic	SPE
Other-Idiosyncratic	Joint (venture|ventures)	Other-Idiosyncratic	S.P.E.
Other-Idiosyncratic	Keep and retain top management	Other-Idiosyncratic	Special purpose entity
Other-Idiosyncratic	Key personnel	Other-Idiosyncratic	Strike
Other-Idiosyncratic	Labor (cost|costs)	Other-Idiosyncratic	(Supplier|suppliers)
Other-Idiosyncratic	Labor relations	Other-Idiosyncratic	Supply chain
Other-Idiosyncratic	Labor (union|unions)	Other-Idiosyncratic	(Synergy|synergies)
Other-Idiosyncratic	(license|licenses)	Other-Idiosyncratic	Systems
Other-Idiosyncratic	Limited operating history	Other-Idiosyncratic	(Tariff|tariffs)
Other-Idiosyncratic	Maintenance	Other-Idiosyncratic	Technological obsolescence
Other-Idiosyncratic	Management retention	Other-Idiosyncratic	Technologies
Other-Idiosyncratic	Market acceptance	Other-Idiosyncratic	Technology
Other-Idiosyncratic	Marketing	Other-Idiosyncratic	Trade

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Other-Idiosyncratic	Material (weakness|weaknesses)	Other-Idiosyncratic	(Trademark|trademarks)
Other-Idiosyncratic	MBS	Other-Idiosyncratic	Training
Other-Idiosyncratic	M.B.S.	Other-Idiosyncratic	Union election
Other-Idiosyncratic	Merger	Other-Idiosyncratic	Variable interest entity
Other-Idiosyncratic	Mortgage backed securities	Other-Idiosyncratic	(Vendor|vendors)
Other-Idiosyncratic	Mortgage servicing rights	Other-Idiosyncratic	VIE
Other-Idiosyncratic	MSR	Other-Idiosyncratic	V.I.E.
Other-Idiosyncratic	M.S.R.	Other-Idiosyncratic	Weather
Other-Idiosyncratic	Natural disasters	Other-Idiosyncratic	Web security
Other-Idiosyncratic	New Construction	Other-Idiosyncratic	(website|websites)
Other-Idiosyncratic	New product acceptance	Legal and Regulatory	Adverse judgment
Other-Idiosyncratic	New product development	Legal and Regulatory	Anti-trust
Other-Idiosyncratic	No current operations	Legal and Regulatory	Casualty
Other-Idiosyncratic	Online	Legal and Regulatory	Charged
Other-Idiosyncratic	Orders	Legal and Regulatory	Class action
Other-Idiosyncratic	Patent	Legal and Regulatory	Compliance
Other-Idiosyncratic	Personnel	Legal and Regulatory	Comply
Other-Idiosyncratic	Preclinical	Legal and Regulatory	(Conflict|conflicts) of interest
Other-Idiosyncratic	Product	Legal and Regulatory	Contamination
Other-Idiosyncratic	Product development	Legal and Regulatory	Defendant
Other-Idiosyncratic	Product mix	Legal and Regulatory	Deregulation
Other-Idiosyncratic	Product performance	Legal and Regulatory	Effects of implementing new (standard|standards)
Other-Idiosyncratic	Production	Legal and Regulatory	Effects of implementing new (method|methods)

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Other-Idiosyncratic	Proprietary	Legal and Regulatory	Enforceability of judgments
Other-Idiosyncratic	Publicity	Legal and Regulatory	Enforcement
Other-Idiosyncratic	Redundancy	Legal and Regulatory	Environmental
Other-Idiosyncratic	Reliance on key (customer|customers)	Legal and Regulatory	FDA approval
Other-Idiosyncratic	Reliance on key (supplier|suppliers)	Legal and Regulatory	Federal
Other-Idiosyncratic	Reporting controls	Legal and Regulatory	Fines
Other-Idiosyncratic	Research and development	Legal and Regulatory	Fraud
Other-Idiosyncratic	Restructuring	Legal and Regulatory	Government investigation
Other-Idiosyncratic	Restructuring implementation	Legal and Regulatory	Government policy
Other-Idiosyncratic	Sarbanes–Oxley	Legal and Regulatory	Governmental approval
Other-Idiosyncratic	SARS	Legal and Regulatory	Hazardous
Legal and Regulatory	IFRS	Other-Systematic	Economic uncertainties
Legal and Regulatory	I.F.R.S.	Other-Systematic	Economy
Legal and Regulatory	Infringe	Other-Systematic	Electricity
Legal and Regulatory	Injury	Other-Systematic	Energy
Legal and Regulatory	Inquiries	Other-Systematic	EU
Legal and Regulatory	Inquiry	Other-Systematic	E.U.
Legal and Regulatory	Intellectual property	Other-Systematic	Euro
Legal and Regulatory	(Investigation|investigations)	Other-Systematic	European Union
Legal and Regulatory	Legislation	Other-Systematic	Exchange (rate|rates)
Legal and Regulatory	Litigation	Other-Systematic	Financial crisis
Legal and Regulatory	Pay damages	Other-Systematic	Fiscal policy
Legal and Regulatory	(penalty|penalties)	Other-Systematic	Foreign currency

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Legal and Regulatory	Pending (lawsuit|lawsuits)	Other-Systematic	Foreign exchange
Legal and Regulatory	Plaintiff	Other-Systematic	(Forward|forwards)
Legal and Regulatory	Possibility of (restatement|restatements)	Other-Systematic	Fuel
Legal and Regulatory	Potential (lawsuit|lawsuits)	Other-Systematic	Future
Legal and Regulatory	Product liability	Other-Systematic	Gas
Legal and Regulatory	(Regulation|regulations)	Other-Systematic	Gasoline
Legal and Regulatory	Regulatory	Other-Systematic	GDP
Legal and Regulatory	Regulatory approval	Other-Systematic	G.D.P.
Legal and Regulatory	Regulatory change	Other-Systematic	GNP
Legal and Regulatory	Regulatory compliance	Other-Systematic	G.N.P.
Legal and Regulatory	Regulatory environment	Other-Systematic	General business risks
Legal and Regulatory	Related (party|parties)	Other-Systematic	General conditions
Legal and Regulatory	Remediation	Other-Systematic	General economic conditions
Legal and Regulatory	(Restatement|restatements)	Other-Systematic	Gold
Legal and Regulatory	Safety	Other-Systematic	Growth (rate|rates)
Legal and Regulatory	Superfund	Other-Systematic	Hedge
Legal and Regulatory	Uncertainties regarding accounting estimates	Other-Systematic	Hedging
Other-Systematic	Afghanistan	Other-Systematic	Housing
Other-Systematic	Aggregate demand	Other-Systematic	Housing Starts
Other-Systematic	Asian crisis	Other-Systematic	Industry (condition|conditions)
Other-Systematic	Business conditions	Other-Systematic	Industry environment
Other-Systematic	Call	Other-Systematic	Inflation
Other-Systematic	Capacity	Other-Systematic	Iraq

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Other-Systematic	Coal	Other-Systematic	(Market|markets)
Other-Systematic	(Commodity|commodities)	Other-Systematic	Market demand
Other-Systematic	Competition	Other-Systematic	Market supply
Other-Systematic	(Competitor|competitors)	Other-Systematic	Marketplace
Other-Systematic	Complement	Other-Systematic	Materials
Other-Systematic	Concentration	Other-Systematic	(Metal|Metals)
Other-Systematic	Consumer confidence	Other-Systematic	Middle East
Other-Systematic	Consumer spending	Other-Systematic	(mineral|minerals)
Other-Systematic	Consumption	Other-Systematic	Mining
Other-Systematic	Currency collapse	Other-Systematic	Monetary policy
Other-Systematic	Currency (fluctuation|fluctuations)	Other-Systematic	Mortgage
Other-Systematic	Cyclical	Other-Systematic	Natural gas
Other-Systematic	Demand	Other-Systematic	Obsolescence
Other-Systematic	(Derivative|Derivatives)	Other-Systematic	Oil
Other-Systematic	Discounting	Other-Systematic	Operating environment
Other-Systematic	(Economic|Economics)	Other-Systematic	Option
Other-Systematic	Economic (condition|conditions)	Other-Systematic	Ore
Other-Systematic	Economic (downturn|downturns)	Other-Systematic	Overstocked
Other-Systematic	Economic growth	Other-Systematic	Peso
Other-Systematic	Petroleum	Tax	Uncertain tax (position|positions)
Other-Systematic	Political climate	Tax	VAT
Other-Systematic	Political instability	Tax	VAT
Other-Systematic	Pound	Tax	Value added tax

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Other-Systematic	Price pressure		
Other-Systematic	Prices		
Other-Systematic	Pricing power		
Other-Systematic	Raw (material|materials)		
Other-Systematic	Real		
Other-Systematic	Real estate investment trust		
Other-Systematic	Recession		
Other-Systematic	REIT		
Other-Systematic	R.E.I.T.		
Other-Systematic	Renmenbi		
Other-Systematic	RMB		
Other-Systematic	Ruble		
Other-Systematic	Rupee		
Other-Systematic	Saving		
Other-Systematic	Seasonal		
Other-Systematic	September (11|11th)		
Other-Systematic	Short		
Other-Systematic	Silver		
Other-Systematic	Steel		
Other-Systematic	(Substitute|Substitutes)		
Other-Systematic	Swap		
Other-Systematic	Terrorism		
Other-Systematic	U.S. dollar		

Table 9 continued

Risk category	Keyword	Risk category	Keyword
Other-Systematic	Underlying		
Other-Systematic	Unsalable inventory		
Other-Systematic	War		
Other-Systematic	Yen		
Other-Systematic	Yuan		
Tax	Aggressive tax (position|positions)		
Tax	Back taxes		
Tax	Deferred tax (asset|assets)		
Tax	Deferred tax (liability|liabilities)		
Tax	Excise (tax|taxes)		
Tax	FIN 48		
Tax	Internal Revenue Service		
Tax	IRS		
Tax	I.R.S.		
Tax	IRS audit		
Tax	IRS judgment		
Tax	Loss (carryback|carrybacks)		
Tax	Loss (carryforward|carryforwards)		
Tax	Property (tax|taxes)		
Tax	Provision for income (tax|taxes)		
Tax	State (tax|taxes)		
Tax	(Tax|Taxes)		
Tax	Tax audit		
Tax	Tax (authority|authorities)		
Tax	Tax (liability|liabilities)		
Tax	Tax (penalty|penalties)		
Tax	Taxable		

References

Amihud, Y., & Mendelson, H. (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics, 17*, 223–249.

Ball, R., & Brown, R. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research, 6*, 159–178.

Ball, R., & Kothari, S. P. (1989). Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics, 25*, 51–74.

Barry, C., & Brown, S. (1985). Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis, 20*, 407–422.
Beaver, W. (1968). The information content of annual earnings announcements. *Journal of Accounting Research, 6*, 67–92.
Bernard, V., & Stober, T. (1989). The nature and amount of information in cash flows and accruals. *The Accounting Review, 64*, 624–652.
Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.
Botosan, C. (1997). Disclosure level and the cost of capital. *The Accounting Review, 72*, 323–350.
Botosan, C., & Plumlee, M. (2002). A re-examination of disclosure level and the expected cost of equity capital. *Journal of Accounting Research, 40*, 21–41.
Brown, S., & Tucker, J. (2011). Large sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research, 49*, 309–346.
Bryan, S. (1997). Incremental information content of required disclosures contained in management discussion and analysis. *The Accounting Review, 72*, 285–301.
CFO (2010). SEC pushes companies for more risk information. CFO Magazine, August 2, 2010.
Chan, K. (1988). On the contrarian investment strategy. *Journal of Business, 61*, 147–163.
Clarkson, P., Kao, J., & Richardson, G. (1999). Evidence that management discussion and analysis (MD&A) is a part of a firm's overall disclosure package. *Contemporary Accounting Research, 16*, 111–134.
Corporate Counsel (2006). MD&A Risk Factors (Nelson Rocks Preserve-Style. By Broc Romanek and Dave Lynn. www.thecorporatecounsel.net, November 30, 2006.
Demsetz, H. (1986). Corporate control, insider trading, and rates of return. *American Economic Review, 76*, 313–316.
Diamond, D., & Verrecchia, R. (1991). Disclosure, liquidity, and the cost of capital. *Journal of Finance, 66*, 1325–1355.
Drake, M., Roulstone, D., Thornock, J. (2011). The demand for mandatory disclosure: evidence from investors' use of the EDGAR database. Working paper, Ohio State University.
Easley, D., & O'Hara, M. (2004). Information and the cost of capital. *Journal of Finance, 59*, 1553–1583.
Fama, E., & French, K. (1993). Common risk factors in the returns on stock and bonds. *Journal of Financial Economics, 33*, 3–56.
Fields, T., Lys, T., & Vincent, L. (2001). Empirical research on accounting choice. *Journal of Accounting and Economics, 31*, 255–307.
Guedhami, O., & Pittman, J. (2008). The importance of IRS monitoring to debt pricing in private firms. *Journal of Financial Economics, 90*, 38–58.
Hamada, R. S. (1972). The effect of the firm's capital structure on the systematic risk of common stocks. *Journal of Finance, 27*, 435–452.
Han, B., Jennings, R., & Noel, J. (1992). Communication of nonearnings information at the financial statement release date. *Journal of Accounting and Economics, 15*, 63–86.
Healy, P., & Palepu, K. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics, 31*, 405–440.
Jayaraman, S. (2008). Earnings volatility, cash flow volatility and informed trading. *The Accounting Review, 46*, 809–851.
Jiang, G., Lee, C., & Zhang, Y. (2005). Information uncertainty and expected returns. *Review of Accounting Studies, 10*, 185–221.
Jones, C., & Weingram, S. (1996). *The determinants of 10b-5 litigation risk*. Stanford Law School: Working paper.
Ke, B., Huddart, S., & Petroni, K. (2003). What insiders know about future earnings and how they use it: Evidence from insider trades. *Journal of Accounting and Economics, 35*, 315–346.
Khan, M., & Watts, R. (2009). Estimation and empirical properties of a firm-year measure of accounting conservatism. *Journal of Accounting and Economics, 48*, 132–150.
Kim, O., & Verrecchia, R. (1994). Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics, 17*, 41–68.
Klein, R., & Bawa, V. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics, 3*, 215–231.
Kothari, S. P. (2001). Capital markets research in accounting. *Journal of Accounting and Economics, 31*, 105–231.

Kothari, S. P., Li, X., & Short, J. (2009a). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: a study using content analysis. *The Accounting Review, 84*, 1639–1670.
Kothari, S. P., Shu, S., & Wysocki, P. (2009b). Do managers withhold bad news? *Journal of Accounting Research, 47*, 241–276.
Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*. doi:10.1007/s11142-013-9228-9.
Kutner, M., Nachsheim, C., & Neter, J. (2004). *Applied linear statistical models* (5th ed.). New York, NY: McGraw-Hill Irwin.
Kyle, A. (1985). Continuous auctions and insider trading. *Econometrica, 6*, 1315–1335.
Lambert, R., Leuz, C., & Verrecchia, R. (2007). Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research, 45*, 385–420.
Lambert, R., Leuz, C., & Verrecchia, R. (2012). Information asymmetry, information precision, and the cost of capital. *Review of Finance, 16*, 1–29.
Lang, M., Lins, K., & Maffett, M. (2012). Transparency, liquidity, and valuation: International evidence on when transparency matters most. *Journal of Accounting Research, 50*, 729–774.
Lang, M., & Lundholm, R. (1996). Corporate disclosure policy and analyst behavior. *The Accounting Review, 71*, 467–492.
Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review, 86*, 1087–1115.
Lennox, C. (1999). Audit quality and auditor size: An evaluation of reputation and deep pockets hypotheses. *Journal of Business, Finance, and Accounting, 26*, 779–805.
Leuz, C., & Verrecchia, R. (2000). The economic consequences of increased disclosure. *Journal of Accounting Research, 38*, 91–124.
Li, F. (2008a). Do stock market investors understand the risk sentiment of corporate annual reports? Working paper, University of Michigan.
Li, F. (2008b). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics, 45*, 221–247.
Li, F. (2010). The information content of forward-looking statements in corporate filings: a naïve Baysian machine learning approach. *Journal of Accounting Research, 48*, 1049–1102.
Linsmeier, T., Thornton, D., Venkatachalam, M., & Welker, M. (2002). The effect of mandated market risk disclosures on trading volume sensitivity to interest rate, exchange rate, and commodity price movements. *The Accounting Review, 77*, 343–377.
Mansi, S., Maxwell, W., & Miller, D. (2004). Does auditor quality and tenure matter to investors? Evidence from the bond market. *Journal of Accounting Research, 42*, 755–793.
Modigliani, F., & Miller, M. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review, 48*, 261–297.
Nelson, K., & Pritchard, A. C. (2007). Litigation risk and voluntary disclosure: The use of meaningful cautionary language. Working paper, Rice University.
Reuters (2005). "Refco risks boiler-plate disclosure." By Scott Malone. Friday, October 21, 2005. http://w4.stern.nyu.edu/news/news.cfm?doc_id=5094.
Roulstone, D. (1999). Effect of SEC financial reporting release No. 48 on derivative and market risk disclosures. *Accounting Horizons, 13*, 343–363.
Schrand, C. (1997). The association between stock-price interest rate sensitivity and disclosures about derivative instruments. *The Accounting Review, 72*, 87–109.
Scott, T. (1994). Incentives and disincentives for financial disclosure: Voluntary disclosure of defined benefit pension plan information by Canadian Firms. *The Accounting Review, 69*, 26–43.
SEC (2005). Securities and exchange commission final rule, release no. 33–8591 (FR-75). http://www.sec.gov/rules/final/33-8591.pdf.
SEC (2009). Remington Arms Company, Inc. Form 10-K Filing for the year ended December 31, 2009, SEC EDGAR database.
SEC (2010). Form 10-K instructions. www.sec.gov/about/forms/form10-k.pdf.
Skinner, D. (1994). Why firms voluntarily disclose bad news. *Journal of Accounting Research, 32*, 38–60.
Spindler, J. (2006). Is it time to wind up the Securities Act of 1933? *Regulation*, Winter 2006–2007.
Stoll, H. (2000). Friction. *Journal of Finance, 55*, 1479–1514.
Tetlock, P. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance, 62*, 1139–1168.

Tufano, P. (1998). Who manages risk? An empirical examination of risk management practices in the gold mining industry. *Journal of Finance, 51*, 1097–1137.
Vuolteenaho, T. (2002). What drives firm-level stock returns? *Journal of Finance, 57*, 233–264.
Wahlen, J. (1994). The nature of information in commercial bank loan loss disclosures. *The Accounting Review, 69*, 455–478.
Watts, R. (1977). Corporate financial statements: Product of market and political processes. *Australian Journal of Management, 2*, 53–75.
Watts, R., & Zimmerman, J. (1986). *Positive accounting theory*. Englewood Cliffs, NJ: Prentice Hall.
Wong, M. H. F. (2001). The association between SFAS No. 119 derivatives disclosures and the foreign exchange risk exposure of manufacturing firms. *Journal of Accounting Research, 38*, 387–417.
You, H., & Zhang, X. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting Studies, 14*, 559–586.