

Why Are 10-K Filings so Long?

April 2015

Rick Cazier*
University of Michigan

Ray Pfeiffer**
Texas Christian University

Abstract. This study provides evidence on relative magnitudes of factors that appear to drive 10-K length. We employ statistical analysis and text analysis software to partition 10-K length into the portions explained by each of three fundamental determinants: 1) firms' operating complexity, 2) disclosure redundancy, and 3) residual disclosure. Our primary analyses shed light on the relative magnitudes of each of these components and on the extent to which each varies across firms. Disclosure redundancy and operating complexity explain roughly equal amounts of variation in 10-K length within our sample. However, 10-K length unexplained by redundancy or firms' operating complexity (residual disclosure) accounts for the largest degree of variation in 10-K length. Our results are consistent with the notion that a substantial amount of disclosure volume contained in 10-K reports is attributable to managerial discretion in how firms respond to mandatory disclosure requirements. Our study expands prior literature that has focused largely on the consequences of 10-K length and provides important insights for policy makers and regulators seeking to improve disclosure requirements.

1. INTRODUCTION

An ongoing discussion among regulators and practitioners centers on what can and should be done to address the increasing volume of firms' financial disclosures. Little disagreement exists among observers that the volume of financial disclosure has expanded considerably. In their 2011 analysis of 25 FORTUNE 500 companies' annual reports, the Financial Executives Research Foundation and KPMG document a 16 percent increase in 10-K length and a 28 percent increase in the length of the notes to financial statements over the period 2004-2010 (KPMG 2011). In this paper, we provide empirical evidence regarding the determinants of 10-K disclosure volume to help inform the discussion of how regulators might address this issue.

Various financial reporting stakeholders increasingly decry the current state of mandatory financial disclosure on the part of business enterprises.¹ In its July 12, 2012 "FASB in Focus" publication, the FASB summarizes the concerns of stakeholders as follows:

*"Many stakeholders have expressed concerns about the relevance and **sheer volume** of information in notes to financial statements, and that some information is either missing or difficult to find." (Emphasis added.)*

In a 2013 speech to the NACD Board Leadership Conference, SEC Chairman Mary Jo White, in calling for a "meaningful review of our disclosure requirements," similarly pointed to disclosure volume and its potentially harmful effects on investors:

"I am raising the question here and internally at the SEC as to whether investors need and are optimally served by the detailed and lengthy disclosures about all of

¹Examples noted by the FASB in their July 12 Discussion Paper, "Disclosure Framework," paragraph 1.38, include the 2001 report by the FASB's Business Reporting Research Project, http://www.fasb.org/cs/ContentServer?c=Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1175801856648, the Investors Technical Advisory Committee's 2007 Agenda Request, http://www.fasb.org/cs/ContentServer?c=Document_C&pagename=FASB%2FDocument_C%2FDocumentPage&cid=1175801635556, the SEC's Advisory Committee on Improvements to Financial Reporting's 2008 "Final Report," the Financial Executives Research Foundation's 2011 report, "Disclosure Overload and Complexity: Hidden in Plain Sight," <http://www.kpmginstitutes.com/financial-reporting-network/insights/2011/pdf/kpmg-fei-dis-rep-disclose-overload-complexity.pdf>, as well as others.

the topics that companies currently provide in the reports they are required to prepare and file with us. When disclosure gets to be “too much” or strays from its core purpose, it could lead to what some have called “information overload” – a phenomenon in which ever-increasing amounts of disclosure make it difficult for an investor to wade through the volume of information she receives to ferret out the information that is most relevant.”

Chairman White goes on to ask,

Are our rules the sole or primary cause of potential disclosure overload or do other sources contribute to it? Or said another way, are changes to our disclosure requirements the only way to improve the quality of disclosure? We should consider all sources that may be contributing to the length and complexity of disclosure.

Questions such as those posed by Chairman White and those that underlie the FASB’s Disclosure Framework project are vitally important in order to determine what is to be done to address the issue of disclosure volume. Although prior research documents a detrimental effect of 10-K length on investors’ ability to process relevant information (e.g., Miller 2010; You and Zhang 2009), little systematic evidence exists on the factors that contribute to 10-K length. Our study focuses squarely on providing such evidence so that policy makers can better assess the likely impact various regulatory initiatives may have on 10-K length.

We begin by framing the factors that contribute to 10-K length as:

- (1) The complexity of the firm’s business transactions, which we label ‘operating complexity’²;
- (2) Redundancy in SEC and GAAP disclosure requirements that may result in the same information being repeated in various sections of the 10-K; and
- (3) Firms’ provision of “residual” disclosure, which is unassociated with the first two factors above.³

² In this study, the term “operating complexity” refers to all business transactions including those classified as operating, investing, and financing activities in the financial statements.

³ Increases in disclosure volume over time can also be affected by changes in disclosure regulation. However, the analysis in our study is focused upon comparisons of disclosure volume across firms. In the cross-section, we interpret the relationship between operating complexity, redundancy, and disclosure volume to be conditioned on existing disclosure regulation. One extension of our work would be to consider the effects of changes in disclosure regulation over time.

We examine a sample of 10-Ks representing 620 firm-years over the ten-year period 2003 through 2012 and find firms' underlying operating complexity, disclosure redundancy, and residual disclosure each contribute significantly to variation in 10-K length. Our analysis suggests disclosure redundancy and operating complexity explain a roughly equal amount of variation in 10-K length within our sample. Specifically, they each account for a difference of approximately 15,000 words in length between the median firm in the lowest quintile of 10-K length and the median firm in the highest quintile of 10-K length in our sample. However, 10-K length unexplained by redundancy or firms' operating complexity (i.e., 'residual' disclosure) appears responsible for a slightly larger degree of variation in 10-K length across the top and bottom quintiles (about 18,000 words) than do either operating complexity or redundancy.

We next provide further evidence on the extent of redundancy in 10-K filings. Given that the SEC permits a significant amount of disclosure to be incorporated by reference to the notes to the financial statements, providing redundant information is at least partially a discretionary reporting choice by management. We test which firm characteristics are associated with the provision of more redundancy in 10-K disclosures. We find redundant disclosure has a strong positive association with firm size and leverage. Redundancy is decreasing in R&D intensity but increasing in special items. We find some evidence that total redundant disclosure has increased over the sample period. However, increases in total 10-K disclosure have outpaced increases in redundant content, such that redundant disclosures have declined as a proportion of total 10-K disclosure.

Finally, we seek to determine which components of 10-K length—operating complexity, redundancy, or residual disclosure—are responsible for investors' information processing problems documented by prior research (e.g., You and Zhang 2009; Miller 2010; Lawrence

2013). Specifically, we look for evidence of return drift following the filing of longer 10-K reports (You and Zhang 2009). Although we are unable to replicate You and Zhang's (2009) overall result that longer 10-Ks are associated with subsequent return drift, we do find some evidence of return drift after partitioning 10-K length into its three components. We find that 10-K length attributable to residual disclosure is associated with return drift six months following the 10-K filing date. However, we find no evidence that 10-K length attributable to operating complexity or disclosure redundancy is associated with returns subsequent to the 10-K filing. This suggests that the residual disclosure component of 10-K length can indeed impair market participants' pricing of securities.

This paper is among the first to provide descriptive evidence on the determinants of 10-K length. This is important because although prior research documents that 10-K length reflects complexity (e.g., Lehavy et al 2011; You and Zhang 2009; Miller 2010), no prior evidence exists regarding whether that complexity is primarily a result of the firm's underlying operating complexity or simply its disclosure complexity. Our results indicate that cross-sectional variation in 10-K length is likely more determined by managerial discretion than by firms' operating characteristics. In addition, we are among the first to provide meaningful descriptive evidence on the impact of disclosure redundancy on total 10-K length.

We believe these results provide potentially useful insights on an issue of current relevance for financial reporting policy. Enhanced understanding of the relative importance of various drivers of length will allow the SEC and FASB to take a more targeted and likely more effective approach to improving existing disclosure requirements. For instance, our evidence on the extent to which disclosure volume varies in the cross-section after holding firm characteristics constant highlights the concern that this disclosure may be excessive in many

cases. Standard setters may want to investigate whether the significant variation in disclosure practices in the 10-K reflects a need for greater standardization in disclosure requirements.

In addition, our evidence on the extent to which redundancy impacts overall length provides further impetus for standard setters to respond to calls for streamlined disclosure. Market participants have proposed several ways in which standard setters may be able to reduce redundancy, such as providing more concrete guidance on the extent to which cross-referencing between the notes and other 10-K sections is permissible. In addition, our research supports calls for greater coordination between the FASB and the SEC to identify and eliminate overlapping disclosure requirements to reduce the volume of redundant disclosures in the 10-K. However, we believe further research examining the content and determinants of redundant 10-K disclosures is necessary to shed greater light on the extent to which these redundancies enhance or inhibit the effectiveness of financial communication in the 10-K.

2. INSTITUTIONAL BACKGROUND AND RELATED RESEARCH

The 10-K is only one firm-issued disclosure among many, including those seen as often conveying more timely information such as earnings announcements and press releases. However, a substantial body of research confirms that 10-Ks contain price-relevant information that investors trade on (e.g., Huddart et al 2007; Griffin 2003; You and Zhang 2009). A recent stream of research finds that investors' ability to process relevant information in the 10-K is impaired by disclosure complexity (e.g., Lehavy et al. 2011; Rennekamp 2012). Disclosure volume is an important component of that complexity (e.g., You and Zhang 2009; Miller 2010). We frame 10-K length as a function of a firm's underlying operational complexity, voluntary or mandated disclosure redundancy, and "residual" disclosure. We discuss each of these three components below.

2.1 Operational Complexity and 10-K Disclosure Volume

The Securities and Exchange Commission regulates the format and content of firms' annual reports (10-K) under Regulation S-K. The SEC's Regulation S-X prescribes rules related specifically to the financial statements and related notes. In addition, the FASB requires mandatory footnote disclosure of items relating to a reporting entity's operations. These reporting requirements mandate the disclosure of information related to material information in various parts of the 10-K. For instance, Item 1 of the 10-K requires firms to describe their business and report on material risk factors the business faces. Item 2 requires firms to report on their properties, Item 3 requires reporting on legal proceedings, etc. Because firms with greater operational complexity are expected to have more transactions, subsidiaries, and events on which to report, a reasonable inference is that firm operating complexity is positively correlated to 10-K length. In addition, certain relatively complex operating, financing, and investing choices by firms involve accounting issues for which there tends to be more extensive required disclosures, such as derivatives, complex tax arrangements, share-based compensation arrangements, pensions and post-retirement benefits, and other financial instruments.

2.2 Redundant Disclosure Requirements and 10-K Disclosure Volume

Critics of the current reporting regime point out that significant overlap exists between the disclosure requirements mandated by the SEC and those mandated by the FASB. For instance, significant accounting policies are required to be reported both in the financial statement notes and in Item 7 of the 10-K (Management's Discussion and Analysis). Similarly, redundancies exist in the reporting for income tax disclosures, related party transactions, segment information, and a variety of other areas (FASB 2001). Information regarding risk factors,

business description, legal proceedings, are often discussed in both the notes and in other sections of the 10-K (KPMG 2011).

Regulators and practitioners have expressed concern that redundant disclosure requirements add unnecessary length and complexity to the 10-K (e.g., White 2013). The SEC reports that its staff is identifying ways to eliminate redundant disclosure requirements (SEC 2012).⁴ However, the SEC also permits significant cross-referencing between the notes to the financial statements contained in Item 8 and other sections of the 10-K in lieu of presenting duplicate information (Code of Federal Regulations, title 17, sec. 229). Our reading of 10-Ks in our sample suggests that firms vary significantly in the extent to which they present redundant information rather than cross-referencing other sections of the 10-K. We provide a representative example in Appendix A, which displays MD&A disclosure of recent accounting pronouncements for two sample firms. The first uses three paragraphs to repeat verbatim information found in the notes to the financial statements. We emphasize this represents only a small fraction of the total amount of redundancy in this firm's 10-K. The second company uses a single sentence to refer the reader to Note 1 of the financial statements to find the relevant information. Thus, it appears at least some of the redundancy found in 10-K reports is due to managerial discretion.

The 2011 KMPG survey on disclosure overload found that several respondents wanted an interpretive release from the SEC providing further guidance on the permissibility of cross-referencing redundant disclosures. This suggests that one reason for variation in firm's reporting

⁴ Paragraphs BC4-BC6 of the FASB's March 4, 2014 Exposure Draft on the Conceptual Framework, "Chapter 8: Notes to Financial Statements," discuss respondents' calls for collaboration with the SEC in reducing overlapping disclosure requirements. However, the exposure draft does not directly reflect such a collaboration. Rather this is largely a FASB-only effort. As explained in these paragraphs, the FASB is concerned about preparers who are not governed by the SEC and about information that might be less complete or less carefully scrutinized in an SEC disclosure than in financial statement notes.

of redundant information is uncertainty regarding exactly how much cross-referencing is permissible.

Redundancy may be helpful to investors if providing the same information in different sections of the 10-K provides context for proximate information to be communicated more clearly. Also, to the extent that differentially skilled financial statement users might look primarily to the notes to the financial statements versus the MD&A, presenting information in both places (redundantly) potentially benefits users by having information accessible where they expect to find it. Li (2014) reports that 10-Ks with greater disclosure redundancy are associated with larger stock price reactions around the filing date. She concludes that disclosure redundancy improves the transparency of relevant financial information. However, casual observation of redundant content in our sample indicates that potentially less material disclosures are repeated in many cases. Repetition of less material disclosures may make the most important information more difficult to discern.

2.3 Residual Disclosure

Disclosure volume unassociated with firms' operating structure or redundant disclosures stem primarily from discretionary managerial choices, which may be associated with management's perception of the firm's litigation or other risks, shareholder demand, or simply management's idiosyncratic disclosure "style" (e.g., Bertrand and Schoar 2003). In many cases, these residual disclosures likely include information that critics allege is irrelevant to shareholders (e.g., KPMG 2011; FASB Discussion Paper (ITC) 2012). For instance, concerns regarding shareholder litigation or regulatory intervention may lead firms to over-disclose to ensure they can't be held liable for not disclosing sufficient information to investors or regulators (ICAEW 2013). In addition, academic literature suggests some disclosure length may be

attributable to managers purposefully seeking to obfuscate information in the 10-K, possibly to mask poor performance (e.g., Li 2008; Rogers, Van Buskirk, and Zechman 2011). In addition, firms may vary the level of useful voluntary disclosure depending on factors such as proprietary costs of disclosure.

Examples of generic disclosures are commonly found in risk factor disclosures and in firms' disclosure of significant accounting policies. In their 2011 report on disclosure issues, KPMG identifies a representative example in one company's disclosure of risk factors:

"Declining economic conditions could adversely affect our results of operations and financial condition" (KPMG 2011, p.13). Similarly, companies often report on recent FASB pronouncements or policy changes that apply to all companies and that are not specific to the company's own accounting choices.

Finally, because reporting requirements are generally uniform across all industries, and because reporting requirements are seldom scaled back after they have outlived their usefulness, many corporations report on issues that may no longer be or never were relevant for their own business. Examples may include mandatory reporting of the historical low and high closing price of a company's common stock, the dilution disclosure, and the ratio of earnings to fixed charges (White 2013).

Empirically, we capture this additional disclosure as the part of total 10-K length not represented by our estimates of complexity or redundancy. We label this component of total length "residual" disclosure. In addition to any irrelevant or immaterial disclosures, residual disclosure may also capture the presence of any voluntary disclosures firms provide that are intended to be useful to investors. Thus, we do not assume that all disclosures unexplained by operating complexity or disclosure redundancy are necessarily problematic for investors.

3. SAMPLE CONSTRUCTION

To provide evidence on the determinants of 10-K length, we start with the population of all Compustat firms-years covering the 10-year period from 2003 to 2012. Because some of our analyses replicate findings of You and Zhang (2009), we incorporate their sample selection criteria by eliminating 65,736 observations with closing share prices less than one dollar or total market cap less than \$200 million. We exclude 4,606 observations missing firm identifiers on Compustat, and another 3,626 observations that are lost in the merge between Compustat and CRSP. We eliminate 3,758 observations with insufficient return data on CRSP, 2,636 observations with missing earnings announcement dates, 247 observations where the earnings announcement date is more than 120 days after the fiscal year end, and 241 observations missing SIC data. Finally, we exclude 5,032 foreign issuers because these firms do not file 10-K reports. These sample selection criteria result in 24,462 firm-years remaining during our sample period.

Because our analysis involves a substantial amount of hand collection of 10-K information, we further narrow our sample by randomly selecting 650 firm-years from the pool of 24,462 remaining firm-years. We find that 21 of these observations are either missing 10-Ks on the SEC website or issued a non-standard 10-K that complicates the measurement of 10-K length. An additional nine of these firm-year observations are missing Compustat data required for constructing the variables in our study. Thus, our final sample consists of 620 firm-years representing 570 unique firms from the sample period 2003-2012. Table 1 depicts the effect of each of our sample selection criteria on our final sample.

[Insert Table 1 here]

Although this sample is smaller than those examined in studies that parse 10-Ks using computer programs such as PERL or Python, we note that our hand-collected method helps

eliminate measurement errors these programs may make when working with 10-Ks with non-standard formats. Non-standard formats include 10-Ks incorporating information by reference to the annual report and 10-Ks with information typically contained in Item 8 included in Item 15 instead.

4. MODELING THE COMPONENTS OF 10-K LENGTH

Following prior studies (e.g., Leheavy et al 2011; You and Zhang 2009; Loughran and McDonald 2010; Miller 2010), we measure 10-K disclosure volume using the number of words in the 10-K filing (*TOTAL_LENGTH*). We decompose 10-K length into the portion attributable to operating complexity, disclosure redundancy, and residual disclosure as described in the following subsections.

4.1 Operating Complexity and 10-K Length

To gain an understanding of how operating complexity affects 10-K length, we conducted a preliminary search through a broad sample of 10-Ks from 2003-2012. Based on our analysis of these 10-Ks, we identified a number of factors that appeared to be significant drivers of 10-K length. This analysis, coupled with our review of the reporting requirements in Regulations S-K and S-X, led to our development of the following model of the factors that impact 10-K length.

$$\begin{aligned} TOTAL_LENGTH_{i,t} = & \beta_0 + \beta_1 REDUNDANCY_{i,t} + \beta_2 DELTA_ROA_{i,t} + \beta_3 DELTA_REV_{i,t} \\ & + \beta_4 ACQUISITION_{i,t} + \beta_5 FY_RET_{i,t} + \beta_6 FY_VOLAT_{i,t} + \beta_7 SPI_D_{i,t} + \beta_8 CAP_LEASE_{i,t} + \\ & \beta_9 OP_LEASE_{i,t} + \beta_{10} RD_{i,t} + \beta_{11} INTANG_{i,t} + \beta_{12} SIZE_{i,t} + \beta_{13} DEBT_RATIO_{i,t} + \beta_{14} FCF_{i,t} \\ & + \beta_{15} DERIVATIVE_USE_{i,t} + \beta_{16} NUM_OPSEG_{i,t} + \beta_{17} NUM_GEOSSEG_{i,t} + \\ & \beta_{18} REF_AR_{i,t} + \beta_{19} DEF14_{i,t} + Industry\ controls + year\ controls + \varepsilon_{i,t} \end{aligned} \quad (1)$$

REDUNDANCY is a measure of the degree to which a given firm's 10-K and financial statements contain identical words, which we compute as described below in section 4.2.

DELTA_ROA is the annual change in income before extraordinary items scaled by lagged assets.

DELTA_REV is the annual percentage change in sales revenue. *ACQUISITION* is an indicator variable equal to 1 if the company made an acquisition during the fiscal year, and zero otherwise.⁵ *FY_RET* is equal to the raw annual return during the reporting period. *FY_VOLAT* is equal to the standard deviation of monthly returns for the reporting period. *SPI_D* is an indicator variable equal to one if Compustat indicates the company has reported a special item in their income statement, and zero otherwise. We expect these variables to reflect significant fiscal-year events that tend to be discussed in the MD&A or in the notes to the financial statements. We predict these variables to be positively associated with 10-K length.

CAP_LEASE is an indicator variable set equal to one if the company has reports a capital lease on its balance sheet, and zero otherwise. *OP_LEASE* is an indicator variable if the value of operating lease payments due in one year is greater than 1 percent of assets, and zero otherwise. We expect the use of leases to be positively associated with 10-K length, given required footnote disclosures and their possible inclusion in discussion of the firm's business in Item 1 of the 10-K.

RD is a continuous variable equal to research and development expenditures scaled by lagged assets. Because material research and development activities are typically discussed in the 10-K, particularly in the business description, we expect *RD* to be positively associated with 10-K length. To account for purchased intangibles, we also include the continuous variable *INTANG*, which is equal to the unamortized value of purchased intangible assets scaled by lagged assets. *SIZE* is equal to the natural logarithm of the market value of equity. We expect size to reflect operating complexity that increases the volume of 10-K disclosures.

DEBT_RATIO is equal to total liabilities scaled by total assets. Because firms discuss the composition of their debt obligations in the notes to their statements, we expect a higher debt

⁵ The variable *ACQUISITION* is set equal to 1 for firm-years with non-missing, non-zero values of *ACQCSHI* and *AQC* as reported by COMPUSTAT (shares issued for acquisition and cash outflows for acquisition, respectively).

ratio to be associated with greater 10-K length. We also include a proxy for free cash flows (*FCF*) measured as the average of operating cash flows scaled by total assets over the prior and current years (Blouin and Krull 2009). Because liquidity concerns are negatively associated with free cash flows and companies discuss their liquidity position in the MD&A and other section of the 10-K, we expect *FCF* to be negatively associated with 10-K length. *DERIVATIVE_USE* is an indicator variable equal to one if the company reports any current or accumulated gains or losses on derivative transactions, and zero otherwise. We expect derivative use to increase the length of the 10-K.

We capture the number of operating and geographic segments the company reports with the variables *NUM_OPSEG* and *NUM_GEOSEG*, respectively. We expect 10-K length to be increasing in both of these variables.

Some firms incorporate certain sections of the 10-K by reference to the annual report, the relevant portion of which is then included as Exhibit 13 to the 10-K. We include the referenced Exhibit 13 material in our total word count (*TOTAL_LENGTH*). To control for the possibility that information primarily intended for the annual report may be written differently (e.g., longer or shorter) than disclosures intended primarily for the 10-K, we create an indicator variable *REF_AR* equal to one when 10-K sections are incorporated by reference and zero otherwise. Also, a small number of firms include proxy statement information (items 10-14) directly into the 10-K rather than following the more standard practice of incorporating this information only by reference. To control for the impact of this information on the 10-K, we also include an indicator variable *DEF14*, which is equal to one for 10-Ks that include proxy statement information directly, and zero otherwise.

4.2 Disclosure Redundancy and 10-K Length

We use two alternative software programs to measure the amount of redundancy contained in a 10-K filing. Both programs provide output based on a comparison of the text within two documents. Because concerns regarding 10-K redundancy generally have to do with overlap between the notes to the financial statements and other sections of the 10-K, we split each 10-K into two documents: one document containing the financial statement notes and a separate one containing all preceding items of the 10-K (Business and Risk Factors, Management's Discussion and Analysis, etc.). We retain all tables in the 10-K, so that our assessment of redundancy across 10-K sections includes disclosures made in tabular form.

The first measure of redundancy we use is the similarity score computed by Ferret Copy Detector, a program designed to detect plagiarism by comparing word sequences in two documents. Ferret examines the number of distinct three-word sequences, or trigrams, in two documents and computes the number of shared trigrams occurring in both documents as a percentage of the total number of distinct trigrams occurring in both documents. The resulting similarity score (*REDUNDANCY*) is computed as a decimal between 0 and 1 with higher scores reflecting a higher degree of shared language between the two documents.

Our second measure of redundancy, *MATCHED_WORDS*, is the output from a plagiarism detection software program called WCopyfind. This program compares pairs of documents and outputs the total number of distinct words that occur in six-word sequences existing in both documents.⁶ Thus, more redundant disclosures result in higher values of *MATCHED_WORDS*.

⁶ WCopyfind allows for considerable latitude in comparison tests. Following McMullin (2014), we choose to compare 6-word sequences with WCopyfind. We also require the sequences to be identically matched.

Strong theoretical and empirical support exists in prior research for using trigrams to identify similar text because independently written text typically share few matching trigrams. Gibbon et al. (1997) note that 77 percent of all trigrams contained in a large sample of Wall Street Journal articles occurred only once. Lyon et al (2001) finds that three-word sequences are optimal for detecting similar text, and trigrams are a common choice in related research (e.g., Nelson and Pritchard 2014). Research on the optimal length of word-sequences to use to identify similar text indicates shorter sequences (two or three-words) provide the best results (Barrón-Cedeño and Rosso 2009). Thus, we are confident our three and six-word sequences are of sufficient length to capture disclosure redundancies in our context.⁷

Both of our measures of redundancy are subject to drawbacks and limitations. As a percentage measure, *REDUNDANCY* does an arguably better job of capturing the importance of duplicative disclosure in explaining the proportion of a given firm's 10-K length, relative to other determinants. However, *REDUNDANCY* likely is a worse measure for our tests modelling total 10-K length measured in words (*TOTAL_LENGTH*). For instance, if a complex firm and a simple firm both have *REDUNDANCY* values of 0.15, the impact of redundancy on total word count is likely to be much greater for the complex firm because it has a longer 10-K to begin with. Thus, we rely primarily on *MATCHED_WORDS* to make inferences regarding the impact of redundant disclosure on total 10-K length, and we use *REDUNDANCY* primarily to make inferences in tests examining the determinants of redundant disclosure. We find that our primary results are qualitatively similar using either measure.

⁷ The majority of content identified as redundant by three and six-word sequences in our sample come from longer sequences of words that often reflect redundant sentences or even paragraphs. We provide the following insight into the distributional features of redundant content in our sample when redundant output is based on sequences of six-word sequences: 80 percent of all words contained in redundant output are part of redundant sequences at least 11 words long. Two-thirds of all words in our redundant output are part of redundant sequences at least 21 words long. Redundant output based on trigrams vs. six-word sequences is highly similar. In sensitivity analyses, we also find results using redundant ten-word sequences to produce very similar results to those using six-word sequences.

4.3 Residual Disclosure and 10-K Length

We proxy for residual disclosure volume as the portion of 10-K length unexplained by operating complexity and disclosure redundancy, or the residual from equation 1. By construction, this measurement choice constrains mean residual disclosure to be equal to zero. Thus, we interpret observations with negative residuals as having more parsimonious disclosures, and firms with positive residuals as having abnormally long disclosures.

5. RESULTS

5.1 Descriptive Statistics and Correlation Matrix

Panel A of Table 2 presents descriptive statistics for the primary firm-level variables in our study. Panel A displays distributional statistics for our final sample of 620 firm-years as well as for the broader sample of 24,462 firm-years from which the final sample was randomly selected (of which 23,407 have sufficient data for Panel A of Table 2). T-tests indicate that the random sample is not significantly different from the population it was drawn from except for *DELTA_ROA*, which difference is barely significant ($p=0.046$). Thus, our sample is representative of the larger Compustat population.

Among the findings of interest in Panel A, we note that nearly 44 percent of sample firms made some type of acquisition during the year, 64 percent reported special items, and the mean firm has 2.3 geographic reporting segments.

[Insert Table 2 here]

Panel B reports descriptive statistics for variables related to sample 10-K filings. Mean 10-K length is approximately 55,000 words. The median of approximately 52,000 words is close to the mean, suggesting this distribution is only mildly skewed. The inter-quartile range is

24,322 words, indicating substantial variation in 10-K length.⁸ The longest section of the 10-K tends to be the financial statements and accompanying notes (*ITEM8_LENGTH*), followed by the MD&A (*ITEM7_LENGTH*) and the business description and risk factors (*ITEM1_LENGTH*). These three sections account for over 85 percent of total 10-K length. We also find substantial variation in the amount of redundancy incorporated in firms' 10-Ks. *REDUNDANCY* has an inter-quartile range equal to 50 percent of its median value, and *MATCHED_WORDS* has an inter-quartile range of nearly 83 percent of its median value.

Panel C provides a breakdown of the sample by industry. Similar to the larger Compustat population, the sample has a particularly high representation among firms in the Finance and Business Equipment industries. Panel D indicates that the sample has roughly equal representation across sample years, and that the both the mean and median length of 10-Ks has increased dramatically over our ten-year sample period. Mean *TOTAL_LENGTH* in 2012 is 45 percent greater than mean *TOTAL_LENGTH* in 2003. This result is consistent with evidence elsewhere of significant increases in 10-K length. In its 2011 study of disclosure trends, KPMG reports a 28 percent increase in mean 10-K length (measured in pages) from the slightly shorter time period covering 2004 to 2010.

Table 3 shows the Pearson correlations among the redundancy measures, the length of the 10-K and key individual items, and firm size. *REDUNDANCY* and *MATCHED_WORDS* are 57 percent correlated, indicating that these two measures are largely capturing the same construct, though slightly differently as the former is a percentage number and the latter an absolute measure of redundant text. Interestingly, *REDUNDANCY* is positively correlated with total 10-K length and with the length of the MD&A and notes to the financial statements (*ITEM7_LENGTH*

⁸ To provide some context, this inter-quartile range corresponds to approximately 30 pages of single-spaced text on a standard page.

and *ITEM8_LENGTH*), but negatively correlated with *ITEM1_LENGTH*. This suggests there is greater redundancy between the content of the MD&A and financial statement notes, and less redundancy between the content of the Business and Risk Factors section and the notes. Not surprisingly, larger firms have longer 10-Ks and higher values of *MATCHED_WORDS*, but there is no correlation between firm size and *REDUNDANCY*.

[Insert Table 3 here]

5.2 Regression Results

Table 4 displays results from estimating equation (1). The first column shows a strong positive coefficient estimate on *REDUNDANCY* (p-value = 0.010). To test for a change in disclosure volume over our sample period, we next estimate the regression after suppressing the year fixed effects and instead substituting a variable (*TIME_TREND*) equal to the fiscal year minus 2003. Results for this alternative specification are displayed in the second column of Table 4 and provide robust evidence that 10-K length is increasing over the sample period, consistent with related research (e.g., Loughran and McDonald, 2010).

As discussed in section 4.3, *MATCHED_WORDS* likely provides a better measure of the impact of disclosure redundancy on total word length than does our *REDUNDANCY* measure. Replacing *REDUNDANCY* with *MATCHED_WORDS* results in coefficient estimates presented in column 3 of Table 4. Consistent with our expectation, the coefficient estimate on *MATCHED_WORDS* is strongly significant (p-value <0.001). Finally, to test whether our results are sensitive to the mild skewness in the dependent variable *TOTAL_LENGTH*, we rerun our regression after replacing the dependent variable with its log transformation and present the results in the last column of Table 4. We find that inferences are virtually unchanged.

Across the four columns of Table 4, we find that the incidence of acquisitions, capital leases, leverage, free cash flows, size, return volatility, special items, and operating segments all contribute significantly to 10-K length. There is some evidence that fiscal year return (*FY_RET*) is negatively related to 10-K length. This may be attributable to managers using more words to present bad news than good news (e.g., Li 2008). Unsurprisingly, we also find a strong positive coefficient estimate on *DEF14*, consistent with firms incorporating information in the 10-K that is typically reserved for the proxy statement having longer 10-Ks. We find no evidence that revenue growth, operating leases, R&D expense, or geographic segments are associated with 10-K length, possibly because those factors have been subsumed by other regressors in the model.

[Insert Table 4 here]

We use the parameter estimates from column 3 of Table 4 Panel A to compute the portion of total 10-K length attributable to operating complexity, disclosure redundancy, and residual disclosure length. We use the fitted values from all parameter estimates and independent variables except for *MATCHED_WORDS* to construct *OP_COMPLEX_LENGTH*, which we use to measure the portion of 10-K length explained by operating complexity. We use the coefficient estimate on *MATCHED_WORDS* to compute a fitted value, *REDUNDANT_LENGTH*, which we use to measure the portion of 10-K length explained by disclosure redundancy. We use the residual as a proxy for residual disclosure, which we name *RESIDUAL_DISCLOSURE*.

Panel B displays distributional statistics for the three components of *TOTAL_LENGTH*. We find that *OP_COMPLEX_LENGTH* is the largest, with a mean value equal to 39,776 words. This represents approximately 72 percent of the mean value of *TOTAL_LENGTH* as reported in Table 2. In contrast, the portion attributable to disclosure redundancy is much smaller, as

indicated by the mean value of *REDUNDANT_LENGTH*, which is only 15,697 words. By construction, the mean value of *RESIDUAL_LENGTH* is close to zero.

Arguably a more important result from Panel B of Table 4 is the inter-quartile range across the three components of total 10-K length because it reveals the degree to which each component of length explains variation in total 10-K length. *REDUNDANT_LENGTH* has an inter-quartile range of 11,656 words, which is over 89 percent of the inter-quartile range of *OP_COMPLEX_LENGTH* (13,073 words). This surprising result indicates that nearly as much cross-sectional variation in 10-K length is explained by variation in disclosure redundancy as by operating complexity. A slightly greater amount of variation in 10-K length is unexplained by either operating complexity or disclosure redundancy, as indicated by the inter-quartile range of *RESIDUAL_LENGTH* (13,933 words).

To provide a sense of how operating complexity, disclosure redundancy, and residual disclosure comprise 10-K length over relatively shorter versus relatively longer 10-Ks, we partition our sample into quintiles based on 10-K length. We then compute the median values of *OP_COMPLEX_LENGTH*, *REDUNDANT_LENGTH*, and *RESIDUAL_LENGTH* in each quintile and present this information in pictorial form in Figure 1. Figure 1 indicates that variation in 10-K length due to operating complexity across the five quintiles are relatively modest. The median firm in the top quintile has *OP_COMPLEX_LENGTH* only 15,078 words (47,013 – 31,935) longer than the median firm in the bottom quintile, an increase of only 47 percent ($15,078 \div 31,935$). However, the median firm in the top quintile has *REDUNDANT_LENGTH* that is 16,509 words (23,894 – 7,355) longer than the median firm in the bottom quintile, representing an increase of 224 percent ($16,509 \div 7,355$), suggesting disclosure redundancy explains more of the difference between the longest and the shortest 10-

Ks than does operating complexity. Finally, we note that the portion of unexplained (residual) 10-K length accounts for even more variation from the top to the bottom quintiles ($11,179 - (-7,629) = 18,808$ words) than either operating complexity or disclosure redundancy.

These findings provide strong support for the assumptions made in prior research that a significant amount of variation in disclosure volume, even in documents as highly regulated as the 10-K, is attributable to managerial discretion. To the best of our knowledge, this is the first systematic evidence of the extent to which disclosure redundancy influences 10-K length. Although we cannot infer from these results the extent to which disclosure redundancy is driven by managerial opportunism versus other factors, our analysis in section 6.2 provides some insight the degree each of these three components is associated with mispricing subsequent to the 10-K filing date.

6. SUPPLEMENTAL ANALYSES

6.1 Where is the Redundancy in the 10-K?

We next examine each of the three primary sections of the typical 10-K (Business and Risk Factors, MD&A, and financial statement notes) to pinpoint which sections are affected most by redundant disclosures. We re-estimate the regression in equation (1) after substituting alternately *ITEM1_LENGTH*, *ITEM7_LENGTH*, and *ITEM8_LENGTH* as the dependent variable. For brevity of exposition, we report only the results using *REDUNDANCY* as our redundancy measure in this analysis, though we note that inferences are similar if we use *MATCHED_WORDS* instead. We tabulate the results in Table 5.

[Insert Table 5 here]

Table 5 indicates that the coefficient estimate on *REDUNDANCY* is significantly negative in the first column explaining Item 1 length (Business and Risk Factors), but

significantly positive in the second column explaining Item 7 length (MD&A). Since *REDUNDANCY* is measured as a percentage, this indicates that 10-Ks marked with a higher proportion of redundant disclosures are associated with shorter disclosures in Item 1 and longer disclosures in Item 7. Although anecdotal evidence indicates that some discussion of business description and risk factors, the primary content of Item 1, can also be found in the notes to the financial statements, the results in Table 5 suggest that the 10-K redundancy is more associated with the content of the MD&A. This finding is interesting in light of the SEC's counsel to not report information in the MD&A section that is redundant with the financial statement notes (US Securities and Exchange Commission, 2003). The third column of Table 5 indicates that a greater level of 10-K redundancy is associated with longer Item 8 disclosures (financial statements and notes), though this result is expected in this test because *REDUNDANCY* is based specifically on a comparison of the financial statement notes to the earlier sections of the 10-K.

One interesting takeaway from an examination of the R-squared in each column of Table 5 is that firm operating characteristics and disclosure redundancy explain relatively less variation in Item 1 disclosure length (0.342), and relatively more variation in Item 8 disclosure length (0.472). This suggests that the drafting of the Business and Risk Factors section is more heavily influenced by factors that extend beyond the firm's operating structure, such as management's assessment of litigation and other risks, shareholder composition, or simply management's idiosyncratic "style". Prior research shows individual managers impose their style on a host of corporate policies and practices even after controlling for firm-level characteristics (e.g., Bertrand and Schoar 2003). In contrast, the length of the financial statement notes (Item 8) appear to be more heavily determined by firm operating characteristics and disclosure redundancies, consistent with note disclosures being audited and thus likely more regimented

than other sections of the 10-K. The variation in the length of Management's Discussion and Analysis (Item 7) explained by firm operating complexity and redundancy appears to lie midway between that of Business and Risk Factors and the Financial Statement Notes.

6.2 What are the Determinants of Disclosure Redundancy?

Although we have documented that disclosure redundancy is an important contributor to overall 10-K length, an unanswered question is why some 10-Ks incorporate significantly more redundancy than do others. The SEC permits but does not necessarily require firms to cross-reference between the notes and other sections of the 10-K in lieu of presenting duplicate information. No established theory in prior accounting literature predicts which accounting topics will be reported more redundantly in firms' financial reports. However, we provide some exploratory analysis into which firm characteristics are associated with greater levels of redundant reporting.

We regress *MATCHED_WORDS* and *REDUNDANCY* on the other right-hand side variables in equation (1) and report the results in first and second columns, respectively, of Table 6. We find that firm size and leverage exhibit a consistent, positive relationship with both measures of disclosure redundancy. Disclosure redundancy is strongly negatively associated with R&D intensity in both columns. In addition, firms are more likely to provide redundant disclosure when they report special items, but less redundancy when they have made an acquisition.

Two other interesting findings emerge from Table 6. First is the negative coefficient estimate on *REF_AR* in both columns, indicating that firms are less likely to provide redundant information when significant sections of the 10-K are incorporated into the filing only by reference to the annual report to shareholders. This may be an indication that managers are more

likely to cut back on redundant reporting for disclosures designed specifically for the annual report, because the annual report tends to be a more “reader-friendly” publication than the 10-K. Second is the positive and significant coefficient estimate on *TIME_TREND* in column 1, which becomes negative and significant in column 2. This suggests that levels of redundancy as measured by the number of redundant words is increasing over time, but that increases in redundancy over time fail to keep up with increases in total 10-K disclosure, resulting in redundant disclosures declining as a proportion of total 10-K length. However, the economic magnitude of *TIME_TREND* coefficient appears to be fairly small, suggesting changes in redundancy in the 10-K over the sample period are not large.

[Insert Table 6 here]

6.3 Components of 10-K Length and Investors’ Ability to Process Relevant Information

Prior research finds that longer 10-Ks are associated with lower trading volume around the 10-K filing date and subsequent return drift after the filing date (e.g., You and Zhang 2009; Miller 2010). However, previous research does not indicate what source of 10-K length might be the driver of impaired information processing around 10-K filing dates. Is it operating complexity, disclosure redundancy, residual disclosure, or some combination of the three?

To address this issue, we build on You and Zhang’s (2009) empirical method to test for a relationship between post-filing-date return drift and 10-K length in our more recent sample. We re-estimate their model after decomposing total 10-K length among the components explained by operating complexity, redundancy, and residual disclosure. We tabulate the results of this analysis in Table 7.

The first column of Table 7 is similar to You and Zhang’s (2009) equation (4), which regresses post-filing date returns on the size-adjusted returns during the three days centered on

the 10-K filing date (FDR), a measure of 10-K length, and the interaction between the two. Size, book-to-market, beta, and momentum are included to control for confounding risk factors. As indicated by the results in Column 1 of Table 7, we are unable to find evidence of significant post-filing-date price drift using You and Zhang's (2009) methodology in our sample.⁹ Table 7 reports results for 6-month subsequent returns as the dependent variable, though we similarly find no results for the 3-month or 12-month return windows. A possible explanation is that investor sophistication has increased in our sample period relative to the period examined in their study (1995 through 2005).

[Insert Table 7 here]

In the second column of Table 7, however, we find that partitioning total 10-K length produces some evidence that post-filing date return drift is associated with the portion of 10-K length attributable to residual disclosure, as evidenced by the significantly positive coefficient estimate on *FDR*RESIDUAL_LENGTH* (p-value = 0.023). The insignificant coefficient estimates on *FDR*OP_COMPLEX_LENGTH* and *FDR*REDUNDANT_LENGTH*, which are significantly smaller in magnitude and far from statistically significant, provide no support for the notion that 10-K length associated with operating complexity or redundancy impair price discovery around 10-K filing dates.¹⁰

6.4 What Accounts for the Increase in 10-K Disclosure from 2003-2012?

The primary objective of our study is to shed light on the extent to which cross-sectional variation in 10-K disclosure volume is driven by variation in firms' operating complexity,

⁹ You and Zhang (2009) use an indicator variable to capture whether a given filing is longer or shorter than the annual sample median. Although our reported regression results are based on a continuous measure of 10-K length, we are unable to replicate You and Zhang's (2009) main result using either measure.

¹⁰ Although our evidence suggests that overall prices are not influenced by redundant disclosures or operating complexity around 10-K filing dates, it is possible that individual retail investors are harmed by lengthy disclosures associated with redundancy or operating complexity.

reporting redundancy, and residual disclosure volume. Determining which time-series factors explain the evolution of disclosure volume over time is beyond the scope of this study. However, we note a marked (45%) increase in mean 10-K length even over our relatively short and sample period. We next provide some insight into the how much of the increase in disclosure volume during our sample period is attributable to increases in firm operating complexity versus increases in required or voluntary disclosure.

We first estimate the baseline relationship between 10-K length and firm operating complexity variables in the early part of our sample period (2003-2006). We use the resulting coefficient estimates to create fitted values for expected 10-K length for observations in the last six years of our sample. If required and voluntary disclosure over our sample period has remained constant after controlling for changes in firm operating complexity, then these fitted values of 10-K length should approximate actual 10-K length (*TOTAL_LENGTH*) from 2007-2012.

Figure 2 plots mean actual and fitted (predicted) 10-K length for each sample year. Figure 2 indicates that changes in disclosure length over the sample period attributable solely to changes in firm complexity are relatively modest. Mean *TOTAL_LENGTH* in 2003 is 44,277 words and increases by 45 percent to 64,343 words in 2012. Mean predicted length in 2003 is 47,133 words, and increases by only five percent to 49,561 words in 2012. Thus, the increase in disclosure volume during our sample period appears to be driven primarily by increases in required or voluntary disclosure rather than increases in firm complexity.

6.5 Other Robustness Tests

Our model of the determinants of 10-K length in equation (1) is necessarily truncated to include only a subset of all operational factors that could potentially affect 10-K length. The

variables included in our model are based primarily on our own reading of numerous 10-K documents. One potential concern is that we have omitted important components of operating complexity from equation (1) that explain significant variation in 10-K length, leading to an overestimate of the amount of variation in 10-K length that should be considered “residual” disclosure. To test the sensitivity of our inferences to this concern, we add three additional variables to our model of 10-K length suggested by Li (2008). We find that including firm age, earnings volatility, and an indicator variable reflecting whether the firm is incorporated in Delaware changes the regression R-squared by less than 1 percent and the coefficient estimate on *MATCHED_WORDS* by less than 2 percent. We conclude that our allocation of 10-K length to the portions explained by operating complexity, redundancy, and residual disclosure are largely robust to concerns about omitted variable bias.

Finally, we address the concern that macroeconomic events may affect 10-K length in ways not captured by our firm-specific variables. Although we expect that macroeconomic events will largely be accounted for by our year and industry fixed-effects, as well as through their effects on variables measured at the firm level, it is possible that external shocks induce industry-year variation in 10-K length in ways that are identified as residual disclosure under our current models. For instance, if banks increased their 10-K disclosures during the recent financial crisis, it is possible that this increased 10-K length would be categorized as *RESIDUAL_LENGTH*, which is intended to capture only firm-year specific effects. If this concern is problematic for our analysis, we should find values of *RESIDUAL_LENGTH* in the financial services industry to be particularly large around the financial crisis.

Although we find that *TOTAL_LENGTH* is significantly greater in 2008 for financial services firms, we find no evidence that mean or median *RESIDUAL_LENGTH* is significantly

larger during the financial crisis than at other times during our sample period. Mean and median *RESIDUAL_LENGTH* are modestly positive in 2008, though smaller in magnitude than in four other years during the sample period. The mean and median are both negative in 2009, and relatively close to zero in 2010. Thus, our model appears to do a fair job of allocating the effects of macroeconomic shocks on 10-K length to our operating complexity variable and not to *RESIDUAL_LENGTH*.

7. CONCLUSION

Although several studies have examined the consequences of 10-K length on various corporate outcomes, we are the first to provide systematic descriptive evidence on the structural components that contribute to cross-sectional variation in 10-K length. We argue that 10-K length can be thought of as having three drivers: (1) the portion attributable to operating complexity; (2) the portion attributable to disclosure redundancies; and (3) residual disclosure. We find that disclosure redundancy accounts for as much variation in 10-K length as does operating complexity. We also find that residual disclosure accounts for more variation in 10-K length across firms than do either operating complexity or disclosure redundancy.

Results from this study should be of interest to both regulators and academics. Both the SEC and the FASB have expressed recent interest in enhancing the quality of disclosure, with an emphasis on reducing disclosure volume. In calling for a meaningful review of disclosure requirements, SEC Chairman Mary Jo White questioned whether “[SEC] rules [are] the primary cause of potential disclosure overload or [whether] other sources contribute to it.” Our study suggests there is a significant discretionary component to disclosure volume. We are the first to provide systematic evidence of the significant role of disclosure redundancy in determining 10-K

length. We also highlight the extent to which firms' idiosyncratic levels of disclosure contribute to variation in 10-K length after controlling for firms' operating characteristics.

We believe several interesting and important research questions remain regarding firms' disclosure practices within the 10-K. Further research on the nature and content of disclosure redundancy may be able to shed further light on whether those redundancies are explained primarily by strategic considerations or simply by managers' efforts to comply with overlapping disclosure requirements. In addition, future research should examine whether 10-K disclosure attributes vary with internal or external governance mechanisms, including discretionary components of CEO pay such as bonuses paid out at the discretion of the board rather than as a strict function of earnings levels.

REFERERNCES

- Barrón-Cedeño, A. and P. Rosso. 2009. On Automatic Plagiarism Detection Based on n-Grams Comparison. Boughanem et al. (Eds.), ECIR 2009, Lecture Notes in Computer Science, vol. 5478, pp. 696-700.
- Bertran, M. and A. Schoar (2003). Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics* 118 (4): 1169-1208.
- Blouin, J. and L. Krull. 2009. Bringing It Home: A Study of the Incentives Surrounding the Repatriation of Foreign Earnings Under the American Jobs Creation Act of 2004. *Journal of Accounting Research* 47 (4): 1027-1059.
- Financial Accounting Standards Board (FASB). 2001. *GAAP-SEC Disclosure Requirements*. Steering Committee Report Series. Norwalk, CT: FASB
- Financial Accounting Standards Board (FASB). 2012. *Discussion Paper: Disclosure Framework*. Norwalk, CT: FASB
- Financial Accounting Standards Board (FASB). 2014. *Exposure Draft: Chapter 8: Notes to Financial Statements*. Norwalk, CT: FASB.
- Gibbon, D., R. Morre, and R. Winski. 1997. *Handbook of Standards and Resources for Spoken Language Systems*. Berlin: Mouton de Gruyter.
- Griffin, P.A. 2003. Got information? Investor response to Form 10-K and Form 10-Q EDGAR filings. *Review of Accounting Studies* 8 pp. 433-460.
- Huddart, S., Ke, B., and C. Shi. 2007. Jeopardy, non-public information, and insider trading around SEC 10-K and 10-Q filings. *Journal of Accounting and Economics* 43 (1): 3-36.
- Institute of Chartered Accountants in England and Wales (ICAEW). 2013, December. *Financial Reporting Disclosures: Market and Regulatory Failures: Information for Better Markets Initiative*. Retrieved from: <http://www.ion.icaew.com/ClientFiles/a42b9c80-6acd-4dca-980a-bac45d9a324d/FRD%20final.pdf>
- KPMG. 2011. Disclosure overload and complexity: hidden in plain sight. <http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/disclosure-overload-complexity.pdf>.
- Lawrence, A. 2013. Individual investors and financial disclosure. *Journal of Accounting and Economics* 56 (1): 130-147.

- Lehavy, R., F. Li, and K. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review* 86 (3): 1087-1115.
- Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2): 221-247.
- Loughran, T. and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35-65.
- Lyon, C., J. Malcolm, and B. Dickerson. 2001. Detecting Short Passages of Similar Text in Large Document Collections, in *Proceedings of Conference on Empirical Methods in Natural Language Processing*.
- Miller, B. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6): 2107-2143.
- Nelson, K. and A. Pritchard. 2014. Carrot or stick? The shift from voluntary to mandatory disclosure of risk factors. Working paper, University of Michigan.
- Rennekamp, K. 2012. Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research* 50: 1319-1354.
- Rogers, J., Van Buskirk, A., and S. Zechman. 2011. Disclosure Tone and Shareholder Litigation. *The Accounting Review* 86 (6): 2155-2183.
- Rogers, J., Schrand, C., and S. Zechman. 2014. Do managers tacitly collude to withhold industry-wide bad news?. Working paper, University of Colorado.
- Securities and Exchange Commission (SEC). December 29, 2003. Release No. 33-8350: Commission Guidance Regarding Management's Discussion and Analysis of Financial Condition and Results of Operations.
- Securities and Exchange Commission (SEC). March 27, 2012. Center for Audit Quality SEC Regulations Committee Joint Meeting. SEC Offices, Washington D.C.
- Standard Instructions for Filing Forms Under Securities Act of 1933, Securities Exchange Act of 1934 And Energy Policy and Conservation Act of 1975—Regulation S-K. *Code of Federal Regulations*. Title 17, Pt 229, 2014 ed.
- White, M. (2013, October). *The Path Forward on Disclosure*. Speech presented at the National Association of Corporate Directors, National Harbor, MD. Accessed online at <http://www.sec.gov/News/Speech/Detail/Speech/1370539878806#.Uzx2VPIdV8E>

You, H., and X. Zhang. 2009. Financial reporting complexity and investor underreaction to 10-K Information. *Review of Accounting Studies* 14 (4): 559-586.

Appendix A: Examples of Redundant and non-Redundant 10-K Disclosures

The following two sample disclosures are pulled from the Management Discussion and Analysis section of the 2012 10-K filings for sample companies. Both disclosures relate to recent accounting pronouncements. The first disclosure, from AG Mortgage Investment Trust, is repeated verbatim in the financial statement notes section of the company's 10-K. The second, from Ford Motor Company, simply incorporates this information by reference to the financial statement notes.

AG Mortgage Investment Trust:

Recent accounting pronouncements

In April 2011, the Financial Accounting Standards Board (the "FASB") issued Accounting Standards Update 2011-03, "Reconsideration of Effective Control for Repurchase Agreements" (ASU 2011-03). ASU 2011-03 simplifies the accounting for financial assets transferred under repurchase agreements and similar arrangements, by eliminating the transferor's ability criterion from the assessment of effective control over those assets. The guidance is effective for fiscal years and interim periods beginning after December 15, 2011. We do not believe that the adoption of the amended guidance will have a significant effect on our consolidated financial statements now or in the future.

In May 2011, the FASB issued Accounting Standards Update 2011-04, "Amendments to Achieve Common Fair Value Measurement and Disclosure Requirements in U.S. GAAP and IFRS" (ASU 2011-04). ASU 2011-04 amends Topic 820 and does not modify the requirements for when fair value measurements apply; rather, it clarifies the Board's intent about the application of existing fair value measurement requirements, and changes to a particular principle or requirement for measuring fair value or for disclosing information about fair value measurements. For public entities, this guidance is effective for fiscal years and interim periods beginning after December 15, 2011. Early application by public entities is not permitted. We do not anticipate any material impact from this guidance now or in the future.

In December 2011, the FASB issued Accounting Standards Updated 2011-11, "Disclosures about Offsetting Assets and Liabilities" (ASU 2011-11). ASU 2011-11 amends Topic 210 to require additional disclosure information about offsetting and related arrangements. Entities will be required to disclose both gross information and net information about both instruments and transactions eligible for offset in the statement of financial position and instruments and transactions subject to an agreement similar to a master netting arrangement. This scope would include derivatives, sale and repurchase agreements and reverse sale and repurchase agreements. The objective of this disclosure is to facilitate comparison between those entities that prepare their financial statements on the basis of US GAAP and those entities that prepare their financial statements on the basis of International Financial Reporting Standards (IFRS). The guidance is effective for periods beginning on or after January 1, 2013, and interim periods within those annual periods. The Company does not anticipate any material impact from this guidance.

Ford Motor Company

Accounting Standards Issued But Not Yet Adopted

For information on accounting standards issued but not yet adopted, see Note 1 of our Notes to the Financial Statements.

Appendix B: Primary Variable Definitions

<i>Variable Name</i>	Definition
<i>ACQUISITION</i>	Indicator variable set equal to one if the company made an acquisition during the fiscal year, and zero otherwise.
<i>CAP_LEASE</i>	Indicator variable equal to one if the company reports a capital lease on its balance sheet, and zero otherwise.
<i>DEBT_RATIO</i>	Total liabilities scaled by total assets.
<i>DEF14</i>	Indicator variable equal to one for 10-Ks that include proxy statement information directly, and zero for firms incorporating that information only by reference to the proxy statement.
<i>DELTA_REV</i>	The annual percentage change in sales.
<i>DELTA_ROA</i>	The annual change in income before extraordinary items, scaled by lagged assets.
<i>DERIVATIVE_USE</i>	Indicator variable equal to one if the company reports any current or accumulated gains or losses on derivative transactions, and zero otherwise.
<i>FCF</i>	The average operating cash flows scaled by total assets over the current and prior years.
<i>FY_RET</i>	Raw annual return over the 12-month fiscal period.
<i>INTANG</i>	The unamortized value of purchased intangible assets, scaled by lagged assets.
<i>ITEM1_LENGTH</i>	The total number of words contained in Item 1 of the 10-K (Business and Risk Factors)
<i>ITEM7_LENGTH</i>	The total number of words contained in Item 7 of the 10-K (Management's Discussion and Analysis)
<i>ITEM8_LENGTH</i>	The total number of words contained in Item 8 of the 10-K (Financial Statements and Accompanying Notes)
<i>MATCHED_WORDS</i>	The number of distinct words that occur in six-word sequences that occur in six-word sequences in both the financial statement footnotes and earlier sections of the 10-K.
<i>NUM_GEOSEG</i>	The number of reported geographic segments.
<i>NUM_OPSEG</i>	The number of reported operating segments.
<i>OP_LEASE</i>	Indicator variable equal to one if the value of operating lease payments due in one year is greater than one percent of assets, and zero otherwise.
<i>RD</i>	Research and development expenditures scaled by lagged assets.
<i>REDUNDANCY</i>	The percentage of total three-word sequences (trigrams) contained in the 10-K that occur in both the financial statement footnotes and earlier sections of the 10-K.
<i>REF_AR</i>	Indicator variable equal to one if certain 10-K items are incorporated by reference and zero otherwise.
<i>SIZE</i>	The natural logarithm of market value of equity.
<i>SPI_D</i>	Indicator variable equal to one if the company reports any special items, and zero otherwise.
<i>TOTAL_LENGTH</i>	The total number of words in the 10-K.

Figure 1: Median *OP_COMPLEX_LENGTH*, *REDUNDANT_LENGTH*, and *RESIDUAL_LENGTH* Across Quintiles of Total 10-K Length

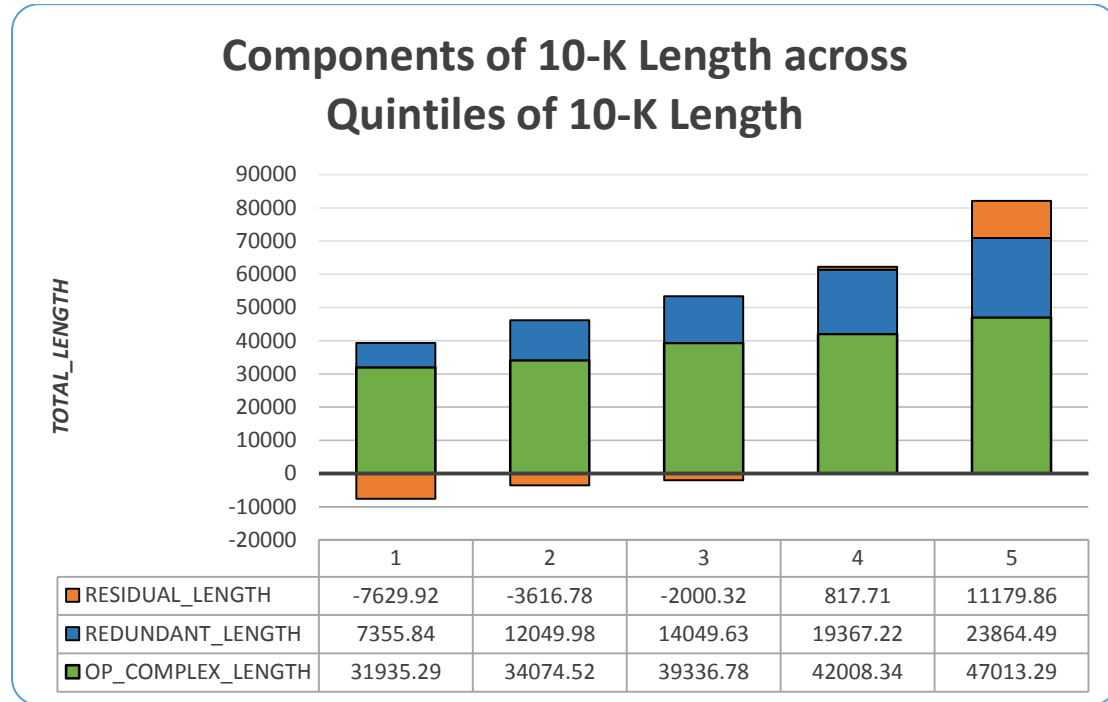


Figure 1 shows the three components of total 10-K length (*RESIDUAL_LENGTH*, *REDUNDANT_LENGTH*, *OP_COMPLEX_LENGTH*) for the median firm in each of the quintiles of total 10-K length measured in words (*TOTAL_LENGTH*). *RESIDUAL_LENGTH* is the residual from estimating equation (1) and represents 10-K disclosure length unexplained by firms' operating complexity or disclosure redundancy. *REDUNDANT_LENGTH* is the fitted value from the coefficient estimate on *MATCHED_WORDS* in Table 4 and represents the impact of redundant disclosures on total 10-K length. *OP_COMPLEX_LENGTH* is the fitted-value from equation (1) after subtracting the estimated impact of disclosure redundancy (*REDUNDANT_LENGTH*).

Figure 2: Are Recent Increases in 10-K Length Attributable to Increases in Firm Complexity?

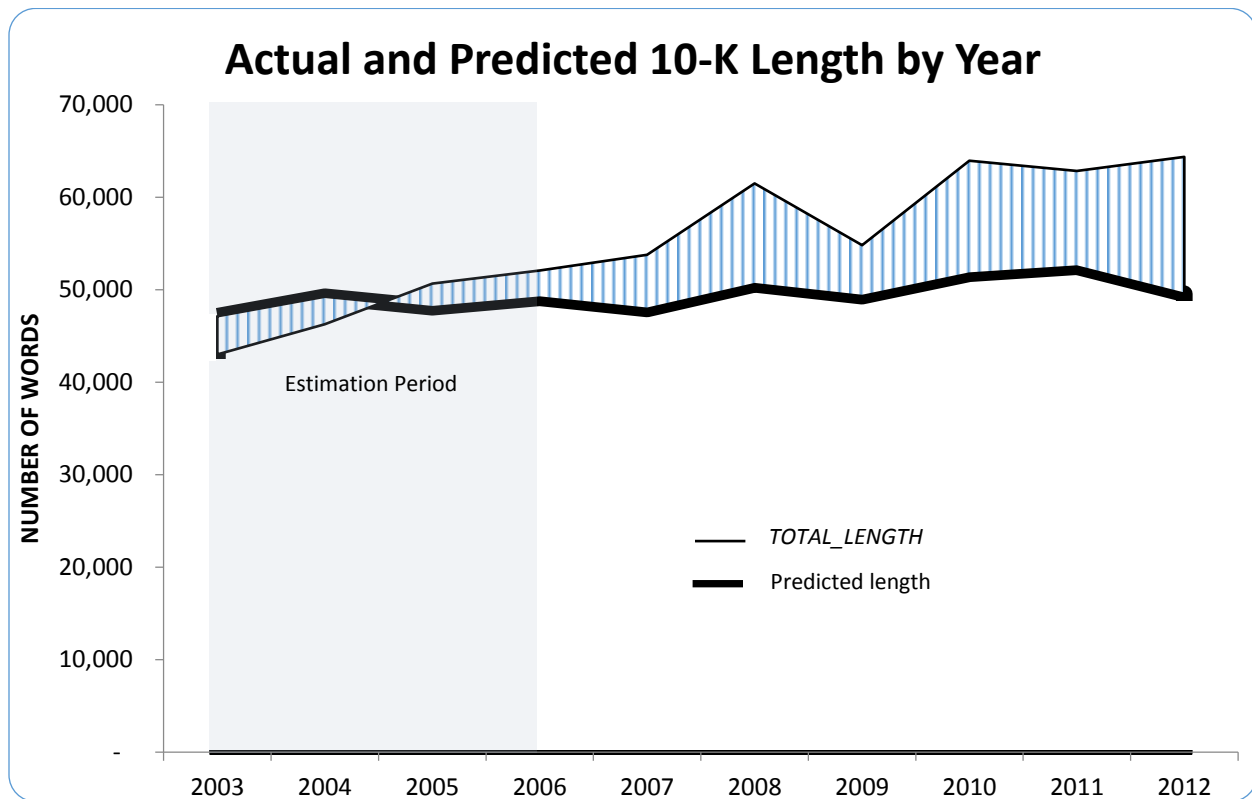


Figure 2 contrasts changes in mean *TOTAL_LENGTH* (number of words in the 10-K) over the sample period with the changes in 10-K length that would be predicted solely by changes in firm complexity during the sample period. The bold line is the fitted value of 10-K length based on coefficient estimates resulting from regressing *TOTAL_LENGTH* on firm operating complexity variables for the period 2003-2006. The relatively flat trend in the bold line suggests growth in 10-K length attributable solely to changes in firm characteristics is relatively modest. The thinner line represents actual mean *TOTAL_LENGTH* for each year. *TOTAL_LENGTH* increases by 45 percent from 2003 through 2012, whereas length predicted solely by increases in firms' operating complexity increases by only five percent over the sample period.

Table 1: Sample Selection

	Firm-years omitted	Firm- years remaining
Start with all Compustat firm-years between 2003 and 2012, inclusive.		110,344
Eliminate:		
Observations with closing share price less than \$1 or total market cap less than \$200 million	(65,736)	44,608
Observations with missing firm identifiers (CIK, GVKEY, or CUSIP)	(4,606)	40,002
Observations lost in merge between Compustat and CRSP	(3,626)	36,376
Observations with insufficient CRSP return data	(3,758)	32,618
Observations missing earnings announcement date	(2,636)	29,982
Observations missing SIC data	(241)	29,741
Observations where the earnings announcement date is more than 120 days after fiscal year end	(247)	29,494
Foreign issuers	(5,032)	<u>24,462</u>
Random selection of 650 observations from the 24,462 observations remaining		650
Missing 10-K or 10-K data is embedded in document with non-10-K data*	(21)	
Missing Compustat data required for variable construction	(9)	
Final Sample		620

Table 1 depicts the impact of each sample selection criterion on our total sample. *10-K data embedded in a document with non-10-K data generally refers to cases where certain sections of the 10-K are incorporated by reference to Exhibit 13, where Exhibit 13 (portions of the annual report to shareholders) includes the President's letter. We discard these observations because measurement of 10-K length becomes ambiguous in these cases.

Table 2: Descriptive Statistics

Panel A: Firm-level variables													
Sample							Broader Population						
Variable	N	25th Pctl	50th Pctl	Mean	75th Pctl	Std Dev	Variable	N	25th Pctl	50th Pctl	Mean	75th Pctl	Std Dev
DELTA_ROA	620	-0.008	0.007	0.020	0.033	0.093	DELTA_ROA	23,407	-0.009	0.006	0.013	0.030	0.096
DELTA_REV	620	-0.004	0.084	0.158	0.203	0.387	DELTA_REV	23,407	0.005	0.092	0.150	0.216	0.328
ACQUISITION	620	0.000	0.000	0.435	1.000	0.496	ACQUISITION	23,407	0.000	0.000	0.454	1.000	0.498
CAP_LEASE	620	0.000	0.000	0.255	1.000	0.436	CAP_LEASE	23,407	0.000	0.000	0.270	1.000	0.444
OP_LEASE	620	0.000	0.000	0.391	1.000	0.488	OP_LEASE	23,407	0.000	0.000	0.429	1.000	0.495
INTANG	620	0.008	0.070	0.172	0.274	0.214	INTANG	23,407	0.009	0.083	0.185	0.285	0.235
RD	620	0.000	0.000	0.038	0.031	0.087	RD	23,407	0.000	0.000	0.036	0.030	0.083
DEBT_RATIO	620	0.176	0.366	0.368	0.515	0.236	DEBT_RATIO	23,407	0.188	0.364	0.365	0.507	0.222
FCF	620	0.035	0.080	0.076	0.129	0.105	FCF	23,407	0.036	0.080	0.080	0.131	0.100
DERIVATIVE_USE	620	0.000	0.000	0.451	1.000	0.498	DERIVATIVE_USE	23,407	0.000	0.000	0.420	1.000	0.494
SIZE	620	6.139	7.066	7.281	8.078	1.430	SIZE	23,407	6.174	7.034	7.273	8.086	1.389
FY_RET	620	-0.112	0.114	0.187	0.376	0.494	FY_RET	23,379	-0.089	0.133	0.212	0.382	0.527
FY_VOLAT	620	0.218	0.326	0.377	0.471	0.220	FY_VOLAT	23,354	0.223	0.323	0.370	0.466	0.204
SPI_D	620	0.000	1.000	0.641	1.000	0.480	SPI_D	23,407	0.000	1.000	0.679	1.000	0.467
NUM_OPSEG	620	0.000	0.000	0.436	0.000	1.516	NUM_OPSEG	23,407	0.000	0.000	0.467	0.000	1.554
NUM_GEOSEG	620	1.000	1.000	2.331	3.000	1.975	NUM_GEOSEG	23,407	1.000	1.000	2.405	3.000	2.054

Table 2: Descriptive Statistics (*continued*)

Panel B: 10-K Characteristics							
Variable	N	25th Pctl	50th Pctl	Mean	75th Pctl	Std Dev	Q3-Q1
TOTAL_LENGTH	620	41,269	52,004	55,335	65,534	19,942	24,322
MATCHED_WORDS	620	2,375	3,679	4,014	5,367	2,481	3,042
REDUNDANCY	620	0.090	0.117	0.122	0.148	0.045	0.058
DEF14	620	0.000	0.000	0.076	0.000	0.265	-
REF_AR	620	0.000	0.000	0.086	0.000	0.280	-
ITEM1_LENGTH	620	7,614	12,122	12,980	16,775	7,122.95	9,161
ITEM7_LENGTH	620	9,586	12,403	14,019	1,6640	7,011.16	7,060
ITEM8_LENGTH	620	14,323	18,428	20,116	2,4457	8,449.6	10,167
OTHER_10K_LENGTH	620	5,560	6,948	8,209.4	9,496	5,008.8	3,936

Panel C: Sample Composition by industry			
Fama French Industry classification	N	Industry Representation in Sample	Industry representation in population
Consumer Nondurables	25	4%	5%
Consumer Durables	17	3%	2%
Manufacturing	55	9%	10%
Energy	21	3%	5%
Chemicals	15	2%	3%
Business Equipment	115	19%	17%
Telecommunications	20	3%	3%
Utilities	23	4%	4%
Wholesale/Retail Shops	53	9%	10%
Healthcare	55	9%	9%
Finance	152	24%	22%
Other	69	11%	12%
Total	620	100%	100%

Table 2: Descriptive Statistics (*continued*)

Panel D: Sample Composition by year				
Year	N	% of total	Mean <i>TOTAL_LENGTH</i>	Median <i>TOTAL_LENGTH</i>
2003	49	8%	44,345	38,764
2004	59	10%	46,259	44,139
2005	71	11%	50,579	50,515
2006	67	11%	51,524	47,370
2007	64	10%	53,768	49,135
2008	58	9%	61,390	55,685
2009	65	11%	54,455	52,657
2010	59	10%	63,277	61,467
2011	71	11%	62,285	60,576
2012	57	9%	64,239	61,882
Total	620	100%		

Panel A of Table 2 provides descriptive statistics for our random sample of 620 firm-year observations and for the larger population of 23,407 Compustat firm-years from which the sample was drawn. Variable definitions are provided in Appendix B. Panel B provides descriptive statistics regarding the 10-Ks of firms in our sample. Panel C provides a breakdown of our sample by industry. Industry definitions are based on Fama/French's 12-industry classification scheme available on Ken French's website:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_12_ind_port.html. Panel D illustrates the composition of our sample across each of the 10 years included in our sample, along with the mean and median number of words in the 10-K (*TOTAL_LENGTH*) for each year.

Table 3: Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
REDUNDANCY (1)	1.00						
MATCHED_WORDS (2)	0.57	1.00					
TOTAL_LENGTH (3)	0.12	0.63	1.00				
ITEM1_LENGTH (4)	-0.19	0.33	0.58	1.00			
ITEM7_LENGTH (5)	0.17	0.51	0.79	0.21	1.00		
ITEM8_LENGTH (6)	0.18	0.58	0.84	0.25	0.73	1.00	
SIZE (7)	0.07	0.17	0.26	0.00	0.29	0.35	1.00

Table 3 presents Pearson correlations for variables related to 10-K length, redundancy, and firm size. Variable definitions are provided in Appendix B. Correlations significant at the 5% level or lower are in bold.

Table 4: Modelling Total 10-K Length

Panel A: Regression results								
Dependent variable =			<i>TOTAL_LENGTH</i>				<i>Ln(TOTAL_LENGTH)</i>	
Independent Variables	Parameter Est	P-value	Parameter Est	P-value	Parameter Est	P-value	Parameter Est	P-value
Intercept	-3,456.38	0.582	-1,520.752	0.791	7,320.77	0.111	9.859	<.001***
REDUNDANCY	38,041.000	0.010**	39,085.000	0.008***			1.003	0.000***
MATCHED_WORDS					3.860	<.001***		
DELTA_ROA	-156.737	0.860	108.912	0.901	16.522	0.982	-0.002	0.889
DELTA_REV	-647.428	0.496	-250.002	0.792	-111.075	0.890	-0.008	0.637
ACQUISITION	2,692.081	0.074*	3,059.107	0.042**	2,523.783	0.047**	0.059	0.029**
CAP_LEASE	2,987.452	0.064*	2,770.379	0.087*	1,463.953	0.286	0.048	0.098*
OP_LEASE	-2,359.632	0.142	-2,410.793	0.135	-1,595.162	0.243	-0.047	0.102
INTANG	-3,022.347	0.375	-3,635.702	0.286	-2,379.020	0.410	-0.015	0.811
RD	-2,555.674	0.806	-1,832.526	0.860	6,437.094	0.462	0.065	0.728
DEF14	11,811.000	<.001***	11,666.000	<.001***	10,968.000	<.001***	0.191	<.001***
REF_AR	-1,188.779	0.637	-2,191.817	0.376	435.552	0.836	-0.084	0.058*
DEBT_RATIO	5,025.656	0.087*	6,008.778	0.040**	1,398.787	0.574	0.145	0.006***
FCF	-22,824.000	0.002***	-22,397.00	0.002***	-15,662.00	0.010***	-0.414	0.001***
DERIVATIVE_USE	1,140.141	0.442	1,024.634	0.488	1,840.376	0.142	0.020	0.443
SIZE	3,398.110	<.001***	3,424.719	<.001***	2,158.337	<.001***	0.049	<.001***
FY_RET	-139.258	0.916	-1,374.626	0.251	-1,697.574	0.095*	-0.019	0.369
FY_VOLAT	12,445.000	0.001***	12,654.000	0.000***	6,802.187	0.015**	0.181	0.002***
SPI_D	4,177.114	0.006***	4,086.281	0.007***	2,503.600	0.048**	0.084	0.002***
NUM_OPSEG	1,017.885	0.020**	1,014.599	0.020**	938.698	0.011**	0.021	0.007***
NUM_GEOSEG	237.775	0.548	190.460	0.630	188.894	0.574	0.008	0.251
TIME_TREND			2,065.816	<.001***	1,062.774	<.001***	0.038	<.001***
Industry fixed effects included?	Yes		Yes		Yes		Yes	
Year fixed effects included?	Yes		No		No		No	
R-squared	0.448		0.434		0.595		0.427	
N	620		620		620		620	

Table 4 (continued)

Panel B: Distributional statistics for the three components of 10-K length							
Variable	N	25th Pctl	Median	Mean	75th Pctl	Std Dev	Q3-Q1
OP_COMPLEX_LENGTH	620	33,092	38,694	39,776	46,165	9,660	13,073
REDUNDANT_LENGTH	620	9,168	14,416	15,697	20,824	9,660	11,656
RESIDUAL_LENGTH	620	-8,029	-997	0	5,904	20,622	13,933

Panel A of Table 4 presents regression results from estimating equation (1). Panel B presents distributional statistics for *OP_COMPLEX*, *REDUNDANT_LENGTH*, and *RESIDUAL_LENGTH*. *RESIDUAL_LENGTH* is the residual from estimating equation (1) and represents 10-K disclosure length unexplained by firms' operating complexity or disclosure redundancy. *REDUNDANT_LENGTH* is the fitted value from the coefficient estimate on *MATCHED_WORDS* in Table 4 and represents the impact of redundant disclosures on total 10-K length. *OP_COMPLEX_LENGTH* is the fitted-value from equation (1) after subtracting the estimated impact of disclosure redundancy (*REDUNDANT_LENGTH*).

Table 5: Modeling 10-K Section Length

Panel A: Regression results						
Dependent variable =						
Independent Variables	<i>ITEM1_LENGTH</i>		<i>ITEM7_LENGTH</i>		<i>ITEM8_LENGTH</i>	
	Parameter Est	P-value	Parameter Est	P-value	Parameter Est	P-value
Intercept	4,971.252	0.028**	-4,778.595	0.021**	-5,162.661	0.034**
REDUNDANCY	-23,987.00	<.001***	23,175.000	<.001***	24,313.000	0.000***
DELTA_ROA	-636.892	0.065*	316.909	0.315	475.557	0.200
DELTA_REV	203.350	0.584	-336.430	0.323	85.881	0.830
ACQUISITION	-99.700	0.866	935.715	0.083*	1,426.562	0.025**
CAP_LEASE	1,354.369	0.033**	891.390	0.125	1,049.407	0.125
OP_LEASE	-1,117.261	0.078*	-645.099	0.266	-688.489	0.312
INTANG	992.083	0.458	-1,977.940	0.107	-2,103.812	0.145
RD	14,190.000	0.001***	-7,892.491	0.034**	-10,172.000	0.021**
DEF14	2,764.903	0.006***	-561.874	0.536	970.233	0.364
REF_AR	-1,573.835	0.106	1,186.774	0.183	897.643	0.392
DEBT_RATIO	1,213.918	0.289	-2,696.225	0.010**	3,309.649	0.008***
FCF	-1,070.400	0.704	-9,697.554	0.002***	-12,571.000	<.001***
DERIVATIVE_USE	-511.612	0.378	278.655	0.600	1,118.715	0.074*
SIZE	300.841	0.179	1,491.530	<.001***	1,758.957	<.001***
FY_RET	-295.807	0.530	-453.287	0.293	-729.846	0.150
FY_VOLAT	2,607.830	0.041**	5,458.367	<.001***	4,282.247	0.002***
SPI_D	-367.442	0.533	2,053.787	0.002***	2,448.225	0.000***
NUM_OPSEG	176.307	0.304	32.269	0.837	567.261	0.002***
NUM_GEOSEG	-138.015	0.375	55.590	0.696	161.113	0.336
TIME_TREND	967.643	<.001***	313.931	0.000***	720.213	<.001***
Industry fixed effects included?	Yes		Yes		Yes	
Year fixed effects included?	Yes		No		No	
R-squared	0.342		0.410		0.472	
N	620		620		620	

Table 5 re-estimates equation (1) after replacing the dependent variable *TOTAL_LENGTH* with the word count of 10-K Items 1, 7, and 8, respectively. Independent variables are identical to those included in Table 4 and defined in Appendix B. *, **, and *** denote statistical significance at the 0.01, 0.05, and 0.10 levels, respectively.

Table 6: Correlating Redundancy Measures with Other Disclosure and Reporting Attributes

Dependent variable =	<i>MATCHED_WORDS</i>		<i>REDUNDANCY</i>	
Independent variable	Parameter Est	Pr > t	Parameter Est	Pr > t
Intercept	-1,106.249	0.156	0.098	<.001***
DELTA_ROA	55.052	0.653	0.003	0.209
DELTA_REV	-52.237	0.693	-0.002	0.544
ACQUISITION	32.268	0.877	-0.011	0.012**
CAP_LEASE	392.060	0.083*	0.005	0.240
OP_LEASE	-203.149	0.367	0.001	0.858
INTANG	-301.523	0.527	0.002	0.803
RD	-3,079.841	0.033**	-0.093	0.001***
DEF14	267.048	0.449	0.009	0.227
REF_AR	-861.927	0.013**	-0.018	0.010***
DEBT_RATIO	1,336.996	0.001***	0.014	0.083*
FCF	-2,041.993	0.042**	-0.029	0.143
DERIVATIVE_USE	-233.598	0.258	-0.002	0.594
SIZE	360.894	<.001***	0.003	0.041**
FY_RET	128.498	0.443	0.004	0.185
FY_VOLAT	1,518.976	0.001***	0.000	0.973
SPI_D	587.260	0.005***	0.018	<.001***
NUM_OPSEG	7.310	0.905	-0.001	0.317
NUM_GEOSEG	1.974	0.972	0.000	0.889
TIME_TREND	240.535	<.001***	-0.002	0.006***
Industry fixed effects included?	Yes		Yes	
Year fixed effects included?	No		No	
N	620		620	
R-square	0.25		0.12	

Table 6 presents from examining two measures of disclosure redundancy (*MATCHED_WORDS* and *REDUNDANCY*) as the dependent variable in equation (1). *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively. All variable descriptions are displayed in Appendix B.

Table 7: Testing for Stock Price Drift Following 10-K Filings

Dependent variable =		Six-month Size-adjusted Returns		
<u>Independent Variable</u>	<u>Parameter Est</u>	<u>Pr > t </u>	<u>Parameter Est</u>	<u>Pr > t </u>
Intercept	0.058	0.419	0.04226	0.574
FDR	-1.443	0.126	0.79629	0.395
TOTAL_LENGTH	3.88E-07	0.520		
FDR*TOTAL_LENGTH	2.22E-05	0.154		
OP_COMPLEX_LENGTH			1.11E-06	0.442
REDUNDANT_LENGTH			-9.21E-07	0.501
RESIDUAL_LENGTH			9.74E-07	0.289
FDR*REDUNDANT_LENGTH			1.19E-05	0.697
FDR*OP_COMPLEX_LENGTH			-2.88E-05	0.205
FDR*RESIDUAL_LENGTH			5.02E-05	0.023**
BETA	0.011	0.644	0.017	0.491
SIZE	-0.011	0.204	-0.011	0.223
BM	-0.015	0.408	-0.014	0.436
MOMENTUM	-0.176	<.001***	-0.177	<.001***
Industry fixed effects included?		No		No
Year fixed effects included?		No		No
N		581		576
R-squared		0.055		0.066

Table 7 presents results from regressing six-month size-adjusted stock returns on the determinants used in You and Zhang's (2009) Table 7. FDR is the size-adjusted returns for the three trading days starting on the 10-K filing date. Beta is estimated using daily returns from a value-weighted market model over three years. SIZE is the natural logarithm of market value of equity. BM is the book-to-market ratio as of the end of the fiscal year. MOMENTUM is the 6-month stock return ending in the filing month. *, **, and *** denote statistical significance at the 0.10, 0.05, and 0.01 levels, respectively.