Measuring Firm Complexity

Tim Loughran
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556-5646
Loughran.9@nd.edu

Bill McDonald
Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556-5646
mcdonald.1@nd.edu

July 9, 2020

ABSTRACT

In business research, firm size is both ubiquitous and readily measured. In contrast, complexity, another firm-related construct, is frequently relevant, but difficult to measure and not well defined. As a result, complexity is seldom incorporated in empirical designs. Measures such as the number of firm segments or the readability of a firm's financial filings are often used as proxies for some aspect of complexity. We argue that most extant measures of complexity are misspecified, one-dimensional, and/or not widely available. We propose a text-based solution as a widely available, omnibus measure of this multidimensional concept and use audit fees—which are well established as being largely driven by size and complexity—as the primary empirical framework for evaluation. Because this is a new measure, we also consider alternative contexts, including returns around 10-K filings, initial public offerings, unexpected earnings, and the COVID-19 crisis.

JEL codes: D82; D83; G14; G18; G30; M40; M41

Key words: Firm complexity; audit fees; textual analysis; Form 10-K; unexpected earnings; IPO returns; COVID-19.

We thank Brad Badertscher, Jeffrey Burks, Tony Cookson, Nan Da, Hermann Elendner, Margaret Forster, Jerry Langley, Andrew Imdieke, Mikaela McDonald, Jamie O'Brien, Marcelo Ortiz, Jay Ritter, Bill Schmuhl, and seminar participants at the Digital Innovation in Finance Conference, Humboldt University Summer Camp, Future of Financial Information Conference, University of Notre Dame, University of Connecticut, Chinese University, Georgia State University, University of Colorado, Swiss Accounting Research Alpine Camp, International Research Symposium for Accounting Academics, and Baylor University for helpful comments.

1. Introduction

Joseph Blitzstein's mantra, in his popular statistics course at Harvard, emphasizes that "conditioning is the soul of statistics." In business research, company size is almost always used as a control variable to condition regressions examining some firm-related dependent variable of economic interest. In most applications, the theoretical basis for including size is neither explicit nor precise; it is self-evident that the economic magnitude of a firm is likely to affect most posited relations between various characteristics of a business. Lacking a specific theoretical basis, size is typically measured either as the market capitalization of a firm's publicly traded stock or as total assets, with both measures log-transformed due to their power-law like distributions.

Complexity, although falling in the penumbra of size, measures a distinct and important aspect of a firm. Because a firm's complexity can be considered from many different perspectives and because it is difficult to measure, complexity is not a prominent variable in regression specifications. Complexity at the firm level can be viewed, for example, in the context of organizational structure, product logistics, accounting ambiguities, financial engineering, or information dissemination.

After attempting to more precisely define our use of the term complexity, we propose a text-based measure as a proxy of this broad construct. The measure addresses the multidimensional aspects of complexity and can be readily produced with current technologies for all firms filing Form 10-Ks (i.e., annual reports) with the Securities and Exchange Commission (SEC).[1] Extending the methods of Loughran and McDonald [2011], we create a list of 374 words most likely to increase the complexity of firms and use unique counts of these words in annual reports

[1] Throughout the paper we will use "10-K" to refer to 10-K, 10-K405, 10KSB, and 10KSB40 SEC form types. We do not include amended filings. See footnote 4 of Loughran and McDonald [2011] for an explanation of 405 types.

as our complexity proxy. Examples of our complexity words include *lease, merger, foreign, patent,* and *contract*. Of the 374 complexity words, a typical firm has a unique count of 81 (i.e., they have used 81 of the 374 words at least once in their 10-K).

Firm size and complexity are two first-order characteristics determining audit fees (Hay, Knechel, and Wong [2006]). Thus, we use audit fees as our main arbiter of success for the measure. Because "readability" is frequently used as a measure of one aspect of complexity, we spend some time discussing the limitations of this approach. We also consider other variables historically used as complexity measures in the context of predicting audit fees. In our primary results, we find that our text-based measure of complexity is statistically and economically significant and dominates alternative measures when considering all desirable aspects of an ideal metric.

Additionally, because we are proposing a novel measure, we briefly consider four alternative empirical contexts. Our measure of complexity shows a strong positive relation with absolute stock returns both on the filing date of the 10-K and in periods surrounding the filing. Complexity also exhibits a strong positive relation with the absolute value of unexpected earnings, which we would expect to the extent that complexity makes valuation more challenging. The role of complexity in initial public offerings (IPOs) is less clear, and we find a significantly negative relation in this case, which likely reflects, after conditioning on industry, the relation between complexity and the maturity of a firm. Finally, we examine the stock returns around the COVID-19 "collapse period" from Fahlenbrach, Rgeeth, and Stulz [2020] and find that more complex firms were more negatively impacted by the 2020 pandemic. Collectively and consistently in the results we are able to show that complexity is not simply a redundant measure of size.

2. Background

Attempts to measure the complexity of a publicly traded firm are numerous in prior accounting and finance research. Typically, "firm complexity" can be thought of in terms of characteristics such as accounting, business, operations, information, and reporting complications. As a proxy of firm-level complexity, previous papers have used items like the number of Compustat segments, whether or not the firm has foreign sales, the Fog Index, the number of words contained in the annual report, initialization of derivative usage, the number of methods listed in the revenue recognition disclosure, and the intangible asset percentage. Most of these measures are reasonable proxies for some aspect of firm complexity. Yet increasingly, many publicly traded U.S. firms have global sales and engage in derivative usage, making these attributes less differentiating. A transparent, omnibus measure of firm-level complexity is missing from the literature.

Our paper takes a different approach in measuring firm-level complexity. We create a list of 374 words that proxy for the complexity of various firm attributes. Like the word lists created by Loughran and McDonald [2011], our complexity list is produced by examining actual word usage in U.S. annual reports. Any word most likely implying business or information complexity is included in the word list. For our measure, we tabulate the number of complexity words that occur at least once in a firm's 10-K filing and then normalize the unique counts by the total number of complex words (374). In our audit fee sample of 52,658 of firms from 2001-2018, the average and median complexity measure is 0.22.

A nice feature of our word list is that it combines many of the prior attempts to measure various aspects of firm complexity into a single multifaceted metric. Thus, firm-level discussion in the annual report relating to M&A activity (*acquired, merger,* and *takeovers*), corporate events (*bankruptcies, partnership,* and *restructure*), legal issues (*lawsuit, litigation*, and *contract*), accounting terms (*accrete*, *carryforwards,* and *leaseback*), international operations (*foreign*, *global,* and *worldwide*), derivatives (*derivatives, hedge,* and *unexercised*), and intangibles (*patents, trademarks,* and *copyrights*) are included in our word list. In addition, we do not need to parse specific items from the 10-K such as the Management, Discussion and Analysis section—a process that can generate substantive errors—since we want the measure to capture all possible instances of the words.

The nature of the term weighting in our measure—i.e., tabulating unique occurrences—means the presence of the most commonly occurring words, such as *acquire, lease*, or *contract*, are not the defining aspects of the measure because they appear in almost all documents, although the contrapositive case (i.e., a common word does not appear) would differentiate a firm. Instead, for our proposed measure, it is the diversity of words from the complexity lexicon that accumulate to signal the overall complexity of the firm. In a subsequent section, we explain why using unique counts is more appropriate than the approach typically used in sentiment measures of tabulating proportional occurrences.

The essence of our measure is that higher distinct occurrences of words associated with complex events, transactions, and intricate business practices should be linked with larger levels of firm-level complexity. To establish the effectiveness of our word list in capturing firm-level complexity, we have two sets of tests. Our primary tests focus on audit fees, because both firm size and complexity are considered dominant explanatory variables among an endless list of candidates that have been found to be significant in that context. Second, because this is a new measure and we want to establish its relevance in a broader context, we examine a few alternative empirical settings including stock return volatility around the 10-K filing date, unexpected earnings, the initial return of IPOs, and stock returns during a critical period of the COVID-19 crisis. The multiple views of the variable also provide strong evidence that our complexity measure is not simply another manifestation of size.

Our specific prediction for the audit fee sample is that a higher proportion of uniquely used complexity words in lagged annual reports should be positively associated with higher subsequent reported audit fees. Consistent with our expectation, we find that firms with more complex business, information, and reporting complexities, after controlling for client size and other audit attributes, do indeed pay significantly higher audit fees. The relations between the unique count of complexity words and return patterns around the 10-K filing date, around earnings announcements, on the IPO date, and during the COVID-19 collapse, also support the efficacy of our proposed measure.

Within the audit fee literature, a weakness of some prior complexity measures is their non-public nature. For example, some papers have used the subjective rating of firm-level complexity provided by the actual or an experimental audit team.[2] This information is obviously unavailable to the general public. Another common empirical measure of complexity is a firm's number of subsidiaries, which is not readily available for U.S. data (but is available for European firms). Similarly, the number of business segments is available for only some firms on Compustat. In all these cases, any sample using a traditional measure of complexity as a control variable will be constrained by data availability.

For example, in our audit fee data, the inclusion of the number of segments as a control variable reduces our sample by more than 35%. This point is one we emphasize as a positive

[2] Using data for a sample of U.S. audits conducted by a large accounting firm, O'Keefe, Simunic, and Stein [1994] find a strong linkage between audit fees and perceived complexity of the client firm. Prawitt [1995], in a survey experiment, uses environmental complexity manipulation to gauge how supervisors assign specific auditors in more challenging situations.

attribute of using our measure since it is based on filings made by all firms with publicly traded securities (and certain large private firms), and these filings are required to be disclosed on a timely basis. In addition, number of segments captures only one aspect of complexity.

The prior literature has lacked a readily available, omnibus measure of the various dimensions of complexity associated with a firm. Using the language appearing in annual reports, we fill this gap in the literature by creating a list of 374 complexity words. This lexicon should capture the various aspects of business complexity that critically impact, for example, investors attempting to forecast future cash flows or auditors preparing the financial statements. Unlike some of the previous proxies for complexity, our measure is available for all U.S. firms filing on the SEC's Electronic Data Gathering, Analysis, and Retrieval (EDGAR) website.

3. Complexity and Its Measure

Many disciplines in both the natural sciences and social sciences consider complexity as an important attribute of systems they study. In some cases, such as computational complexity theory, the term's definition is relatively precise (see, for example, Goldreich [2010]), whereas in others, such as management (see, for example, Snowden and Boone [2007]), the definition is more descriptive. To better delineate complex systems, the term is frequently juxtaposed with "complicated" systems. Although there is not a bright line separating complex from complicated systems, complicated systems are ones where, despite having many layers, the layers themselves are capable of being understood to a degree of reasonable precision. Complex systems are more so characterized by unpredictability and nonlinear interactions, making them much more difficult to separate out for a more comprehensive understanding.

A car is complicated, as it can be understood primarily as the sum of its components (e.g., engine, drive train, suspension, steering, etc.), whereas traffic, because it involves interactions dictated by the diversity of human behavior, is complex. The Latin derivatives of the two terms provide additional insight, with complicated coming from "complicare" which means "to fold together", while complex comes from "cum plectere" which means "to intertwine together." Unfolding a system to better understand its components is far easier than unbraiding.

Whether the perspective is management or analyst, a complicated system can be broken down into potentially predictable components and this makes the mapping of forward-looking strategies more straightforward. Alternatively, the more complex a system, the more difficult it is to disentangle its components, and because the interaction between the components can be chaotic, predicting outcomes is much more challenging. Although many dimensions of the firm will impact its complexity and the form and magnitude of valuation-related interactions, prior measures of complexity have been relatively confined to specific characteristics of the firm. We attempt to provide an omnibus measure of complexity by defining a set of words that most likely signal a layer of complexity in the firm and then counting the number of distinct occurrences in the company's Form 10-K filing, which describes the firm's operations, risks, strategies, and finances.

Specifically, we will define the variable *Complexity* as follows. Let $c(i)_{j,t}$ represent the count in document j and time period t for word i in a lexicon of complexity words of length $C(n)$. *Complexity* is then:

$$Complexity(j,t) = \frac{\sum_i \mathbf{1}_{[c(i)_{j,t} > 0]}}{C(n)} \quad (1)$$

When tabulating word counts, a critical but often overlooked input is how the terms are weighted in the counts.[3] Most frequently, textual analytic papers in accounting and finance use the proportion of words—that is, a targeted word's count normalized by the total number of words in the document—as the term weighting method. Such applications are typically measuring sentiment, where it is unlikely that, for example, a negative word will be repeated innumerably simply as an artifact of 10-K composition. Clearly, in the case of sentiment, relative frequency is important. As previously defined, *Complexity* uses the unique counts of words from our list normalized by the length of the list, versus the proportion of complex words. That is, the number of words from the list that occur at least once in a given document. We believe that using the more common weighting scheme of proportions of total counts relative to total words would overemphasize firms that chose, for example, to itemize their leases, franchisees, or employee contracts in their filings.

For example, if we adopted the proportional count measure using the number of complexity words normalized by the total number of words, based on their fiscal year 2001 10-K report, California First National Banc would be ranked in the top 1% of the sample (high complexity), whereas, using our unique count measure it is ranked in the bottom 5% (low complexity). The word *lease* and its variants are included in our complexity lexicon. Since California First National Banc is engaged in leasing technology to mid-cap firms in the U.S., the normalized count for complex words for its 10-K is large. However, this is not a firm that most would consider complex relative to a typical firm. At the other end of the spectrum, AT&T's fiscal-year 2008 filing was unusually large, containing 264,492 words (sample average is 41,492), with much of the filing size attributable to a large number of exhibits. Their proportional measure of complex words is in the

[3] See, for example, the discussion of term weighting in Manning and Schütze [2003].

bottom 5% of the sample (low complexity), while their unique complexity count measure is in the top 5%. With AT&T's rich history of mergers and acquisitions and changing markets, we would expect them to be measured as relatively complex.

4. Literature Review

4.1 MEASURES OF FIRM-LEVEL COMPLEXITY

As a simple proxy of firm-level informational complexity, numerous papers have used the word count in the annual report. Obviously, as managers provide more text describing their company's future or past operations, investors should have increased difficulty incorporating all the annual report disclosures into stock prices. For example, You and Zhang [2009] use the median 10-K word count to categorize companies into low/high complexity groups. Bloomfield [2008] argues that firms facing adversity will have lengthier annual reports to explain their losses or other difficulties to investors. Other papers using the number of words in an annual report as a proxy for informational complexity include Lehavy, Li, and Merkley [2011], Loughran and McDonald [2014], and Dyer, Lang, and Stice-Lawrence [2017].

The number of Compustat business segments and a dummy variable equaling one if the firm has foreign sales have also been used to identify complex firms (see Doyle, Ge, and McVay [2007], Ge and McVay [2005], Ashbaugh-Skaife, Collins, and Lafond [2009], and Cohen and Lou [2012]). Hoitash and Hoitash [2018] report that a simple count of 10-K accounting items disclosed in eXtensible Business Reporting Language (XBRL) is a good measure for a firm's accounting reporting complexity. We discuss this variable in more detail in our empirical results.

The fractional percentage of intangible assets relative to total assets is sometimes used as a measure of complexity (Gomes, Gorton, and Madureira [2007]) as is the initiation of derivative usage (Chang, Donohoe, and Sougiannis [2016]). Our proposed measure of complexity attempts to improve on this diversity of measures by providing a construct that is not sample limiting due to its availability and one that is multidimensional in its purview.

4.2 READABILITY

Another firm specific variable related to complexity and used in the literature is the Fog Index. The Fog Index is a combination of two variables: average sentence length (in words) and complex words (fraction of words with more than two syllables). This readability measure estimates the number of years of formal education needed to comprehend the text in an initial reading. Li [2008] reports that the median Fog Index value for annual reports is 19.24, which implies that the reader needs slightly more than an MBA level of education to understand the document in a first reading. Although Jones and Shoemaker [1994] sharply criticize, and Loughran and McDonald [2014] empirically discredit, the use of the Fog Index, a number of accounting papers have continued to use it as a readability/complexity measure (see Li and Zhang [2015], Guay, Samuels, and Taylor [2016], and Lo, Ramos, and Rogo [2017]).

Even if we ignore the empirical results of Loughran and McDonald [2014], where the dominant words driving readability scores are virtually all relatively common business words, the objective of the most frequently used measure—the Fog Index—is not at all clear.[4] Any reading of a sample of 10-Ks makes evident that writing style, in terms of vocabulary and density, is not something that varies much at all in the cross-section of firms. And, if it did, it would still not be clear what the objective was for readability, i.e., surely you would not want to minimize the score.

[4] Word counts have a power-law distribution, much like market capitalization, where a small subset of words account for a major portion of the total counts. Table IV of Loughran and McDonald [2014] shows that 52 words from the approximately 48,000 complex words appearing in 10-Ks account for more than 25% of the total complex word count in the Fog Index. All of the words are relatively common business terms, with the first five being *financial, company, interest, agreement,* and *including*.

Attempts to use alternative readability measures such as Flesch-Kincaid or the Bog Index do not overcome this fundamental criticism.

Leuz and Wysocki [2016] emphasize that it is impossible to disentangle the documents from the business, leading Loughran and McDonald [2016] to conclude that the broader topic of complexity might be a more appropriate way of addressing the attribute readability measures intend to capture.

4.3 COMPLEXITY AND AUDIT FEES

Hay, Knechel, and Wong [2006] provide a survey of published studies on auditing and note that empirical research has clearly identified size, complexity, and risk as central components in determining audit fees. They consider 147 papers with 186 distinct independent variables. In their meta-analysis, size is the dominant factor in determining audit fees, typically accounting for around 70% of the variation in fees. Obviously, the higher are total assets, the more effort the auditor would likely expend to prepare the financial statements, thus the higher the audit fees should be. Another common measure of firm size is a dummy variable indicating membership in the S&P 500 Index (Chaney and Philipich [2002]). The empirical auditing literature clearly verifies that larger firms pay more in audit fees.

Second in their discussion of fee attributes is complexity. They identify 33 metrics in prior research used to proxy complexity, with two of the most common being the number of subsidiaries or segments. They conclude that complexity is clearly relevant and the strongest results are for the number-of-subsidiaries proxy.

Risk as assayed in Hay, Knechel, and Wong [2006] focuses on the risk of error or specialized audit procedures, consistent with the models of Simunic [1980] and Stice [1991]. The most common attributes used to measure this concept are relative levels of inventories and receivables, and Hay, Knechel, and Wong note that the combination of the two accounts seems to be more effective than considering them separately.[5]

Although early work suggests that top-tier auditors charge less in fees due to economies of scale (Simunic [1980]), more recent evidence finds that the top 4, 5, 6, or 8 auditors are associated with significantly higher fees (Palmrose [1986] and Hogan and Wilkins [2008]). The reputation of auditors should have significant value that warrants increased compensation for their services (Balvers, McDonald, and Miller [1988]). Since auditors expose themselves to increased litigation risk if their client goes bankrupt, numerous papers have included a dummy variable for negative net income (Carcello, Hermanson, Neal, and Riley [2002], Whisenant, Sankaraguruswamy, and Raghunandan [2003], and Hogan and Wilkins [2008]). Hay, Knechel, and Wong [2006, page 171] note that "… the most recent results suggest that the existence of a loss for a client has become an increasingly important driver of audit fees."

Some of the prior evidence finds that financial institutions tend to pay less in audit fees than other industries. Part of the lower fees is driven by banks having limited receivables, inventory, and intellectual-based assets (Hay, Knechel, and Wong [2006]). However, the financial meltdown of 2008 dramatically exposed bank auditors to enormous client risk and substantially increased the average audit fee in this sector. Thus, regressions with audit fees as the dependent variable should incorporate both industry and time dummies as controls.

In sum, a large number of variables have been shown to be relevant in some context for predicting audit fees. For independent variables such as profitability, leverage, and ownership form, the results are mixed, with the significance of these candidates varying across samples and applications. Undoubtedly, at the margin, myriad variables affect the dollar amount auditing firms

[5] Of the 129 analyses considered in Hay, Knechel, and Wong [2006], more than 71% use some combination of inventory and/or receivables as the proxy for risk.

charge, but empirical studies to date document size, complexity, and risk as the three dominant factors influencing audit fees.

5. Audit Fee Sample Creation

5.1 DATA SOURCES AND VARIBLE DEFINITIONS FOR THE AUDIT FEE SAMPLE

The auditing literature provides a clear framework where the concept of complexity is a critical component in explaining audit fees. Thus, in our empirical analysis we will focus initially on our audit fee sample and in subsequent sections consider alternative empirical contexts.

To create the audit fee sample, we download and parse all 10-K, 10-K405, 10KSB, 10-KSB, and 10KSB40 filings, excluding amended documents, from the SEC's EDGAR website (www.sec.gov) and combine them with firm-level audit data from Audit Analytics. Although the EDGAR data is available and updated on a daily basis since 1994, the Audit Analytics data limits our sample to the filing years 2001-2018. Compustat is the source of income statement and balance sheet control variables and we use their "All Segments" measure for segments. We use the Center for Research in Security Prices (CRSP) for stock returns and for cases where we use market capitalization as an alternate measure of size.

Table 1 shows how our various data screens affect the original sample of audit data and SEC filings. The initial sample from Audit Analytics includes private firms and is reduced from more than 200,000 observations to approximately 126,000 in the process of matching with the EDGAR filings. Notice that the sample of publicly traded firms from EDGAR is reduced by more than 35% by the data requirement for segment data, which we consider as a competing measure of complexity. The final audit sample with complete data is 52,658 firm-year observations from 2001-2018.

Consistent with prior research, we will use the natural log of audit fees as the dependent variable in our initial regressions and, for size—the dominant control variable in audit fee studies—we use the log of total assets. Some of the additional control variables and data sources are described below.

All of our independent variables are known to investors *before* the disclosure of the audit fees. The well-established control variables available from Audit Analytics include: *Top 5 Auditor Dummy* (a dummy variable set to one if the auditor is either PricewaterhouseCoopers, Ernst & Young, Deloitte & Touche, KPMG, or Arthur Andersen, else zero), and *S&P 500 Dummy* (a dummy variable set to one if the firm is listed in the Standard & Poor's 500 Index, else zero). From Compustat we obtain, *Total Assets* (as of the fiscal year end); *Loss Dummy* (a dummy variable set to one if the firm's net income is less than zero, else zero); *% Leverage* ((short-term debt + long-term debt)/total assets); *% Goodwill* (goodwill/total assets); *% Inventory + Receivables* ((inventory + receivables)/total assets); and the *Segments* variable ("all segments").

We also include the total number of words in the 10-K as one of our control variables (*Total Words*), since it has been used as a measure of complexity in the past, and to assure that *Complexity* is not simply capturing the same construct in the data. Because of varying time and industry effects during the course of our sample period, we incorporate calendar year and industry dummies in our regressions. Detailed definitions for variables used in the alternative tests appear in appendix A.

6. The Complexity Word List and Measured Outcomes

6.1 THE COMPLEXITY WORD LIST

The complexity word list is created in a manner similar to the word lists created by Loughran and McDonald [2011]. Only tokens appearing in the 2018 Loughran and McDonald Master

Dictionary (https://sraf.nd.edu/textual-analysis/resources/) can potentially be included in the list. Their Master Dictionary excludes proper nouns, single character letters, and acronyms.

Words the typical reader of the annual report would view as adding to the firm-level complexity for forecasting subsequent cash flows or creating audited financial disclosures are included in our complexity word list. For example, annual report language describing *leases, intangible* assets, *international* operations, or *acquisitions* would make forecasting operating performance or the auditing of financial statements more challenging. To facilitate the ability of other researchers to use the complexity lexicon, the entire list of 374 complexity words is reported in Table 2 and is available at https://sraf.nd.edu.

In the next two subsections, we will first consider the raw counts of the words from our lexicon to better understand how the measure works and then look at the measure's extremes to determine if the intended metric effectively maps into actual outcomes.

6.2 MOST FREQUENTLY APPEARING COMPLEXITY WORDS

In text-based research, showing the reader which words have a disproportionate impact on the results is a critical part of exposition. As argued by Loughran and McDonald [2016], transparency is critical to ensure the results from new word lists are not driven by misclassifications.

In column (1) of Table 3, we report the most and least frequently occurring lemmas based on our word list. Note that for purposes of this tabulation we use lemmas, the root form of the words, so that the counts are all tabulated under the same root word. For example, we count the 17 forms of the word *lease* under the root term *lease.* Since we are initially trying to provide some general sense of the use of these words, these tabulations are not based on unique occurrences and simply represent the raw word counts.

There are 69 unique lemmas in our list of 374 complexity words. The most frequently occurring complexity lemmas are all common business terms that reflect some layer of complexity in a firm. As might be expected, the top tokens, such as *acquire*, *lease*, *contract*, and *subsidiary* all appear in more than 90% of all annual reports. Conversely, words like *nonmarketable*, *conglomerate*, *subtenant,* and *sovereign* appear in less than 10% of the filings. Misclassification of tokens does not appear to be an issue with our proposed word list.

These statistics highlight an important characteristic of our proposed measure. Because we use unique counts to calculate *Complexity* (see equation (1)), the absence of common words will be more important than their presence. That is, firms that do not use the token *lease* or *contract* are unusual since these are relatively common business terms. Our measure attempts to distinguish the layers of complexity by the diversity of occurrence of complex words.

We argued before that using proportional measures (raw counts divided by total words) is not appropriate in this context, however note that by using all inflections of the complexity terms in our lexicon (versus lemmas), we implicitly allow terms to take on at least some additional weight to the extent they are used in their various forms.

Column (2) of Table 3 looks at the dynamics of the word frequencies from the first third of the time-series to the last (2001-2006 to 2013-2018). In this case, we tabulate a given lemma's proportion of the total count (not unique count) of complexity words. Not surprisingly, two of the lemmas increasing in use pertain to globalization (*global* and *foreign*). The bottom 10 in this case, show the words with the most significant decline in relative usage; for example, the token *bankrupt* was down in the later third of the sample by approximately 30%.

6.3 WHICH FIRMS ARE MEASURED AS COMPLEX OR NOT COMPLEX USING OUR PROXY?

As designed, our measure should reflect the diversity of word usage from our lexicon, and this, in turn, should reflect the various layers of complexity in a firm. As with all text-based measures, checking the outcomes for reasonableness is critical. Unconditionally in the audit fee sample, the three firms with the highest measure of complexity (using a little more than 47% of the complexity terms) are BearingPoint (2006), Inamed (2002), and Horizon Therapeutics (2016). BearingPoint is an international management and technology-consulting firm spun off from KPMG. After going public in 2001, the company went through a number of acquisitions, was losing money in most years, and was late filing in 2006 due to material weaknesses in their financial reporting. Their 10-K's description of the Audit Committee and SEC investigation of the firm is replete with issues relating to employee misconduct and difficulties with clients. This filing had 59 exhibits, while their prior filing had only 15. The second case, Inamed, is a global medical device manufacturer with 160 patents and 90 trademarks in a regulated industry subject to substantial litigation risk. Their primary product is breast implants and this year was when legal proceedings claiming bodily injury were beginning to explode. In the third example, Horizon Therapeutics is a biopharmaceutical company formed in Ireland. Although, as recorded in Compustat and their 10-K, they have only three primary business units, they are actively involved in acquisitions to develop a pharma portfolio. Their 10-K has an entire segment explaining the complications caused by being subject to Irish laws. And, by far, Pharmaceuticals is the industry with the highest average complexity score.

The lowest scoring company across the entire audit sample is Cagles (2003), a domestic poultry processing firm that owns its own hatcheries and produces its own food. They reported zero research and development over the past three years, had a 10-K with only 4,218 words (the 10-K is relatively primitive in its format), and a market capitalization of about $23 million.

These extreme examples from the full sample are relatively small firms in specific years where their idiosyncratic filings could just happen to reasonably reflect the intended measure. If we focus instead on the average score across the time series for S&P 500 firms (with a minimum of five observations), the results also resonate with the intent of the measure and indicate that it can differentiate even among a sample of the very largest firms. For the S&P 500 firms, Exelon, PPL, and Coty are ranked as the most consistently complex based on our measure. Exelon is a utility services holding company whose 10-K is filed for nine entities. They are actively involved in mergers and are dominant players in the gas and electric markets, and, more notably, nuclear energy. Its 2018 10-K is in the top 1% of filings in terms of total number of words and contains a five-page glossary of terms and abbreviations. PPL is another utility holding company whose 10-K is similar to Exelon in its complexity. Coty is a global portfolio of 75 brands primarily in the beauty industry, acquisitive, and involved in patents and trademarks. The lowest two firms based on the time-series average in the S&P 500 are Chipotle and Netflix, both highly focused firms in relatively well-defined markets. In all cases, the sample extremes seem to reasonably align with the targeted attributes.

7. Empirical Results

7.1 AUDIT FEE SAMPLE SUMMARY STATISTICS AND CORRELATIONS

Panel A of Table 4 reports the summary statistics for our variables while Panel B of Table 4 presents the correlations. In the audit fee sample, for consistency we work with the dataset derived in Table 1, where all of the observations have complete data, producing 52,658 observations from filing years 2001-2018. For this sample, *Complexity* has a mean and median of 0.22, indicating that a typical 10-K uses 81 of the 374 complexity words. Not surprisingly, *Complexity* has trended upward from 0.18 in 2001 to a high of 0.25 in 2018.

The mean of *Audit Fees* is $1.77 million with a median of $0.73 million. The rate of inflation over the entire 2001-2018 sample period was approximately 43%. Even in real terms, the average audit fee has gone up substantially, with the inflation-adjusted mean increasing by about 300% and the inflation-adjusted median increasing by a little more than 200%. *Total Assets* averaged about $3.7 billion with a median of $380 million over the sample period. Although both *Total Assets* and *Complexity* increased over the 2001-2018 time period (median *Total Assets* increased in real terms by a little less than 300%), the increase in audit fees is also likely attributable to the increases in regulation during this period.

We include *Total Words* as another variable that has been historically used to measure complexity. This variable also increased over the sample period, increasing in both mean and median by about 100%, with an overall mean (median) of 49,224 (41,837). The dominance of the top-tier auditors, captured by *Top-5 Auditor Dummy*, is apparent with almost three quarters of the sample using one of the top auditing firms. The average number of *Segments* based on the Compustat data is 5.26, with 80% of the sample ranging from 1 to 10 segments.

In Panel B of Table 4, the correlations among the key variables are reported. *Audit Fees* and *Total Assets* are transformed using natural logs. Consistent with our assertions, *Complexity* and *log(Audit Fees)* are positively correlated (0.48). More unique usage of complexity words like *merger*, *leases, hedged*, and *global* appearing in the annual report is associated with higher subsequently reported auditor fees. One concern in measuring complexity is its ability to discern this attribute from size. The correlation of 0.37 between *log(Total Assets)* and *Complexity* suggests that we are not simply re-measuring size.

As previously discussed, *Segments* is frequently used as a measure of complexity in audit fee research. Notice that our omnibus measure has a relatively low correlation (0.22) with *Segments*. Typically used as a measure of risk, *% Invent. + Rec.* has a small negative correlation with *log(Audit Fees)*, however we will see in the regressions that the variable takes on a significantly positive value in the context of other control variables. Hay, Knechel, and Wong [2006] note that total assets, on average, explains about 70% of the variation in audit fees. Not surprisingly, in our sample, the correlation between *log(Audit Fees)* and *log(Total Assets)* is relatively high at 0.85.

7.2 REGRESSIONS WITH AUDIT FEES AS THE DEPENDENT VARIABLE

Can the text contained in an annual report capture the various aspects of firm-level complexity? Are auditor fees higher for firms using a broader array of the complexity lexicon and is this effect distinguishable from firm size? In Table 5, we estimate regressions with *log(Audit Fees)* as the dependent variable:

$$\begin{aligned} log(Audit\ Fees)_{i,t+1} = \alpha &+ \beta_1 Complexity_{i,t} + \beta_2\ log(Total\ Assets)_{i,t} + \beta_3 log(Total\ Words)_{i,t} \\ &\beta_4 Top\text{-}5\ Auditor\ Dummy_{i,t} + \beta_5 S\&P\ 500\ Dummy_{i,t} + \\ &\beta_6 Loss\ Dummy_{i,t} + \beta_7 \%\ Leverage_{i,t} + \beta_8 \%\ Goodwill_{i,t} + \\ &\beta_9 (\%\ Invent._{i,t} + Rec._{i,t}) + \beta_{10} Segments_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{2}$$

where $Complexity_{i,t}$ is the unique count of words from our complexity word list divided by 374 (see equation (1)). The dependent variable, $log(Audit\ Fees)_{i,t+1}$, is the natural log of the dollar amount of audit fees reported after the Form 10-K filing date. As indicated by the subscripts, our independent variables are all known at the time of the audit fee disclosure date.

As noted before, there is a strong upward trend in audit fees during our sample period. The banking sector went from having typically cheaper average auditor fees to being the most expensive industry following the 2008 financial meltdown. Thus, regressions include an intercept, Fama and French [1997] 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with standard errors clustered by year and industry.

In the first column of Table 5, as a baseline test, the only independent variable is *Complexity* along with the industry and year dummies. The coefficient on *Complexity* is positive (11.14) and statistically significant at the 1% level (*t*-statistic of 16.87). Higher unique complexity language usage in the annual report is associated with higher subsequent auditor fees. The R-squared value in the regression is more than 40%. The second column appends *log(Total Assets)* to the regression as a control variable. Not surprisingly, *log(Total Assets)* has a positive coefficient (0.49) with a *t*-statistic of 43.62, and the R-squared value increases to about 82%. The coefficient on *Complexity* declines to 3.57 while the *t*-statistic increases to 18.73.

In column (3), we include the eight other control variables in the regression. Both *log(Total Assets)* and *Complexity* remain positive and significant. The R-squared increases to about 85%. Of the traditional accounting variables, all have the expected sign and, with the exception of *% Leverage,* are significant. Firms with relatively higher goodwill and inventories + receivables, a top-5 auditor, S&P 500 Index membership, more segments, and negative net income are significantly related to higher auditor fees. These results are consistent with the vast majority of the past audit fee literature (see Hay, Knechel, and Wong [2006]).

Of the alternative complexity measures, *log(Total Words)* is not significant, however *Segments* remains positive and significant. Thus *Segments* appears to still capture some aspect of complexity, however we would argue that its measure is far more one-dimensional. To the extent segments are in diverse locations, their impact on the cost of auditing might overwhelm the actual increase in complexity associated with segment counts.[6]

---

[6] As in prior studies, we use the natural log of audit fees to account for the skewness in the variable. The sample contains some extraordinary cases, e.g., AIG paid PWC $97 million in audit fees in 2008 (see AIG's DEF 14A filed

In Table 6 we consider some alternative views of the audit fee sample and examine the robustness of the results to alternative assumptions. In the prior regressions, we used audit fees as reported by Audit Analytics as the dependent variable. Although many firms reported audit-related fees prior to 2003, the requirement for this information became more refined with the SEC's Financial Reporting Release in 2003. In column (1) of Table 6, we substitute *log(Audit Fees + Audit-Related Fees)* for the dependent variable. The mean (median) audit-related fee is $186,999 ($9,600). The regression is otherwise identical to the one in column (3) of Table 5, however we only present the *Complexity* and *log(Total Assets)* coefficient and *t*-statistic values. The results remain essentially identical.

One of the previous points about the relative merits of our complexity measure is that it is available at no cost for all publicly traded firms in the U.S. In column (2) of Table 6, we show a simple regression of *Complexity* along with the industry and year dummies on the full SEC sample of 10-Ks from the same period. The sample in this case increases from 52,658 to 126,286. *Complexity* remains positive and significant and the regression explains more than half of the variation in audit fees.

Finally, in column (3) we consider a recently proposed measure of accounting reporting complexity by Hoitash and Hoitash [2018]. Their accounting reporting complexity (*ARC*) measure counts the number of distinct monetary XBRL tags in Item 8 (Financial Statements and Supplementary Data) of a firm's SEC filing.[7] Clearly, their measure should provide a reasonable proxy for one dimension of complexity. It does not, however, account for the differences in complexity associated with various accounts that are typically detailed in footnote disclosures.

on 2008-04-04) and Bank of America paid PWC about $96 million in both 2011 and 2012 (see their DEF 14A filings on 2012-03-28 and 2011-03-30). If we run the regressions in Table 5 and exclude the top 5% of the sample, the results remain essentially identical.

[7] The Hoitash and Hoitash [2018] XBRL data is obtained from http://www.xbrlresearch.com/.

They document that their measure more consistently captures their outcome variables in comparison to other measures of complexity such as segments or readability. The downside to *ARC* is that the implementation for XBRL occurred over a three-year period beginning in 2009, thus eliminating much of the EDGAR time series and their measure focuses on only one aspect of complexity.

In addition, it is not clear what the impact of the integration of inline-XBRL—which is being phased in beginning in 2019 but also appears as early as 2016 for some firms—will have on this measure. In the column (3) regression, we append *log(ARC)* to the regression including all of the control variables. Although the coefficients and *t*-statistics on *Complexity* and *log(Total Assets)* decline, they remain positive and significant. As expected, the coefficient on *log(ARC)* is positive and significant. *ARC* provides an interesting alternative measure of complexity, however, we would argue that it substantially limits the sample and is relatively one-dimensional. If the objective of a study were simply to measure accounting reporting complexity, then this measure would certainly be one of the inputs.

7.3 FILING DATE STOCK RETURNS

We also use the audit fee sample to consider the impact of complexity and size on stock returns surrounding the filing date. Prior research (see, for example, Griffen [2003]) finds that the immediate impact of 10-K filings on stock returns is surprisingly modest. The concept of complexity does not suggest any hypotheses concerning directional stock returns; however we would expect the volatility of returns, as measured by the average absolute return, to be higher for complex firms. We also include *log(Total Assets)* in the regression and the industry and time dummies. We would expect larger firms to have less stock price variability, however to the extent size is capturing complexity, this relation is less clear. Relative to the 10-K filing date (t=0) we consider three periods: [t-20, t-1], [t, t+1], and [t+2, t+10] and measure the absolute value of the stock returns over each period. In the last two periods, where the prior-period volatility is available we include this as a measure of lagged volatility.

Over the three periods, *Complexity* is consistently positive and statistically significant. The *log(Total Assets)* variable is, over the three periods, negative and statistically significant. The signs and significance of the variables are encouraging to the extent this shows that *Complexity* is not simply another proxy for size. Interestingly, however, the relation is not notably different on the filing date, versus the weeks surrounding that date.

7.4 ALTERNATIVE CONTEXTS

Because we are proposing a new measure, we consider alternative sample contexts to empirically support its application. The results are intended to be descriptive and we do not formally develop the regression models. In Table 8, we examine standardized unexpected earnings, initial public offerings, and recent returns around the COVID-19 pandemic.

In column (1) of Table 8, we use the absolute value of analyst forecast error (i.e., standardized unexpected earnings) as the dependent variable with complexity and size measured from data available prior to the earnings announcement date. We use the programs and data available on Wharton Research Data Services (WRDS) to create the measure of unexpected earnings. Consistent with research on this topic, we use price to normalize the difference between announced earnings and analysts' estimates. To avoid explosive values in the estimate, typically these samples are filtered based on some minimal price deflator. We only include observations where the price is greater than or equal to $2, however our results remain essentially identical without this filter.

We would expect complex firms to have a wider range of unexpected outcomes in terms of analysts' ability to estimate earnings and we would expect, in general, larger firms to have relatively more predictable earnings. Consistent with these priors, the coefficient on *Complexity* is positive and statistically significant and the coefficient on *log(Total Assets)* is negative and significant. Complexity, as measured here, decreases the precision with which earnings can be estimated, and once again is clearly measuring something distinct from size.

In column (2) of Table 8, we consider the sample context of initial public offerings using data obtained from SDC Platinum with additional items and corrections supplied by Professor Jay Ritter. Initial returns is the percentage change from the IPO offer price to the CRSP closing price. During 2001-2018, the average initial return is 14.6% for our sample of 1,813 IPOs. We select as control variables in this regression variables typically included in estimating the initial return of IPOs (see Loughran and Ritter [2004] and Lowry, Officer, Schwert [2010].

For IPOs, there are potentially two different effects related to complexity. As suggested by the prior unexpected earnings regression, estimating the value of the firm should be more challenging as complexity increases, suggesting a positive coefficient. However, firms typically evolve to become more complex. IPOs tend to be young firms in an early stage of development. The median age from founding date to IPO date of our sample is 11 years. Thus, complexity in this case also reflects the maturity of the firm, in which case we would expect a negative coefficient.

Our results suggest that the latter hypothesis dominates, which is not surprising given the importance of a firm's maturity, beyond simply age, in affecting the ability to project future cash flows. The relation between initial returns and our control variables are generally consistent with the prior literature for the time period examined. For example, the variable *Up Revision* (upward revision in the IPO offer price from the mid-point of the filing range) has consistently been linked with higher first-day returns. Likewise, the coefficient on *log(Total Assets)* is not significant, which was shown by Loughran and Ritter [2004] for the post-internet bubble period. If we use the natural log of the IPO proceeds or sales as the measure of size in the regression (versus total assets), the coefficients on *Complexity* in both cases becomes more negative (-0.43 and -0.40, respectively) as does the *t*-statistics (-2.78 and -2.92, respectively).

Finally, in column (3) of Table 8 we consider stock returns for S&P 500 firms during the recent COVID-19 pandemic using the framework of Fahlenbrach, Rageth, and Stulz [2020].[8] We include some of the variables most relevant from their analysis: market capitalization, long-term debt-to-total assets ratio, and cash-to-total assets ratio. The dependent variable includes the period 2020-02-02 through 2020-03-23, which they label as the "collapse period." Because the pandemic's effects were highly varied across industries, we also include industry dummy variables. Consistent with their results, larger firms had less negative returns during the collapse, as did firms with more cash. Firms with more debt were negatively impacted, however this variable was not statistically significant in our sample. After controlling for these variables, we would expect more complex firms to be more negatively impacted. Our results indicate that the coefficient on *Complexity* is negative and statistically significant. The COVID-19 crisis created a situation where all of the uncertainties of complex interactions were magnified, resulting in more negative returns for complex firms.

8. Conclusion

Our paper provides the literature with an omnibus measure of firm-level complexity. The complexity word list is created by selecting words from management's description of their

[8] Due to dual listings, the S&P 500 consists of 505 stocks. For dual listed securities, we use the listing that is most liquid. Companies being added or deleted to the index or experiencing major merger activity were excluded, leaving a sample of 497 stocks.

business, as detailed in a 10-K filing, that would typically be associated with greater complexity of a firm. The data required for the measure is available at no cost for all publicly traded firms in the U.S. Some of the most frequently occurring words from the list are *acquire*, *lease, contract,* and *subsidiary.* However, for our measure it is not the frequency of use that counts, but the diversity of words used from the list.

The first setting selected to gauge the complexity word list is in the prediction of audit fees, where a rich literature has clearly identified complexity to be a critical component. The secondary tests consider the relation of complexity to stock returns around the 10-K filing date, unexpected earnings, initial public offerings, and the COVID-19 crisis.

We find a strong association between the unique count of complexity language in the annual reports and subsequent audit fees. Increased discussion of *intangible* assets, *acquisitions*, *foreign* operations, or *subsidiaries* by managers is linked with significantly higher fees charged by the auditors. The variable also is shown to be significant in the other sample contexts and consistently seems to be differentiated from firm size.

Complexity is, and will likely remain, an amorphous yet important attribute of the firm. Similar to firm size, when examining firm-related economic phenomena, complexity is a characteristic that frequently merits inclusion in a regression specification, typcially as a control variable. It is not unrelated to size, but it is a distinctly different attribute affecting the inputs and outputs of corporations. At the same time, complexity is multidimensional and not precisely prescribed by a specific economic theory. Traditional quantitative measures of complexity are limited in the breadth of what they measure and in most cases the availability of data. A firm's 10-K report discusses in detail the business, operations, accounting, strategies, and other aspects of the firm, which, in turn, provides a collection of terms that potentially capture the varied dimensions of complexity. Any attempt to measure variables such this will be imperfect, but our proposed measure is widely available, multidimensional, and, importantly, appears to be empirically valid.

Perhaps this is an example where textual analysis provides a clear advantage over traditional quantitative metrics. That is, textual analysis, in this case, provides a means of measuring an important yet not precisely defined multidimensional attribute that is not readily measured with observable quantitative data.

APPENDIX A

Definitions of Variables for Alternative Tests

Panel A: Unexpected earnings data

Abs(unexpected earnings)	Defined as absolute value of (Actual EPS minus median IBES EPS estimate) scaled by stock price. Stock price must be at least $2.00 for the observation to be included in the sample.
log(Total Assets)	Defined as the natural logarithm of total assets as of the prior fiscal year.

Panel B: IPO data

log(1+IPO Initial Return)	Defined as the natural logarithm of 1 plus the percentage change from the IPO offer price to the CRSP closing price on the IPO date.
Up Revision	Defined as the percentage upward revision in the IPO offer price from the mid-point of the filing range if the offer price is greater than the mid-point, (i.e., ((offer price-mid-point)/mid-point) x 100), if offer price > mid-point, else zero.
Venture Capital	Dummy variable set to one if IPO is backed by venture capital, else zero.
Top Tier IB	Dummy variable set to one if the lead underwriter of the IPO has an updated Carter and Manaster [1990] rank of eight or more, else zero.
NASDAQ	Dummy variable set to one if IPO lists initially on NASDAQ, else zero.
NASDAQ15	The buy-and-hold returns of the CRSP NASDAQ value-weighted index on the 15-trading days prior to the IPO date, ending on day t-1.
log(1+Age)	Defined as the natural logarithm of 1 plus the number of years since the firm's founding date as of the IPO.

Panel C: COVID-19 data

COVID-19 Stock Return	Defined as the firm's percentage stock returns from February 2, 2020 to March 23, 2020.
log(Mkt Cap)	Defined as the natural logarithm of number of shares outstanding times stock price as of December 31, 2019.
Long-term Debt Ratio	Defined as long-term debt/total assets as of the prior fiscal year.
Cash Ratio	Defined as (cash and cash equivalents + short-term investments) divided by total assets as of the prior fiscal year.

REFERENCES

Ashbaugh-Skaife, H.; D. W. Collins; and R. Lafond. "The Effect of SOX Internal Control Deficiencies on Firm Risk and Cost of Equity." *Journal of Accounting Research* 47 (2009): 1-43.

Balvers, R. J.; B. McDonald; and R. E. Miller. "Underpricing of New Issues and the Choice of Auditor as a Signal of Investment Banker Reputation." *Accounting Review* 63 (1988): 605-622.

Bloomfield, R. "Discussion of "Annual Report Readability, Current Earnings, and Earnings Persistence"." *Journal of Accounting and Economics* 45 (2008): 248-252.

Carcello, J. V.; D. R. Hermanson; T. L. Neal; and R. A. Riley Jr. "Board Characteristics and Audit Fees." *Contemporary Accounting Research* 19 (2002): 365-384.

Carter, R.B., and S. Manaster. "Initial Public Offerings and Underwriter Reputation," *Journal of Finance* 45 (1990): 1045-1068.

Chaney, P. K., and K. L. Philipich. "Shredded Reputation: The Cost of Audit Failure." *Journal of Accounting Research* 40 (2002): 1221-1245.

Chang, H. S.; M. Donohoe; and T. Sougiannis. "Do Analysts understand the Economic and reporting Complexities of Derivatives?" *Journal of Accounting and Economics* 61 (2016): 584-604.

Cohen, L., and D. Lou. "Complicated Firms." *Journal of Financial Economics* 104 (2012): 383-400.

Doyle, J.; W. Ge; and S. McVay, "Determinants of Weaknesses in Internal Control over Financial Reporting." *Journal of Accounting and Economics* 44 (2007): 193-223.

Dyer, T.; M. Lang; and L. Stice-Lawrence. "The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation." *Journal of Accounting and Economics* 64 (2017): 221-245.

Fahlenbrach, R.; K. Rageth; and R. Stulz. "How Valuable is Financial Flexibility when Revenue Stops? Evidence from the COVID-19 crisis." Working paper, Ohio State University, 2020.

Fama, E. F., and K. R. French. "Industry Costs of Equity." *Journal of Financial Economics* 43 (1997): 153-193.

Ge, W. and S. McVay, "The Disclosure of Material Weaknesses in Internal Control after the Sarbanes-Oxley Act." *Accounting Horizons* 19 (2005): 137-158.

Goldreich, O. P, *Np, and Np-Completeness: The Basics of Computational Complexity*, (2010), Cambridge University Press.

Gomes, A.; G. Gorton; and L. Madureira. "SEC Regulation Fair Disclosure, Information, and the Cost of Capital." *Journal of Corporate Finance* 13 (2007): 300-334.

Griffin, P. A. "Got Information? Investor Response to Form 10-K and Form 10-Q EDGAR Filings." *Review of Accounting Studies* 8 (2003), 433-460.

Guay, W.; D. Samuels; and D. Taylor. "Guiding through the Fog: Financial Statement Complexity and Voluntary Disclosure." *Journal of Accounting and Economics* 62 (2016): 234-269.

Hay, D. C.; W. R. Knechel; and N. Wong. "Audit Fees: A Meta-Analysis of the Effect of Supply and Demand Attributes." *Contemporary Accounting Research* 23 (2006): 141-191.

Hogan, C. E., and M. S. Wilkins. "Evidence on the Audit Risk Model: Do Auditors Increase Audit Fees in the Presence of Internal Control Deficiencies?" *Contemporary Accounting Research* 25 (2008): 219-242.

Hoitash, R., and U. Hoitash. "Measuring Accounting Reporting Complexity with XBRL." *Accounting Review* 93 (2018): 259-287.

Jones, M. J. and P. A. Shoemaker. "Accounting Narratives: A Review of Empirical Studies of Content and Readability." *Journal of Accounting Literature* 13 (1994): 142-184.

Lehavy, R.; F. Li; and K. Merkley. "The Effect of Annual Report Readability on Analyst following and the Properties of their Earnings Forecasts." *Accounting Review* 86 (2011): 1087-1115.

Leuz, C., and P. Wysocki, "The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research." *Journal of Accounting Research* 54 (2016): 525-622.

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

Li, Y., and L. Zhang. "Short Selling Pressure, Stock Price Behavior, and Management Forecast Precision: Evidence from a Natural Experiment." *Journal of Accounting Research* 53 (2015): 79-117.

Lo, K; F. Ramos; and R. Rogo. "Earnings Management and Annual Report Readability." *Journal of Accounting and Economics* 63 (2017): 1-25.

Loughran, T. and B. McDonald. "When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks." *Journal of Finance* 66 (2011): 35-65.

Loughran, T. and B. McDonald. "Measuring Readability in Financial Disclosures." *Journal of Finance* 69 (2014): 1643-1671.

Loughran, T. and B. McDonald. "Textual Analysis in Accounting and Finance: A survey." *Journal of Accounting Research* 54 (2016): 1187-1230.

Loughran, T. and Ritter, J. "Why has IPO Underpricing Changed over time?" *Financial Management* 33 (2004): 5-37.

Lowry, M.; M. S. Officer; and G. W. Schwert. "The Variability of IPO Initial Returns." *Journal of Finance* 65 (2010): 425-465.

Manning, C. and H. Schütze, *Foundations of Statistical Natural Language Processing*: (2003), MIT Press, Cambridge, MA.

O'Keefe, T. B.; D. A. Simunic; and M. T. Stein. "The Production of Audit Services: Evidence from a Major Public Accounting Firm." *Journal of Accounting Research* 32 (1994): 241-261.

Palmrose, Z. "Audit Fees and Auditor Size: Further Evidence." *Journal of Accounting Research* 24 (1986): 97-110.

Prawitt, D. F. "Staffing Assignments for Judgment-Oriented Audit Tasks: The Effects of Structured Audit Technology and Environment." *Accounting Review* 70 (1995): 443-465.

Simunic, D. A. "The Pricing of Audit Services: Theory and Evidence." *Journal of Accounting Research* 18 (1980): 161-190.

Snowden, D. and M. Boone, "A Leader's Framework for Decision Making." *Harvard Business Review* (2007): 68-77.

Stice, J.D. "Using Financial and Market Information to Identify Pre-Engagement Factors Associated with Lawsuits against Auditors." *The Accounting Review* 66 (1991): 516-533.

Whisenant, S; S. Sankaraguruswamy, and K. Raghunandan. "Evidence on the Joint Determination of Audit and Non-Audit Fees." *Journal of Accounting Research* 41 (2003): 721-744.

You, H. and Zhang, X.J. "Financial Reporting Complexity and Investor Underreaction to 10-K Information." *Review of Accounting Studies* 14 (2009): 559-586.

TABLE 1

Sample Creation

	Dropped (1)	Sample Size (2)
Audit Analytics:		
Initial Wharton Research Data Services		224,246
Duplicate [cik, fiscal year], keep observation with higher fee	9,553	214,693
Drop firm/filings with duplicate data	2	214,691
Drop if *Audit Fees* is zero or missing	378	214,313
SEC EDGAR 10-K, 10-K405, 10KSB, 10KSB40 files:		
Filing date 2001-2018 and matched with Audit Analytics	86,641	127,672
Firms filing both 10-K and 10-KSB, retain 10-K observation	5	127,667
Drop if (10-K filing date - fiscal-year-end) > 180 days	1,309	126,358
Drop if 10-K total word count < 3,000	72	126,286
Compustat data:		
Matching segment data	44,621	81,665
Matching financial data and total assets > $1 million	15,984	65,681
CRSP data:		
Matching return and market capitalization data	13,023	52,658
Final Audit Fee Sample		52,658

This table reports the derivation of the primary sample used in the analysis. The sample time period covers 10-K filing years 2001 through 2018.

TABLE 2

List of 374 Complexity Words

ACCRUABLE	CONVERTIBILITY	LEASEBACK	PATENTING	SEGMENTING
ACCRUAL	CONVERTIBLE	LEASEBACKS	PATENTS	SEGMENTS
ACCRUALS	CONVERTIBLES	LEASED	REACQUIRE	SOVEREIGN
ACCRUE	COPYRIGHT	LEASEHOLD	REACQUIRED	SOVEREIGNS
ACCRUED	COPYRIGHTABLE	LEASEHOLDER	REACQUIRES	SOVEREIGNTIES
ACCRUES	COPYRIGHTED	LEASEHOLDERS	REACQUIRING	SOVEREIGNTY
ACCRUING	COPYRIGHTING	LEASEHOLDS	REACQUISITION	SUBCONTRACT
ACQUIRE	COPYRIGHTS	LEASER	REACQUISITIONS	SUBCONTRACTED
ACQUIRED	COUNTERPARTIES	LEASES	RECAPITALIZATION	SUBCONTRACTING
ACQUIREE	COUNTERPARTY	LEASING	RECAPITALIZATIONS	SUBCONTRACTOR
ACQUIREES	COVENANT	LESSEE	RECAPITALIZE	SUBCONTRACTORS
ACQUIRER	COVENANTED	LESSEES	RECAPITALIZED	SUBCONTRACTS
ACQUIRERS	COVENANTING	LESSOR	RECAPITALIZES	SUBLEASE
ACQUIRES	COVENANTS	LESSORS	RECAPITALIZING	SUBLEASED
ACQUIRING	DERIVATIVE	LICENCE	RECLASSIFICATION	SUBLEASEE
ACQUIROR	DERIVATIVES	LICENCED	RECLASSIFICATIONS	SUBLEASEHOLD
ACQUIRORS	EMBEDDED	LICENCES	RECLASSIFIED	SUBLEASES
ACQUISITION	ENTITIES	LICENCING	RECLASSIFIES	SUBLEASING
ACQUISITIONS	EXERCISABILITY	LICENSABLE	RECLASSIFY	SUBLESSEE
ACQUISITIVE	EXERCISABLE	LICENSE	RECLASSIFYING	SUBLESSEES
AFFILIATE	EXERCISEABILITY	LICENSED	REISSUANCE	SUBLESSOR
AFFILIATED	EXERCISEABLE	LICENSEE	REISSUANCES	SUBLESSORS
AFFILIATES	EXERCISED	LICENSEES	REISSUE	SUBLET
AFFILIATING	FLOATING	LICENSES	REISSUED	SUBLETS
AFFILIATION	FOREIGN	LICENSING	REISSUES	SUBLETTING
AFFILIATIONS	FRANCHISE	LICENSOR	REISSUING	SUBLETTINGS
ALLIANCE	FRANCHISED	LICENSORS	REORGANISATION	SUBLICENSABLE
ALLIANCES	FRANCHISEE	LIEN	REORGANIZATION	SUBLICENSE
BANKRUPT	FRANCHISEES	LIENHOLDER	REORGANIZATIONAL	SUBLICENSEABLE
BANKRUPTCIES	FRANCHISER	LIENHOLDERS	REORGANIZATIONS	SUBLICENSED
BANKRUPTCY	FRANCHISERS	LIENS	REORGANIZE	SUBLICENSEE
BANKRUPTED	FRANCHISES	LIQUIDATE	REORGANIZED	SUBLICENSEES
CARRYBACK	FRANCHISING	LIQUIDATED	REORGANIZES	SUBLICENSES
CARRYBACKS	FRANCHISOR	LIQUIDATES	REORGANIZING	SUBLICENSING
CARRYFORWARD	FRANCHISORS	LIQUIDATING	REPATRIATE	SUBLICENSOR
CARRYFORWARDS	FUTURES	LIQUIDATION	REPATRIATED	SUBSIDIARIES
COLLABORATE	GLOBAL	LIQUIDATIONS	REPATRIATES	SUBSIDIARY
COLLABORATED	GLOBALIZATION	LIQUIDATOR	REPATRIATING	SUBSIDIES
COLLABORATES	GLOBALIZE	LIQUIDATORS	REPATRIATION	SUBSIDING
COLLABORATING	GLOBALIZED	LITIGATE	REPATRIATIONS	SUBSIDIZATION
COLLABORATION	GLOBALIZING	LITIGATED	RESTRUCTURE	SUBSIDIZE
COLLABORATIONS	GLOBALLY	LITIGATES	RESTRUCTURED	SUBSIDIZED
COLLABORATIVE	HEDGE	LITIGATING	RESTRUCTURES	SUBSIDIZERS
COLLABORATIVELY	HEDGED	LITIGATION	RESTRUCTURING	SUBSIDIZES
COLLABORATOR	HEDGES	LITIGATIONS	RESTRUCTURINGS	SUBSIDIZING
COLLABORATORS	HEDGING	LITIGIOUS	REVALUATION	SUBSIDY
COLLATERAL	IMBEDDED	MERGE	REVALUATIONS	SUBTENANCIES
COLLATERALIZATION	INFRINGE	MERGED	REVALUE	SUBTENANCY
COLLATERALIZE	INFRINGED	MERGER	REVALUED	SUBTENANT
COLLATERALIZED	INFRINGEMENT	MERGERS	REVALUES	SUBTENANTS
COLLATERALIZES	INFRINGEMENTS	MERGES	REVALUING	SWAP
COLLATERALIZING	INFRINGER	MERGING	REVOCABILITY	SWAPS
COLLATERALS	INFRINGERS	NATIONALIZATION	REVOCABLE	SWAPTION
COMPLEX	INFRINGES	NATIONALIZATIONS	REVOCATION	SWAPTIONS
COMPLEXITIES	INFRINGING	NATIONALIZE	REVOCATIONS	TAKEOVER
COMPLEXITY	INSOLVENCIES	NATIONALIZED	REVOKE	TAKEOVERS
COMPLEXLY	INSOLVENCY	NATIONALIZING	REVOKED	TRADEMARK
CONGLOMERATE	INSOLVENT	NONMARKETABLE	REVOKES	TRADEMARKED
CONGLOMERATES	INTANGIBLE	OUTSOURCE	REVOKING	TRADEMARKING
CONTINGENCIES	INTANGIBLES	OUTSOURCED	ROYALTIES	TRADEMARKS
CONTINGENCY	INTERCONNECT	OUTSOURCER	ROYALTY	UNEXERCISABLE
CONTINGENT	INTERCONNECTED	OUTSOURCERS	SECURITIZABLE	UNEXERCISED
CONTINGENTLY	INTERCONNECTEDNESS	OUTSOURCES	SECURITIZATION	UNRECOGNIZED
CONTRACT	INTERCONNECTING	OUTSOURCING	SECURITIZATIONS	UNREMITTED
CONTRACTED	INTERCONNECTION	PARTNER	SECURITIZE	UNREPATRIATED
CONTRACTHOLDER	INTERCONNECTIONS	PARTNERED	SECURITIZED	VENTURE
CONTRACTHOLDERS	INTERCONNECTS	PARTNERING	SECURITIZER	VENTURES
CONTRACTING	INTERNATIONAL	PARTNERS	SECURITIZERS	WARRANTEES
CONTRACTS	INTERNATIONALIZATION	PARTNERSHIP	SECURITIZES	WARRANTIED
CONTRACTUAL	INTERNATIONALLY	PARTNERSHIPS	SECURITIZING	WARRANTIES
CONTRACTUALLY	LAWSUIT	PATENT	SEGMENT	WARRANTING
CONTRACTUALS	LAWSUITS	PATENTABILITY	SEGMENTAL	WARRANTOR
CONTRACTURAL	LEASABLE	PATENTABLE	SEGMENTATION	WARRANTY
CONVERSION	LEASE	PATENTED	SEGMENTATIONS	WORLDWIDE
CONVERSIONS	LEASEABLE	PATENTEE	SEGMENTED	

This table presents the 374 words included in the *COMPLEXITY* lexicon.

TABLE 3

Rankings of Complexity Word Lemmas in Annual Reports, 2001-2018

	Total Frequency (1)	% Change in Relative Occurrence of Complexity Lemmas from 2001-2006 to 2013-2018 (2)
Top 10		
1	ACQUIRE	UNRECOGNIZED
2	LEASE	GLOBAL
3	CONTRACT	DERIVATIVE
4	SUBSIDIARY	COLLABORATE
5	LICENSE	COMPLEX
6	PARTNER	FOREIGN
7	SEGMENT	PARTNER
8	FOREIGN	CONTINGENT
9	PATENT	PATENT
10	AFFFILIATE	LITIGATE
Bottom 10		
1	NONMARKETABLE	EXERCISED
2	IMBEDDED	REORGANIZATION
3	CONGLOMERATE	RESTRUCTURE
4	NATIONALIZE	BANKRUPT
5	CARRYBACK	LEASE
6	UNREMITTED	CONVERSION
7	SOVEREIGN	COLLATERAL
8	REISSUE	LIQUIDATE
9	REVALUE	SUBSIDAIRY
10	SUBSIDY	WARRANTY

For the 69 lemmas based on the 374 complexity words, this table presents the top 10 and bottom 10, ranked using the column's criterion. A "lemma" is the root form of the various inflections included in the complexity lexicon (e.g., hedge: hedged, hedges, hedging). The bottom 10 lemmas are listed beginning with the least frequent or most negative values. Column (1) reports the ranking based on the simple raw count for each complexity lemma. The column (2) ranking begins with a lemma's proportional occurrence relative to the total complexity count (across all documents in that period) and calculates the percentage change in this proportion from the 2001-2006 period to the 2013-2018 period. For words to be included in the column (2) ranking, they must account for at least 0.5% of the total complexity count in one of the two periods.

T A B L E 4

Audit Fee Sample Summary Statistics and Correlations

Panel A: Summary Statistics

Variable Name	Mean	Median	Standard Deviation	10th Percentile	90th Percentile
Complexity	0.22	0.22	0.06	0.14	0.29
Audit Fees	$1.77 MM	$0.73 MM	$3.67 MM	$0.12 MM	$4.00 MM
Total Assets	$3,743 MM	$380 MM	$18,833 MM	$27 MM	$6,326 MM
Total Words	49,224	41,837	32,303	21,879	83,019
Top-5 Auditor Dummy	0.74	1.00	0.44	0.00	1.00
S&P Dummy	0.10	0.00	0.30	0.00	1.00
Loss Dummy	0.37	0.00	0.48	0.00	1.00
% Leverage	0.22	0.17	0.23	0.00	0.53
% Goodwill	0.11	0.04	0.15	0.00	0.34
% Inventory + Receivables	0.24	0.21	0.19	0.03	0.52
Segments	5.26	4.00	3.89	1.00	10.00

TABLE 4 — *Continued*

Panel B: Correlations										
	Complexity	*log (Audit Fees)*	*log (Total Assets)*	*log (Total Words)*	*Top-5 Auditor Dummy*	*S&P 500 Dummy*	*Loss Dummy*	*% Leverage*	*% Goodwill*	*% Invent. + Rec.*
log(Audit Fees)	0.482									
log(Total Assets)	0.369	0.845								
log(Total Words)	0.865	0.499	0.442							
Top-5 Auditor Dummy	0.228	0.468	0.475	0.235						
S&P 500 Dummy	0.154	0.428	0.497	0.153	0.190					
Loss Dummy	0.124	-0.238	-0.364	0.061	-0.145	-0.190				
% Leverage	0.206	0.221	0.293	0.246	0.095	0.065	0.034			
% Goodwill	0.161	0.303	0.268	0.112	0.125	0.148	-0.144	0.113		
% Invent. + Rec.	-0.242	-0.054	-0.104	-0.265	-0.099	-0.053	-0.148	-0.086	-0.098	
Segments	0.221	0.468	0.406	0.173	0.184	0.229	-0.174	0.038	0.168	0.123

The summary statistics and correlations are across 52,658 10-K filings for filing years 2001-2018. *Complexity* is the total number of unique complexity words in a firm's 10-K divided by the total number of words in the complexity lexicon (374). *Total Words* is the total word count from a firm's parsed 10-K filing. *Audit Fees* are the audit fees for the firm in the fiscal year of the 10-K as reported by Audit Analytics. Also taken from Audit Analytics are: *Top-5 Auditor Dummy* which is set to one if the auditor is among the top 5, else zero; *S&P 500 Dummy* which is set to one if the firm is in the S&P 500 Index, else zero; and, *Loss Dummy* which is set to one if net income has a negative value, else zero. The remaining variables are taken from Compustat: *Total Assets* are end-of-fiscal-year; *% Leverage* is (short-term debt + long-term debt)/total assets; *% Goodwill* is goodwill/total assets; *% Inventory + Receivables* is (inventory + receivables)/total assets; and, *Segments* is the number of reported "all segments". *% Leverage*, *% Goodwill*, and *% Invent. + Rec.* are winsorized at the 99th percentile.

T A B L E 5

Audit Fee Regressions

	log (Audit Fees) (1)	*log (Audit Fees)* (2)	*log (Audit Fees)* (3)
Complexity	11.141	3.572	3.139
	(16.87)	(18.73)	(9.49)
log(Total Assets)		0.493	0.423
		(43.62)	(29.01)
log(Total Words)			-0.029
			(-0.95)
Top-5 Auditor Dummy			0.436
			(17.05)
S&P 500 Dummy			0.106
			(2.56)
Loss Dummy			0.162
			(9.57)
% Leverage			-0.041
			(-0.66)
% Goodwill			0.364
			(3.59)
% Invent.+ Rec.			0.449
			(6.09)
Segments			0.038
			(10.30)
Intercept	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
R-Squared	40.8%	81.9%	84.6%
Sample Size	52,658	52,658	52,658

This table examines the role of *Complexity* in predicting *Audit Fees*. The dependent variable in each of the regressions is *log(Audit Fees). Audit Fees* are the auditor fees as reported by Audit Analytics. *Complexity* is the unique count of complexity words appearing in a firm's annual report (i.e., Form 10-K) divided by the total number of words in the complexity lexicon (374). *Total Assets* are as of the end of the fiscal year. *Total Words* is the total word count from the parsed 10-K filing. *Top-5 Auditor Dummy* is set to one if the auditor is among the top 5 (includes Arthur Andersen in the earlier portion of the sample), else zero. *S&P 500 Dummy* is set to one if the firm is in the S&P 500 Index, else zero. *Loss Dummy* is set to one if net income has a negative value, else zero. *% Leverage* is (short-term debt + long-term debt)/total assets. *% Goodwill* is goodwill/total assets. *% Inventory + Receivables* is (inventory + receivables)/total assets. *Segments* is the sum of all segment types for a firm reported in the Compustat Historical Segments file. All the regressions include an intercept, Fama and French [1997] 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with standard errors clustered by year and industry. The sample period is 2001-2018.

T A B L E 6

Regressions for Alternative Specifications

	log (Audit Fees + Audit-Related Fees) (1)	*log (Audit Fees)* (2)	*log (Audit Fees)* (3)
Complexity	3.157	19.581	3.000
	(9.61)	(21.56)	(6.41)
log(Total Assets)	0.430		0.396
	(30.27)		(22.66)
log(ARC)			0.267
			(2.84)
Intercept	Yes	Yes	Yes
Control Variables	Yes	No	Yes
Industry Dummies	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
R-Squared	84.8%	56.7%	85.6%
Sample Size	52,658	126,286	21,629
Sample Period	2001 - 2018	2001 – 2018	2009 - 2018

This table considers alternative specifications of the audit fee models. Variables are defined in Table 5. *ARC* is the XBRL measure documented in Hoitash and Hoitash [2018] and downloaded from their website. In column (1) the dependent variable is the sum of the *Audit Fee* variable from Table 4 and *Audit-Related Fees* reported by Audit Analytics. Column (2) is the full sample of all firms with *Audit Fee* data from Audit Analytics and 10-K word counts greater than or equal to 3,000 (i.e., not constrained by the other data requirements) which provides 126,286 observations. Column (3) merges the primary dataset from Table 5 with the Hoitash and Hoitash [2018] data leaving a sample of 21,629 observations. The control variable coefficients not reported in the table are those for *Top-5 Auditor Dummy, S&P 500 Dummy, Loss Dummy, % Leverage, % Goodwill,* and *% Invent. + Rec*, and *Segments*. All the regressions include an intercept, Fama and French [1997] 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with standard errors clustered by year and industry. The sample period is 2001-2018.

T A B L E 7

Complexity and the Absolute Value of Stock Returns around 10-K Filing Dates

	Average Absolute Return [t-20, t-1] (1)	*Average Absolute Return [t, t+1]* (2)	*Average Absolute Return [t+2, t+10]* (3)
Complexity	4.839	2.252	1.882
	(8.47)	(5.42)	(8.25)
log(Total Assets)	-0.397	-0.150	-0.138
	(-13.07)	(-10.27)	(-9.19)
Average Absolute Return [t-20, t-1]		0.491	0.549
		(20.39)	(20.74)
Intercept	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes
Year Dummies	Yes	Yes	Yes
R-Squared	30.1%	12.3%	45.2%
Sample Size	52,658	52,658	52,658

This table compares the absolute value of stock returns around the filing date of the 10-K with the complexity and size proxies. The dependent variable in column (1) is the average absolute stock returns for the 20 days prior to the filing date (a minimum of 10 valid returns is required to be included in the sample). Column (2) uses the average absolute value of the filing date (t=0) and day t=1 returns as the dependent variable. Similarly, in Column (3) the dependent variable is average absolute return over days t+2 to days t+10. All the regressions include an intercept, Fama and French [1997] 48-industry dummies, and calendar year dummies. The *t*-statistics are in parentheses with standard errors clustered by year and industry. The sample period is 2001-2018.

T A B L E 8

Complexity: Alternative Tests

	Abs(Unexpected Earnings) (1)	*log(1+ IPO Initial Return)* (2)	*COVID-19 Stock Return (2/3-3/23)* (3)
Complexity	3.626	-0.399	-41.307
	(9.23)	(-2.42)	(-2.36)
log(Total Assets)	-0.270	-0.001	
	(-20.24)	(-0.31)	
IPO Control Variables			
Up Revision		0.010	
		(6.18)	
Venture Capital		0.017	
		(1.13)	
Top Tier IB		0.024	
		(2.07)	
NASDAQ		0.011	
		(1.08)	
NASDAQ15		0.002	
		(1.42)	
log(1+Age)		0.005	
		(1.42)	
COVID Control Variables			
log(Mkt Cap)			3.325
			(3.00)
Long-term Debt Ratio			-5.653
			(-1.01)
Cash Ratio			23.064
			(3.96)
Intercept	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes
Year Dummies	Yes	Yes	No
R-Squared	13.2%	36.1%	41.0%
Sample Size	35,479	1,813	497

This table presents alternative tests for the Complexity measure. Column (1) uses the absolute value of unexpected earnings as the dependent variable. Column (2) uses the log of one plus the initial return of an IPO as the dependent variable. The word counts for the *Complexity* calculation are based on the last S-1 or 424 filing that occurs before five business days following the IPO date. The sample period in both Columns (1) and (2) is 2001-2018. Column (3) uses the returns for the S&P 500 for the COVID-19 "collapse period" period defined in Fahlenbrach, Rageth, and Stulz [2020], 2020-02-02 to 2020-03-23, as the dependent variable. The IPO and COVID-19 control variables are defined in the appendix. The *t*-statistics are in parentheses with standard errors clustered by year and industry in columns (1) and (2) and by industry in column (3).