

UNCOMMON INFORMATION IN FIRM DISCLOSURES

By

Matthew David DeAngelis

A DISSERTATION

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

Business Administration – Doctor of Philosophy

2014

## **ABSTRACT**

### **UNCOMMON INFORMATION IN FIRM DISCLOSURES**

By

Matthew David DeAngelis

This study investigates uncommon information in firm disclosures. Using a method from computational linguistics called Latent Semantic Analysis (LSA), I measure the difference in topics discussed in the Management's Discussion and Analysis (MD&A) section of a firm's 10-K filing relative to its peers. I hypothesize that the presence of uncommon information signals to investors that the firm's valuation function differs from its peers'. Consistent with this hypothesis, I find that uncommon information is associated with higher firm-specific stock returns. I also hypothesize that investors will process uncommon information slowly because uncommon information is textually and conceptually complex and investors have little data with which to estimate its value. Consistent with this hypothesis, I find that the market response to uncommon information is delayed. Drawing on the information theory and linguistics literature, I further hypothesize and find that text containing uncommon information is longer and less readable, suggesting that readability is at least partially a function of the information presented in the text. This study makes several contributions to the literature. First, I identify a mechanism through which investors identify firm-specific components of value. Second, I link the information content of a disclosure to higher information processing costs. In doing so, I identify a cost-benefit tradeoff faced by managers when crafting their disclosures and by investors when they analyze those disclosures. Third, I introduce Latent Semantic Analysis to the accounting literature as a method for comparing firm disclosures.

Copyright by  
MATTHEW DAVID DEANGELIS  
2014

## **ACKNOWLEDGEMENTS**

I am grateful to my dissertation committee, Marilyn Johnson, Brian Pentland, Alan Munn, and especially Kathy Petroni, my chair, for their unfailing support and guidance. I also gratefully acknowledge the helpful comments of Andrew Acito, Ramji Balakrishnan, Colleen Boland, Tony Bucaro, Ted Christensen, Elizabeth Connors, Richard Crowley, Paul Demere, Nick Dopuch, Susanna Gallani, Chris Hogan, Andy Imdieke, Raffi Indjejikian, John Jiang, Rong Jin, Ranjani Krishnan, Dan Lynch, Xiumin Martin, Miles Romney, Joe Schroeder, Karen Sedatole, Mike Shields, Amy Swaney, Inna Voytsekhivska, Isabel Wang, Philip Wang, Dan Wangerin, participants at my brownbag seminar and dissertation proposal at Michigan State, doctoral students at the University of Illinois and workshop participants at Washington University in St. Louis, Georgia State University, the University of Illinois and the University of Michigan.

## TABLE OF CONTENTS

<b>LIST OF TABLES . . . . .</b>	<b>vi</b>
<b>CHAPTER 1 INTRODUCTION . . . . .</b>	<b>1</b>
<b>CHAPTER 2 LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT . . .</b>	<b>5</b>
2.1 Uncommon Information and Firm Valuation . . . . .	5
2.1.1 Hypothesis 1 . . . . .	7
2.1.2 Hypothesis 2 . . . . .	8
2.2 Difficult-to-Estimate Information . . . . .	9
2.3 Difficult-to-Process Information . . . . .	11
2.3.1 Hypothesis 3 . . . . .	13
2.3.2 Hypothesis 4 . . . . .	13
<b>CHAPTER 3 MEASURING UNCOMMON INFORMATION . . . . .</b>	<b>15</b>
3.1 Measuring the Characteristics of Language in the MD&A . . . . .	16
3.2 Calculating <i>Uncommon</i> . . . . .	19
3.3 Examples of Uncommon Information . . . . .	22
<b>CHAPTER 4 TESTS OF HYPOTHESES AND EMPIRICAL RESULTS . . . . .</b>	<b>24</b>
4.1 Uncommon Information and Firm-Specific Returns . . . . .	24
4.2 Uncommon Information and Information Processing Costs . . . . .	33
4.2.1 Uncommon Information and Delayed Market Response . . . . .	33
4.2.2 Uncommon Information and Readability . . . . .	39
4.3 Summary of Results . . . . .	44
<b>CHAPTER 5 FUTURE WORK AND CONCLUSION . . . . .</b>	<b>46</b>
<b>APPENDICES . . . . .</b>	<b>49</b>
A Variables . . . . .	50
B Excerpts from Most and Least Common MD&As . . . . .	55
<b>BIBLIOGRAPHY . . . . .</b>	<b>59</b>

## LIST OF TABLES

Table 3.1	Sample Selection Comparison . . . . .	20
Table 3.2	Top 10 Words for the Top 5 Dimensions in the Software Industry, Fiscal Year 2007	22
Table 3.3	Top 10 Words for the Most and Least Common MD&As in the Software Industry, Fiscal Year 2007 . . . . .	23
Table 4.1	Descriptive Statistics for Determinants Analysis . . . . .	26
Table 4.2	Correlations between Determinants . . . . .	27
Table 4.3	Regression of <i>Uncommon</i> on Determinants . . . . .	29
Table 4.4	Descriptive Statistics for <i>ReturnSync</i> Regressions . . . . .	31
Table 4.5	Correlations between <i>ReturnSync</i> Variables . . . . .	32
Table 4.6	Regression of $\Delta ReturnSync$ on <i>Uncommon</i> . . . . .	34
Table 4.7	Descriptive Statistics for CAR Regressions . . . . .	36
Table 4.8	Correlations between CAR Variables . . . . .	37
Table 4.9	Regression of <i>CAR</i> and <i>PCAR</i> on <i>Uncommon</i> . . . . .	40
Table 4.10	Descriptive Statistics for Readability Regressions . . . . .	42
Table 4.11	Correlations between Readability Variables . . . . .	43
Table 4.12	Regression of Readability Variables on <i>Uncommon</i> . . . . .	45
Table A.1	Variable List . . . . .	50

# **CHAPTER 1**

## **INTRODUCTION**

In this study, I examine uncommon information in firm disclosures. Uncommon information is present in a firm's disclosure if the disclosure discusses or emphasizes different valuation parameters, or different properties of shared parameters, than are discussed in the disclosures of the firm's peers. For instance, a growth firm may devote substantial space to a discussion of sales and little to a discussion of earnings, whereas the firm's peers predominantly discuss earnings. Or, a firm operating in a geographic market with few competitors may provide information about the relationship between sales and earnings in that market. I predict that the presence of uncommon information in a disclosure signals to investors that the firm's valuation function differs from its peers'. As such, uncommon information is useful to investors both in choosing the parameters of their valuation models and in populating those models.

Although uncommon information is useful, prior research on investors' categorical and contextual thinking from accounting and finance (Lee et al., 2006; Hopkins, 1996; Peng & Xiong, 2006) and psychology and linguistics (Zipf, 1949; Becker, 1980; Pierce, 1980; Goodman et al., 1981; Halgren et al., 2002) suggests that uncommon information may be difficult for investors to process. First, investors may have difficulty judging the value relevance and precision of uncommon information due to a lack of data from comparable firms. Second, the collection and processing of uncommon information may be impaired by increased cognitive and textual complexity of this information. Since the market processes information less completely when information processing costs are higher (Bloomfield, 2002; Li, 2008; Lee, 2012), I predict that the market response to uncommon information is delayed. As such, while uncommon information is useful for identifying firm-specific components of value, the market incorporates that information slowly compared with common information.

I use textual analysis to identify uncommon information in the Management's Discussion and

Analysis (MD&A) section of the 10-K filing. I analyze the MD&A for several reasons. First, prior research has found that the MD&A contains decision-relevant information for investors (Barron et al., 1999; Clarkson et al., 1999; Brown & Tucker, 2011; Davis & Tama-Sweet, 2012). Second, identifying uncommon information requires a significant amount of cross-sectional data. Since the MD&A is a mandatory disclosure, it enables me to observe the information disclosed by a large number of firms. Third, the MD&A is specifically designed to enable “investors to see through the eyes of management” and to “provide the context within which financial information should be analyzed” (Securities and Exchange Commission, 2003). The MD&A is therefore a rich setting for comparing the information content of disclosures in a cross-section of firms.

In comparing MD&As across firms, it is useful to distinguish the words that managers use in their discussion from the information conveyed by those words.<sup>1</sup> If a firm manager spends 90% of the MD&A discussing “earnings”, while the manager of another firm talks about “profits”, a fair measure of information content would consider these MD&As to be very similar. In order to “see through” these word choices, I use a method from computational linguistics, Latent Semantic Analysis (LSA), to identify the topics that managers are discussing when they use particular words. Related to factor analysis, LSA examines co-occurrences between words to determine whether words are likely to be related to one another. I use LSA to group related words into topics and measure the relative differences in topics discussed between a firm’s MD&A and the MD&As of its peer firms to identify MD&As that contain uncommon information.

I measure uncommon information across a sample of 35,585 MD&As across 7,984 firms from 1994 to 2012 and examine the relationship between uncommon information and stock returns. Specifically, I investigate whether firms with MD&As that contain uncommon information have higher firm-specific stock returns. I hypothesize that uncommon information is related to firm-specific returns in two ways. First, the firm may possess unusual characteristics or experience unusual events that result in both higher firm-specific returns and a higher level of disclosed uncommon information. Second, the disclosure of uncommon information provides incremental informa-

---

<sup>1</sup>Bloomfield (2002) distinguishes between “data”, or marks on a page, and “statistics”, the information gleaned from processing data.



tion to investors about firm value. To test the first hypothesis, I regress uncommon information on measures of firm characteristics and returns and find that firm-specific returns and characteristics in the year prior to the disclosure are significant determinants of uncommon information. I also find that firms are more likely to provide uncommon information when the differences between the firm and its peers are not reflected in earnings, consistent with managers disclosing uncommon information as a supplement to earnings information. To test whether uncommon information in the MD&A is incrementally informative, I regress the change in return synchronicity, or the change in the portion of a firm's returns explained by market and industry returns (Durnev et al., 2003; Piotroski & Roulstone, 2004), on uncommon information. Consistent with uncommon information in the MD&A providing incremental information, I find that uncommon information is associated with an increase in firm-specific returns in the year following the disclosure.

I further hypothesize that the high information processing costs of uncommon information lead to an incomplete market response. In order to investigate the timing of the market response to uncommon information, I regress the absolute value of market returns prior to, around, and following the disclosure on uncommon information. Consistent with a delayed market response to uncommon information in the MD&A, I find a significant positive association between uncommon information and market returns in the two months following the disclosure, but no association between uncommon information and stock returns in the three-day window around the disclosure. Inconsistent with the market responding to information disclosed prior to the MD&A filing, I find no significant relationship between uncommon information and stock returns prior to the disclosure. These results suggest that the market responds to uncommon information disclosed in the MD&A, but that this information is difficult to process. Also consistent with high processing costs, I find that text containing uncommon information is less readable than text containing common information. This finding supports my market tests and provides insight into sources of textual complexity in firm disclosures.

This study contributes to the literature in several ways. First, while prior research has shown that the inclusion of firm-specific information improves the informativeness of stock prices, lit-

tle is known about how investors learn about and incorporate this information into their decision models. My study suggests that uncommon information helps investors identify variables in their decision models that they may not have otherwise considered. Second, prior studies on the information processing costs of financial reporting focus on the structural elements of disclosure, such as readability (Li, 2008; Lee, 2012) and report presentation (Lee et al., 2006; Hodder et al., 2008; Bloomfield et al., 2010), and largely attribute these costs to managerial opportunism. In contrast, this study is the first to my knowledge to directly link the information *content* of a disclosure to the associated processing costs. As such, my study suggests that investors face a cost-benefit tradeoff in choosing to utilize uncommon information. These tradeoffs may also have implications for managers' reporting strategies. Third, this study contributes to our understanding of how information processing constraints impact investors and market efficiency. Whereas prior research has shown that limited attention negatively impacts market efficiency (Hirshleifer et al., 2009; Dellavigna & Pollet, 2009; Aboody et al., 2010), my results suggest that investors may attend to information and still process it slowly. My results may also help to shed light on why investors revisit previously disclosed information in making their trading decisions (Tetlock, 2011; Drake et al., 2012).

The remainder of this study proceeds as follows. Section 2 reviews literature relevant to my research questions and develops my hypotheses. Section 3 discusses the calculation of my measure of uncommon information. In Section 4, I discuss the results of my study and in Section 5, I discuss future work and conclude.

## CHAPTER 2

### LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

In the next three Sections, I review literature relevant to examining the role of uncommon information in financial reporting and develop my hypotheses.

#### 2.1 Uncommon Information and Firm Valuation

Ohlson (1995) articulates a model of firm valuation that incorporates accounting information. In this model, a firm's value,  $P_t$ , is expressed as a function of book value,  $y_t$ , and expected future residual income,  $E[\tilde{x}_{t+\tau}]$ . Specifically:

$$P_t = y_t + \sum_{\tau=1}^{\infty} R_f^{-\tau} E[\tilde{x}_{t+\tau}] \quad (2.1)$$

where  $R_f$  represents the risk-free rate. Since book value is (at least in theory) readily available, an investor's valuation task centers on the second term in this linear equation, the investor's expectation of future residual income. Ohlson expresses this expectation as:

$$E[x_{t+1}] = (R_f - 1)y_t + \omega x_t + v_t \quad (2.2)$$

where  $y_t$ ,  $x_t$  and  $v_t$  represent book value, residual income and information other than residual income, respectively, at time  $t$ . Together these equations suggest that value is a function of observable accounting information, namely book value and residual income, observable other information, and unobservable expectations of future residual income and other information that depend on observable information. In Ohlson's model, the coefficient  $\omega$  on  $x_t$  and  $v_t$  above are assumed to be fixed and known. However, in practice investors must form these expectations based on the information available to them. As a result, an individual investor faces estimation risk as he chooses the functional form of his estimation model and collects the relevant data to estimate the model.

While there is some debate in the literature about whether or not estimation risk is priced (Francis et al., 2005; Core et al., 2008), there is considerable theoretical work suggesting that, if the market model for a firm must be estimated using incomplete information, then estimation risk impacts investors' estimates of firm value and risk (Handa & Linn, 1993; Clarkson et al., 1996). Lambert et al. (2007) model a firm's cost of capital under this framework and show that disclosure unambiguously reduces the cost of capital by reducing investors' estimates of the covariance of the firm's cash flows with the market as a whole. In contrast to prior studies, which view estimation risk as being primarily resolved over time (as observations of a firm's returns increase), Lambert et al. (2007) show analytically that any information about a firm's future cash flows reduces estimation risk by reducing investors' estimates of the firm's exposure to systematic risk.

In Lambert et al. (2007), an investor estimates this exposure by estimating the covariance between a firm's cash flows and other firms in the market. This estimate is a function of two variables: the estimated covariance prior to the disclosure and the precision of the disclosure. In a market of two firms, this relationship is expressed as:

$$\text{Cov}(\tilde{V}_j, \tilde{V}_k | Z_j) = \text{Cov}(\tilde{V}_j, \tilde{V}_k) \frac{\text{Var}(\tilde{\epsilon}_j)}{\text{Var}(\tilde{Z}_j)} \quad (2.3)$$

where  $\tilde{V}_j$  is the expected future cash flows of firm  $j$ ,  $\tilde{V}_k$  is the expected future cash flows of firm  $k$ ,  $\tilde{Z}_j$  is noisy information about firm  $j$ 's cash flow, and  $\tilde{\epsilon}_j$  is the noise in  $\tilde{Z}_j$ . In terms of the model in Ohlson (1995), an investor's expectation of the future cash flows of firms  $j$  and  $k$ ,  $\tilde{V}_j$  and  $\tilde{V}_k$ , depend on the functional form he uses to estimate Equation 2.2, or the variables that he chooses to include in forming his expectation. In updating these expectations, an investor uses observed financial reports,  $\tilde{Z}_j$ , to modify and populate his estimation model and update his estimated covariance. In general, higher quality financial reporting reduces the estimated covariance between firms as investors incorporate more precise information into their valuation models.

Empirical studies support the idea that better information about a firm leads to a lower covariance of returns between firms. Durnev et al. (2003) find that current returns contain more information about future earnings when the proportion of a firm's returns explained by market and industry returns (called *return synchronicity*) is lower. Piotroski & Roulstone (2004) find that the

number of analyst forecasts are related to higher return synchronicity, indicating that analysts tend to provide information that applies to all firms in an industry rather than precise firm-specific information. Since analysts are often used as a proxy for sophisticated outside investors (see, for example, Bamber et al., 2011), this suggests that collecting firm-specific information is often too costly for outsiders. Public disclosure of this information could reduce the cost of collection and increase the proportion of investors who are informed (Kim & Verrecchia, 1994).

I predict that the presence of uncommon information in a firm's MD&A signals to investors that the firm's valuation model differs from those of its peers. In particular, it suggests variables for inclusion that the investor may not otherwise have considered. The resulting revision of investors' decision models creates differences between the firm's returns and the returns of peer firms. As a result, I predict that uncommon information is associated with a lower portion of firm returns explained by the returns of the market and the firm's industry.

### **2.1.1 Hypothesis 1**

I examine the relationship between uncommon information and firm-specific returns both prior to and following the release of the MD&A. MD&A disclosure is endogenously determined by the economics of the firm, so if the firm experiences unusual economic events or possesses unusual characteristics, those firms will disclose more uncommon information. In addition, because accounting information tends to lag stock prices (Ball & Shivakumar, 2008), the market should respond to these events and characteristics prior to the disclosure. This leads to my first hypothesis, stated in alternative form:

**Hypothesis 1** *Firm-specific returns and characteristics in year  $t$  are determinants of uncommon information disclosed at the end of year  $t$ .*

### 2.1.2 Hypothesis 2

In addition, the presence of uncommon information in a firm's disclosure may provide incremental information to the market. First, prior studies have found that the MD&A is incrementally informative over other disclosures (Davis & Tama-Sweet, 2012; Lee, 2012). Second, the presence of uncommon information in a firm's disclosures may signal to investors that the firm's valuation function differs from that of other firms. Managers have superior information to outside investors (Healy & Palepu, 2001) and securities regulations require that managers report relevant and material information in their financial reports (Financial Accounting Standards Board, 2010). As such, a decision by managers to include uncommon information in a financial report likely increases investors' estimates of its importance (Arya & Mittendorf, 2005). This is particularly true for certain sections of the report, such as the MD&A, in which managers are expected to "present their disclosure so that the most important information is most prominent" (Securities and Exchange Commission, 2003). The fact that managers of other firms decline to include this information suggests that this information is relevant only for the reporting firm.

If uncommon information in the MD&A is incrementally informative to the market, then uncommon information should be associated with an increase in the firm-specific portion of returns following the disclosure. This leads to my second hypothesis, stated in alternative form:

**Hypothesis 2** *Uncommon information disclosed at the end of year  $t$  is associated with an increase in firm-specific returns between year  $t$  and year  $t + 1$ .*

The model in Lambert et al. (2007) suggests that the valuation effect of a particular piece of information,  $\tilde{Z}_j$ , is a function of both its value relevance and precision. However, these properties are not known and must be estimated. In Section 2.2, I discuss characteristics of information that impose constraints on estimation. In addition, prior literature has found that, all else equal, investors are less likely to collect and process information when the costs of doing so are high (Grossman & Stiglitz, 1980; Bloomfield, 2002; Li, 2008; Lee, 2012). In Section 2.3, I discuss characteristics of information that increase processing costs.

## 2.2 Difficult-to-Estimate Information

As discussed in Section 2.1, an investor faces *parameter uncertainty* in estimating firm value: he is uncertain about the relevant parameters to include in his model and their relationship to value. Hong et al. (2007) argue that, faced with a large amount of information and parameter uncertainty, an investor chooses a parsimonious model that includes a subset of the most useful information. This investor behaves like an econometrician, using a theoretical model to predict what information is most relevant and then testing the model against economic outcomes. He incorrectly excludes relevant parameters (assuming he is aware of them) under two conditions. First, the investor lacks a strong theoretical reason to link a parameter with firm value. I discuss this condition in greater detail in Section 2.3. Second, the investor lacks sufficient data to generate a high-quality estimate of the parameter (Hong et al., 2007; Branch & Evans, 2006, 2010). Sufficient data can be generated using a time-series of observations (Barry & Brown, 1985; Coles et al., 1995; Lewellen & Shanken, 2002) or a cross-section (Lambert et al., 2007). Branch & Evans (2010) show analytically that even fully rational investors (i.e. not subject to cognitive limitations that might constrain model selection) exclude relevant parameters from their decision models when they possess insufficient data. When an investor is not confident in the estimates produced by including a particular variable, he is more likely to exclude that variable from his model.

There is already substantial evidence in the accounting literature that investors incorporate information less completely when the reliability of that information is difficult to estimate. The accrual anomaly may be the most well-documented market inefficiency in accounting research. Sloan (1996) and Richardson et al. (2005) argue that accruals are less persistent than cash flows because they are more subjectively determined and thus less reliable. However, investors seem to respond slowly to this lack of reliability. While the accounting literature has generally viewed this as an investor *overreaction* to the persistence of earnings, Bloomfield (2002) notes that it is fully consistent with an *underreaction* to information in accruals that is difficult to estimate. In other words, the low reliability of accruals suppresses investor learning about and the market response to the information contained in them.

Lewellen & Shanken (2002) demonstrate analytically that investor learning causes return patterns that are predictable *ex post* but are not visible to investors during the learning process. The empirical specification of accruals studies supports this explanation for the accrual anomaly. Richardson et al. (2005) estimate their model over a period of almost forty years. Xie (2001) shows that the accrual anomaly is most likely due to abnormal accruals, which are usually calculated over a sizable cross-section of firms or a time-series. Xie (2001) uses all firms in the same two-digit SIC code per industry, for instance, while Francis et al. (2005) use the Fama-French 48 industry groups and require at least 20 firms per industry. Francis et al. (2008) require that a firm have 11 years of data to be included in their model of accruals quality. If it takes over a decade of data or 20 comparable firms to estimate the reliability of accruals, it is unsurprising that investors process this information slowly.

Experimental research in accounting confirms that individuals' use of accounting information depends on a critical mass of comparable data. Lipe & Salterio (2000) find that managers tend to overweight (underweight) performance measures that are common (unique) across subordinates when evaluating their performance. This allows the manager to reduce cognitive effort by focusing only on the measures most useful for comparison (Slovic & MacPhillamy, 1974). Libby et al. (2004) find that an additional reason that individuals underweight the unique measures is that they are more uncertain about the quality of those measures. When individuals are provided with assurance about the quality of the unique measures, or are motivated to expend a greater amount of cognitive effort, they increase their weights on unique measures. More broadly, experiments in a variety of settings consistently show that the ability to compare alternatives significantly reduces the cognitive costs of choosing between them (Heneman, 1986; Zhang & Markman, 2001). This suggests that a rare item of information is likely to impose significant cognitive costs on investors.

Even in the presence of sufficient data to estimate a parameter, an investor may exclude it from his decision model if he lacks *ex ante* justification for its inclusion or the expected costs of collecting and processing the information exceed the expected benefits. I review these conditions in the next Section.



## 2.3 Difficult-to-Process Information

Research in judgment and decision-making finds that individuals economize on effort when forming their decision models, often settling for a “good enough” model rather than a complete one (Simon, 1955). This leads to differential search strategies depending on the decision at hand, with extra effort made to simplify the decision as the number of possible relevant parameters grows (Shields, 1980; Karelaia & Hogarth, 2008). Simplification strategies often involve ranking the importance of the available parameters subjectively (i.e., without calculating exact weights) and then including only those parameters judged to be most important in the decision model (Payne et al., 1993). Shields (1980, pg. 432) suggests that an individual’s choice of simplification strategy and the subjective weights placed on parameters depends on the individual’s internal representation of the decision task, with higher weights placed on parameters that are more closely “grouped” with other available parameters and the task.

An individual’s internal representation of the world in which he lives, sometimes referred to as a “semantic net” (Kintsch, 1988) or “concept map” (Sowa, 1984), represents knowledge as a network of ideas with linkages of varying strength between them. When an individual encounters a new idea, he creates a new concept in his map and forms linkages to related ideas based on the context. On each subsequent occasion that the individual encounters the same idea, he notes the context and augments his concept map to strengthen existing linkages (if they are present) and adds new linkages if necessary. Concepts themselves are defined in part by these linkages (Sowa, 1984, p. 76). For example, the individual can represent the concept of a “nurse” by noting the similarities and differences between the nurse’s role to the doctor’s in a medical context. When the individual retrieves a particular idea, he activates the portion of this concept map in which that idea resides, and stronger linkages lead to stronger activation of related concepts. If the individual has to make a decision involving a set of related concepts, the decision is less costly to make if the relevant concepts are closely linked to one another. If not, he has to search the remainder of his concept map and form new linkages between previously unlinked concepts.<sup>1</sup>

---

<sup>1</sup>For instance, if you see a television commercial advertising a “rug doctor”, you have to think

In this model, information processing costs occur both in the creation of new linkages and concepts (“learning”) and the retrieval of concepts from the map (“memory”). As such, information processing costs increase when the individual is presented with unfamiliar concepts or combinations of concepts that are not closely linked within the individual’s concept map. Consistent with this expectation, experimental research in psychology suggests that unfamiliar or unexpected words are more costly to process (Becker, 1976; Forster & Chambers, 1973; Glanzer & Ehrenreich, 1979; Forster, 1981). These studies find that words that occur with low probability in English text are harder for individuals to retain (Postman, 1970; Hulme et al., 1997), learn (Hall, 1954), and recall from long-term memory (Scarborough et al., 1977; Rayner & Duffy, 1986). In addition, familiar words are more costly to process when they are placed in an unfamiliar context. Halgren et al. (2002) finds that brain activity associated with language processing increases to a greater degree when individuals encounter contextually unexpected words in sentences than when they encounter words with low frequency in English text. Controlling for word frequency, Becker (1980) finds that individuals recall words more quickly when they are accompanied by related words (“nurse” and “doctor”), and less quickly when they are accompanied by unrelated words (“nurse” and “furniture”), than when words are presented with a neutral stimulus such as a non-sense string of characters. Tweedy et al. (1977) adds that this effect is more pronounced when the number of related words is increased. Context is also important in establishing the *meaning* of a word: Schvaneveldt et al. (1976) finds that individuals are able to decide on the meaning of an ambiguous word more quickly when it is presented with related words. On the whole, it seems that a decrease in the conditional probability of observing a word in a particular context leads to a significant increase in information processing.

---

about what a doctor is (someone who fixes physical problems that people have) and what kind of physical problems a rug might have (worn out, dirty) to decide that a rug doctor is probably a carpet cleaner or a person who replaces carpeting.

### 2.3.1 Hypothesis 3

If uncommon information imposes greater data collection and cognitive costs on investors, then uncommon information should be less completely incorporated into market prices (Bloomfield, 2002; Li, 2008). This leads to my third hypothesis:

**Hypothesis 3** *Uncommon information is associated with a delayed market response.*

### 2.3.2 Hypothesis 4

Examining uncommon information also provides an opportunity to examine the effect of conceptual complexity on textual complexity. Prior research finds that less readable language is associated with higher processing costs (Baddeley et al., 1975; Lee, 2012; Miller, 2010; Lehavy et al., 2011; Tan et al., 2013; Rennekamp, 2012a). However, there is little evidence on whether textual complexity in financial reports is due to the complexity of the information presented (Bloomfield, 2008; Rennekamp, 2012b) or managers' incentives to obfuscate negative information (Li, 2008). Information theory suggests that the cost of transmitting and processing a message is dependent on its information content (Shannon & Weaver, 1949; Pierce, 1980). A key feature of an efficient communication system is that messages are transmitted over a channel using the smallest number of informative units. As demonstrated by Pierce (1980, p. 94-96), this means that commonly (uncommonly) used messages will be encoded using shorter (longer) signals. In this way, the system will transmit shorter signals with higher probability than it transmits longer signals and the average signal length will be minimized. Thus, if we assume that it is costly to both transmit and receive messages, an efficient communication system reserves more costly signals for less common messages.

Research in information theory and linguistics suggests that common concepts are also more likely to be represented by words that are shorter, and thus more readable, than less common concepts. Zipf (1949) extends this feature to linguistic communication. Using an analogy of an artisan (language user) selecting her tools (words) from a workbench (vocabulary), Zipf notes that

the artisan will minimize her effort by keeping the tools that she uses most closest to her. Since smaller tools are easier to use, as well as easier to pack tightly so that they reduce the distance to the *next* tool in line, Zipf hypothesizes that the artisan will seek to minimize the size of those tools that she uses most frequently. Zipf then predicts and finds that there will be an inverse relationship between the lengths of words and the frequencies of their usage (Zipf, 1949, p. 63). This so-called Zipf's Law has subsequently been used heavily in computational linguistics, and is also one of the foundations of the Gunning Fog Index used in Li (2008). This leads to my fourth hypothesis:

**Hypothesis 4** *Text containing uncommon information is longer and less readable than text containing common information.*

The above hypotheses are predicated on identifying information that is perceived as uncommon by the average investor. Although investor perceptions are not directly observable, Landauer & Dumais (1997) find that Latent Semantic Analysis, the statistical technique used in this study, identifies concepts and relationships in text that agree with human judgments. Landauer et al. (1998) provides a comprehensive review of experimental research using LSA, including the use of computer models to approximate human performance on the Test of English as a Foreign Language. I discuss my use of LSA to measure uncommon information in the next section.<sup>2</sup>

---

<sup>2</sup>I use LSA to measure uncommon information because of its use in prior research to simulate human knowledge. Another method, Latent Dirichlet Allocation (Blei et al., 2003), is gaining currency in search engine applications to identify related documents. This method is used in concurrent accounting studies (Ball et al., 2013; Huang et al., 2014) and is worthy of further exploration.

## CHAPTER 3

### MEASURING UNCOMMON INFORMATION

My measure of uncommon information uses text of the Management's Discussion and Analysis (MD&A) section of the annual report to determine differences between a firm's MD&A and its industry peers. I use a textual measure for its power in discriminating between the meanings of messages: prior literature has shown that text has power to convey the tone (Loughran & McDonald, 2011b), uncertainty (Loughran & McDonald, 2013), riskiness Li (2006) and time interval Li (2010). Since numerical financial information is continuous in nature, it is difficult to formulate an interpretation of a particular realization. What constitutes a material change in a line item, and thus a distinct message? Even if a change is certainly material, the meaning of that change is highly contingent on the movements of other line items. Text, on the other hand, is discrete: although theoretically language choices are infinite, in practice word choice is highly constrained by context. Zipf (1949) finds that you need only 1,000 unique words to characterize most of the linguistic variation in a work as subtle as James Joyce's *Ulysses*; the average MD&A should be far more limited in its word choice. In addition, textual analysis provides tools to map the words to particular topics, thus providing a fair representation of the meaning of the text itself. Using this representation also allows for quantification of what is omitted: to the extent that what managers do *not* say is informative (Bloomfield, 2012), my measure can capture those choices as well. I focus on the MD&A because prior research shows that the MD&A is useful for users' decision-making (Barron et al., 1999; Clarkson et al., 1999; Brown & Tucker, 2011; Davis & Tama-Sweet, 2012). In addition, the MD&A provides a useful setting for quantifying the relative probabilities of observing particular messages. The MD&A is a mandatory component of the 10-K filing, and thus the decision to *have* an MD&A is not voluntary. However, although the SEC has provided some guidance on how the MD&A should be used, there remains considerable managerial discretion in what to discuss (Brown & Tucker, 2011). As a result, the MD&A provides a reasonably *complete*

characterization of what manager's discuss while also providing sufficient variation in disclosure choice.

### 3.1 Measuring the Characteristics of Language in the MD&A

The first paper to measure differences in language across MD&As was Brown & Tucker (2011)'s measure of year-over-year MD&A modifications. Brown & Tucker (2011) use a vector space model to represent the text of the MD&A as a vector in  $n$ -dimensional space, where  $n$  represents the number of unique words in the document and each entry in the vector represents the frequency of that word in the MD&A. Each entry is then weighted according to its "inverse document frequency" (IDF), or the number of documents in which it occurs; words that occur less frequently across all documents are more useful in discriminating between documents. IDF weighting is applied by multiplying the entry by  $\log\left(\frac{M}{m}\right)$ , where  $M$  represents the total number of documents and  $m$  is the number of documents in which that word appears. They then measure the cosine similarity between a firm's MD&A and the same firm's MD&A in the prior year. Cosine similarity represents the distance between the two MD&As in this  $n$ -dimensional space, and is calculated as:

$$\frac{v_1 \cdot v_2}{||v_1|| ||v_2||} \quad (3.1)$$

where  $v_1$  and  $v_2$  are the two document vectors,  $\cdot$  is the dot product operator, and  $||v_1||$  is the vector norm of  $v_1$ . Since document-term vectors are in positive space, cosine similarity is bounded between 0 and 1.

Brown & Tucker (2011) find that MD&As are quite sticky from year to year, with a mean (median) cosine similarity of 84% (89%) (Brown, 2013). However, they find that modifications, when they occur, are significantly associated with economic changes in the firm and with the market response to the 10-K filing, suggesting that modifications are informative about changes in firm value. Contrary to prior research (Barron et al., 1999; Clarkson et al., 1999), however, they find that analysts do not use changes in the MD&A in their earnings forecasts. Brown and Tucker

present evidence that MD&A modifications are more associated with changes in liquidity and capital resources than changes in operations, and suggest that these changes may be uninformative for annual earnings forecasts.

While Brown & Tucker (2011)’s measure works well for year-over-year changes within the same firm, it is not well suited for measuring differences between firms. MD&As within firms tend to be sticky, not just in terms of the information they convey, but also the words they use to convey it. Since most of the MD&A may be copied from year to year, observing new words is very likely to indicate that new information is being conveyed. Across firms, however, managers may use different words to express the same meaning (known as *synonymy* in linguistics) or the same word to express different meanings (*polysemy*). This tendency can cause similar MD&As to look different and different MD&As to look similar, in terms of raw word frequencies. The profusion of word choice across firms also results in a technical problem, common to data analysis, called the “curse of dimensionality” (Bellman, 2003). Measures of similarity tend to shrink as the number of dimensions in the measured space gets large: in essence, there are so many directions in which documents can be pushed that they all end up far away from one another. This results in little differentiation between similar and different documents. In computational linguistics, it is common to reduce the size of the space from one of words to one of meanings: from word space to *semantic* space, or topic space. I use Latent Semantic Analysis (LSA) for this purpose.

LSA is a form of factor analysis that identifies sources of variation in text. A purely mathematical technique, LSA infers the relationships between words by their co-occurrences: if two words occur in documents together often, they are likely to be related. Like other vector space models, LSA disregards the effects of word order or syntax. However, like factor analysis, LSA is able to uncover complex relationships from simple co-occurrences.

As an example of how LSA identifies relationships, take a collection of documents (corpus) that contains only two topics: guns and butter. LSA would examine documents in the corpus and identify the two main sources of variation, the topics guns and butter, and identify words that are most highly related to each: perhaps “shoot”, “aim”, “fire”, “metal”, and “barrel” for guns

and “dairy”, “milk”, “melt”, “churn” and “fattening” for butter. Since LSA is evaluating the co-occurrences for all words simultaneously, however, it also forms indirect linkages between words that do not co-occur but share co-occurrences with other words. For instance, if the word “heat” commonly co-occurs with the words “barrel” and “melt”, “heat” will also be linked to guns and butter, even if the word “heat” does not directly co-occur with the word “gun” or “butter” in any of the documents. In this way, LSA constructs a representation of linkages between words that is akin to the “concept map” discussed in Section 2.3.

LSA accomplishes this complex representation using Singular Value Decomposition (SVD), which is a form of eigendecomposition. Provided with a matrix of documents (often represented in rows) and counts of the words in those documents (represented in columns), SVD projects word co-occurrences into a high-dimensional space and determines the dimensions that best represent that space. Each word and document is then expressed as a coordinate within this high-dimensional space, and a distance measure (usually cosine similarity, as described in Equation 3.1 above) can be used to measure the strength of the relationship between particular words or documents. SVD ranks the dimensions in the order of the variation they explain, so that the first singular value corresponds to the dimension that explains the highest degree of variation. Each singular value is accompanied by a left and right singular vector that represents the dimension in row and column space, respectively. The original data is then projected on a subset of the singular vectors to obtain a representation of the data in the new semantic space, and distances can be calculated between words and documents in that space.

The right singular vectors (representing column space or word space) represent topics in the document collection. A researcher can determine the meaning of each topic by examining the words that have the highest association with the given topic in making his determination. Note, however, that assigning meaning to the topics is not necessary for determining similarities between documents and words in the collection.



### 3.2 Calculating *Uncommon*

I construct my sample of MD&As as follows. I download all 10-Ks posted on EDGAR between 1994 and 2012.<sup>1</sup> There were 137,516 10-Ks posted during this period. A little over 2,000 of these are duplicate 10-Ks filed across multiple subsidiaries of the same entity; eliminating these reduces the sample to 135,384. I extract the Management's Discussion and Analysis section from each 10-K using a Perl procedure. My extraction procedure is not perfect: due to formatting inconsistencies and the decision by some firms to include the MD&A by reference, I successfully identify and extract about 62% of the MD&As from this time period, for a sample of 83,457. I remove all 10-Ks for firms that do not occur in Compustat and group the 10-Ks by industry, eliminating 10-Ks for industry-years with fewer than 20 firms. This gives a sample on which *Uncommon* can be calculated of 35,385.

Table 3.1 shows descriptive statistics for annual data available in Compustat from 1994 to 2012 (Panel A) and for my final sample (Panel B). The differences between my sample and Compustat are statistically significant for most variables but few are economically significant. Unsurprisingly given my criterion that firms have at least 20 firms in their industry in a given year, the number of firms per industry and industry concentration are higher in my sample than in Compustat. In addition, MD&As are longer for firms in my final sample, which may be due in part to a slightly higher average number of segments (Li, 2008). Since length tends to increase mathematical measures of similarity between documents (Brown & Tucker, 2011), increased length in my final sample may make it more difficult for me to identify firms disclosing uncommon information. My sample also has a similar range and standard deviation for beta and return synchronicity, suggesting that I have sufficient variation to test my hypotheses. Overall, it does not appear that my study suffers from selection bias.

I construct my measure of uncommon information, *Uncommon*, using the following procedure. First, I remove all HTML and section headers from the text, in order to capture only the narrative

---

<sup>1</sup>Due to an oversight, my sample does not contain filings 10-K405, a special designation prior to 2003 indicating that the company did not file timely insider trading information.

Table 3.1: Sample Selection Comparison

<b>Panel A: Compustat Sample</b>							
Statistic	N	Mean	St. Dev.	Min	Median	Max	
<i>Beta</i>	36,892	1.064	0.712	−0.240	0.950	3.620	
<i>ReturnSync</i>	99,809	−2.194	1.542	−14.438	−2.053	2.103	
<i>EarnSync</i>	70,477	−2.323	2.382	−27.875	−1.925	5.701	
<i>Age</i>	99,654	13.137	11.987	1	9	67	
<i>Reg</i>	115,586	0.045	0.208	0	0	1	
<i>Size</i>	112,814	1,615.361	4,154.987	1.361	219.505	35,718.200	
<i>BTM</i>	112,339	0.660	0.530	0.0003	0.532	3.862	
<i>Leverage</i>	114,410	0.201	0.189	0.000	0.159	0.777	
<i>NBUSSEG</i>	85,093	3.948	3.813	1	3	33	
<i>NGEOSEG</i>	85,093	4.620	5.047	1	3	79	
<i>RevConc</i>	98,961	0.315	6.483	0.020	0.232	1,703.142	
<i>IndConc</i>	114,387	0.043	0.076	0.001	0.005	1.000	
<i>ReturnVol</i>	105,169	0.138	0.084	0.029	0.116	0.521	
<i>EarnVol</i>	72,762	0.059	0.062	0.001	0.038	0.349	
<i>AnaFollow</i>	65,043	8.056	7.739	1	5	68	
<i>Instown</i>	41,452	0.433	0.289	0.001	0.405	1.000	
<i>Length</i>	53,262	6,170.301	4,493.175	200	5,175	171,467	
<i>NumFirms</i>	98,576	109.955	74.970	20	93	326	
<b>Panel B: Uncommon Information Sample</b>							
Statistic	N	Mean	St. Dev.	Min	Median	Max	
<i>Uncommon</i>	35,385	−0.775	0.141	−0.998	−0.787	−0.289	
<i>Beta</i>	24,524	1.095	*** 0.728	−0.240	0.981	*** 3.620	
<i>ReturnSync</i>	31,827	−2.006	*** 1.504	−14.438	−1.820	*** 2.003	
<i>EarnSync</i>	24,646	−2.333	2.373	−21.043	−1.929	5.509	
<i>Age</i>	31,824	12.941	*** 11.337	1	10	*** 67	
<i>Reg</i>	35,385	0.040	*** 0.195	0	0	*** 1	
<i>Size</i>	35,020	1,313.116	*** 3,386.329	1.364	254.317	*** 35,503.660	
<i>BTM</i>	34,508	0.669	*** 0.549	0.0004	0.530	*** 3.861	
<i>Leverage</i>	34,969	0.196	*** 0.195	0.000	0.145	*** 0.777	
<i>NBUSSEG</i>	25,879	4.298	*** 3.701	1	3	*** 30	
<i>NGEOSEG</i>	25,879	5.128	*** 5.029	1	3	*** 73	
<i>RevConc</i>	30,840	0.315	9.734	0.030	0.167	*** 1,703.142	
<i>IndConc</i>	35,385	0.128	*** 0.088	0.025	0.107	*** 1.000	
<i>ReturnVol</i>	32,616	0.145	*** 0.086	0.029	0.124	*** 0.521	
<i>EarnVol</i>	23,589	0.064	*** 0.065	0.001	0.042	*** 0.349	
<i>AnaFollow</i>	23,209	8.284	*** 7.630	1	6	*** 68	
<i>Instown</i>	27,897	0.454	*** 0.290	0.001	0.434	*** 1.000	
<i>Length</i>	35,385	6,818.828	*** 4,491.641	229	5,946	*** 70,950	
<i>NumFirms</i>	35,373	117.355	*** 75.603	20	102	*** 326	

Note: \*\*\* indicates that mean or median in the Uncommon sample is significantly different from the Compustat sample at less than the 1% level.

All variables defined in Table A.1

portion of the MD&A. Second, I remove all stopwords (the, and, of, etc.) and stem the remaining words to their roots.<sup>2</sup> Third, to reduce error from typos and proper nouns, I remove all word stems that occur in fewer than 1% of the MD&As in each Fama-French industry and in the MD&A of only one CIK code per industry, for all years. Fourth, I construct a document-term matrix for each industry-year. The document-term matrix contains the number of times that each word appears in each MD&A. Following Brown & Tucker (2011), I apply a normalized TF-IDF weighting to each entry. I modify this weighting in two ways. First, as recommended by Loughran & McDonald (2011b), I scale the raw count of each word stem in each MD&A by the total number of words in that MD&A. Second, I multiply the result by the inverse document frequency of that word stem within each industry-year instead of the full sample, to be consistent with calculating distances on an industry-year basis. Formally, my normalized TF-IDF weighting is calculated as:

$$(WordCount/Length) * \log \left( \frac{M}{m} \right) \quad (3.2)$$

where *WordCount* is the number of times a particular word stem occurs in a particular MD&A, *Length* is the total word count in that MD&A, *M* is the total number of MD&As in the same industry-year and *m* is the number of MD&As in the same industry-year in which this particular word stem occurs. I also center and scale each column to have a mean of zero and a variance of one to normalize the data. Third, I perform SVD on the matrix. Like factor analysis, SVD produces as many dimensions in the projected space as there are in the original data. In order to obtain meaningful distances, it is necessary to discard dimensions that have low explanatory power. However, there is no clear rule of thumb for selecting these dimensions. I find that distances are well-behaved when the dimensions in column space are about  $\frac{1}{10}$  of the size of the sample to be distanced, which provides enough observations per dimension to be meaningful while still capturing a substantial portion of the variation in the dimensions. As such, I select a number of dimensions equal to 10% of the number of firms in the industry-year. Prior studies have shown

---

<sup>2</sup>Stemming removes the suffix of words in order to group related words by their root. For example, “represent”, “represented”, and “representing” are all reduced to “repres”.

that, since there is considerable variation within industries, comparisons are best made between a firm and a few comparable firms instead of the entire industry (Bhojraj & Lee, 2002; DeFranco et al., 2011; Gong et al., 2013). As such, I calculate the average cosine similarity between each MD&A and the 4 most similar MD&As in the same industry-year. Unlike Brown & Tucker (2011), my transformed matrix contains both positive and negative values, so cosine similarity is bounded between one (completely similar) and negative one (completely dissimilar) instead of one and zero. Since uncommon information should be contained in those MD&As that are *dissimilar*, *Uncommon*, is this average cosine similarity multiplied by negative one.

### 3.3 Examples of Uncommon Information

The following examples illustrate how *Uncommon* represents the contents of documents and distances between them. Table 3.2 shows the top 5 dimensions for the Software Industry for Fiscal Year 2007.

Table 3.2: Top 10 Words for the Top 5 Dimensions in the Software Industry, Fiscal Year 2007

	Competitive Advantage	Labor Force	Benchmarking	Growth/ Change	Contracting
1	skill	group	excel	frequent	defer
2	context	aim	innov	requir	arrang
3	expertis	worker	window	rapid	evid
4	sap	stoppag	compel	intens	valuat
5	cultur	colleg	portal	rigor	recogn
6	rigor	council	highest	expertis	consist
7	solv	serv	mda	sap	establish
8	creat	vice	lowest	bargain	compens
9	problem	degre	reader	solv	undeliv
10	analyst	timeli	year	secret	residu

To aid in the interpretation of this example, I have provided my interpretation of the topic represented by each dimension, based on the word list and observing these words in context. As you can see from the first and second dimensions, the software industry is focused on solving problems and attracting top talent. The focus on words associated with growth and meeting benchmarks

suggest that it is also a rapidly changing and competitive industry.

Table 3.3 shows the top 10 word stems (by TF-IDF weighting) for the most and least common MD&As in this industry year. The most common MD&A belongs to NCI Information Systems, Inc. This company provides tech support to mostly government agencies. NCI is particularly concerned with software contracts, discussing “backlog” and “subcontractor” a great deal. To see many of these words in context, see B. It is not surprising that a company that produces primarily stable, commonplace services would discuss common topics in its MD&A.

Table 3.3: Top 10 Words for the Most and Least Common MD&As in the Software Industry, Fiscal Year 2007

	NCI	Borland
1	backlog	vice
2	client	serv
3	subcontractor	counsel
4	percent	degre
5	civilian	interoper
6	fring	stoppag
7	subcontract	document
8	contract	hold
9	predetermin	colleg
10	unfund	council

The least common MD&A in the Software Industry for 2007 belongs to Borland Software Corporation. Borland provides services to support software developers. The prominence of the word stem “serv” in its MD&A makes this apparent. Borland states in its MD&A that it is “a leading vendor of Open Application Lifecycle Management solutions, or ALM... a *new*, customer-centric approach to helping IT organizations transform software delivery” (emphasis mine). It “differentiate[s] our products and solutions from those of our competitors based on cross-platform interoperability”, hence the word stem “interoper” in their top words list. Presumably to give shareholders confidence, they spend a significant amount of time discussing the qualifications of their top management, leading to word stems such as “vice”, “counsel”, “degree” and “college”. Borland appears to be pioneering in a niche market, leading to a highly uncommon MD&A. For an excerpt of this MD&A, see B.

## CHAPTER 4

### TESTS OF HYPOTHESES AND EMPIRICAL RESULTS

#### 4.1 Uncommon Information and Firm-Specific Returns

In Hypothesis 1, I predict that firm-specific returns and characteristics are determinants of uncommon information. Observing an association between uncommon information and firm-specific returns and characteristics prior to the disclosure is consistent with economic events affecting both variables. To this end, I examine the relationship between *Uncommon* and two measures of firm-specific returns in the year prior to the disclosure. I examine market beta (*Beta*) because the model in Lambert et al. (2007) predicts that investors respond to the presence of firm-specific information by reducing their estimates of beta. I also examine return synchronicity, which measures the portion of firm returns explained by both the market (as does beta) and firms in the same industry. I calculate return synchronicity, *ReturnSync*, following Piotroski & Roulstone (2004):  $\log\left(\frac{R^2}{1-R^2}\right)$ , where  $R^2$  is the the R-squared from a regression of weekly firm returns on concurrent and lagged value-weighted market and industry returns over the fiscal year. I hypothesize that uncommon information is associated with the prior incorporation into stock prices of a higher amount of firm-specific information, leading to a negative association between *Uncommon* and the market beta and return synchronicity in the year prior to the disclosure.

I also examine other factors that may cause a firm to differ economically from other firms. Young firms (*Age*) tend to have higher product differentiation than other firms (Khan & Manopichetwattana, 1989), which might lead to uncommon information, while regulated firms (*Reg*), tend to be more homogenous. In addition to market beta, I examine firm size (*Size*), book-to-market ratio (*BTM*) and leverage (*Leverage*). These common sources of risk may be associated with *Uncommon* if risky firms are more likely to provide uncommon information.

In Section 1, I suggest that managers make a cost-benefit tradeoff in deciding whether to pro-

vide uncommon information. As such, firms might be more likely to provide uncommon information if the benefits of doing so are higher. As such, I predict that uncommon information is more likely to be present when other sources of information, such as the firm's earnings, do not fully reflect the economic differences between the firm and its peers (for instance, if the differences are forward-looking). I include earnings synchronicity, *EarnSync*, and interact it with *Beta* and *ReturnSync*, to see if uncommon information is more prevalent among firms with economic differences that are not reflected in earnings. Summary earnings is also a poorer reflection of a firm's economics when the firm's income is more diverse (Piotroski & Roulstone, 2004), so I include the number of business and geographical segments (*NBUSSEG* and *NGEOSEG*); *RevConc*, a firm-specific, revenue-based Herfindahl index to reflect the diversification of the firm across multiple segments; and earnings volatility, *EarnVol*. Analysts and institutional owners are also significant sources of information about the firm, so firms might make more uncommon information when they have lower analyst coverage and institutional ownership. On the other hand, analysts and institutional owners are attracted to better disclosures, so the reverse could also be true. Since firms are sensitive to proprietary costs (Verrecchia, 1983), they may be less likely to provide uncommon information when their industry is less concentrated (*IndConc*).

I also examine factors that might mechanically impact *Uncommon*: the number of words in the MD&A (*Length*) and the number of firms in the industry (*NumFirms*). Brown & Tucker (2011) show that the cosine similarity measure causes documents that are longer to be more similar to all other documents, reducing differences between them. Although I adjust my similarity measure for document length, I include *Length* to see if any of this mechanical relationship remains. *NumFirms* might impact *Uncommon* in two ways. More firms in the industry might make it easier to find close peers for the firm, reducing *Uncommon*. On the other hand, if industry classification is noisy, a larger industry might indicate diverse firms being grouped together, increasing *Uncommon*.

Table 4.1 shows descriptive statistics for *Uncommon* and the above determinants and Table 4.2 shows univariate Pearson (above diagonal) and Spearman (below diagonal) correlations. The average (median) firm has a highly similar MD&A to its peers, with a cosine similarity of 0.795

(0.813) (keeping in mind that *Uncommon* is cosine similarity multiplied by  $-1$ ). In the univariate case, *Uncommon* is positively associated with *Beta* and unassociated with *ReturnSync*, contrary to predictions. Results on riskiness are mixed: firms providing uncommon information are lower growth and less leveraged, but also younger and in less competitive industries.

Table 4.1: Descriptive Statistics for Determinants Analysis

Statistic	N	Mean	St. Dev.	Min	Median	Max
<i>Uncommon</i>	8,276	-0.795	0.138	-0.998	-0.813	-0.290
<i>Beta</i>	8,276	1.186	0.727	-0.217	1.079	3.691
<i>ReturnSync</i>	8,276	-1.539	1.365	-11.839	-1.341	1.699
<i>EarnSync</i>	8,276	-2.364	2.386	-20.829	-1.957	4.605
<i>Age</i>	8,276	17.495	12.114	3	13	66
<i>Reg</i>	8,276	0.068	0.252	0	0	1
<i>Size</i>	8,276	2,070.291	4,299.256	2.681	527.195	35,412.500
<i>BTM</i>	8,276	0.590	0.429	0.001	0.494	3.798
<i>Leverage</i>	8,276	0.201	0.191	0.000	0.168	0.777
<i>NBUSSEG</i>	8,276	5.000	4.050	1	3	30
<i>NGEOSEG</i>	8,276	5.981	5.604	1	3	66
<i>RevConc</i>	8,276	0.209	0.369	0.037	0.163	31.699
<i>EarnVol</i>	8,276	0.065	0.063	0.001	0.044	0.349
<i>Anafollow</i>	8,276	8.049	7.261	1	6	47
<i>Instown</i>	8,276	0.595	0.254	0.001	0.624	1.000
<i>IndConc</i>	8,276	0.124	0.080	0.024	0.109	0.716
<i>Length</i>	8,276	7,140.683	4,305.407	280	6,417	57,467
<i>NumFirms</i>	8,276	95.674	61.095	20	78	326

All variables defined in Table A.1



Table 4.2: Correlations between Determinants

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. <i>Uncommon</i>		<b>0.11</b>	-0.02	0.02	<b>-0.10</b>	<b>-0.07</b>	<b>0.04</b>	<b>-0.08</b>	<b>-0.15</b>	-0.01	<b>0.06</b>	-0.01	<b>0.12</b>	<b>0.03</b>	<b>-0.03</b>	<b>-0.17</b>	<b>0.08</b>	<b>0.58</b>
2. <i>Beta</i>	<b>0.11</b>		<b>0.15</b>	<b>0.06</b>	<b>-0.18</b>	<b>-0.19</b>	<b>-0.07</b>	0.00	<b>-0.15</b>	<b>-0.12</b>	<b>0.21</b>	<b>-0.02</b>	<b>0.33</b>	<b>0.09</b>	<b>0.08</b>	<b>0.04</b>	<b>0.08</b>	<b>0.21</b>
3. <i>ReturnSync</i>	-0.02	<b>0.18</b>		0.02	<b>0.22</b>	<b>0.07</b>	<b>0.26</b>	<b>-0.06</b>	<b>0.04</b>	<b>0.17</b>	<b>0.15</b>	<b>-0.08</b>	<b>-0.14</b>	<b>0.32</b>	<b>0.42</b>	<b>-0.10</b>	<b>0.31</b>	0.01
4. <i>EarnSync</i>	0.02	<b>0.07</b>	<b>0.03</b>		<b>-0.04</b>	<b>-0.07</b>	<b>0.06</b>	-0.01	<b>-0.04</b>	0.00	<b>0.07</b>	-0.01	<b>0.11</b>	<b>0.05</b>	<b>0.04</b>	<b>0.03</b>	0.00	0.00
5. <i>Age</i>	<b>-0.09</b>	<b>-0.17</b>	<b>0.23</b>	<b>-0.04</b>		<b>0.32</b>	<b>0.27</b>	0.01	<b>0.09</b>	<b>0.27</b>	<b>0.09</b>	<b>-0.03</b>	<b>-0.25</b>	<b>0.11</b>	<b>0.10</b>	<b>-0.09</b>	<b>0.17</b>	<b>-0.20</b>
6. <i>Reg</i>	<b>-0.06</b>	<b>-0.20</b>	<b>0.07</b>	<b>-0.07</b>	<b>0.24</b>		<b>0.12</b>	0.02	<b>0.16</b>	<b>0.16</b>	<b>-0.13</b>	0.01	<b>-0.16</b>	0.02	<b>-0.11</b>	<b>-0.20</b>	<b>0.16</b>	<b>-0.09</b>
7. <i>Size</i>	0.01	<b>-0.08</b>	<b>0.49</b>	<b>0.03</b>	<b>0.25</b>	<b>0.15</b>		<b>-0.16</b>	0.02	<b>0.22</b>	<b>0.15</b>	<b>-0.05</b>	<b>-0.14</b>	<b>0.62</b>	<b>0.23</b>	-0.01	<b>0.28</b>	-0.01
8. <i>BTM</i>	<b>-0.11</b>	<b>-0.04</b>	-0.01	<b>-0.03</b>	<b>0.08</b>	<b>0.07</b>	<b>-0.35</b>		<b>0.11</b>	0.01	<b>-0.07</b>	0.01	<b>-0.12</b>	<b>-0.19</b>	<b>-0.11</b>	-0.02	<b>0.05</b>	<b>-0.11</b>
9. <i>Leverage</i>	<b>-0.17</b>	<b>-0.19</b>	<b>0.05</b>	<b>-0.05</b>	<b>0.10</b>	<b>0.18</b>	<b>0.13</b>	<b>0.13</b>		<b>0.07</b>	<b>-0.16</b>	0.01	<b>-0.24</b>	-0.01	0.01	-0.01	<b>0.12</b>	<b>-0.15</b>
10. <i>NBUSSEG</i>	-0.01	<b>-0.10</b>	<b>0.18</b>	0.01	<b>0.19</b>	<b>0.13</b>	<b>0.26</b>	<b>0.05</b>	<b>0.09</b>		<b>0.25</b>	<b>-0.19</b>	<b>-0.19</b>	<b>0.10</b>	<b>0.11</b>	<b>-0.07</b>	<b>0.30</b>	<b>-0.07</b>
11. <i>NGEOSEG</i>	<b>0.08</b>	<b>0.19</b>	<b>0.18</b>	<b>0.07</b>	<b>0.07</b>	<b>-0.15</b>	<b>0.18</b>	<b>-0.08</b>	<b>-0.18</b>	<b>0.40</b>		<b>-0.19</b>	<b>0.04</b>	<b>0.14</b>	<b>0.18</b>	0.00	<b>0.13</b>	<b>0.07</b>
12. <i>RevConc</i>	<b>-0.05</b>	<b>-0.06</b>	<b>-0.17</b>	<b>-0.05</b>	<b>-0.12</b>	<b>0.05</b>	<b>-0.22</b>	<b>0.04</b>	<b>0.07</b>	<b>-0.76</b>	<b>-0.79</b>		0.01	<b>-0.04</b>	<b>-0.07</b>	0.01	<b>-0.08</b>	<b>-0.03</b>
13. <i>EarnVol</i>	<b>0.13</b>	<b>0.38</b>	<b>-0.19</b>	<b>0.15</b>	<b>-0.28</b>	<b>-0.25</b>	<b>-0.30</b>	<b>-0.18</b>	<b>-0.33</b>	<b>-0.19</b>	<b>0.10</b>	0.02		<b>-0.04</b>	<b>-0.17</b>	<b>0.10</b>	<b>-0.11</b>	<b>0.20</b>
14. <i>Anafollow</i>	<b>0.02</b>	<b>0.12</b>	<b>0.39</b>	<b>0.07</b>	<b>0.08</b>	<b>0.04</b>	<b>0.73</b>	<b>-0.23</b>	<b>0.03</b>	<b>0.11</b>	<b>0.15</b>	<b>-0.13</b>	<b>-0.06</b>		<b>0.42</b>	<b>-0.03</b>	<b>0.26</b>	<b>0.04</b>
15. <i>Instown</i>	-0.02	<b>0.09</b>	<b>0.42</b>	<b>0.04</b>	<b>0.10</b>	<b>-0.12</b>	<b>0.52</b>	<b>-0.07</b>	<b>0.02</b>	<b>0.14</b>	<b>0.22</b>	<b>-0.21</b>	<b>-0.15</b>	<b>0.50</b>		<b>-0.04</b>	<b>0.25</b>	0.00
16. <i>IndConc</i>	<b>-0.22</b>	<b>0.04</b>	<b>-0.10</b>	<b>0.03</b>	<b>-0.08</b>	<b>-0.30</b>	<b>-0.05</b>	<b>-0.05</b>	-0.01	<b>-0.07</b>	<b>0.03</b>	0.01	<b>0.12</b>	<b>-0.03</b>	0.01		<b>-0.10</b>	<b>-0.24</b>
17. <i>Length</i>	<b>0.09</b>	<b>0.12</b>	<b>0.37</b>	0.01	<b>0.08</b>	<b>0.10</b>	<b>0.34</b>	<b>0.05</b>	<b>0.11</b>	<b>0.30</b>	<b>0.22</b>	<b>-0.26</b>	<b>-0.13</b>	<b>0.29</b>	<b>0.31</b>	<b>-0.08</b>		<b>0.07</b>
18. <i>NumFirms</i>	<b>0.64</b>	<b>0.18</b>	0.01	0.01	<b>-0.17</b>	<b>-0.08</b>	0.00	<b>-0.14</b>	<b>-0.19</b>	<b>-0.06</b>	<b>0.07</b>	<b>-0.03</b>	<b>0.18</b>	0.02	0.01	<b>-0.31</b>	<b>0.11</b>	

All variables defined in Table A.1

Pearson correlations are above the diagonal, Spearman below.

Correlations in bold are significant at below the 5% level.

Table 4.3 shows the results of the regression of *Uncommon* on its determinants. Consistent with expectations and contrary to the univariate results, *Uncommon* is negatively associated with both *Beta* and *ReturnSync* at less than the 1% level. This suggests that it is important to control for other factors in evaluating these relationships. In addition, *Uncommon* is significantly positively associated with *EarnSync* in one of the two specifications, indicating that firms make more uncommon information when earnings do not reflect differences between firms. Also consistent with this prediction, the interactions between *Beta* and *EarnSync* and between *ReturnSync* and *EarnSync* have negative coefficients that are significant at less than the 1% level, suggesting that a firm's disclosures are more uncommon when there is a mismatch between return synchronicity and earnings synchronicity. Firms with uncommon information also have more business and geographic segments, as predicted. However, these firms are also less diversified across their segments, inconsistent with predictions. *Uncommon* is negatively associated with analyst following and institutional ownership, adding to the evidence that uncommon information is more likely when investors have fewer sources for firm-specific information.<sup>1</sup>

It does not appear that firms with more uncommon information are riskier along the traditional dimensions of risk. *Uncommon* is negatively associated with *Beta*, positively associated with *Size*, unassociated with *BTM* and negatively associated with leverage. In addition, *Uncommon* is positively associated with *Length*, allaying fears that the measure is mechanically affected by document length and providing preliminary evidence for Hypothesis 4. However, as is clear from Table 4.2 and the coefficient on *NumFirms*, *Uncommon* is highly positively associated with *NumFirms*, suggesting that it is important to control for *NumFirms* in all regressions that contain *Uncommon*.

The results above suggest that disclosures of uncommon information are affected by the underlying economics and characteristics of the firm. To test Hypothesis 2, I examine the association between uncommon information and the change in return synchronicity following the disclosure, controlling for firm characteristics. Finding an association in this analysis suggests that the disclosure of uncommon information affects the firm's information environment. I regress  $\Delta ReturnSync$ ,

---

<sup>1</sup>It is also possible that analysts are more likely to follow firms with common information.

Table 4.3: Regression of *Uncommon* on Determinants

	Dependent variable:	
	<i>Uncommon</i>	
<i>Beta</i>	−0.008*** (0.002)	
<i>ReturnSync</i>		−0.004*** (0.001)
<i>EarnSync</i>	0.003*** (0.001)	−0.0004 (0.001)
<i>Age</i>	0.0002 (0.0002)	0.0002 (0.0002)
<i>Reg</i>	0.023 (0.018)	0.024 (0.018)
<i>Size</i>	0.007*** (0.002)	0.008*** (0.002)
<i>BTM</i>	0.002 (0.003)	0.002 (0.003)
<i>Leverage</i>	−0.024** (0.010)	−0.025** (0.010)
<i>NBUSSEG</i>	0.014*** (0.004)	0.014*** (0.004)
<i>NGEOSEG</i>	0.010*** (0.003)	0.010*** (0.003)
<i>RevConc</i>	0.025*** (0.009)	0.025*** (0.009)
<i>EarnVol</i>	0.006 (0.028)	−0.008 (0.028)
<i>Anafollow</i>	−0.007*** (0.002)	−0.008*** (0.002)
<i>Instown</i>	−0.022*** (0.007)	−0.021*** (0.007)
<i>IndConc</i>	−0.007 (0.005)	−0.007 (0.005)
<i>Length</i>	0.012** (0.006)	0.012* (0.006)
<i>NumFirms</i>	0.111*** (0.012)	0.111*** (0.012)
<i>Beta</i> × <i>EarnSync</i>	−0.002*** (0.001)	
<i>ReturnSync</i> × <i>EarnSync</i>		−0.001** (0.0003)
Constant	−1.314*** (0.042)	−1.325*** (0.040)
Year fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
Observations	8,276	8,276
Adjusted R <sup>2</sup>	0.391	0.391

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

All variables defined in Table A.1

Standard errors in parentheses are clustered at the firm and year level.

the change in return synchronicity from the year prior to the disclosure to the year following, on *Uncommon* and  $\Delta Uncommon$ , the level and change in uncommon information, and control variables. All variables in this analysis are defined in detail in Table A.1. The coefficient on *Uncommon* ( $\Delta Uncommon$ ) reflects the association between the level of (change in) uncommon information and a future decline in return synchronicity. To control for other characteristics of the MD&A, I include the year-over-year change in the MD&A, *Change*, after Brown & Tucker (2011) obtained from Stephen Brown's website (Brown, 2013), and  $\Delta Length$ , the change in the total number of words in the MD&A. Following Piotroski & Roulstone (2004), I control for known determinants of return synchronicity. I include  $\Delta EarnSync$ , the change in earnings synchronicity (calculated similarly to return synchronicity using return on assets in place of returns and calculated over the last 12 quarters); *Reg*, whether the firm is in a regulated industry (two-digit SIC code of 49 or 62);  $\Delta Instown$ , the change in the percentage of the firm owned by institutions from the 13-F filings; *IndConc*, industry concentration (the log of an industry-specific revenue-based Herfindahl index);  $\Delta RevConc$ , the change in the diversification of the firm (the log of a firm-specific revenue-based Herfindahl index);  $\Delta Anafollow$ , the change in the number of analysts issuing a forecast for the firm's earnings during the year;  $\Delta Size$ , the change in the size of the firm (the log of the market value of equity at year-end);  $\Delta EarnVol$ , the change in operating earnings volatility (the standard deviation of earnings over the last five years). I also include  $\Delta NumFirms$ , the change in the log of the number of firms in the industry, since it is highly correlated with *Uncommon*.

Table 4.4 shows descriptive statistics for these variables. Consistent with the determinants analysis, MD&As across firms in the same industry are quite similar. In addition, *Uncommon* changes little from year-to-year, with a mean (median) absolute value of 0.085 (0.061) and a standard deviation of 0.081. There is, however, significant variation, with absolute changes up to 0.586 relative to a mean level of -0.791 for *Uncommon*. The year-over-year change in MD&A is consistent with Brown & Tucker (2011), with a mean and median between 10% and 15%.

Table 4.5 shows univariate Pearson (Spearman) correlations above (below) the diagonal for the regression variables. Consistent with Hypothesis 2, both *Uncommon* and  $\Delta Uncommon$  are

Table 4.4: Descriptive Statistics for *ReturnSync* Regressions

Statistic	N	Mean	St. Dev.	Min	Median	Max
$\Delta ReturnSync$	3,272	0.070	1.421	-10.549	0.071	9.953
<i>Uncommon</i>	3,272	-0.791	0.140	-0.998	-0.809	-0.299
$\Delta Uncommon$	2,567	0.002	0.118	-0.586	0.001	0.576
$ \Delta Uncommon $	2,567	0.085	0.081	0.0001	0.061	0.586
<i>Change</i>	3,272	0.148	0.138	0.002	0.102	0.858
$\Delta Length$	3,272	533.586	2,271.784	-16,427	433.5	16,160
$\Delta EarnSync$	3,272	0.008	2.937	-16.160	0.008	14.501
<i>Reg</i>	3,272	0.057	0.232	0	0	1
$\Delta Instown$	3,272	0.033	0.143	-0.612	0.025	0.672
$\Delta IndConc$	3,272	-0.018	0.079	-0.575	-0.005	0.371
$\Delta RevConc$	3,272	0.0001	0.105	-0.831	0.000	0.678
$\Delta Ana follow$	3,272	0.273	2.517	-14	0	19
$\Delta Size$	3,272	226.259	1,412.057	-9,459.574	34.196	23,132.040
$\Delta EarnVol$	3,272	-0.004	0.031	-0.281	-0.001	0.172
$\Delta NumFirms$	3,272	-0.762	18.293	-53	-2	128

All variables defined in Table A.1

significantly associated with a decline in return synchronicity in the year following the disclosure.

In addition, the MD&A is more likely to contain a higher level and an increase in uncommon information as it increases in length over the prior year.

Table 4.5: Correlations between *ReturnSync* Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\Delta ReturnSync$		<b>-0.04</b>	<b>-0.05</b>	-0.02	<b>-0.05</b>	-0.01	0.01	<b>0.04</b>	-0.01	<b>-0.06</b>	0.01	-0.01	<b>0.04</b>	-0.01
<i>Uncommon</i>	<b>-0.05</b>		<b>0.44</b>	<b>0.05</b>	<b>0.07</b>	-0.02	<b>-0.09</b>	-0.01	<b>0.08</b>	<b>0.04</b>	0.00	<b>0.06</b>	-0.03	<b>-0.12</b>
$\Delta Uncommon$	<b>-0.05</b>	<b>0.40</b>		0.03	<b>0.10</b>	-0.03	-0.01	<b>-0.06</b>	0.02	<b>-0.04</b>	-0.02	0.00	0.02	0.02
<i>Change</i>	-0.02	<b>0.05</b>	0.03		<b>0.06</b>	-0.01	<b>-0.05</b>	0.01	<b>-0.06</b>	<b>-0.06</b>	0.03	-0.03	0.02	<b>0.10</b>
$\Delta Length$	<b>-0.07</b>	<b>0.09</b>	<b>0.11</b>	<b>0.07</b>		<b>-0.06</b>	0.03	0.00	0.03	-0.02	-0.03	0.01	0.01	0.03
$\Delta EarnSync$	-0.02	-0.01	-0.03	-0.01	-0.02		0.00	0.02	-0.01	0.00	-0.02	0.00	0.00	-0.01
<i>Reg</i>	0.01	<b>-0.09</b>	0.00	<b>-0.04</b>	0.03	0.00		-0.03	0.03	-0.03	<b>-0.04</b>	0.02	0.03	0.02
$\Delta Instown$	<b>0.04</b>	0.00	<b>-0.05</b>	-0.01	-0.03	0.02	-0.03		-0.01	0.01	<b>0.08</b>	0.01	0.02	0.00
$\Delta IndConc$	-0.02	<b>0.09</b>	0.04	<b>-0.05</b>	<b>0.06</b>	0.00	0.01	-0.02		<b>0.04</b>	0.00	0.02	<b>-0.06</b>	<b>-0.19</b>
$\Delta RevConc$	-0.03	<b>0.05</b>	-0.01	-0.03	-0.01	0.01	-0.03	<b>0.04</b>	0.00		0.02	-0.02	<b>-0.08</b>	<b>-0.07</b>
$\Delta AnaFollow$	0.03	-0.01	-0.03	0.00	<b>-0.04</b>	-0.01	-0.03	<b>0.10</b>	-0.03	<b>0.05</b>		0.00	0.03	0.01
$\Delta Size$	<b>0.04</b>	0.02	-0.01	<b>-0.06</b>	0.01	-0.01	<b>0.05</b>	<b>0.15</b>	<b>0.07</b>	-0.01	<b>0.07</b>		0.00	<b>-0.04</b>
$\Delta EarnVol$	<b>0.04</b>	<b>-0.04</b>	0.00	0.03	-0.02	0.01	0.02	0.03	<b>-0.06</b>	<b>-0.04</b>	0.03	-0.01		<b>0.11</b>
$\Delta NumFirms$	0.03	<b>-0.17</b>	0.01	<b>0.07</b>	0.01	-0.03	<b>0.04</b>	0.00	<b>-0.29</b>	<b>-0.06</b>	0.02	<b>-0.06</b>	<b>0.06</b>	

All variables defined in Table A.1

Pearson correlations are above the diagonal, Spearman below.

Correlations in bold are significant at below the 5% level.

Table 4.6 shows the multivariate results. Consistent with Hypothesis 2, both the level and the change in *Uncommon* are significantly associated (at the 10% level or greater) with a decline in return synchronicity. With the exception of the change in earnings volatility, which is significantly associated with an increase in return synchronicity, the other variables in the regression are not significant.

The above results are consistent with disclosed uncommon information being both a result of uncommon firm economics and characteristics and a source of information about the firm. However, an alternative explanation for my findings is that uncommon information released prior to the MD&A disclosure is processed slowly, leading to a decline in return synchronicity following the disclosure. In the next Section, I examine the timing of the market response to uncommon information to determine whether it precedes or follows disclosure.

## **4.2 Uncommon Information and Information Processing Costs**

### **4.2.1 Uncommon Information and Delayed Market Response**

In this Section, I examine the effect of uncommon information on the firm's returns around the MD&A disclosure. In Hypothesis 3, I predict that a significant portion of the market response to uncommon information is delayed, leading to abnormal returns in the window following its release. To test Hypothesis 3, I regress the absolute value of the firm's cumulative abnormal returns in various windows on *Uncommon* and control variables. I calculate cumulative abnormal returns using a market model estimated over 255 days ending 63 days prior to the disclosure. I choose 63 days because 95% of firms report earnings fewer than 63 days prior to the 10-K filing. I take the absolute value of returns because I do not know the sign of the news that is reported in uncommon information. This is consistent with theoretical research (e.g. Kim & Verrecchia, 1991) that uses the absolute value of returns as a measure of the absolute value of news. I measure the absolute value of cumulative abnormal returns (henceforth *CAR*) over four windows. The first two windows examine the market response around the disclosure:  $CAR_{-1,1}$  for the day before to the day after the

Table 4.6: Regression of  $\Delta ReturnSync$  on *Uncommon*

	Dependent variable:	
	$\Delta ReturnSync$	
<i>Uncommon</i>	−0.297* (0.177)	
$\Delta Uncommon$		−0.421** (0.211)
<i>Change</i>	0.122 (0.212)	0.116 (0.149)
$\Delta Length$	−0.00002* (0.00001)	−0.00001 (0.00001)
$\Delta EarnSync$	−0.006 (0.009)	−0.006 (0.009)
<i>Reg</i>	−0.237 (0.182)	−0.243 (0.161)
$\Delta Instown$	0.245 (0.260)	0.290 (0.254)
$\Delta IndConc$	0.270 (0.358)	0.336 (0.450)
$\Delta RevConc$	−0.724 (0.561)	−0.694 (0.525)
$\Delta Anafollow$	0.007 (0.007)	0.007 (0.007)
$\Delta Size$	−0.00002 (0.00002)	−0.00003 (0.00003)
$\Delta EarnVol$	1.871** (0.788)	1.791** (0.832)
$\Delta NumFirms$	−0.002 (0.002)	−0.003 (0.002)
Constant	−1.440*** (0.343)	−1.230
Year fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
Observations	3,272	2,567
Adjusted R <sup>2</sup>	0.065	0.056

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01  
All variables defined in Table A.1

Standard errors in parentheses are clustered at the firm and year level.



filing and  $CAR_{2,60}$  for the post-disclosure period from 2 to 60 days after the disclosure. In order to provide evidence that an observed market response is not due to unobserved characteristics of the firm, I also examine a pseudo-event period prior to the earnings announcement. This examination also provides evidence that the market is responding to uncommon information in the MD&A, as opposed to previously disclosed information.  $PCAR_{-1,1}$  measures  $CAR$  over the three days starting 63 days prior to the earnings announcement (-63 to -61) and  $PCAR_{2,60}$  measures  $CAR$  over the 58 days leading up to the earnings announcement window (-60 to -2). Results for  $CAR_{-1,1}$  or  $CAR_{2,60}$  that are not mirrored in the corresponding pseudo-event period increases confidence that the result is due to a characteristic of the MD&A disclosure and not the firm. All variables in this analysis are defined in detail in Table A.1.

I control for the year-over-year change (*Change*) and length (*Length*) of the MD&A in order to separate from *Uncommon* the effect of additional information or information processing costs that might be contained in changed or long MD&As. I also control for other factors that impact returns. Since prior research has found significant drift associated with earnings announcements (i.e. Ball & Brown, 1968), I include  $CAR$  calculated during the earnings announcement window ( $EA_{-1,1}$ ) and the log of the absolute value of unexpected earnings ( $UE$ ) to control for news contained in earnings. I also include known sources of risk: market beta (*Beta*) and the firm's size (*Size*), book-to-market ratio (*BTM*) and leverage (*Leverage*). To control for uncertainty surrounding the firm's economics, I also control for the volatility of returns (*ReturnVol*) and earnings (*EarnVol*). Following Brown & Tucker (2011), I also include *FileLate*, an indicator variable equal to one if the 10-K was filed more than 90 days after the fiscal year-end and zero otherwise, and *NumItems*, the log of the number of non-missing Compustat items. To control for the effects of transaction costs, I include trading volume for each window (*Volume*) and the stock price at the end of the fiscal year (*Price*). Finally, consistent with the findings in the determinants analysis, I control for the impact of the number of firms per industry on *Uncommon*.

Tables 4.7 and 4.8 provide descriptive statistics and univariate correlations between these variables. In the univariate case, *Uncommon* has significant Pearson correlations with  $CAR$  in all

windows, emphasizing the importance of controlling for sources of risk and other firm-specific factors in the multivariate case.

Table 4.7: Descriptive Statistics for CAR Regressions

Statistic	N	Mean	St. Dev.	Min	Median	Max
<i>CAR</i> <sub>-1,1</sub>	5,767	0.034	0.043	0.00002	0.021	0.516
<i>CAR</i> <sub>2,60</sub>	5,767	0.169	0.173	0.0001	0.118	1.960
<i>PCAR</i> <sub>-1,1</sub>	5,767	0.036	0.043	0.00000	0.023	0.579
<i>PCAR</i> <sub>2,60</sub>	5,767	0.159	0.177	0.00000	0.108	3.635
<i>Volume</i> <sub>-1,1</sub>	5,767	2.585	1.324	-31.841	2.691	6.335
<i>Volume</i> <sub>2,60</sub>	5,767	5.697	0.977	0.935	5.754	9.343
<i>PVolume</i> <sub>-1,1</sub>	5,767	2.407	1.471	-33.964	2.504	6.955
<i>PVolume</i> <sub>2,60</sub>	5,767	5.578	0.989	0.660	5.627	8.971
<i>Uncommon</i>	5,767	-0.779	0.143	-0.998	-0.791	-0.288
<i>Change</i>	5,767	0.144	0.136	0.002	0.098	0.883
<i>Length</i>	5,767	6,390.743	3,830.657	259	5,689	37,819
<i>EA</i> <sub>-1,1</sub>	5,767	0.056	0.061	0.00001	0.037	0.746
<i>VolumeEA</i> <sub>-1,1</sub>	5,767	3.001	1.292	-3.326	3.111	6.591
<i>UE</i>	5,767	-6.061	1.722	-11.864	-6.202	2.258
<i>Beta</i>	5,767	1.057	0.737	-0.217	0.919	3.683
<i>Size</i>	5,767	1,714.738	3,521.883	2.786	508.615	34,929.220
<i>BTM</i>	5,767	0.565	0.408	0.0004	0.480	3.770
<i>Leverage</i>	5,767	0.205	0.184	0.000	0.176	0.776
<i>ReturnVol</i>	5,767	0.126	0.073	0.029	0.110	0.521
<i>EarnVol</i>	5,767	0.066	0.065	0.001	0.044	0.349
<i>FileLate</i>	5,767	0.168	0.374	0	0	1
<i>NumItems</i>	5,767	292.687	43.494	133	296	391
<i>Price</i>	5,767	24.113	24.656	0.320	19.550	769.400
<i>NumFirms</i>	5,767	108.145	71.251	20	89	326

All variables defined in Table A.1

Table 4.8: Correlations between CAR Variables

	1	2	3	4	5	6	7	8	9	10	11	12
1. $CAR_{-1,1}$		<b>0.23</b>	<b>0.17</b>	<b>0.26</b>	<b>0.19</b>	<b>0.10</b>	<b>0.06</b>	<b>0.09</b>	<b>0.04</b>	<b>0.11</b>	<b>-0.12</b>	<b>0.29</b>
2. $CAR_{2,60}$	<b>0.20</b>		<b>0.21</b>	<b>0.27</b>	<b>0.08</b>	<b>0.16</b>	<b>0.06</b>	<b>0.11</b>	<b>0.07</b>	<b>0.14</b>	<b>-0.10</b>	<b>0.19</b>
3. $PCAR_{-1,1}$	<b>0.16</b>	<b>0.15</b>		<b>0.21</b>	<b>0.07</b>	<b>0.10</b>	<b>0.10</b>	<b>0.14</b>	<b>0.06</b>	<b>0.08</b>	<b>-0.05</b>	<b>0.14</b>
4. $PCAR_{2,60}$	<b>0.21</b>	<b>0.24</b>	<b>0.19</b>		<b>0.12</b>	<b>0.16</b>	<b>0.16</b>	<b>0.25</b>	<b>0.03</b>	<b>0.12</b>	<b>-0.10</b>	<b>0.24</b>
5. $Volume_{-1,1}$	<b>0.19</b>	<b>0.10</b>	<b>0.08</b>	<b>0.13</b>		<b>0.78</b>	<b>0.52</b>	<b>0.70</b>	<b>0.04</b>	<b>-0.03</b>	<b>0.16</b>	<b>0.18</b>
6. $Volume_{2,60}$	<b>0.11</b>	<b>0.16</b>	<b>0.11</b>	<b>0.16</b>	<b>0.87</b>		<b>0.63</b>	<b>0.88</b>	<b>0.04</b>	<b>-0.02</b>	<b>0.18</b>	<b>0.22</b>
7. $PVolume_{-1,1}$	<b>0.08</b>	<b>0.11</b>	<b>0.12</b>	<b>0.18</b>	<b>0.71</b>	<b>0.78</b>		<b>0.71</b>	0.03	0.00	<b>0.14</b>	<b>0.15</b>
8. $PVolume_{2,60}$	<b>0.10</b>	<b>0.13</b>	<b>0.14</b>	<b>0.22</b>	<b>0.80</b>	<b>0.88</b>	<b>0.88</b>		<b>0.04</b>	0.00	<b>0.17</b>	<b>0.20</b>
9. <i>Uncommon</i>	0.02	<b>0.05</b>	<b>0.06</b>	<b>0.04</b>	<b>0.05</b>	<b>0.05</b>	<b>0.05</b>	<b>0.06</b>		<b>0.03</b>	<b>0.14</b>	<b>0.04</b>
10. <i>Change</i>	<b>0.11</b>	<b>0.13</b>	<b>0.08</b>	<b>0.11</b>	<b>-0.06</b>	<b>-0.04</b>	<b>-0.02</b>	<b>-0.02</b>	<b>0.04</b>		<b>-0.20</b>	<b>0.10</b>
11. <i>Length</i>	<b>-0.13</b>	<b>-0.10</b>	<b>-0.07</b>	<b>-0.10</b>	<b>0.21</b>	<b>0.21</b>	<b>0.18</b>	<b>0.19</b>	<b>0.16</b>	<b>-0.24</b>		<b>-0.06</b>
12. $EA_{-1,1}$	<b>0.21</b>	<b>0.18</b>	<b>0.15</b>	<b>0.19</b>	<b>0.22</b>	<b>0.24</b>	<b>0.20</b>	<b>0.23</b>	<b>0.04</b>	<b>0.11</b>	<b>-0.06</b>	
13. $VolumeEA_{-1,1}$	<b>0.09</b>	<b>0.11</b>	<b>0.10</b>	<b>0.17</b>	<b>0.78</b>	<b>0.84</b>	<b>0.74</b>	<b>0.84</b>	<b>0.06</b>	<b>-0.03</b>	<b>0.19</b>	<b>0.39</b>
14. <i>UE</i>	<b>0.12</b>	<b>0.17</b>	<b>0.07</b>	<b>0.14</b>	<b>-0.14</b>	<b>-0.15</b>	<b>-0.13</b>	<b>-0.15</b>	<b>-0.04</b>	<b>0.08</b>	<b>-0.08</b>	<b>0.10</b>
15. <i>Beta</i>	<b>0.20</b>	<b>0.21</b>	<b>0.18</b>	<b>0.22</b>	<b>0.36</b>	<b>0.41</b>	<b>0.38</b>	<b>0.42</b>	<b>0.11</b>	<b>0.05</b>	<b>0.04</b>	<b>0.25</b>
16. <i>Size</i>	<b>-0.19</b>	<b>-0.24</b>	<b>-0.13</b>	<b>-0.20</b>	<b>0.39</b>	<b>0.40</b>	<b>0.36</b>	<b>0.39</b>	<b>0.03</b>	<b>-0.15</b>	<b>0.36</b>	<b>-0.10</b>
17. <i>BTM</i>	-0.01	0.01	-0.02	<b>-0.03</b>	<b>-0.31</b>	<b>-0.33</b>	<b>-0.31</b>	<b>-0.34</b>	<b>-0.10</b>	-0.01	<b>-0.03</b>	<b>-0.07</b>
18. <i>Leverage</i>	<b>-0.09</b>	<b>-0.08</b>	<b>-0.08</b>	<b>-0.06</b>	<b>-0.07</b>	<b>-0.08</b>	<b>-0.08</b>	<b>-0.10</b>	<b>-0.18</b>	<b>-0.04</b>	<b>0.08</b>	<b>-0.10</b>
19. <i>ReturnVol</i>	<b>0.35</b>	<b>0.40</b>	<b>0.34</b>	<b>0.44</b>	<b>0.22</b>	<b>0.28</b>	<b>0.29</b>	<b>0.32</b>	<b>0.07</b>	<b>0.18</b>	<b>-0.17</b>	<b>0.33</b>
20. <i>EarnVol</i>	<b>0.26</b>	<b>0.30</b>	<b>0.25</b>	<b>0.30</b>	<b>0.27</b>	<b>0.31</b>	<b>0.33</b>	<b>0.35</b>	<b>0.05</b>	<b>0.16</b>	<b>-0.11</b>	<b>0.29</b>
21. <i>FileLate</i>	<b>0.15</b>	<b>0.14</b>	<b>0.10</b>	<b>0.13</b>	<b>-0.08</b>	<b>-0.10</b>	<b>-0.07</b>	<b>-0.09</b>	0.02	<b>0.16</b>	<b>-0.17</b>	<b>0.09</b>
22. <i>NumItems</i>	<b>-0.10</b>	<b>-0.03</b>	<b>-0.07</b>	<b>-0.08</b>	<b>0.35</b>	<b>0.37</b>	<b>0.33</b>	<b>0.33</b>	<b>-0.05</b>	<b>-0.15</b>	<b>0.39</b>	<b>0.06</b>
23. <i>Price</i>	<b>-0.24</b>	<b>-0.28</b>	<b>-0.19</b>	<b>-0.26</b>	<b>0.24</b>	<b>0.24</b>	<b>0.19</b>	<b>0.22</b>	-0.02	<b>-0.13</b>	<b>0.18</b>	<b>-0.16</b>
24. <i>NumFirms</i>	0.01	<b>0.04</b>	0.02	<b>0.03</b>	<b>0.04</b>	<b>0.04</b>	<b>0.04</b>	<b>0.05</b>	<b>0.66</b>	0.00	<b>0.16</b>	0.01

Continued on next page

Table 4.8 (cont'd)

	13	14	15	16	17	18	19	20	21	22	23	24
1. <i>CAR</i> <sub>-1,1</sub>	<b>0.08</b>	<b>0.18</b>	<b>0.15</b>	<b>-0.10</b>	<b>0.07</b>	<b>-0.05</b>	<b>0.33</b>	<b>0.22</b>	<b>0.17</b>	<b>-0.07</b>	<b>-0.14</b>	0.01
2. <i>CAR</i> <sub>2,60</sub>	<b>0.08</b>	<b>0.21</b>	<b>0.19</b>	<b>-0.13</b>	<b>0.11</b>	<b>-0.05</b>	<b>0.40</b>	<b>0.26</b>	<b>0.16</b>	-0.02	<b>-0.17</b>	<b>0.05</b>
3. <i>PCAR</i> <sub>-1,1</sub>	<b>0.09</b>	<b>0.13</b>	<b>0.16</b>	<b>-0.07</b>	<b>0.06</b>	<b>-0.05</b>	<b>0.36</b>	<b>0.21</b>	<b>0.12</b>	<b>-0.04</b>	<b>-0.11</b>	<b>0.04</b>
4. <i>PCAR</i> <sub>2,60</sub>	<b>0.17</b>	<b>0.17</b>	<b>0.19</b>	<b>-0.12</b>	<b>0.04</b>	<b>-0.04</b>	<b>0.46</b>	<b>0.27</b>	<b>0.13</b>	<b>-0.06</b>	<b>-0.16</b>	<b>0.03</b>
5. <i>Volume</i> <sub>-1,1</sub>	<b>0.70</b>	<b>-0.14</b>	<b>0.31</b>	<b>0.16</b>	<b>-0.25</b>	<b>-0.04</b>	<b>0.18</b>	<b>0.19</b>	<b>-0.07</b>	<b>0.31</b>	<b>0.15</b>	0.00
6. <i>Volume</i> <sub>2,60</sub>	<b>0.84</b>	<b>-0.15</b>	<b>0.40</b>	<b>0.18</b>	<b>-0.29</b>	<b>-0.06</b>	<b>0.25</b>	<b>0.25</b>	<b>-0.11</b>	<b>0.38</b>	<b>0.16</b>	0.00
7. <i>PVolume</i> <sub>-1,1</sub>	<b>0.61</b>	<b>-0.10</b>	<b>0.31</b>	<b>0.14</b>	<b>-0.21</b>	<b>-0.04</b>	<b>0.23</b>	<b>0.22</b>	<b>-0.08</b>	<b>0.28</b>	<b>0.10</b>	0.01
8. <i>PVolume</i> <sub>2,60</sub>	<b>0.84</b>	<b>-0.15</b>	<b>0.41</b>	<b>0.18</b>	<b>-0.29</b>	<b>-0.07</b>	<b>0.30</b>	<b>0.29</b>	<b>-0.09</b>	<b>0.35</b>	<b>0.15</b>	0.02
9. <i>Uncommon</i>	<b>0.05</b>	<b>-0.03</b>	<b>0.10</b>	<b>0.04</b>	<b>-0.08</b>	<b>-0.16</b>	<b>0.07</b>	<b>0.08</b>	0.03	<b>-0.08</b>	-0.02	<b>0.61</b>
10. <i>Change</i>	-0.01	<b>0.09</b>	<b>0.04</b>	<b>-0.07</b>	0.02	<b>-0.04</b>	<b>0.17</b>	<b>0.13</b>	<b>0.15</b>	<b>-0.11</b>	<b>-0.07</b>	-0.01
11. <i>Length</i>	<b>0.17</b>	<b>-0.07</b>	<b>0.05</b>	<b>0.24</b>	<b>-0.05</b>	<b>0.10</b>	<b>-0.14</b>	<b>-0.07</b>	<b>-0.14</b>	<b>0.26</b>	<b>0.13</b>	<b>0.14</b>
12. <i>EA</i> <sub>-1,1</sub>	<b>0.38</b>	<b>0.12</b>	<b>0.21</b>	<b>-0.07</b>	<b>0.03</b>	<b>-0.08</b>	<b>0.29</b>	<b>0.21</b>	<b>0.11</b>	<b>0.04</b>	<b>-0.11</b>	<b>0.03</b>
13. <i>VolumeEA</i> <sub>-1,1</sub>		<b>-0.14</b>	<b>0.38</b>	<b>0.17</b>	<b>-0.29</b>	<b>-0.09</b>	<b>0.22</b>	<b>0.24</b>	<b>-0.09</b>	<b>0.38</b>	<b>0.15</b>	0.01
14. <i>UE</i>	<b>-0.13</b>		<b>0.07</b>	<b>-0.27</b>	<b>0.35</b>	<b>0.14</b>	<b>0.26</b>	<b>0.19</b>	<b>0.20</b>	<b>-0.04</b>	<b>-0.35</b>	<b>-0.07</b>
15. <i>Beta</i>	<b>0.40</b>	<b>0.06</b>		-0.02	<b>-0.13</b>	<b>-0.22</b>	<b>0.43</b>	<b>0.42</b>	0.02	<b>0.25</b>	<b>-0.14</b>	<b>0.12</b>
16. <i>Size</i>	<b>0.36</b>	<b>-0.48</b>	0.01		<b>-0.20</b>	0.01	<b>-0.18</b>	<b>-0.10</b>	<b>-0.13</b>	<b>0.15</b>	<b>0.38</b>	-0.02
17. <i>BTM</i>	<b>-0.33</b>	<b>0.31</b>	<b>-0.18</b>	<b>-0.40</b>		<b>0.11</b>	<b>0.06</b>	<b>-0.11</b>	<b>0.15</b>	<b>-0.16</b>	<b>-0.28</b>	<b>-0.12</b>
18. <i>Leverage</i>	<b>-0.12</b>	<b>0.11</b>	<b>-0.24</b>	<b>0.10</b>	<b>0.15</b>		<b>-0.13</b>	<b>-0.22</b>	0.01	<b>-0.08</b>	0.02	<b>-0.16</b>
19. <i>ReturnVol</i>	<b>0.26</b>	<b>0.25</b>	<b>0.49</b>	<b>-0.35</b>	<b>-0.04</b>	<b>-0.17</b>		<b>0.48</b>	<b>0.20</b>	-0.01	<b>-0.26</b>	<b>0.07</b>
20. <i>EarnVol</i>	<b>0.31</b>	<b>0.21</b>	<b>0.49</b>	<b>-0.23</b>	<b>-0.19</b>	<b>-0.29</b>	<b>0.60</b>		<b>0.14</b>	<b>0.13</b>	<b>-0.19</b>	<b>0.07</b>
21. <i>FileLate</i>	<b>-0.08</b>	<b>0.19</b>	<b>0.04</b>	<b>-0.29</b>	<b>0.09</b>	0.01	<b>0.22</b>	<b>0.16</b>		<b>-0.17</b>	<b>-0.16</b>	0.00
22. <i>NumItems</i>	<b>0.37</b>	<b>-0.06</b>	<b>0.23</b>	<b>0.27</b>	<b>-0.18</b>	<b>-0.14</b>	<b>-0.06</b>	<b>0.14</b>	<b>-0.22</b>		<b>0.05</b>	<b>-0.20</b>
23. <i>Price</i>	<b>0.22</b>	<b>-0.54</b>	<b>-0.21</b>	<b>0.76</b>	<b>-0.40</b>	<b>0.07</b>	<b>-0.46</b>	<b>-0.37</b>	<b>-0.26</b>	<b>0.11</b>		<b>-0.03</b>
24. <i>NumFirms</i>	<b>0.04</b>	<b>-0.05</b>	<b>0.11</b>	0.00	<b>-0.14</b>	<b>-0.19</b>	<b>0.05</b>	<b>0.03</b>	0.00	<b>-0.11</b>	<b>-0.04</b>	

All variables defined in Table A.1

Pearson correlations are above the diagonal, Spearman below.

Correlations in bold are significant at below the 5% level.

Table 4.9 shows the results of the regression. Consistent with the MD&A being a source of uncommon information, *Uncommon* is unassociated with market returns prior to the disclosure. In addition, consistent with a delayed market response to uncommon information, the coefficient on *Uncommon* is positive and significant at less than the 1% level in the post-disclosure window but insignificant in the three-day window around the 10-K filing. In addition, the coefficient on *Change* is positive and significant in the post-disclosure window, suggesting that a change in the MD&A also imposes significant information processing costs on investors. *Change* is likewise insignificant in the pre-disclosure period. Contrary to the findings of Brown & Tucker (2011), *Change* is insignificant in the disclosure window. However, due to data limitations of calculating *Uncommon* and including additional control variables, my sample size is significantly smaller (5,767 firm years to their 23,083), which may limit the power of my test to detect an effect. Note, however, that significance in the disclosure window for either *Change* or *Uncommon* does not contradict the hypothesis that changes in the MD&A and uncommon information impose significant information processing costs on investors. It would simply indicate that some investors are faster at processing this information than others.

#### **4.2.2 Uncommon Information and Readability**

Hypothesis 4 predicts that uncommon information will be longer and less readable than common disclosure. To test this hypothesis, I examine two common readability measures from the linguistics and accounting literature: the Gunning Fog index (*Fog*) and the total number of words (*Length*). Following Li (2008), I test if *Uncommon* has explanatory power for these two measures after controlling for known determinants. Since the Fog index is calculated as the summary of two components, I also examine whether *Uncommon* explains each of these components: percentage of complex words (where a complex word has three syllables or more) and words per sentence. All variables in this analysis are defined in detail in Table A.1.

I calculate the Fog index following Li (2008), with one modification. Prior literature has calculated the Fog index using the Fathom package, which counts the syllables per word and words

Table 4.9: Regression of *CAR* and *PCAR* on *Uncommon*

	Dependent variable:			
	<i>CAR</i> <sub>-1,1</sub>	<i>CAR</i> <sub>2,60</sub>	<i>PCAR</i> <sub>-1,1</sub>	<i>PCAR</i> <sub>2,60</sub>
<i>CAR</i> <sub>-1,1</sub>		0.158*** (0.057)		
<i>Volume</i> <sub>-1,1</sub>	0.011*** (0.002)	-0.009 (0.007)		
<i>Volume</i> <sub>2,60</sub>		0.071*** (0.012)		
<i>PCAR</i> <sub>-1,1</sub>				0.052 (0.096)
<i>PVolume</i> <sub>-1,1</sub>			0.001** (0.001)	-0.004 (0.003)
<i>PVolume</i> <sub>2,60</sub>				0.063*** (0.013)
<i>Uncommon</i>	0.004 (0.005)	0.042*** (0.016)	0.006 (0.005)	-0.017 (0.019)
<i>Change</i>	0.004 (0.004)	0.036** (0.016)	-0.003 (0.003)	0.010 (0.017)
<i>Length</i>	0.0004 (0.001)	0.004 (0.004)	0.003*** (0.001)	0.003 (0.003)
<i>EA</i> <sub>-1,1</sub>	0.148*** (0.021)	0.161** (0.067)	0.004 (0.007)	0.249* (0.142)
<i>VolumeEA</i> <sub>-1,1</sub>	-0.006*** (0.002)	-0.024*** (0.007)	0.002*** (0.0005)	-0.012*** (0.004)
<i>UE</i>	0.001** (0.0003)	0.002* (0.001)	0.001 (0.001)	0.002 (0.002)
<i>Beta</i>	-0.001 (0.001)	-0.007* (0.004)	-0.002 (0.001)	-0.015* (0.008)
<i>Size</i>	-0.001** (0.001)	-0.016*** (0.003)	-0.0005 (0.001)	-0.010*** (0.002)
<i>BTM</i>	-0.004** (0.002)	-0.005 (0.006)	-0.001 (0.002)	-0.013*** (0.004)
<i>Leverage</i>	-0.005 (0.004)	0.006 (0.013)	-0.002 (0.004)	0.00002 (0.010)
<i>ReturnVol</i>	0.065*** (0.013)	0.384*** (0.052)	0.143*** (0.024)	0.639*** (0.092)
<i>EarnVol</i>	0.010 (0.013)	0.075* (0.040)	0.019 (0.016)	0.001 (0.050)
<i>FileLate</i>	0.004* (0.003)	0.012** (0.006)	0.003 (0.002)	0.0001 (0.003)
<i>NumItems</i>	0.009 (0.012)	0.051 (0.037)	0.011 (0.015)	-0.0004 (0.036)
<i>Price</i>	-0.008*** (0.002)	-0.024*** (0.005)	-0.005*** (0.002)	-0.027*** (0.006)
<i>NumFirms</i>	-0.001 (0.004)	0.011 (0.017)	0.005 (0.005)	0.004 (0.018)
Constant	-0.004 (0.069)	-0.340 (0.223)	-0.070 (0.084)	-0.109 (0.201)
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	5,767	5,767	5,767	5,767
Adjusted R <sup>2</sup>	0.245	0.251	0.152	0.288

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

All variables defined in Table A.1

Standard errors in parentheses are clustered at the firm and year level.

per sentence and then calculates the relevant summary statistics. However, the Fathom package is designed for general use in (often noisy) machine-readable English text and makes certain assumptions about the structure of the text in calculating its metrics. One of these assumptions is that a period (.) ends a sentence if it occurs between two word characters (word characters in Perl are capital and lowercase letters, the digits 0 through 9 and the underscore) with an *optional* space. This means that the Fathom package identifies a new sentence in almost all cases where it observes a period, including periods that represent a decimal point in a number. As a result, text that contains more decimalized numbers is calculated as having more, shorter sentences and is considered more readable. I modify the Fathom package to exclude periods that precede digits in my calculation of *Fog*. I calculate *Length* following the prior literature, since this calculation excludes numbers and thus is not affected by this issue.

To test Hypothesis 4, I regress *Fog*, percentage of complex words (*%Complex*), words per sentence (*WPS*), and *Length* on uncommon information, the year-over-year change in the MD&A and known determinants of *Fog* and *Length*. These determinants include the log of the market value of equity (*Size*), the book-to-market ratio (*BTM*), the ratio of special items to total assets (*Special*), return and earnings volatility (*ReturnVol* and *EarnVol*), the number of business and geographic segments (*NBUSSEG* and *NGEOSEG*), the financial complexity of the firm, calculated as the log of the number of non-missing items in Compustat (*NumItems*), an indicator variable if the firm is incorporated in Delaware (*DE*), and indicator variables if the firm has had a seasoned equity offering or merger/acquisition during the year (*SEO* and *MA*). Consistent with the determinants analysis for *Uncommon*, I also include the log of the number of firms in the industry (*NumFirms*).

Table 4.10 shows descriptive statistics for these variables. The mean (median) of *Fog* for the MD&A is 21.447 (21.389). This is somewhat higher than that reported in Li (2008), which may be attributable to the issue with the Fathom package mentioned above. The mean (median) of *Length* is 5,837 (4,971), higher than but in line with Li's results. *Uncommon* and *Change* are also in line with prior analyses. Table 4.11 shows univariate correlations between these variables. The univariate results are consistent with predictions: *Uncommon* is significantly positively correlated

with *Fog* and *Length*. The relationship between uncommon information and the Fog index appears to be due to uncommon information containing longer sentences rather than longer words. In addition, *Change* is negatively correlated with *Fog* , *WPS* and *Length* but positively correlated with *%Complex*.

Table 4.10: Descriptive Statistics for Readability Regressions

Statistic	N	Mean	St. Dev.	Min	Median	Max
<i>Fog</i>	7,842	21.447	1.263	18.078	21.389	25.799
<i>%Complex</i>	7,842	27.867	1.645	24.309	27.830	31.549
<i>WPS</i>	7,842	25.751	2.634	20.743	25.569	33.497
<i>Length</i>	7,842	5,837.828	3,719.285	259	4,971	44,672
<i>Uncommon</i>	7,842	-0.783	0.142	-0.998	-0.796	-0.288
<i>Change</i>	7,842	0.157	0.143	0.001	0.110	0.956
<i>Size</i>	7,842	1,300.269	3,184.431	1.769	281.058	34,884.640
<i>BTM</i>	7,842	0.644	0.510	0.002	0.516	3.848
<i>Special</i>	7,842	-0.017	0.113	-5.056	0.000	1.032
<i>ReturnVol</i>	7,842	0.139	0.079	0.029	0.121	0.521
<i>EarnVol</i>	7,842	0.074	0.067	0.002	0.053	0.349
<i>NBUSSEG</i>	7,842	4.973	3.814	1	3	30
<i>NGEOSEG</i>	7,842	5.732	5.284	1	3	65
<i>NumItems</i>	7,842	294.547	34.738	196	294	391
<i>DE</i>	7,842	0.561	0.496	0	1	1
<i>SEO</i>	7,842	0.863	0.344	0	1	1
<i>MA</i>	7,842	0.328	0.469	0	0	1
<i>Age</i>	7,842	15.742	10.808	1	12	66
<i>NumFirms</i>	7,842	101.958	65.428	20	87	326

All variables defined in Table A.1



Table 4.11: Correlations between Readability Variables

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. <i>Fog</i>		<b>0.55</b>	<b>0.85</b>	<b>0.22</b>	<b>0.07</b>	<b>-0.06</b>	<b>0.03</b>	<b>-0.06</b>	-0.02	-0.01	<b>0.06</b>	0.02	0.01	<b>0.10</b>	<b>0.07</b>	<b>0.04</b>	<b>0.04</b>	<b>-0.08</b>	<b>0.11</b>
2. <i>%Complex</i>	<b>0.54</b>		<b>0.04</b>	<b>-0.05</b>	0.02	<b>0.03</b>	<b>0.03</b>	-0.02	0.01	<b>-0.05</b>	<b>-0.05</b>	0.00	<b>-0.07</b>	<b>-0.05</b>	0.00	<b>-0.03</b>	<b>0.06</b>	-0.02	<b>0.05</b>
3. <i>WPS</i>	<b>0.84</b>	<b>0.04</b>		<b>0.29</b>	<b>0.08</b>	<b>-0.09</b>	<b>0.02</b>	<b>-0.07</b>	<b>-0.03</b>	<b>0.02</b>	<b>0.09</b>	0.02	<b>0.05</b>	<b>0.15</b>	<b>0.09</b>	<b>0.06</b>	0.01	<b>-0.09</b>	<b>0.10</b>
4. <i>Length</i>	<b>0.24</b>	<b>-0.05</b>	<b>0.32</b>		<b>0.12</b>	<b>-0.21</b>	<b>0.22</b>	<b>-0.10</b>	<b>-0.04</b>	<b>-0.06</b>	0.01	<b>0.33</b>	<b>0.26</b>	<b>0.46</b>	<b>0.05</b>	<b>0.12</b>	<b>0.13</b>	<b>0.04</b>	<b>0.16</b>
5. <i>Uncommon</i>	<b>0.08</b>	0.02	<b>0.08</b>	<b>0.15</b>		0.02	<b>0.05</b>	<b>-0.05</b>	<b>-0.05</b>	<b>0.12</b>	<b>0.15</b>	0.01	<b>0.07</b>	<b>0.05</b>	<b>0.04</b>	<b>0.03</b>	0.02	<b>-0.07</b>	<b>0.58</b>
6. <i>Change</i>	<b>-0.06</b>	<b>0.06</b>	<b>-0.11</b>	<b>-0.27</b>	<b>0.03</b>		<b>-0.07</b>	<b>0.03</b>	<b>-0.10</b>	<b>0.16</b>	<b>0.11</b>	<b>-0.08</b>	<b>-0.05</b>	<b>-0.23</b>	0.00	-0.01	<b>0.03</b>	0.02	<b>0.02</b>
7. <i>Size</i>	<b>0.10</b>	<b>0.03</b>	<b>0.11</b>	<b>0.39</b>	0.01	<b>-0.18</b>		<b>-0.20</b>	0.01	<b>-0.17</b>	<b>-0.09</b>	<b>0.20</b>	<b>0.19</b>	<b>0.17</b>	<b>0.05</b>	<b>0.06</b>	<b>0.14</b>	<b>0.17</b>	<b>0.03</b>
8. <i>BTM</i>	<b>-0.08</b>	-0.01	<b>-0.09</b>	<b>-0.11</b>	<b>-0.08</b>	0.02	<b>-0.45</b>		-0.01	<b>0.06</b>	<b>-0.12</b>	0.02	<b>-0.09</b>	<b>-0.20</b>	<b>-0.04</b>	<b>-0.29</b>	<b>-0.08</b>	<b>0.03</b>	<b>-0.10</b>
9. <i>Special</i>	<b>-0.06</b>	0.01	<b>-0.07</b>	<b>-0.15</b>	<b>-0.07</b>	<b>-0.07</b>	<b>-0.03</b>	0.00		<b>-0.12</b>	<b>-0.10</b>	0.01	<b>-0.03</b>	-0.01	<b>-0.03</b>	0.01	<b>-0.02</b>	<b>0.03</b>	<b>-0.04</b>
10. <i>ReturnVol</i>	-0.01	<b>-0.06</b>	<b>0.03</b>	<b>-0.08</b>	<b>0.14</b>	<b>0.16</b>	<b>-0.34</b>	<b>-0.03</b>	<b>-0.13</b>		<b>0.43</b>	<b>-0.12</b>	<b>0.07</b>	<b>-0.16</b>	<b>0.14</b>	<b>0.04</b>	<b>-0.08</b>	<b>-0.22</b>	<b>0.18</b>
11. <i>EarnVol</i>	<b>0.03</b>	<b>-0.08</b>	<b>0.08</b>	-0.01	<b>0.17</b>	<b>0.13</b>	<b>-0.22</b>	<b>-0.18</b>	<b>-0.11</b>	<b>0.54</b>		<b>-0.15</b>	<b>0.09</b>	-0.01	<b>0.16</b>	<b>0.10</b>	<b>-0.06</b>	<b>-0.22</b>	<b>0.21</b>
12. <i>NBUSSEG</i>	0.01	-0.02	<b>0.02</b>	<b>0.38</b>	0.01	<b>-0.09</b>	<b>0.21</b>	<b>0.04</b>	<b>-0.03</b>	<b>-0.11</b>	<b>-0.14</b>		<b>0.23</b>	<b>0.28</b>	<b>-0.06</b>	0.02	<b>0.16</b>	<b>0.22</b>	-0.02
13. <i>NGEOSEG</i>	<b>0.02</b>	<b>-0.07</b>	<b>0.08</b>	<b>0.39</b>	<b>0.10</b>	<b>-0.04</b>	<b>0.24</b>	<b>-0.11</b>	<b>-0.13</b>	<b>0.08</b>	<b>0.12</b>	<b>0.41</b>		<b>0.39</b>	<b>0.08</b>	<b>0.10</b>	<b>0.13</b>	<b>0.05</b>	<b>0.11</b>
14. <i>NumItems</i>	<b>0.10</b>	<b>-0.06</b>	<b>0.16</b>	<b>0.55</b>	<b>0.07</b>	<b>-0.24</b>	<b>0.32</b>	<b>-0.18</b>	<b>-0.10</b>	<b>-0.17</b>	0.00	<b>0.36</b>	<b>0.50</b>		<b>0.06</b>	<b>0.18</b>	<b>0.13</b>	<b>0.09</b>	<b>0.07</b>
15. <i>DE</i>	<b>0.08</b>	0.00	<b>0.09</b>	<b>0.06</b>	<b>0.05</b>	0.00	<b>0.04</b>	<b>-0.08</b>	<b>-0.07</b>	<b>0.17</b>	<b>0.19</b>	<b>-0.06</b>	<b>0.09</b>	<b>0.06</b>		0.02	<b>0.08</b>	<b>-0.16</b>	<b>0.03</b>
16. <i>SEO</i>	<b>0.03</b>	<b>-0.03</b>	<b>0.07</b>	<b>0.15</b>	<b>0.03</b>	<b>-0.04</b>	<b>0.22</b>	<b>-0.24</b>	<b>-0.03</b>	<b>0.05</b>	<b>0.10</b>	<b>0.03</b>	<b>0.12</b>	<b>0.17</b>	0.02		<b>0.09</b>	<b>-0.07</b>	<b>0.08</b>
17. <i>MA</i>	<b>0.04</b>	<b>0.06</b>	0.01	<b>0.15</b>	0.02	<b>0.04</b>	<b>0.24</b>	<b>-0.07</b>	<b>-0.11</b>	<b>-0.06</b>	<b>-0.05</b>	<b>0.17</b>	<b>0.16</b>	<b>0.13</b>	<b>0.08</b>	<b>0.09</b>		<b>0.04</b>	0.02
18. <i>Age</i>	<b>-0.09</b>	<b>-0.03</b>	<b>-0.10</b>	<b>-0.04</b>	<b>-0.07</b>	<b>0.05</b>	<b>0.09</b>	<b>0.07</b>	<b>0.07</b>	<b>-0.23</b>	<b>-0.22</b>	<b>0.15</b>	<b>0.02</b>	<b>0.06</b>	<b>-0.17</b>	<b>-0.05</b>	0.01		<b>-0.17</b>
19. <i>NumFirms</i>	<b>0.12</b>	<b>0.03</b>	<b>0.12</b>	<b>0.15</b>	<b>0.65</b>	<b>0.03</b>	<b>0.04</b>	<b>-0.13</b>	<b>-0.05</b>	<b>0.15</b>	<b>0.19</b>	-0.01	<b>0.11</b>	<b>0.08</b>	0.02	<b>0.08</b>	-0.01	<b>-0.15</b>	

All variables defined in Table A.1  
 Pearson correlations are above the diagonal, Spearman below.  
 Correlations in bold are significant at below the 5% level.

Table 4.12 shows the results of regressing the readability variables on *Uncommon*, *Change* and previously identified determinants. In the multivariate case, *Uncommon* and *Change* are no longer associated with *Fog*. However, *Uncommon* is significantly positively associated with *WPS* and *Length*, consistent with Hypothesis 4. This result is also consistent with recent work that suggests that the percentage of complex words is a misspecified measure of readability. Loughran & McDonald (2011a) find that most complex words in financial reporting are very familiar and readable to the average user, while words per sentence displays convergent validity with other readability measures. Overall, my results are consistent with uncommon information being longer and less readable than common disclosure. Since uncommon information is also informative, these results suggest that information content, as well as managers' incentives to hide information Li (2008), can decrease readability.

Interestingly, *Change* is significantly associated with shorter MD&As and shorter sentences within the MD&A. This result is consistent both with managers taking greater care in discussing changes in the firm and with managers providing less informative (more common) disclosure when a change in the firm occurs.

### **4.3 Summary of Results**

Overall, my results are consistent with my hypotheses. I find that uncommon information both reflects and provides firm-specific information to investors. However, the provision of that information imposes additional processing costs on investors, decreasing the readability of and delaying the market response to the disclosure.

Table 4.12: Regression of Readability Variables on *Uncommon*

	<i>Dependent variable:</i>			
	<i>Fog</i>	<i>%Complex</i>	<i>WPS</i>	<i>Length</i>
<i>Uncommon</i>	0.278 (0.174)	−0.018 (0.208)	0.028** (0.011)	0.332*** (0.111)
<i>Change</i>	−0.208 (0.185)	0.297 (0.292)	−0.032*** (0.008)	−0.361*** (0.062)
<i>Size</i>	0.056*** (0.017)	0.006 (0.019)	0.005*** (0.001)	0.090*** (0.006)
<i>BTM</i>	0.034 (0.032)	−0.018 (0.039)	0.004 (0.003)	0.082*** (0.013)
<i>Special</i>	−0.137 (0.140)	0.137 (0.127)	−0.019 (0.013)	−0.285*** (0.062)
<i>ReturnVol</i>	0.651*** (0.252)	−0.774** (0.381)	0.095*** (0.018)	0.874*** (0.122)
<i>EarnVol</i>	0.995** (0.417)	−0.983** (0.477)	0.135*** (0.030)	0.531*** (0.146)
<i>NBUSSEG</i>	−0.028 (0.037)	−0.024 (0.063)	−0.002 (0.003)	0.110*** (0.013)
<i>NGEOSEG</i>	0.006 (0.031)	−0.033 (0.047)	0.002 (0.003)	0.065*** (0.013)
<i>NumItems</i>	0.250 (0.432)	1.664*** (0.536)	−0.040 (0.032)	0.888*** (0.185)
<i>DE</i>	0.175*** (0.051)	0.024 (0.062)	0.016*** (0.004)	0.040** (0.020)
<i>SEO</i>	−0.064 (0.053)	−0.176** (0.076)	0.001 (0.004)	0.038** (0.019)
<i>MA</i>	0.072* (0.038)	0.140*** (0.046)	0.002 (0.003)	0.054*** (0.010)
<i>Age</i>	−0.090*** (0.033)	−0.035 (0.046)	−0.008*** (0.003)	−0.068*** (0.015)
<i>NumFirms</i>	19.498*** (2.303)	19.441*** (2.936)	3.380*** (0.171)	2.597** (1.053)
Year fixed effects	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes
Observations	7,842	7,842	7,842	7,842
Adjusted R <sup>2</sup>	0.094	0.107	0.103	0.475

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

All variables defined in Table A.1

Standard errors in parentheses are clustered at the firm and year level.

## CHAPTER 5

### FUTURE WORK AND CONCLUSION

In this study, I provide evidence that uncommon information affects investors' valuation models but imposes significant processing costs on investors. In doing so, I contribute to the literature in several ways. First, I describe one mechanism by which firm-specific components of firm value are communicated between managers and investors. Second, I tie information processing costs directly to the information content of disclosure, as opposed to prior studies that focus on managerial manipulation of word and sentence structure. Third, I provide evidence that uncommon information, and thus the firm-specific component of firm value, is incorporated slowly into market prices. Finally, I introduce Latent Semantic Analysis to the accounting literature.

There are several limitations to this study. First, I assume that Latent Semantic Analysis, a statistical procedure from computational linguistics, is able to represent text in a similar way as it is processed by investors. There may be other methods that would provide a better measure of uncommon information. Second, I only examine similarities across firms in the same industry year. This is a naive, and possibly noisy, measure of uncommon information. To the extent that there is a time-series component to uncommon information or a firm's peers are in a different industry classification, my measure may be misspecified. Third, my data collection procedure for firms' MD&As is imperfect in that it omits a significant percentage of firm-years. Since my analysis in Table 3.1 suggests that my sample differs significantly from Compustat, increasing my sample size may improve tests of my hypotheses. Fourth, many of my variables, such as market beta or return synchronicity, are measured over a wide time period. As a result, there may be events of which I am unaware that bias my results. Fifth, since uncommon information is new to the literature, correlated omitted variable bias is a concern. Sixth, my tests depend on how the market interacts with uncommon information and thus may not capture the effect on individuals. Examining the effect of uncommon information on individuals, such as analysts, could be a fruitful area for future

work.

There are many remaining research opportunities regarding uncommon information. First, a significant unanswered question in this study is whether there is an opportunistic component to uncommon information. Prior research has found that managers manipulate perceptions of firm value in a variety of ways, including methods that delay the incorporation of information into stock price, and are particularly apt to do so when it is difficult for investors to verify provided information (Rogers & Stocken, 2005). While my results regarding firm-specific information suggest that at least some portion of uncommon information is informative, uncommon information is hard to verify and thus represents an opportunity for managerial opportunism. Future research can examine the sensitivity of uncommon information to management incentives.

Second, investors vary in their information sets and their sensitivity to information processing costs (Bamber, 1987; Holthausen & Verrecchia, 1990; Morse et al., 1991; Bushee & Noe, 2000). As such, they should respond differently to the information value and processing costs associated with uncommon information. Future research can examine whether investors who are more sophisticated, have greater firm-specific knowledge or lower information processing costs respond more quickly and more accurately to uncommon information than other investors.

Third, uncommon information provides a rich setting to examine investor opinion formation. Future research can explore how uncommon information interacts with other characteristics of the firm's disclosures and information environment. For instance, while this study examines changes in the MD&A separately from uncommon information, investors should react differently to changes in the firm that increase or decrease uncommon information. Understanding that response can help researchers to understand how investors respond to changes in a firm's economic environment more broadly.

Fourth and finally, this study uses Latent Semantic Analysis (LSA) to identify sources of variation in the text of financial reports. Whereas prior research uses simple word counts to examine qualitative disclosures, LSA allows the researcher to observe the underlying meaning of text. This ability is potentially very powerful for identifying how managers talk about firm value. Future

studies can apply LSA to a variety of settings and research questions.

## **APPENDICES**

## A Variables

Table A.1: Variable List

Name	Description
<b>Dependent Variables:</b>	
$CAR_{-1,1}$	The absolute cumulative abnormal return during the 10-K filing window (the day before to the day after). Abnormal returns are calculated using a market model estimated for the 255 days ending 63 days prior to the 10-K filing date.
$CAR_{2,60}$	The absolute cumulative abnormal return during the post-filing window (2 to 60 days after the filing).
$\%Complex$	The percentage of total words in the MD&A that are three syllables or more in length.
$Fog$	Words per sentence ( $WPS$ ) plus percent complex words ( $\%Complex$ ), multiplied by 0.4.
$Length$	The log of the total number of words in the firm's MD&A.
$PCAR_{-1,1}$	The absolute cumulative abnormal return during the 3-day window 60 days prior to the earnings announcement (the day before to the day after). Abnormal returns are calculated using a market model estimated for the 255 days ending 63 days prior to the selected date.
$PCAR_{2,60}$	The absolute cumulative abnormal return during the 58 days leading up to the earnings announcement window (2 to 60 days prior to the announcement).
Continued on next page	



Table A.1 (cont'd)

Name	Description
<i>ReturnSync</i>	Following Piotroski & Roulstone (2004), $\log \frac{R^2}{1 - R^2}$ , where $R^2$ is the the R-squared from a regression of weekly firm returns on concurrent and lagged value-weighted market and industry returns over the fiscal year.
$\Delta ReturnSync$	<i>ReturnSync</i> minus its own value in the previous year.
<i>Uncommon</i>	The mean cosine dissimilarity between a firm's MD&A and the MD&As of the four closest firms in its industry year. Cosine dissimilarity is measured in a vector space model after performing Latent Semantic Analysis on all reports in the industry year and selecting the top $N$ dimensions, where $N$ is equal to 10% of the total number of firms in that industry year.
$\Delta Uncommon$	<i>Uncommon</i> minus its own value in the previous year.
<i>WPS</i>	The average number of words per sentence in the MD&A.
<b>Independent Variables:</b>	
<i>Age</i>	The number of years since the firm first appeared in CRSP.
<i>AnaFollow</i>	The number of analysts in the IBES database issuing at least one forecast for the firm during the fiscal year.
<i>Beta</i>	Market beta calculated over the 60 months prior to the 10-K filing.
<i>BTM</i>	The book value per share divided by the price per share, calculated at the end of the fiscal year.
Continued on next page	

Table A.1 (cont'd)

Name	Description
<i>Change</i>	The MD&A modification score from Brown & Tucker (2011), obtained from Brown (2013).
<i>DE</i>	An indicator variable equal to one if the firm is incorporated in Delaware.
<i>RevConc</i>	The log of the Herfindahl index, calculated by firm as the proportion of sales per segment, squared and summed across the segments.
<i>EA<sub>-1,1</sub></i>	The absolute cumulative abnormal return in the earnings announcement window (the day before to the day after). Abnormal returns are calculated using a market model estimated for the 255 days ending 63 days prior to the earnings announcement date.
<i>EarnSync</i>	Following Piotroski & Roulstone (2004), $\log \frac{R^2}{1 - R^2}$ , where $R^2$ is the the R-squared from a regression of quarterly return on assets on concurrent and lagged value-weighted industry return on assets over the last 12 quarters.
<i>EarnVol</i>	The standard deviation of earnings over the last five years.
<i>FileLate</i>	An indicator variable equal to one if the 10-K was filed 90 or more days after the end of the fiscal period.
<i>IndConc</i>	The log of the Herfindahl index, calculated by industry as the proportion of industry sales per firm, squared and summed across the industry.
<i>Instown</i>	The percentage of a firm's equity held by institutional owners, as reported in the Thomson-Reuters database .

Continued on next page

Table A.1 (cont'd)

Name	Description
<i>Leverage</i>	Total debt divided by total assets, calculated at the end of the fiscal year.
<i>MA</i>	An indicator variable equal to one if the firm had a merger or acquisition during the year.
<i>NBUSSEG</i>	The log of the firm's number of operating segments.
<i>NGEOSEG</i>	The log of the firm's number of geographic segments.
<i>NumItems</i>	The log of the number of non-missing line items in Compustat.
<i>NumFirms</i>	The number of firms in the industry-year.
<i>Price</i>	The log of the stock price at the end of the fiscal year.
<i>PVolume<sub>-1,1</sub></i>	The log of the proportion of total shares traded over the 3-day window 60 days prior to the earnings announcement
<i>PVolume<sub>2,60</sub></i>	The log of the proportion of total shares traded over the 58 days leading up to the earnings announcement window (2 to 60 days prior to the announcement).
<i>Reg</i>	An indicator variable set equal to one if the firm is in a regulated industry (two-digit SIC code is 49 or 62).
<i>ReturnVol</i>	The standard deviation of monthly returns over the fiscal year.
<i>SEO</i>	An indicator variable equal to one if the firm has had a seasoned equity offering during the year.
<i>Size</i>	The log of the market value of equity at the end of the fiscal year.
<i>Special</i>	The amount of special items scaled by total assets.
Continued on next page	

Table A.1 (cont'd)

Name	Description
<i>UE</i>	IBES reported actual earnings for the year minus the mean estimate from the last consensus period prior to the earnings announcement, scaled by the price at the end of the fiscal year.
<i>Volume<sub>-1,1</sub></i>	The log of the proportion of total shares traded over the three days surrounding the 10-K filing date.
<i>Volume<sub>2,60</sub></i>	The log of the proportion of total shares traded over the 58 days surrounding the 10-K filing date.
<i>VolumeEA<sub>-1,1</sub></i>	The log of the proportion of total shares traded over the three days surrounding the earnings announcement date.

## **B Excerpts from Most and Least Common MD&As**

### **NCI Information Systems, Inc.**

NCI has the most common MD&A in the Software Industry in 2007. In the business description portion of their MD&A, they say:

We are a provider of information technology (IT) and professional services and solutions to federal government agencies. Our technology and industry expertise enables us to provide a full spectrum of services and solutions that assist our clients in achieving their program goals. We deliver a wide range of complex services and solutions by leveraging our skills across seven core service offerings:

- network engineering;
- information assurance;
- systems engineering and integration;
- enterprise systems management;
- engineering and logistics;
- medical transformation/health IT; and
- distance learning and training.

We generate substantially all of our revenue from federal government contracts. We report operating results and financial data as one operating segment. Revenue from our contracts and task orders is generally linked to trends in federal government spending by defense, intelligence and federal civilian agencies.

We believe that our contract base is well diversified. As of December 31, 2007, we had approximately 200 active contracts and 600 task orders. As of December 31, 2007, our total contract backlog was approximately \$756 million of which approximately \$189 million was funded. We define backlog as our estimate of the remaining

future revenue from existing signed contracts over the remaining base contract performance period and from the option periods of those contracts, assuming the exercise of all related options. Our backlog does not include any estimate of future potential delivery orders that might be awarded under our GWAC or other multiple award contract vehicles. We define funded backlog as the portion of backlog for which funding currently is appropriated and obligated to us under a contract or other authorization for payment signed by an authorized purchasing agency, less the amount of revenue we have previously recognized. Our funded backlog does not represent the full potential value of our contracts, as Congress often appropriates funds for a particular program or agency on a quarterly or yearly basis, even though the contract may provide for the provision of services over a number of years. We define unfunded backlog as the total backlog less the funded backlog. Unfunded backlog includes values for contract options that have been priced but not yet funded.

### **Borland Software Corporation**

Borland has the least common MD&A in the Software Industry in 2007. In the business description portion of their MD&A, they say:

Borland is a leading vendor of Open Application Lifecycle Management solutions, or ALM, which represents the segment of the ALM market in which vendors' solutions are flexible enough to support a customer's specific processes, tools and platforms. Open ALM is a new, customer-centric approach to helping IT organizations transform software delivery into a managed, efficient and predictable business process. We offer a combination of software products as well as consulting and education services to help our customers better manage the growing complexity of software development. Our goal is to provide customers with a foundation which will allow them to consistently deliver software on-time, on-budget and with increased business value.

Borland's solutions address five critical ALM processes: Project & Portfolio Management, Requirements Definition & Management, Lifecycle Quality Management, Model Driven Development and Software Change Management. Each solution can play an important role in helping enterprises manage the complexity of software development and delivery, by providing business, development and operational teams with increased visibility and control over all phases of the software delivery lifecycle. We believe this is especially crucial for large enterprises working within heterogeneous and distributed environments.

We have been evolving our business and strategy in recent years in response to the many changes occurring in the software industry and, specifically in our market. In a March 2005 study, IDC forecasted the ALM market to grow to \$3.3 billion in 2009, achieving a 9.2% compound annual growth rate between 2004 and 2009. In order to capitalize on the ALM market growth, we have made changes to our overall product portfolio, our worldwide services organization, our R&D investments, as well as our global sales and marketing models to reflect our Open ALM vision and product strategy. As part of this transformation, we have shifted our focus from selling individual stand-alone products to selling more multi-product, enterprise-class solutions.

Effective January 1, 2007, consistent with how we manage our business, we changed from reporting one segment to reporting two segments: Enterprise and CodeGear. A summary of the types of products and services provided by the Enterprise and CodeGear segments is provided below.

Enterprise. Our Enterprise segment focuses on Open Application Lifecycle Management solutions, or ALM, which includes a combination of software products as well as consulting and education services to help our customers better manage their software development projects. Our ALM portfolio includes products and services for project and portfolio management, requirements definition and management, lifecycle quality management, software configuration and change management and modeling. The

Enterprise segment also includes our Deployment Product Group, or DPG, products.

CodeGear. Our CodeGear segment focuses on developing tools for individual developers and currently offers a number of Integrated Developer Environment, or IDE, and database products for Java, .NET and Windows development. CodeGear products include Delphi, Delphi for PHP, C++Builder, C#Builder, JBuilder, TurboVCL and Interbase. CodeGear also provides worldwide developer support and education services.



## **BIBLIOGRAPHY**

## BIBLIOGRAPHY

- Aboody, David, Reuven Lehavay & Brett Trueman. 2010. Limited attention and the earnings announcement returns of past stock market winners. *Review of Accounting Studies* 15(2). 317–344. doi:10.1007/s11142-009-9104-9.
- Arya, Anil & Brian Mittendorf. 2005. Using disclosure to influence herd behavior and alter competition. *Journal of Accounting and Economics* 40(1–3). 231–246. doi:10.1016/j.jacceco.2005.07.001.
- Baddeley, Alan D., Neil Thomson & Mary Buchanan. 1975. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior* 14(6). 575–589. doi:10.1016/S0022-5371(75)80045-4.
- Ball, Christopher, Gerard Hoberg & Vojislav Maksimovic. 2013. Disclosure informativeness and the tradeoff hypothesis: A text-based analysis. SSRN Scholarly Paper ID 2260371 Social Science Research Network Rochester, NY.
- Ball, Ray & Philip Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6(2). 159–178. doi:10.2307/2490232.
- Ball, Ray & Lakshmanan Shivakumar. 2008. How much new information is there in earnings? *Journal of Accounting Research* 46(5). 975–1016. doi:10.1111/j.1475-679X.2008.00299.x.
- Bamber, Linda Smith. 1987. Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review* 62(3). 510–532. doi:10.2307/247574.
- Bamber, Linda Smith, Orie E. Barron & Douglas E. Stevens. 2011. Trading volume around earnings announcements and other financial reports: Theory, research design, empirical evidence, and directions for future research. *Contemporary Accounting Research* 28(2). 431–471. doi:10.1111/j.1911-3846.2010.01061.x.
- Barron, Orie E., Charles O. Kile & Terrence B. O’Keefe. 1999. MD&A quality as measured by the SEC and analysts’ earnings forecasts. *Contemporary Accounting Research* 16(1). 75–109. doi:10.1111/j.1911-3846.1999.tb00575.x.
- Barry, Christopher B. & Stephen J. Brown. 1985. Differential information and security market equilibrium. *Journal of Financial & Quantitative Analysis* 20(4). 407–422.
- Becker, Curtis A. 1976. Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 2(4). 556–566. doi:http://dx.doi.org.proxy1.cl.msu.edu/10.1037/0096-1523.2.4.556.
- Becker, Curtis A. 1980. Semantic context effects in visual word recognition: An analysis of semantic strategies. *Memory & Cognition* 8(6). 493–512. doi:10.3758/BF03213769.
- Bellman, Richard Ernest. 2003. *Dynamic programming*. Courier Dover Publications.

- Bhojraj, Sanjeev & Charles M. C. Lee. 2002. Who is my peer? a valuation-based approach to the selection of comparable firms. *Journal of Accounting Research* 40(2). 407–439. doi: 10.1111/1475-679X.00054.
- Blei, David M., Andrew Y. Ng & Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3. 993–1022.
- Bloomfield, Robert. 2008. Discussion of "Annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics* 45(2-3). 248–252. doi:10.1016/j.jacceco.2008.04.002.
- Bloomfield, Robert J. 2002. The "Incomplete Revelation Hypothesis" and financial reporting. *Accounting Horizons* 16(3). 233–243.
- Bloomfield, Robert J. 2012. A pragmatic approach to more efficient corporate disclosure. *Accounting Horizons* 26(2). 357–370.
- Bloomfield, Robert J., Frank D. Hodge, Patrick E. Hopkins & Kristina M. Rennekamp. 2010. Does enhanced disaggregation and cohesive classification of financial information help credit analysts identify firms' operating structures? *SSRN eLibrary*.
- Branch, William A. & George W. Evans. 2006. Intrinsic heterogeneity in expectation formation. *Journal of Economic Theory* 127(1). 264–295. doi:10.1016/j.jet.2004.11.005.
- Branch, William A. & George W. Evans. 2010. Asset return dynamics and learning. *Review of Financial Studies* 23(4). 1651–1680. doi:10.1093/rfs/hhp112.
- Brown, Stephen V. 2013. Data - Stephen V. Brown.
- Brown, Stephen V. & Jennifer Wu Tucker. 2011. Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research* 49(2). 309–346. doi:10.1111/j.1475-679X.2010.00396.x.
- Bushee, Brian J. & Christopher F. Noe. 2000. Corporate disclosure practices, institutional investors, and stock return volatility. *Journal of Accounting Research* 38. 171–202. doi: 10.2307/2672914.
- Clarkson, Pete, Jose Guedes & Rex Thompson. 1996. On the diversification, observability, and measurement of estimation risk. *Journal of Financial & Quantitative Analysis* 31(1). 69–84.
- Clarkson, Peter M., Jennifer L. Kao & Gordon D. Richardson. 1999. Evidence that management discussion and analysis (MD&A) is a part of a firm's overall disclosure package. *Contemporary Accounting Research* 16(1). 111–134. doi:10.1111/j.1911-3846.1999.tb00576.x.
- Coles, Jeffrey L., Uri Loewenstein & Jose Suay. 1995. On equilibrium pricing under parameter uncertainty. *The Journal of Financial and Quantitative Analysis* 30(3). 347–364. doi:10.2307/2331345.
- Core, John E., Wayne R. Guay & Rodrigo Verdi. 2008. Is accruals quality a priced risk factor? *Journal of Accounting and Economics* 46(1). 2–22. doi:10.1016/j.jacceco.2007.08.001.

- Davis, Angela K. & Isho Tama-Sweet. 2012. Managers' use of language across alternative disclosure outlets: Earnings press releases versus MD&A. *Contemporary Accounting Research* 29(3). 804–837. doi:10.1111/j.1911-3846.2011.01125.x.
- DeFranco, Gus, S.P. Kothari & Rodrigo S. Verdi. 2011. The benefits of financial statement comparability. *Journal of Accounting Research* 49(4). 895–931. doi:10.1111/j.1475-679X.2011.00415.x.
- Dellavigna, Stefano & Joshua M. Pollet. 2009. Investor inattention and friday earnings announcements. *The Journal of Finance* 64(2). 709–749. doi:10.1111/j.1540-6261.2009.01447.x.
- Drake, Michael S., Darren T. Roulstone & Jacob R. Thornock. 2012. The informativeness of stale financial disclosures. SSRN Scholarly Paper ID 2083812 Social Science Research Network Rochester, NY.
- Durnev, Artyom, Randall Morck, Bernard Yeung & Paul Zarowin. 2003. Does greater firm-specific return variation mean more or less informed stock pricing? *Journal of Accounting Research* 41(5). 797–836.
- Financial Accounting Standards Board. 2010. Conceptual framework for financial reporting. *Statement of Financial Accounting Concepts* No. 8.
- Forster, Kenneth I. & Susan M. Chambers. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12(6). 627–635. doi:10.1016/S0022-5371(73)80042-8.
- Forster, K.I. 1981. Frequency blocking and lexical access: One mental lexicon or two? *Journal of Verbal Learning and Verbal Behavior* 20(2). 190–203. doi:10.1016/S0022-5371(81)90373-X.
- Francis, Jennifer, Ryan LaFond, Per Olsson & Katherine Schipper. 2005. The market pricing of accruals quality. *Journal of Accounting and Economics* 39(2). 295–327. doi:10.1016/j.jacceco.2004.06.003.
- Francis, Jennifer, Dhananjay Nanda & Per Olsson. 2008. Voluntary disclosure, earnings quality, and cost of capital. *Journal of Accounting Research* 46(1). 53–99. doi:10.1111/j.1475-679X.2008.00267.x.
- Glanzer, Murray & S.L. Ehrenreich. 1979. Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior* 18(4). 381–398. doi:10.1016/S0022-5371(79)90210-X.
- Gong, Guojin, Laura Yue Li & Ling Zhou. 2013. Earnings non-synchronicity and voluntary disclosure. *Contemporary Accounting Research* 1560–1589. doi:10.1111/1911-3846.12007.
- Goodman, George O., James L. McClelland & Raymond W. Gibbs. 1981. The role of syntactic context in word recognition. *Memory & Cognition* 9(6). 580–586. doi:10.3758/BF03202352.
- Grossman, Sanford J. & Joseph E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *The American Economic Review* 70(3). 393–408.

- Halgren, Eric, Rupali P. Dhond, Natalie Christensen, Cyma Van Petten, Ksenija Marinkovic, Jeffrey D. Lewine & Anders M. Dale. 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage* 17(3). 1101–1116. doi:10.1006/nimg.2002.1268.
- Hall, John F. 1954. Learning as a function of word-frequency. *The American Journal of Psychology* 67(1). 138–140. doi:10.2307/1418080.
- Handa, Puneet & Scott C. Linn. 1993. Arbitrage pricing with estimation risk. *Journal of Financial & Quantitative Analysis* 28(1). 81–100.
- Healy, Paul M & Krishna G Palepu. 2001. Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31(1–3). 405–440. doi:10.1016/S0165-4101(01)00018-0.
- Heneman, Robert L. 1986. The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology* 39(4). 811–826. doi:10.1111/j.1744-6570.1986.tb00596.x.
- Hirshleifer, David, Sonya Seongyeon Lim & Siew Hong Teoh. 2009. Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance* 64(5). 2289–2325. doi:10.1111/j.1540-6261.2009.01501.x.
- Hodder, Leslie, Patrick E. Hopkins & David A. Wood. 2008. The effects of financial statement and informational complexity on analysts, cash flow forecasts. *Accounting Review* 83(4). 915–956.
- Holthausen, Robert W. & Robert E. Verrecchia. 1990. The effect of informedness and consensus on price and volume behavior. *The Accounting Review* 65(1). 191–208. doi:10.2307/247883.
- Hong, Harrison, Jeremy C. Stein & Jialin Yu. 2007. Simple forecasts and paradigm shifts. *The Journal of Finance* 62(3). 1207–1242. doi:10.1111/j.1540-6261.2007.01234.x.
- Hopkins, Patrick E. 1996. The effect of financial statement classification of hybrid financial instruments on financial analysts' stock price judgments. *Journal of Accounting Research* 34. 33–50. doi:10.2307/2491424.
- Huang, Allen, Reuven Lehav, Amy Y. Zang & Rong Zheng. 2014. A thematic analysis of analyst information discovery and information interpretation roles. SSRN Scholarly Paper ID 2409482 Social Science Research Network Rochester, NY.
- Hulme, Charles, Steven Roodenrys, Richard Schweickert, D. A. Sarah Martin & George Stuart. 1997. Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(5). 1217–1232. doi:10.1037/0278-7393.23.5.1217.
- Karelaia, Natalia & Robin M. Hogarth. 2008. "Determinants of linear judgment: A meta-analysis of lens model studies": Correction. *Psychological Bulletin* 134(5). 741. doi:http://dx.doi.org.proxy1.cl.msu.edu/10.1037/a0013550.

- Khan, Arshad M. & V. Manopichetwattana. 1989. Innovative and noninnovative small firms: Types and characteristics. *Management Science* 35(5). 597–606.
- Kim, Oliver & Robert E. Verrecchia. 1991. Trading volume and price reactions to public announcements. *Journal of Accounting Research* 29(2). 302–321. doi:10.2307/2491051.
- Kim, Oliver & Robert E. Verrecchia. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17(1–2). 41–67. doi:10.1016/0165-4101(94)90004-3.
- Kintsch, Walter. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95(2). 163–182. doi:http://dx.doi.org.proxy1.cl.msu.edu/10.1037/0033-295X.95.2.163.
- Lambert, Richard, Christian Leuz & Robert E. Verrecchia. 2007. Accounting information, disclosure, and the cost of capital. *Journal of Accounting Research* 45(2). 385–420. doi:10.1111/j.1475-679X.2007.00238.x.
- Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 211–240. doi:10.1037/0033-295X.104.2.211.
- Landauer, Thomas K, Peter W. Foltz & Darrell Laham. 1998. An introduction to latent semantic analysis. *Discourse Processes* 25(2-3). 259–284. doi:10.1080/01638539809545028.
- Lee, Yen-Jung. 2012. The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research* 29(4). 1137–1170. doi:10.1111/j.1911-3846.2011.01152.x.
- Lee, Yen Jung, Kathy R. Petroni & Min Shen. 2006. Cherry picking, disclosure quality, and comprehensive income reporting choices: The case of property-liability insurers. *Contemporary Accounting Research* 23(3). 655–700. doi:10.1506/5QB8-PBQY-Y86L-DRYL.
- Lehavy, Reuven, Feng Li & Kenneth Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *Accounting Review* 86(3). 1087–1115.
- Lewellen, Jonathan & Jay Shanken. 2002. Learning, asset-pricing tests, and market efficiency. *The Journal of Finance* 57(3). 1113–1145. doi:10.1111/1540-6261.00456.
- Li, Feng. 2006. Do stock market investors understand the risk sentiment of corporate annual reports? *SSRN eLibrary*.
- Li, Feng. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45(2–3). 221–247. doi:10.1016/j.jacceco.2008.02.003.
- Li, Feng. 2010. The information content of forward-looking statements in corporate filings: A naive bayesian machine learning approach. *Journal of Accounting Research* 48(5). 1049–1102. doi:10.1111/j.1475-679X.2010.00382.x.

- Libby, Theresa, Steven E. Salterio & Alan Webb. 2004. The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review* 79(4). 1075–1094.
- Lipe, Marlys Gascho & Steven E. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75(3). 283–298.
- Loughran, Tim & Bill McDonald. 2011a. Measuring readability in financial disclosures. SSRN Scholarly Paper ID 1920411 Social Science Research Network Rochester, NY.
- Loughran, Tim & Bill McDonald. 2011b. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance* 66(1). 35–65. doi:10.1111/j.1540-6261.2010.01625.x.
- Loughran, Tim & Bill McDonald. 2013. IPO first-day returns, offer price revisions, volatility, and form s-1 language. *Journal of Financial Economics* 109(2). 307–326. doi:10.1016/j.jfineco.2013.02.017.
- Miller, Brian P. 2010. The effects of reporting complexity on small and large investor trading. *Accounting Review* 85(6). 2107–2143. doi:10.2308/accr.00000001.
- Morse, Dale, Jens Stephan & Earl K. Stice. 1991. Earnings announcements and the convergence (or divergence) of beliefs. *The Accounting Review* 66(2). 376–388.
- Ohlson, James A. 1995. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research* 11(2). 661–687.
- Payne, John W., James R. Bettman & Eric J. Johnson. 1993. *The adaptive decision maker*. Cambridge University Press.
- Peng, Lin & Wei Xiong. 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80(3). 563–602. doi:10.1016/j.jfineco.2005.05.003.
- Pierce, John R. 1980. *An introduction to information theory: Symbols, signals and noise*. New York, NY: Dover Publications, Inc. 2nd edn.
- Piotroski, Joseph D. & Darren T. Roulstone. 2004. The influence of analysts, institutional investors, and insiders on the incorporation of market, industry, and firm-specific information into stock prices. *Accounting Review* 79(4). 1119–1151.
- Postman, Leo. 1970. Effects of word frequency on acquisition and retention under conditions of free-recall learning. *Quarterly Journal of Experimental Psychology* 22(2). 185–195. doi:10.1080/00335557043000113.
- Rayner, Keith & Susan A. Duffy. 1986. Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14(3). 191–201. doi:10.3758/BF03197692.
- Rennekamp, Kristina. 2012a. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research* 50(5). 1319–1354. doi:10.1111/j.1475-679X.2012.00460.x.

- Rennekamp, Kristina M. 2012b. The influence of performance and reporting goals on managers' choice of reporting complexity in disclosures. *SSRN eLibrary*.
- Richardson, Scott A., Richard G. Sloan, Mark T. Soliman & Irem Tuna. 2005. Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39(3). 437–485. doi:10.1016/j.jacceco.2005.04.005.
- Rogers, Jonathan L. & Phillip C. Stocken. 2005. Credibility of management forecasts. *Accounting Review* 80(4). 1233–1260.
- Scarborough, Don L., Charles Cortese & Hollis S. Scarborough. 1977. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance* 3(1). 1–17. doi:10.1037/0096-1523.3.1.1.
- Schvaneveldt, Roger W., David E. Meyer & Curtis A. Becker. 1976. Lexical ambiguity, semantic context, and visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance* 2(2). 243–256. doi:10.1037/0096-1523.2.2.243.
- Securities and Exchange Commission. 2003. Interpretation: Commission guidance regarding management's discussion and analysis of financial condition and results of operations.
- Shannon, Claude & Warren Weaver. 1949. The mathematical theory of communication. *University of Illinois Press* 19(7). 1.
- Shields, Michael D. 1980. Some effects on information load on search patterns used to analyze performance reports. *Accounting, Organizations and Society* 5(4). 429–442. doi:10.1016/0361-3682(80)90041-0.
- Simon, Herbert A. 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69(1). 99–118.
- Sloan, Richard G. 1996. Do stock prices fully reflect information in accruals and cash flows about future earnings? *The Accounting Review* 71(3). 289–315. doi:10.2307/248290.
- Slovic, Paul & Douglas MacPhillamy. 1974. Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance* 11(2). 172–194. doi:10.1016/0030-5073(74)90013-0.
- Sowa, John F. 1984. *Conceptual structures: information processing in mind and machine* The Systems programming series. Reading, Mass: Addison-Wesley.
- Tan, Hun-Tong, Elaine Ying Wang & Bo Zhou. 2013. When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research* 52(1). 273–302. doi:10.1111/1475-679X.12039.
- Tetlock, Paul C. 2011. All the news that's fit to reprint: Do investors react to stale information? *Review of Financial Studies* 24(5). 1481–1512. doi:10.1093/rfs/hhq141.



- Tweedy, James R., Robert H. Lapinski & Roger W. Schvaneveldt. 1977. Semantic-context effects on word recognition: Influence of varying the proportion of items presented in an appropriate context. *Memory & Cognition* 5(1). 84–89. doi:10.3758/BF03209197.
- Verrecchia, Robert E. 1983. Discretionary disclosure. *Journal of Accounting and Economics* 5. 179–194. doi:10.1016/0165-4101(83)90011-3.
- Xie, Hong. 2001. The mispricing of abnormal accruals. *Accounting Review* 76(3). 357.
- Zhang, Shi & Arthur B. Markman. 2001. Processing product unique features: Alignability and involvement in preference construction. *Journal of Consumer Psychology* 11(1). 13–27. doi: 10.1207/S15327663JCP1101\_2.
- Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, Mass: Addison-Wesley Press.