**Linguistic Complexity and Cost of Equity Capital[[1]](#footnote-1)**

|  |  |  |
| --- | --- | --- |
| **Alexandre Garel[[2]](#footnote-2)**  Auckland University of Technology, Labex ReFi | **Aaron Gilbert[[3]](#footnote-3)**  Auckland University of Technology | **Ayesha Scott[[4]](#footnote-4)**  Auckland University of Technology |

**Abstract:** In this paper, we investigate how linguistic complexity relates to a company’s cost of equity capital. When management uses linguistic complexity to obfuscate information, we expect shareholders to perceive a greater risk and hence to require a greater return on equity. Our results show that, within a given firm, a decrease in the readability of the annual report is associated with an increase in the cost of equity capital. Greater linguistic complexity may result from managerial obfuscation but can also be a byproduct of greater intrinsic business complexity. Complementary analyses support the influence of linguistic complexity on the cost of equity capital through an obfuscation channel. Finally, we provide evidence that the effect of linguistic complexity is robust to controlling for the quantity of information disclosed.

**JEL:** G10, G14, G32

**Keywords:** Cost of capital; Cost of equity; Implied cost of equity capital; Readability; Fog index; Linguistic complexity; Managerial Obfuscation

1. **Introduction**

Theoretical research shows that by disclosing more information to the market (by releasing more frequent or more informative disclosures) a firm can reduce information asymmetries and thus the risk perceived by its investors (e.g., Barry and Brown 1985; Handa and Linn 1993; Coles et al. 1995). A range of empirical studies show that disclosure of information to financial markets reduces the market risk premium measured by the cost of equity (e.g., Richardson and Welker 2001; Chen et al. 2010; Li 2010b; Fu et al. 2012; Daske et al. 2013)[[5]](#footnote-5). In this paper, we argue that on top of the quantity of information provided to the market, its readability also matters.

Corporate disclosure is presented as accounting numbers, such as financial statements, framed or accompanied by a substantial amount of text. The text can be written in a way that makes it either more or less easy to read and process. We argue that the ease with which investors can process these accompanying passages of text may influence how they perceive the firm’s riskiness. This is because less readable financial information can signal an attempt to hide bad news. Managers have an incentive to obfuscate information when firm performance is bad (Bloomfield 2002; Bloomfield 2008) and prior literature provides evidence of managerial obfuscation (e.g., Lang and Lundholm 1993; Schrand and Walther 2000). To obfuscate financial statements managers may resort to linguistic complexity. Li (2008) and Lo et al. (2017) show that firms issue less readable reports if they are performing poorly or are likely to have managed earnings to beat the prior year’s earnings. Investors should collectively be able to link a decrease in the readability of financial information to an attempt to obfuscate bad news. Rennekamp (2012), using an experiment, finds that investors predict that managers will provide less readable disclosures when news is bad rather than when news is good. As a result, we expect investors to revise their risk perception when they observe less readable firm information and thus, charge a higher cost of equity. More formally, our core hypothesis is that lower annual report readability is associated with higher firm cost of equity capital.

To measure readability, we build on recent works studying the ease with which the text in corporate disclosure documents can be processed (e.g., Miller 2010; Lehavy et al. 2011; Lawrence 2013; Loughran and McDonald 2014)[[6]](#footnote-6). We underpin our analysis using the foundational concept of readability: easy to understand text employs less complicated words and shorter sentences. Therefore, each readability metric we use relies on the number of syllables in a word, and the number of words per sentence. More specifically, we use the Fog Index, Flesch reading ease and Flesch-Kincaid grade level.

We follow prior research and measure a firm’s cost of equity as the average of several implied cost of equity estimates derived from earnings expectations. This approach relies on the simple idea that, as the share price is the present value of expected future cash flows per share, the rate that equates the present value of expected future cash flow per share to the share price must be the true discount rate of the firm. In our main specification, we use earnings expectations from analysts (e.g., Gebhardt et al. 2001; Claus and Thomas 2001; Hail and Leuz 2006; Li 2010b; Dhaliwal et al. 2011). For robustness, we also use earnings predicted by cross-sectional models (e.g., Hou et al. 2012).

The results of our empirical analysis show that, on average, issuing a markedly less readable annual report is associated with a significant increase in the cost of equity. A one standard-deviation increase in the Fog Index is associated with a significantly higher cost of equity, ranging from 0.16% to 0.38% depending on the expected earnings estimation model. This result is robust to controlling for standard determinants of the cost of equity such as size, market-to-book, leverage, whether the firm pays dividends, institutional ownership, analyst coverage, forecast dispersion, long-term growth rate, CAPM beta, and year and industry fixed effects (e.g., Gode and Mohanram 2003; Hail and Leuz 2006; Attig et al. 2008; El Ghoul et al. 2011; Boubakri et al. 2012; Mishra 2014; Dhaliwal et al. 2016; Boubakri and Mishra 2017). We alleviate the potential concern that time-varying differences across industries (e.g., business complexity) or persistent differences across firms (e.g., corporate culture) could explain both a higher cost of equity and a lower annual report readability by showing that our results hold when we control for year-industry and firm fixed effects. Within a given firm, a decrease in readability is still associated with a significant increase in the cost of equity. The time-varying effect of readability on the cost of equity we document lends some support to our hypothesis that management uses linguistic complexity to obfuscate information and that investors react to this behavior by charging a higher cost of equity capital.

While the literature generally interprets complex language in firms’ disclosures as indicative of managerial obfuscation, it can also reflect the provision of complex information (see Bushee et al. (2018)). Complex language can be necessary to convey information about the firm's business transactions and operating strategy (Bloomfield 2002). An important challenge is thus to isolate the effect of readability driven by managerial obfuscation from readability that is driven by the intrinsic complexity of a firm’s business. To identify increases in linguistic complexity most likely to result from managerial obfuscation as opposed to the provision of complex information, we use four complementary strategies.

First, we follow Bushee et al. (2018) and decompose linguistic complexity into two latent components. This first reflects the language complexity inherent to the business activity and the second is more likely to be caused by a willingness to obfuscate information (i.e., cannot be related to language complexity inherent to business activity). To do so, we regress our measures of linguistic complexity on proxies of firm business complexity. We follow prior literature (e.g., Frijns et al. 2016; Boubakri and Mishra 2017) and measure firm complexity using the number of business segment sales and the number of geographical segment sales. Then, we define the residuals of these regressions as the part of linguistic complexity that is more likely to result from managerial obfuscation. We document a strong and positive association between linguistic complexity that is not explained away by business complexity and a firm’s cost of equity.

Second, we build on the first test and examine the relationship between linguistic complexity purged of the effect of business complexity and cost of equity for firms likely to have managed their earnings. If managerial obfuscation is driving the effect of linguistic complexity that is not explained by business complexity on a firm’s cost of equity, it should be much stronger in firms likely to have managed their earnings, i.e., in firms where management has an incentive to obfuscate information[[7]](#footnote-7). We identify firms likely to manage their earnings upwards as firms that meet or just beat their earnings target (e.g., Burgstahler and Dichev 1997; Degeorge et al. 1999; Graham et al. 2005; Bhojraj et al. 2009). For the earnings target, we use the consensus analyst forecasts (e.g., Burgstahler and Dichev 1997; Caskey and Ozel 2017). We document a positive association between the purged linguistic complexity and a firm’s cost of equity, only for firms likely to have managed their earnings to meet or just beat their earnings target.

Third, economic theory suggests that obfuscation (informative disclosure) is associated with greater (lower) information asymmetry between shareholders and managers (Bloomfield 2002). Bushee et al. (2018) show that while the information component of linguistic complexity reduces information asymmetry, the obfuscation component increases it. We use this prior finding to separate the effect of readability attributable to obfuscation from the one attributable to complex information disclosure. We investigate the effect of readability on the cost of equity conditional on whether the within-firm change in linguistic complexity is associated with a greater within-firm change in information asymmetry. To the extent that the effect we document is partially driven by managerial obfuscation, we should find a positive association between linguistic complexity and cost of equity in a subsample of firms for which increases in linguistic complexity increase information asymmetry. Results show that when information asymmetry has increased contemporaneously to an increase in linguistic complexity, the effect of linguistic complexity on the implied cost of equity is significantly positive and sizeable.

Fourth, we focus on a section of the 10-K annual report where managers are more likely to use linguistic complexity to obfuscate information. Prior literature indicates that managers have more room to use their discretion to strategically obfuscate financial results in the Management Discussion and Analysis (MD&A) section of the 10-K annual report (e.g., Li 2008; Lo et al. 2017)[[8]](#footnote-8). We show that, as for the readability of the whole 10-K annual report, the readability of the MD&A section is negatively related to a firm’s cost of equity. We then generate measures of MD&A readability that are orthogonal to the readability of the whole 10-K annual report. The idea is to capture the part of the linguistic complexity of the MD&A section that cannot be explained by the linguistic complexity of the rest of the 10-K document and thus is more likely to be a manifestation of managerial discretion. We find that in the section of the 10-K annual report where managers are more likely to use their discretion to obfuscate information, an abnormal increase in linguistic complexity with respect to the rest of the document is associated with an increase in the cost of equity capital.

Overall, our results hold in settings where linguistic complexity is more likely to result from managerial obfuscation as opposed to the provision of complex information. It supports our core hypothesis that the ease with which investors can process text influences their risk perception when lower readability of the financial information signals an attempt to obfuscate information.

Our empirical analysis mitigates the concern of intrinsic firm complexity rather than managerial obfuscation being the main driving force behind our results. While the obfuscation channel calls for a causal interpretation of our findings, we cannot fully establish causality. As discussed in Leuz and Wysocki (2016), causal effects of disclosure are hard to establish. Guay et al. (2016) identify exogenous sources of increased linguistic complexity in the form of the adoption of complex accountings standards[[9]](#footnote-9). However, because the adoption of such accounting standards are driven by intrinsic business complexity rather than managerial obfuscation, we do not consider them as appropriate sources of exogenous variation to establish causality in our paper. We rather document a set of consistent findings, showing that, within a given firm, an increase in linguistic complexity is associated with an increase in the cost of equity, and that this effect holds in settings where managerial obfuscation is the likely driver of linguistic complexity.

Our findings echo those of Boubakri and Mishra (2017), who show that information overload may affect the cost of equity capital. While Boubakri and Mishra (2017) argue that too much information (information quantity) may overload investors and result in a higher rate of return required by investors, we argue that information readability, which is quantity-adjusted, may change the risk perception of investors and result in a higher rate of return. To further disentangle the effect of information quantity from information readability, we purge our measure of linguistic complexity from the effect of quantity by taking the residuals of a regression of linguistic complexity on file size, a standard proxy for information quantity (e.g., Loughran and McDonald 2014; Boubakri and Mishra 2017). Results show that there is an effect of linguistic complexity on the cost of equity capital above and beyond the measure of information quantity. Both our findings and the Boubakri and Mishra (2017)’s ones highlight that the way a firm frames information disclosure to investors may affect its cost of equity capital.

This paper contributes to the accounting and finance literature in three ways. First, the question of how corporate disclosure affects investors’ perception of firm risk has motivated a significant body of research (e.g., Core 2001; Fields et al. 2001; Lehavy et al. 2011; Beyer et al. 2010). We contribute to this stream of the literature by arguing that disclosing less readable information is likely to increase the risk perceived by investors when it results from managerial obfuscation and by documenting supportive evidence. We show that, on top of the effect of the quantity of information disclosed on a firm’s cost of equity uncovered by Boubakri and Mishra (2017), there is an effect of information readability.

Second, a recent stream of studies document an influence of information readability on a series of corporate outcome variables such as market reaction, analyst dispersion, analyst forecast accuracy, trading volume, and voluntary disclosure (e.g., You and Zhang 2009; Miller 2010; Rennekamp 2012; Lehavy et al. 2011; Lawrence 2013; Guay et al. 2016). In these studies, the relationship between the risk perception of investors and linguistic complexity is latent. Our incremental contribution with respect to this literature is to capture more directly how an increase in investors’ risk perception resulting from a lower 10-K readability may have tangible consequences for a firm. We show that investors may react to an increase in linguistic complexity by requiring a greater rate of return, which affects, for instance, a firm’s ability to finance its projects. Closely related to our paper, Hwang and Kim (2017) show that equity closed-end investment companies issuing financial disclosure documents with low readability causes them to trade at significant discount relative to the value of their fundamentals. Our results are consistent with their findings and contribute to build a body of evidence showing that when disclosures are less readable, market stakeholders revise their risk perception upward.

Third, a recent paper from Bushee et al. (2018) highlights the necessity to disentangle between two drivers of linguistic complexity (business intrinsic complexity and managerial obfuscation) when assessing the effect of readability on a variable of interest. Building on Bushee et al. (2018), we design four tests, which, together, provide concurring evidence that at least part of the effect of linguistic complexity on a firm’ cost of equity we document is driven by managerial obfuscation. These tests offer a reasonable starting point for future researchers intending to disentangle between the effects of readability due to business intrinsic complexity from that due to managerial obfuscation.

Fourth, our findings add to the literature on the determinants of the cost of equity. Recent contributions also highlight the importance of customer concentration (e.g., Dhaliwal et al. 2016), inventory investment (e.g., Jones and Tuzel 2013), geographic location (e.g., Boubakri et al. 2016), property crime (e.g., Brushwood et al. 2016), executive pay differences (e.g., Chen et al. 2013), political connections (e.g., Boubakri et al. 2012; Ben‐Nasr et al. 2012), political rights (e.g., Boubakri et al. 2014), CSR (e.g., Ng and Rezaee 2015), investor horizon (e.g., Attig et al. 2013), and directors’ and officers’ liability insurance (e.g., Chen et al. 2016). In particular, our results add to a narrower set of papers focusing on the links between disclosure and cost of equity. Kothari et al. (2009) show that firms’ disclosure tone volatility is useful in assessing firm risk, while Dhaliwal et al. (2011) finds that CSR disclosure affects the cost of equity. We complement this literature by documenting evidence that disclosure readability matters too.

The remainder of our paper is organized as follows. Section 2 presents our data sources and sample construction, and describes the computation of our readability and cost of equity measures as well as our empirical methodology. Section 3 presents our results and Section 4 concludes.

1. **Data, Measures, Empirical Methodology, and Sample**
   1. *Data Sources*

To examine the association between the readability of corporate disclosures and the cost of equity, we collect annual 10-K filings for all US companies between 2000 and 2016 from the Software Repository for Accounting and Finance (SRAF) provided by Loughran and McDonald. The SRAF provides 10-K files sourced from the EDGAR database that are cleaned such that they are suitable for textual analysis. Specifically, tables, pictures, and all hypertext language are removed from the files[[10]](#footnote-10). To identify the MD&A section of the 10-K report, we follow the procedure set out in Li (2008). Specifically, we identify the beginning of the MD&A section by looking for variations of “item”, “7”, and “management discussion” which does not contain the word “see”. We select the end of that section by looking for variations of “item 7a” or “item 8” where there is no item 7a. Accounting data are from COMPUSTAT, financial data are from CRSP, data on analyst coverage and analysts’ consensus forecasts are from I/B/E/S, and institutional investor data are from Thomson Reuters 13-F Filings.

* 1. *Readability Measures*

Measuring the readability of documents is not new. The most commonly employed measure for readability, the Gunning Fog index (Fog index hereafter) was developed in 1952. More recently, alternatives have been developed. Specifically, we also use the Flesch reading ease measure and the Flesch-Kincaid grade level measure. Each of the three measures rely on a similar understanding of the concept of readability; easy to understand text employs shorter sentences and less complicated words, typically measured by the number of syllables in a word.

The Fog Index is calculated using the following formula:

(1)

In essence, the Fog index defines readability as 0.4 of the average number of words per sentence, the length of the sentences, and the percentage of words defined as complex. Complex words are defined as words with three or more syllables. A lower Fog index value indicates more readable text, and its value can be roughly equated to the level of education required to understand a piece of writing. A score of 6 indicates a sixth grade (U.S.) education, while 17 is a college graduate. Scores in excess of 18 are generally considered to be unreadable. Loughran and McDonald (2014) find that on average, for a sample of 10-K files between 1994 and 2011, the Fog index was 18.44.

The Flesch reading ease measure is the second-most dominant readability measure, after the Fog index (Loughran and McDonald 2014). The reading ease test does away with the simple categorization of words as complex or not, and instead measures the average number of syllables per word. In addition, it places different weights on the contribution of the length of sentences and words to the overall score. Additionally, it differs from the Fog index in that a higher score indicates more readable text, with scores between 80-90 indicating a sixth grade level, and 0-30 indicating a college level graduate and therefore extremely difficult to read for most people. The formula for the reading ease measure is:

(2)

The third measure we employ, the Flesch-Kincaid grade level, is a restructuring and reweighting of the Flesch reading ease measure to equate the results with the U.S. grade level, making the results somewhat easier to interpret. It does, however, place more emphasis on the sentence length than the reading ease measure. As with the Fog index, a lower score indicates a more readable piece of text.

(3)

For all three measures, we define a sentence in line with the method laid out in Loughran and McDonald (2014). One issue in the 10-K reports can be lists or text separated by a comma or semicolon. These are intended to be read as separate sentences. As such we convert these to full stops, along with other sentence terminators such as question marks, and then count the number of full stops in the document. To get the number of syllables for each word, we employ the Loughran and McDonald Master Dictionary, and for the Fog index we define any word with three or more syllables as complex.

* 1. *Implied Cost of Equity Capital*

We follow prior research in accounting and finance and estimate the cost equity capital as the *ex ante* cost of equity implied in current stock prices and earnings expectations[[11]](#footnote-11). The implied cost of equity capital measures rely on the simple idea that as the share price is the present value of expected future cash flows per share, the rate that equates the present value of expected future cash flow per share to the share price must be the true discount rate of the firm (Gebhardt et al. 2001; Claus and Thomas 2001). We follow prior research and measure, in our main specification, a firm’s cost of equity as the average of several implied cost of equity estimates derived from analysts’ earnings forecasts (e.g., Gebhardt et al. 2001; Claus and Thomas 2001; Hail and Leuz 2006; Li 2010b; Dhaliwal et al. 2011).

The first model we use, Gordon and Gordon (1997)’s model, is derived from the basic dividend discount model. Specifically, the model employs specific forecasts of dividend payments for the *T* years before calculating a terminal value based on the earnings per share. The model assumes that return on equity will equal the cost of capital beyond the forecast period, and hence firms will pay out 100% of earnings. Specifically, the model states:

(4)

where is the current share price, is the current dividend per share payout, is the analyst mean earnings per share forecast and is the mean analyst long-term earnings growth forecast for the company. We set *T* to be 5 years as the cross-sectional earnings forecast approach of Hou et al. (2012) allows us to estimate earnings up to 5 years into the future.

Additionally, we use the two residual income models from Claus and Thomas (2001) and Gebhardt et al. (2001). Claus and Thomas (2001) calculate the residual income for five years, using forecasts of earnings, and a terminal value, which employs expected inflation as the long-term growth rate. They proxy expected inflation as the current risk-free rate minus 3%. In addition, they assume a 50% payout ratio and calculate the book value of equity based on the clean-surplus method. Clean-surplus calculates the book value of equity at the end of the year as the book value at the start of the year plus the net income less dividends paid in that year. The Claus and Thomas model calculates the implied cost of capital based on the following equation:

(5)

where is the book value of equity, and . As few firms have five years of forecasts available, we compute the Claus and Thomas (2001) measure of implied cost of equity only for those with analyst earnings forecasts for one- and two-year ahead earnings and for the long-term growth, and then, when missing, we extrapolate three- to five-year ahead earnings as , where is the mean analyst forecast of long-term earnings growth for the firm.

The Gebhardt et al. (2001) model takes a similar approach to the Claus and Thomas (2001) model, albeit using analyst earnings forecasts for only the first two years. The third year’s earnings are grown at the firm’s forecast long-term growth rate, and then the model uses estimated return on equity for years 4 to 11 for estimates of the residual income for those years and the terminal value in year 12. The model estimates the return on equity for years 4- 11 as a steady convergence from the firm’s year 3 return on equity to the median industry return on equity. Essentially the model assumes that firms, in the long-run, converge to the industry norm. The other difference from Claus and Thomas (2001) is that Gebhardt et al. (2001) uses the current payout ratio when calculating clean-surpluses rather than a 50% payout ratio. The Gebhardt’s model is as follows:

(6)

where , and is the median return on equity for all firms in the current year in the same two-digit SIC industry group.

For the DDM and residual income models, we calculate the implied cost of equity using an iterative method that finds the cost of equity that minimizes the distance between the right-hand side of and the left-hand side of equations (4) to (6). Where a solution cannot be found, we exclude the estimate.

The last two models we employ are abnormal earnings growth models, which do not require the use of iterative calculations. Specifically, we employ Easton (2004)’s and Ohlson and Juettner-Nauroth (2005)’s models. Abnormal earnings growth models use estimates of price to earnings ratio, dividend growth and abnormal earnings growth to derive the implied cost of equity. Easton (2004)’s model calculates the cost of equity capital as:

(7)

where *A* =, *E1* and *E2* are expected earnings in forecast years *t+1* and *t+2* respectively, *D1* is the dividend per share in year *t+1*, and *P0* is the price at time *t*.

Ohlson and Juettner-Nauroth (2005)’s model calculates the cost of equity capital as:

(8)

where *GE2*is the short-term earnings growth and *GAEG*is the long-term earnings growth. *GAEG* is calculated as the difference between the risk-free rate less 3%, and *GE2*is calculated as the percentage change between the analyst forecast of EPS for years *t+1* and *t+2*.

As a robustness test, we use Hou et al. (2012)’s cross-sectional earnings estimates to compute our different ICC measures. Hou et al. (2012) offer an alternative, based on a firm’s past information, to find estimates of future earnings[[12]](#footnote-12). They use the following formula to estimate the cross-sectional earnings estimate for -periods ahead:

(9)

where the coefficients to are calculated by estimating the above regression using the last 10 years of data. *Ai,t* is total assets of company *i* in year *t*, *Di,t* is the dividends of company *i* in year *t,* *DDi,t* is a dummy variable that equals 1 if company *i* paid dividends in year *t*, *Ei,t* is the earnings of company *i* in year *t,*  *NegEi,t* is a dummy variable that equals 1 if company *i* had negative earnings in year *t* and *ACi,t* is the accruals of company *i* in year *t.* We do not require a firm to have information available in all of the 10 years prior to the year we are estimating forecast earnings for, in line with Hou et al. (2012).

Finally, using either analyst forecasts or cross-sectional estimates, we calculate all five implied cost of equity capital measures (model 4 to 8) for each firm and year. We then average these measures to create an average implied cost of equity capital that we call ICCA, when we use analyst forecasts, and ICCCSE, when we use cross-sectional earnings estimates[[13]](#footnote-13).

* 1. *Empirical Methodology*

To examine the relationship between readability (measured by the Fog index, reading ease, and grade level measures) and the cost of equity capital (measured by the average implied cost of equity capital (ICCA) predicted by the models specified above) we estimate the following regression:

(10)

We draw from the implied cost of equity literature to select variables that have been shown to influence ICC in prior studies (e.g., Gode and Mohanram 2003; Hail and Leuz 2006; Attig et al. 2008; El Ghoul et al. 2011; Boubakri et al. 2012; Mishra 2014; Dhaliwal et al. 2016; Boubakri and Mishra 2017). Specifically, in our baseline regression (10), we control for size, book-to-market ratio, leverage, whether the firm pays dividends, analyst coverage, institutional ownership, the firm growth rate, the dispersion of analyst forecasts, the CAPM beta, and industry and year fixed effects. We expect the leverage, beta, size, and book-to-market to have a positive effect on ICC because prior literature shows that these factors are positively related with expected stock returns (Sharpe 1964; Lintner 1965; Fama and French 1992). In addition, we control for analyst forecast attributes, where we use both forecast dispersion and the consensus long-term earnings growth forecast. In line with prior findings, we expect these two variables to be positively related to ICC (e.g., Gode and Mohanram 2003; El Ghoul et al. 2011). We also control for institutional ownership and analyst coverage. To the extent that greater analyst coverage and institutional ownership enhance monitoring and reduce information asymmetry, we expect these variables to be negatively related to the cost of equity capital[[14]](#footnote-14). We also control for whether firms pay dividends because prior literature documents a positive relationship between dividend taxes and ICC (e.g., Dhaliwal et al. 2005). In our core specification, we also include firm fixed effects to control for potential time-invariant factors affecting ICC and industry-year fixed effects to control for potential industry time-varying factors affecting ICC.

* 1. *Sample*

Over the period 2000-2016, after removing files that only contain a reference to external documents, we are left with 25,613 10-K files (20,793 MD&A sections) for which we can compute our readability measures and we can find a match in the COMPUSTAT, I/B/E/S and CRPS datasets. From there, we drop observations for which we can neither compute our core measure of implied cost of equity capital (ICCA) nor compute one of our core control variables. It results in a sample of 16,855 firm-year observations. We winsorize the continuous market and accounting variables at the 1st and 99th percentiles[[15]](#footnote-15).

1. **Results** 
   1. *Summary Statistics*

Panel A of Table 1 presents summary statistics on the readability variables and their components, the averaged ICC variables, and the control variables we employ throughout our empirical analysis. The reported statistics for our measures of readability are in line with prior empirical studies (e.g., Li 2008; Loughran and McDonald 2014; Lo et al. 2017). The median 10-K report in our sample consists of 45,535 words (1,404 sentences). The median sentence is of 32 words and 1 word in 4 is deemed to be complex, i.e., contains three or more syllables. As already documented in prior literature, there are important variations in the numbers of words, sentences, sentence length and sentence complexity within the universe of 10-K file of U.S companies. On average, as suggested by the Fog index summary statistics, the 10-K annual reports of our sample firms are extremely hard to read, with a mean value of 22.94. Fog index scores in excess of 18 are generally defined as unreadable (Loughran and McDonald 2014). The descriptive statistics for our two additional readability measures show results that are consistent with the Fog index. Grade Level can be interpreted as the equivalent grade level required to understand information, or the number of years of education required. On this basis, on average, to read the content of an annual report a person would need 19 years of education. While the Fog index and the Grade Level are positively associated with linguistic complexity, i.e., the higher those scores the less readable a document is, Reading Ease is negatively associated with linguistic complexity. The average Reading Ease score is 16.16, which indicates that the text is very difficult to read and requires the reader to be, at the very least, a college graduate (the minimum score possible is 0 and the maximum score possible is 100). Overall, the summary statistics indicate that 10-K annual reports are difficult to read and that some reports are much easier to read than others. We also report the summary statistics of the readability measures for just the MD&A section of the 10-K report. On average, the readability of the MD&A section is in line with that of the full 10-K report.

[*Insert Table 1 about here*]

The reported statistics for our measures of ICC and our main control variables are also in line with recent empirical literature (e.g., El Ghoul et al. 2011; Boubakri et al. 2012; Hou et al. 2012; Mishra 2014; Ng and Rezaee 2015; Chen et al. 2016; Dhaliwal et al. 2016; Goh et al. 2016; Boubakri and Mishra 2017). The median firm in our sample has an average implied cost of equity capital of about 9% based on analyst forecast estimates (ICCA) and of 8% based on cross-sectional estimates (ICCCSE). Our median firm has a leverage of about 21%, a book to market of 0.5, a beta of 0.9, an institutional ownership of 67%, and is covered by 7 analysts. Fifty-eight percent of our sample firms pay dividends. Analyst one-year ahead EPS forecasts have an average dispersion around the consensus forecast of 2 percent. The median long-term earnings growth forecast is about 11%.

Panel B of Table 1 presents the mean of our readability scores, their inputs, and our main ICCmeasure by year over the sample period. We observe an important increase in the size of the 10-K annual report over time as measured by the number of words or sentences. While the aggregate increase in the average number of words from the year 2000 to 2016 is about 60%, the increase in linguistic complexity is less pronounced, about 3.5% based on the Fog index. This observation is not surprising because the Fog index, as our other measures of readability, are size-adjusted. Despite the greater number of words used in the 10-K reports, the number of words per sentences has remained the same over the 2000-2016 period on average. What drives the increase in the aggregate linguistic complexity of the 10-K reports is the increase in the number of complex words used per sentence (about +9%). These results are in line with Dyer et al. (2017), who document 10-K disclosure over the period 1996–2013 increases in length, boilerplate, stickiness, and redundancy, and decreases in specificity, readability, and the relative amount of ‘hard’ information. With regards to the ICCA measure, the values stay relatively consistent between 8.5-10.5% on average over the entire sample period. We observe a surge in the average cost of equity following the GFC, over the 2008-2011 period. Large variations in the mean cost of equity of firms across years further motivate our choice to include year fixed effects in our regressions.

Panel C of Table 1 shows the mean readability and ICC measures by industry (Fama-French 12 Industry Classification). We observe some variations in the cost of equity and readability of 10-K reports across industries. While telecom, chemicals, finance, and utilities industries have the highest linguistic complexity scores of the 10-K annual reports, consumer durables and consumer non-durables have the lowest ones. However, these differences seem to be marginal. The different scores of linguistic complexity across industries fall within the range [22.51:23.54]. The mean ICCA by industry shows more pronounced differences. The lowest mean ICCA is 7.91% for the utilities industry and the highest is 10.88% for the energy industry over our sample period. It suggests persistence difference in the return required by shareholders for firms belonging to different industries and motivates the inclusion of industry fixed effects in our regressions.

*3.2. Univariate Results*

We begin our investigation of the relationship between readability and the cost of equity capital by looking at the average ICC values for the firms belonging to the top and bottom quartile of each of the three readability measures. We present the results in Table 2. We first consider the differences in the mean and median ICCA for the 25% highest and lowest observations in terms of Fog index. We observe a difference of 0.63% (0.36%) in the mean (median) ICCA between firms in the top and bottom quartile of linguistic complexity, as measured by the Fog index. The difference is statistically significant at the 1% level. We find similar results for our two additional measures of readability of the 10-K annual reports (Grade Level and Reading Ease). These preliminary results support our basic proposition that firms with less readable 10-K annual reports have higher costs of equity capital.

[*Insert Table 2 about here*]

*3.3 Main Results*

We next examine the relationship between ICC and readability controlling for variables that prior literature identifies as important determinants of ICC (e.g., Gode and Mohanram 2003; Hail and Leuz 2006; Attig et al. 2008; El Ghoul et al. 2011; Boubakri et al. 2012; Mishra 2014; Dhaliwal et al. 2016; Boubakri and Mishra 2017). We run an OLS regression to estimate the coefficients of equation (10) using standard errors robust to heteroscedasticity and clustered by firm. We start from the specification including industry and year fixed effects. We present our results in Panel A of Table 3. We document a positive association between 10-K Fog index scores and Grade Level scores and a firm’s implied cost of equity capital. We document a negative association between 10-K Reading Ease scores and a firm’s implied cost of equity capital. The positive relationship we find between 10-K linguistic complexity and a firm’s cost of equity is statistically significant at the 1% level. Economically, a one-standard-deviation increase in the Fog index is associated with a significant increase in the cost of equity of 0.16%. The magnitude of the effect compares with other determinants affecting ICC recently uncovered in the literature (e.g., Ben‐Nasr et al. 2012; Boubakri et al. 2012; Chen et al. 2013; Boubakri et al. 2014; Dhaliwal et al. 2016). The control variables have their expected signs. While size and institutional ownership are negatively associated with ICC, leverage, book-to-market, the dividend payer dummy, forecast diversion and long-term growth are positivity related to ICC.

[*Insert Table 3 about here*]

In Panel B of Table 3, instead of using the average ICC derived analyst earnings forecast (ICCA), we estimate model (10) using the average ICC derived from cross-sectional earnings predictions (ICCCSE). We still find a positive and significant association between our various proxies of 10-K linguistic complexity and a firm’s cost of equity capital. Economically, a one-standard-deviation increase in the Fog index is associated with a significant increase in the cost of equity of 0.36%.

These findings support our main hypothesis that a lower readability of the 10-K report increases investors’ risk perception. However, they could also be a byproduct of time-varying differences across industries or persistent differences across firms we fail to control for. In particular, persistent differences across firms (e.g., corporate culture) could explain both a higher cost of equity and a lower annual report readability. Panel C of Table 3 reports our baseline regression augmented by industry-year and firm fixed effects. Controlling for time-varying differences in ICC across industries, within a given firm, we again find that variations in the linguistic complexity of the 10-K are positively related to variations in the cost of equity[[16]](#footnote-16). The time-varying effect of readability on the cost of equity we document lends some support to our hypothesis that management uses linguistic complexity to obfuscate information and that investors react to this behavior by requiring a higher cost of equity. However, the existence of a positive relationship between linguistic complexity and the cost of equity is also open to alternative interpretations.

*3.4. Managerial Obfuscation versus Intrinsic Business Complexity*

While the literature generally interprets complex language in firms’ disclosures as indicative of managerial obfuscation, it can also reflect the provision of complex information (see Bushee et al. (2018)). Complex language can be necessary to convey information about the firm's business transactions and operating strategy (Bloomfield 2002). An important challenge is thus to isolate the effect of readability driven by managerial obfuscation from that of readability that is driven by the intrinsic complexity of a firm’s business. To identify increases in linguistic complexity most likely to result from managerial obfuscation as opposed to the provision of complex information, we use four complementary strategies.

First, we follow Bushee et al. (2018) and decompose linguistic complexity into two latent components. One reflects the language complexity inherent to the business activity and another more likely to be caused by the willingness to obfuscate information (i.e., not related to language complexity inherent to business activity). To do so, we regress our measures of linguistic complexity on proxies of firm business complexity to obtain a measure of linguistic complexity that is purged of the effect of the complexity of business activity (the residuals). We follow prior literature (e.g., Frijns et al. 2016; Boubakri and Mishra 2017) and measure firm complexity by the number of business segment sales and the number of geographical segment sales. Specifically, we use the inverse of the Herfindahl-Hirschman index of business segment sales and the inverse of the Herfindahl-Hirschman index of geographical segment sales as our two complementary proxies of firm complexity. For our three proxies of linguistic complexity, we name the measures of linguistic complexity purged from the effect of business complexity *Fog Index Orthogonal to Business Complexity, Grade Level Orthogonal to Business Complexity, and Reading Ease Orthogonal to Business Complexity*, respectively. We then replace our proxies of readability of the 10-K annual reports in our core regression (Table 3, Panel C) by their purged counterparts. Panel A of Table 4 reports the results. We document a strong and positive association between the purged linguistic complexity and a firm’s cost of equity. The magnitude of the effect is reduced (halved) compared to our specification using proxies of ‘raw’ linguistic complexity, suggesting that part of the effect we document is driven by business complexity. More importantly, the results indicate that there is an effect of linguistic complexity on a firm’s cost of equity that is not attributable to business complexity, which is consistent with a managerial obfuscation interpretation.

[*Insert Table 4 about here*]

Second, we build on the first test and examine the relationship between the purged linguistic complexity and cost of equity for firms likely to have managed their earnings. If managerial obfuscation is driving the effect of the purged linguistic complexity on a firm’s cost of equity, it should be much stronger in firms likely to have managed their earnings, i.e., in firms where management has an incentive to obfuscate information[[17]](#footnote-17). We identify firms likely to manage their earnings upwards as firms that meet or just beat the earnings target. Burgstahler and Dichev (1997) provide evidence that firms manage reported earnings to avoid earnings decrease and losses. Degeorge et al. (1999) show how thresholds induce specific types of earnings management, while Graham et al. (2005) report that some CFOs admit to manage earnings in order to meet earnings benchmarks. Bhojraj et al. (2009) document survey evidence suggesting managers engage in myopic behavior to beat benchmarks. In defining small earnings surprises, we follow prior studies, which use cut-offs ranging anywhere from zero to five cents, and select a cut-off of three cents (e.g., Brown 2001; Dhaliwal et al. 2004; Bhojraj et al. 2009; Haw et al. 2011; Alissa et al. 2013; Call et al. 2014). We focus on analyst forecasts as our measure of earnings expectations (e.g., Burgstahler and Dichev 1997; Caskey and Ozel 2017)[[18]](#footnote-18). More specifically, we use the last known I/B/E/S consensus forecast. For each measure of readability, we run the regression from Panel A of Table 4 in two different subsamples, one consisting of firms likely to have managed earnings to meet their benchmark (earnings surprise between 0 and 3 cents) and another one consisting of firms unlikely to have managed their earnings to meet their benchmark (negative earnings surprise or earnings surprise greater than 3 cents). Panel B of Table 4 report the results. We document a positive association between purged linguistic complexity and firm’s cost of equity only for firms likely to have managed their earnings to meet or just beat their earnings target. This finding is consistent with management using linguistic complexity to obfuscate information in order to beat or just meet their earnings target.

Third, economic theory suggests that obfuscation (informative disclosure) is associated with greater (lower) information asymmetry between shareholders and managers (Bloomfield 2002). Bushee et al. (2018) shows that while the information component of linguistic complexity reduces information asymmetry, the obfuscation component increases it. We use this prior finding to isolate the effect of readability attributable to obfuscation from the readability attributable to complex information disclosure. We investigate the effect of readability on the cost of equity conditional on whether a within-firm change in readability is associated with a greater within-firm change in information asymmetry. We measure information asymmetry using Amihud (2002)’s illiquidity construct (e.g., Lang and Maffett 2011; Bushee et al. 2018). Details of the variable construction are provided in Appendix A. If the effect we document is partially driven by managerial obfuscation, we should find a positive association between linguistic complexity and cost of equity in a subsample of firms for which increases in readability increase information asymmetry. Panel C of Table 4 reports the results. Results show that when information asymmetry has increased contemporaneously to an increase in linguistic complexity, the effect of linguistic complexity on the implied cost of equity is significantly positive and sizeable. When linguistic complexity induces greater information symmetry and is thus more likely to be a manifestation of managerial obfuscation, shareholders require a greater return on capital. These findings are consistent with managerial obfuscation being an important driver of the effect of linguistic complexity on a firm’s cost of equity.

Fourth, we focus on a section of the 10-K annual report where managers are more likely to use linguistic complexity to obfuscate information. Prior literature indicates that managers have more room to use their discretion to strategically obfuscate financial results in the Management Discussion and Analysis (MD&A) section of the 10-K annual report (e.g., Li 2008; Lo et al. 2017)[[19]](#footnote-19). The first three columns of Panel D of Table 4 report our core regression using proxies of readability computed only for the MD&A sections of the annual reports (*MD&A Fog Index, MD&A Grade Level, MD&A Reading Ease*). Results show that, as for the readability of the whole 10-K annual report, the readability of the MD&A section is negatively related to a firm’s cost of equity. We then generate measures of MD&A readability that are orthogonal to the readability of the whole 10-K document. The idea is to capture the component of linguistic complexity of the MD&A section that cannot be explained by the linguistic complexity of the rest of the 10-K document, and thus is more likely to be a manifestation of managerial discretion. We generate *MD&A Fog Index Orthogonal to 10-K Fog Index, MD&A Grade Level Orthogonal to 10-K Fog Index and MD&A Reading Ease Orthogonal to 10-K Fog Index,* by regressing *MD&A Fog Index*, *MD&A Grade Level*, and *MD&A Reading Ease* on *Fog Index*, *Grade Level* and *Reading Ease*, respectively. We include industry-year and firm fixed effects in these regressions. The last three columns of Panel D of Table 4 reports the results of our core regression when we replace our usual proxies of readability by the MD&A ones purged from the effect of the linguistic complexity of the whole 10-K report. We find that in the section of the 10-K annual report where managers are more likely to use their discretion to obfuscate information, an abnormal increase in linguistic complexity with respect to the rest of the document is associated with an increase in the cost of equity capital. The effect is significant and sizeable.

Overall, our results indicate that the effect of linguistic complexity on the cost of equity capital holds in settings where linguistic complexity is more likely to result from managerial obfuscation as opposed to the provision of complex information. It supports our core hypothesis that, when lower readability of financial information signals an attempt to obfuscate information, the ease with which investors can process text influences their risk perception.

*3.5. Information Readability versus Information Quantity*

Our findings echo those of Boubakri and Mishra (2017), who show that the quantity of information disclosed in a 10-K report may affect the cost of equity capital. While Boubakri and Mishra (2017) argue that too much information (information quantity) may overload investors and result in a higher rate of return required, we argue that information readability, which is quantity-adjusted, may change the risk perception of investors and result in a higher rate of return required.

Using file size as our main proxy for information quantity (e.g., Loughran and McDonald 2014; Boubakri and Mishra 2017), we seek to isolate the effect of readability from information overload. We compute the file size of the 10-K annual reports in line with Loughran and McDonald (2014). We remove pictures and html coding, so the file size is a direct measure of the number of characters contained within the document. Therefore, a larger file size represents a document that contains more content, and by extension more information. Although our readability measures are adjusted for the number of words and sentences, prior literature shows that file size can be used as a proxy for readability (Loughran and McDonald 2014). We thus examine how 10-K readability and file size are related for our sample firms. Panel A of Table 5 reports the results of OLS regressions of our proxies of 10-K readability on file size plus industry-year and firm fixed effects. We find a positive association between file size and readability, consistent with Loughran and McDonald (2014). However, we also observe that about 50% of the within-firm variation in readability are not explained away by file size variations. We seek to use this unexplained portion of 10-K readability as the component of linguistic complexity that is not related to information quantity.

[*Insert Table 5 about here*]

We use readability proxies that are orthogonal to file size to examine whether information complexity unrelated to information quantity still has a positive association with a firm’s cost of equity. Panel B of Table 5 reports the results of the OLS regression of ICCA on 10-K readability measures orthogonal to 10-K file size and our usual control variables. Results show that there is an effect of linguistic complexity on the cost of equity capital above and beyond the effect of information quantity. Our findings complement those of Boubakri and Mishra (2017), and, together, show that the way a firm frames its information disclosure to investors may affect its cost of equity capital.

1. **Conclusion**

In this paper, we investigate how linguistic complexity relates to a company’s cost of equity capital. We argue that information readability may influence investors’ perception of a firm’s riskiness. In particular, when management uses linguistic complexity to obfuscate information, we expect investors to perceive a greater risk and hence to require a greater return on equity. Our results show that, within a given firm, an increase in the linguistic complexity of the annual report is positively associated with an increase in the cost of equity capital. Because greater linguistic complexity may result from managerial obfuscation but can also be a byproduct of greater intrinsic business complexity, we run a series of tests to disentangle both effects. Our results show that the linguistic complexity that cannot be explained away by the business intrinsic complexity is still positively related to the cost of equity capital. In addition, this effect is much stronger for firms likely to have managed earnings to meet or just beat their earnings forecast. In a subsample of firms where an increase in linguistic complexity leads to greater information asymmetry, which is suggestive of managerial obfuscation rather than the disclosure of complex information, we still find a positive association with the cost of equity capital. We also show that the linguistic complexity of the MD&A section is positively related to a firm’s cost of equity and that linguistic complexity of the MD&A section that is orthogonal to the linguistic complexity of the rest of the 10-K document is still positively related to a firm’s cost of equity. This set of findings is consistent with managerial obfuscation being an important driver of the positive association between linguistic complexity of the 10-K and a firm’s cost of equity. Finally, prior literature shows that information overload may influence the cost of equity. We address the concern of the cofounding effect of information quantity and readability by documenting an effect of information readability on a firm’s cost of equity that is orthogonal to the one of information quantity.

**References**

Alissa, W., S. B. Bonsall IV, K. Koharki, and M. W. Penn Jr. 2013. Firms' use of accounting discretion to influence their credit ratings. *Journal of Accounting and Economics* 55 (2-3):129-147.

Amihud, Y. 2002. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets* 5 (1):31-56.

Attig, N., S. Cleary, S. El Ghoul, and O. Guedhami. 2013. Institutional investment horizons and the cost of equity capital. *Financial Management* 42 (2):441-477.

Attig, N., O. Guedhami, and D. Mishra. 2008. Multiple large shareholders, control contests, and implied cost of equity. *Journal of Corporate Finance* 14 (5):721-737.

Barry, C. B., and S. J. Brown. 1985. Differential information and security market equilibrium. *Journal of Financial and Quantitative Analysis* 20 (4):407-422.

Ben‐Nasr, H., N. Boubakri, and J. C. Cosset. 2012. The political determinants of the cost of equity: Evidence from newly privatized firms. *Journal of Accounting Research* 50 (3):605-646.

Beyer, A., D. A. Cohen, T. Z. Lys, and B. R. Walther. 2010. The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics* 50 (2):296-343.

Bhojraj, S., P. Hribar, M. Picconi, and J. McInnis. 2009. Making sense of cents: An examination of firms that marginally miss or beat analyst forecasts. *The Journal of Finance* 64 (5):2361-2388.

Bloomfield, R. 2008. Discussion of “annual report readability, current earnings, and earnings persistence”. *Journal of Accounting and Economics* 45 (2-3):248-252.

Bloomfield, R. J. 2002. The “incomplete revelation hypothesis” and financial reporting. *Accounting Horizons* 16 (3):233-243.

Boubakri, N., S. El Ghoul, and W. Saffar. 2014. Political rights and equity pricing. *Journal of Corporate Finance* 27:326-344.

Boubakri, N., O. Guedhami, D. Mishra, and W. Saffar. 2012. Political connections and the cost of equity capital. *Journal of Corporate Finance* 18 (3):541-559.

Boubakri, N., O. Guedhami, and W. Saffar. 2016. Geographic location, foreign ownership, and cost of equity capital: Evidence from privatization. *Journal of Corporate Finance* 38:363-381.

Boubakri, N., and D. R. Mishra. 2017. Information Overload and Cost of Equity Capital.

Brown, L. D. 2001. A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research* 39 (2):221-241.

Brown, L. D., and M. L. Caylor. 2005. A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences. *The Accounting Review* 80 (2):423-440.

Brushwood, J., D. Dhaliwal, D. Fairhurst, and M. Serfling. 2016. Property crime, earnings variability, and the cost of capital. *Journal of Corporate Finance* 40:142-173.

Burgstahler, D., and I. Dichev. 1997. Earnings management to avoid earnings decreases and losses. *Journal of accounting and economics* 24 (1):99-126.

Bushee, B. J., I. D. Gow, and D. J. Taylor. 2018. Linguistic Complexity in Firm Disclosures: Obfuscation or Information? *Journal of Accounting Research* 56 (1):85-121.

Call, A. C., S. Chen, B. Miao, and Y. H. Tong. 2014. Short-term earnings guidance and accrual-based earnings management. *Review of Accounting Studies* 19 (2):955-987.

Caskey, J., and N. B. Ozel. 2017. Earnings expectations and employee safety. *Journal of accounting and economics* 63 (1):121-141.

Chen, Z., D. S. Dhaliwal, and H. Xie. 2010. Regulation fair disclosure and the cost of equity capital. *Review of Accounting Studies* 15 (1):106-144.

Chen, Z., Y. Huang, and K. J. Wei. 2013. Executive pay disparity and the cost of equity capital. *Journal of Financial and Quantitative Analysis* 48 (3):849-885.

Chen, Z., O. Z. Li, and H. Zou. 2016. Directors׳ and officers׳ liability insurance and the cost of equity. *Journal of Accounting and Economics* 61 (1):100-120.

Claus, J., and J. Thomas. 2001. Equity premia as low as three percent? Evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance* 56 (5):1629-1666.

Coles, J. L., U. Loewenstein, and J. Suay. 1995. On equilibrium pricing under parameter uncertainty. *Journal of Financial and Quantitative Analysis* 30 (3):347-364.

Core, J. E. 2001. A review of the empirical disclosure literature: discussion. *Journal of Accounting and Economics* 31 (1-3):441-456.

Daske, H., L. Hail, C. Leuz, and R. Verdi. 2013. Adopting a label: Heterogeneity in the economic consequences around IAS/IFRS adoptions. *Journal of Accounting Research* 51 (3):495-547.

Dechow, P. M., S. A. Richardson, and I. Tuna. 2003. Why are earnings kinky? An examination of the earnings management explanation. *Review of Accounting Studies* 8 (2-3):355-384.

Degeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *The Journal of Business* 72 (1):1-33.

Dhaliwal, D., J. S. Judd, M. Serfling, and S. Shaikh. 2016. Customer concentration risk and the cost of equity capital. *Journal of Accounting and Economics* 61 (1):23-48.

Dhaliwal, D., L. Krull, O. Z. Li, and W. Moser. 2005. Dividend taxes and implied cost of equity capital. *Journal of accounting research* 43 (5):675-708.

Dhaliwal, D. S., C. A. Gleason, and L. F. Mills. 2004. Last‐chance earnings management: Using the tax expense to meet analysts' forecasts. *Contemporary Accounting Research* 21 (2):431-459.

Dhaliwal, D. S., O. Z. Li, A. Tsang, and Y. G. Yang. 2011. Voluntary nonfinancial disclosure and the cost of equity capital: The initiation of corporate social responsibility reporting. *The Accounting Review* 86 (1):59-100.

Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64 (2-3):221-245.

Easton, P. D. 2004. PE ratios, PEG ratios, and estimating the implied expected rate of return on equity capital. *The Accounting Review* 79 (1):73-95.

El Ghoul, S., O. Guedhami, C. C. Kwok, and D. R. Mishra. 2011. Does corporate social responsibility affect the cost of capital? *Journal of Banking & Finance* 35 (9):2388-2406.

Elton, E. J. 1999. Presidential Address: Expected Return, Realized Return, and Asset Pricing Tests. *The Journal of Finance* 54 (4):1199-1220.

Fama, E. F., and K. R. French. 1992. The cross‐section of expected stock returns. *The Journal of Finance* 47 (2):427-465.

———. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1):3-56.

———. 1997. Industry costs of equity. *Journal of Financial Economics* 43 (2):153-193.

Fields, T. D., T. Z. Lys, and L. Vincent. 2001. Empirical research on accounting choice. *Journal of Accounting and Economics* 31 (1-3):255-307.

Frijns, B., O. Dodd, and H. Cimerova. 2016. The impact of cultural diversity in corporate boards on firm performance. *Journal of Corporate Finance* 41:521-541.

Fu, R., A. Kraft, and H. Zhang. 2012. Financial reporting frequency, information asymmetry, and the cost of equity. *Journal of Accounting and Economics* 54 (2-3):132-149.

Gebhardt, W. R., C. Lee, and B. Swaminathan. 2001. Toward an implied cost of capital. *Journal of Accounting Eesearch* 39 (1):135-176.

Gode, D., and P. Mohanram. 2003. Inferring the cost of capital using the Ohlson–Juettner model. *Review of Accounting Studies* 8 (4):399-431.

Goh, B. W., J. Lee, C. Y. Lim, and T. Shevlin. 2016. The effect of corporate tax avoidance on the cost of equity. *The Accounting Review* 91 (6):1647-1670.

Gordon, J. R., and M. J. Gordon. 1997. The finite horizon expected return model. *Financial Analysts Journal* 53 (3):52-61.

Graham, J. R., C. R. Harvey, and S. Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1):3-73.

Guay, W., D. Samuels, and D. Taylor. 2016. Guiding through the fog: Financial statement complexity and voluntary disclosure. *Journal of Accounting and Economics* 62 (2-3):234-269.

Hail, L., and C. Leuz. 2006. International differences in the cost of equity capital: Do legal institutions and securities regulation matter? *Journal of Accounting Research* 44 (3):485-531.

Handa, P., and S. C. Linn. 1993. Arbitrage pricing with estimation risk. *Journal of Financial and Quantitative Analysis* 28 (1):81-100.

Haw, I. M., S. S. Ho, and A. Y. Li. 2011. Corporate governance and earnings management by classification shifting. *Contemporary Accounting Research* 28 (2):517-553.

Hou, K., M. A. Van Dijk, and Y. Zhang. 2012. The implied cost of capital: A new approach. *Journal of Accounting and Economics* 53 (3):504-526.

Hwang, B.-H., and H. H. Kim. 2017. It pays to write well. *Journal of Financial Economics* 124 (2):373-394.

Jones, C. S., and S. Tuzel. 2013. Inventory investment and the cost of capital. *Journal of Financial Economics* 107 (3):557-579.

Kothari, S. P., X. Li, and J. E. Short. 2009. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review* 84 (5):1639-1670.

Lang, M., and R. Lundholm. 1993. Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research*:246-271.

Lang, M., and M. Maffett. 2011. Transparency and liquidity uncertainty in crisis periods. *Journal of Accounting and Economics* 52 (2-3):101-125.

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.

Leuz, C., and P. D. Wysocki. 2016. The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research* 54 (2):525-622.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2-3):221-247.

———. 2010a. Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature* 29:143.

Li, S. 2010b. Does mandatory adoption of International Financial Reporting Standards in the European Union reduce the cost of equity capital? *The Accounting Review* 85 (2):607-636.

Lintner, J. 1965. Security prices, risk, and maximal gains from diversification. *The Journal of Finance* 20 (4):587-615.

Lo, K., F. Ramos, and R. Rogo. 2017. Earnings management and annual report readability. *Journal of Accounting and Economics* 63 (1):1-25.

Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *The Journal of Finance* 69 (4):1643-1671.

———. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54 (4):1187-1230.

Miller, B. P. 2010. The effects of reporting complexity on small and large investor trading. *The Accounting Review* 85 (6):2107-2143.

Mishra, D. R. 2014. The dark side of CEO ability: CEO general managerial skills and cost of equity capital. *Journal of Corporate Finance* 29:390-409.

Ng, A. C., and Z. Rezaee. 2015. Business sustainability performance and cost of equity capital. *Journal of Corporate Finance* 34:128-149.

Ohlson, J. A., and B. E. Juettner-Nauroth. 2005. Expected EPS and EPS growth as determinantsof value. *Review of Accounting Studies* 10 (2-3):349-365.

Pástor, Ľ., M. Sinha, and B. Swaminathan. 2008. Estimating the intertemporal risk–return tradeoff using the implied cost of capital. *The Journal of Finance* 63 (6):2859-2897.

Rennekamp, K. 2012. Processing fluency and investors’ reactions to disclosure readability. *Journal of Accounting Research* 50 (5):1319-1354.

Richardson, A. J., and M. Welker. 2001. Social disclosure, financial disclosure and the cost of equity capital. *Accounting, Organizations and Society* 26 (7-8):597-616.

Schrand, C. M., and B. R. Walther. 2000. Strategic benchmarks in earnings announcements: The selective disclosure of prior-period earnings components. *The Accounting Review* 75 (2):151-177.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19 (3):425-442.

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

Table 1. Descriptive statistics

*Panel A: Main dependent and independent variables*

This table presents descriptive statistics of the variables we use throughout our empirical analysis. The sample period is 2000-2016. Variable definitions are provided in Appendix A.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Obs. | Mean | S.D. | 25% | Median | 75% |
| *Readability Variables* | | | | | | | |
| Words | 16,855 | 55,491.26 | 40,403.86 | 32,219 | 45,535 | 65,926 |
| Sentences | 16,855 | 1,597.27 | 931.58 | 1,029 | 1,404 | 1,910 |
| Sentence Length | 16,855 | 33.54 | 5.41 | 30.12 | 32.35 | 35.53 |
| % Complex | 16,855 | 0.24 | 0.01 | 0.23 | 0.24 | 0.25 |
| Fog Index | 16,855 | 22.94 | 1.95 | 21.68 | 22.62 | 23.8 |
| Reading Ease | 16,855 | 16.16 | 4.73 | 13.59 | 16.51 | 19.19 |
| Grade Level | 16,855 | 19.33 | 1.85 | 18.14 | 18.98 | 20.11 |
| MD&A Fog Index | 13,301 | 22.25 | 1.97 | 20.98 | 22.08 | 23.23 |
| MD&A Reading Ease | 13,301 | 16.63 | 6.03 | 13.44 | 16.95 | 20.44 |
| MD&A Grade Level | 13,301 | 18.91 | 1.93 | 17.67 | 18.71 | 19.84 |
| File Size | 16,855 | 422.55 | 297.31 | 253.06 | 348.54 | 496.13 |
| *ICCA - Analyst Forecast Estimates – in %* | | | | | | | |
| RGG | 10,063 | 8.85 | 3.64 | 6.72 | 8.66 | 10.58 |
| RETN | 11,956 | 10.42 | 4.63 | 7.95 | 9.70 | 11.81 |
| RCT | 13,859 | 8.38 | 3.49 | 6.32 | 8.16 | 9.94 |
| ROJN | 11,472 | 10.51 | 4.55 | 8.06 | 9.78 | 11.83 |
| RGLS | 9,191 | 9.00 | 5.00 | 6.48 | 8.26 | 10.48 |
| ICCA | 16,855 | 9.46 | 3.53 | 7.28 | 9.09 | 10.96 |
| *ICCCSE – Cross-Sectional Earnings Model Estimates – in %* | | | | | | | |
| RGG | 13,115 | 11.05 | 12.09 | 5.14 | 7.73 | 11.98 |
| RETN | 13,577 | 12.36 | 11.17 | 6 | 9.4 | 14.66 |
| RCT | 12,975 | 10.31 | 11.92 | 4.95 | 7.23 | 10.83 |
| ROJN | 13,046 | 7.98 | 8.83 | 3.87 | 5.8 | 8.9 |
| RGLS | 12,196 | 8.75 | 6.43 | 5.46 | 7.72 | 10.35 |
| ICCCSE | 13,604 | 10.65 | 9.74 | 5.77 | 7.89 | 11.46 |
| *Control Variables* | | | | | | | |
| Size | 16,855 | 7.55 | 1.74 | 6.34 | 7.46 | 8.68 |
| Leverage | 16,855 | 0.22 | 0.16 | 0.10 | 0.21 | 0.32 |
| Book to Market | 16,855 | 0.58 | 0.36 | 0.33 | 0.52 | 0.76 |
| Dividend Payer | 16,855 | 0.74 | 0.44 | 0.00 | 1.00 | 1.00 |
| CAPM Beta | 16,855 | 0.90 | 1.30 | 0.26 | 0.91 | 1.57 |
| Institutional Ownership | 16,855 | 0.62 | 0.27 | 0.46 | 0.67 | 0.81 |
| Analyst Coverage | 16,855 | 8.89 | 7.46 | 3.00 | 6.92 | 13.25 |
| Forecast Dispersion | 16,855 | 0.02 | 0.06 | 0.01 | 0.01 | 0.03 |
| Long-Term Growth | 16,855 | 0.12 | 0.07 | 0.08 | 0.11 | 0.15 |
| *Complexity Variables* |  |  |  |  |  |  |
| Complexity - Business | 16,855 | 1.40 | 1.30 | 0.00 | 1.00 | 2.00 |
| Complexity - Geographic | 16,855 | 1.34 | 1.26 | 0.00 | 1.00 | 1.93 |
| *Other Variables* |  |  |  |  |  |  |
| Information Asymmetry | 16,855 | 0.07 | 0.39 | 0.00 | 0.00 | 0.01 |
| Earnings Surprises | 16,706 | 0.01 | 0.31 | -0.01 | 0.01 | 0.05 |

*Panel B: Changes in average 10-K readability and ICC over time*

This table provides the cross-sectional average per year for selected variables. Variable definitions are provided in Appendix A.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Year* | Words | Sentences | WPS | % Complex | Fog Index | Reading Ease | Grade Level | ICCA (%) |
| 2000 | 40,174.17 | 1,122.66 | 33.54 | 0.23 | 22.46 | 18.48 | 19.02 | 9.82 |
| 2001 | 41,881.28 | 1,172.69 | 33.51 | 0.23 | 22.48 | 18.44 | 19.01 | 9.18 |
| 2002 | 47,591.27 | 1,326.92 | 33.65 | 0.23 | 22.62 | 17.90 | 19.12 | 9.21 |
| 2003 | 53,772.27 | 1,477.78 | 34.59 | 0.23 | 23.10 | 16.56 | 19.52 | 8.52 |
| 2004 | 54,349.54 | 1,503.49 | 34.46 | 0.23 | 23.12 | 16.32 | 19.52 | 8.54 |
| 2005 | 52,506.96 | 1,503.98 | 33.87 | 0.24 | 23.05 | 16.19 | 19.41 | 9.34 |
| 2006 | 51,508.77 | 1,514.63 | 33.05 | 0.24 | 22.81 | 16.45 | 19.16 | 9.16 |
| 2007 | 52,294.81 | 1,551.14 | 32.85 | 0.24 | 22.78 | 16.42 | 19.12 | 9.51 |
| 2008 | 58,531.51 | 1,687.31 | 33.29 | 0.24 | 22.93 | 16.06 | 19.28 | 10.46 |
| 2009 | 65,514.98 | 1,852.14 | 34.11 | 0.24 | 23.26 | 15.32 | 19.59 | 10.40 |
| 2010 | 59,620.59 | 1,753.74 | 33.16 | 0.24 | 22.96 | 15.62 | 19.31 | 10.16 |
| 2011 | 60,759.98 | 1,777.30 | 33.15 | 0.24 | 22.96 | 15.56 | 19.32 | 10.63 |
| 2012 | 59,868.52 | 1,759.39 | 33.20 | 0.24 | 23.01 | 15.28 | 19.37 | 10.07 |
| 2013 | 61,146.86 | 1,798.06 | 33.27 | 0.24 | 23.04 | 15.21 | 19.40 | 9.02 |
| 2014 | 61,892.14 | 1,805.38 | 33.33 | 0.24 | 23.08 | 15.04 | 19.44 | 8.92 |
| 2015 | 61,168.96 | 1,798.30 | 33.28 | 0.24 | 23.11 | 14.90 | 19.45 | 9.33 |
| 2016 | 64,413.38 | 1,873.47 | 33.58 | 0.25 | 23.24 | 14.44 | 19.58 | 8.84 |

*Panel C: 10-K Readability and ICC by industry*

This table provides the average readability measures and the ICCA estimate by industry (Fama-French 12 Industry Classification). Variable definitions are provided in Appendix A.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *FF-12 Industries* | Number of  firm-year observations | Fog Index | Grade Level | Reading Ease | ICCA (%) |
| Business Equipment | 1,810 | 22.73 | 15.89 | 19.18 | 8.89 |
| Chemicals | 637 | 23.00 | 16.02 | 19.47 | 9.30 |
| Consumer Durables | 440 | 22.55 | 17.11 | 19.09 | 10.62 |
| Consumer Non-Durables | 917 | 22.51 | 17.94 | 18.99 | 9.33 |
| Energy | 654 | 22.89 | 17.46 | 19.27 | 10.88 |
| Finance | 4,146 | 23.19 | 15.63 | 19.41 | 9.68 |
| Health | 1,117 | 22.81 | 16.00 | 19.26 | 8.39 |
| Manufacturing | 2,178 | 22.76 | 16.31 | 19.25 | 10.08 |
| Other | 1,864 | 22.87 | 16.33 | 19.30 | 9.51 |
| Shops | 1,757 | 22.88 | 17.02 | 19.24 | 9.64 |
| Telecom | 242 | 23.20 | 15.17 | 19.59 | 9.40 |
| Utilities | 1,093 | 23.54 | 14.41 | 19.98 | 7.91 |

Table 2. Difference in ICC between firms with high and low 10-K readability

This table presents the differences in the implied cost of equity based on analyst earnings forecast (ICCA) between the top and bottom 25% of firm-years sorted by the readability scores. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively. The table reports the significance of t-tests of mean differences and of Wilcoxon Rank-Sum tests of median differences.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | |  |  | | --- | --- | | ICCA |  | | | |
| *10-K Readability Scores* |  | Bottom Quartile | Top Quartile | Difference |
| Fog Index | Mean | 9.09 | 9.72 | 0.63\*\*\* |
| Median | 8.83 | 9.19 | 0.36\*\*\* |
| Grade Level | Mean | 9.06 | 9.66 | 0.57\*\*\* |
| Median | 8.86 | 9.14 | 0.28\*\*\* |
| Reading Ease | Mean | 9.58 | 9.22 | -0.36\* |
| Median | 9.11 | 8.91 | -0.20\*\*\* |

Table 3. 10-K Readability and ICC – Regression Analysis

*Panel A: Baseline regressions*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are alternative measures of 10-K readability. We include the control variables of the model (10). The regressions include industry and year fixed effects. The variables are defined in Appendix A. We do not report the constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| ICCA | (1) | (2) | (3) |
|  |  |  |  |
| **Fog Index** | **0.082\*\*\*** |  |  |
|  | **(4.487)** |  |  |
| **Grade Level** |  | **0.086\*\*\*** |  |
|  |  | **(4.512)** |  |
| **Reading Ease** |  |  | **-0.023\*\*\*** |
|  |  |  | **(-2.776)** |
| Size | -0.178\*\*\* | -0.180\*\*\* | -0.181\*\*\* |
|  | (-4.023) | (-4.070) | (-4.078) |
| Leverage | 0.367\*\*\* | 0.366\*\*\* | 0.369\*\*\* |
|  | (11.260) | (11.257) | (11.326) |
| Book to Market | 2.402\*\*\* | 2.398\*\*\* | 2.404\*\*\* |
|  | (14.036) | (14.012) | (14.022) |
| Dividend Payer | 0.849\*\*\* | 0.849\*\*\* | 0.849\*\*\* |
|  | (7.155) | (7.157) | (7.136) |
| Ln(Analyst Coverage) | -0.119 | -0.118 | -0.117 |
|  | (-1.609) | (-1.594) | (-1.575) |
| Institutional Ownership | -0.515\*\* | -0.516\*\* | -0.493\*\* |
|  | (-2.170) | (-2.175) | (-2.075) |
| Forecast Dispersion | 1.692\*\* | 1.691\*\* | 1.717\*\* |
|  | (2.528) | (2.525) | (2.557) |
| Long-Term Growth | 3.213\*\*\* | 3.207\*\*\* | 3.196\*\*\* |
|  | (4.259) | (4.251) | (4.233) |
|  |  |  |  |
| Observations | 16,855 | 16,855 | 16,855 |
| Year Fixed Effects | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes |
| Adjusted R-squared | 0.190 | 0.190 | 0.189 |

*Panel B: Cross-sectional estimates of ICC*

This panel presents the OLS regression results where the dependent variable is the implied cost of equity based on cross-sectional estimates (ICCCSE) and the main independent variables are the alternative measures of the 10-K readability. We include the control variables of the model (10). The regressions include industry and year fixed effects. All the variables are defined in Appendix A. We do not report the constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| ICCCSE | (1) | (2) | (3) |
|  |  |  |  |
| **Fog Index** | **0.187\*\*\*** |  |  |
|  | **(4.006)** |  |  |
| **Grade Level** |  | **0.192\*\*\*** |  |
|  |  | **(3.893)** |  |
| **Reading Ease** |  |  | **-0.053\*\*** |
|  |  |  | **(-2.559)** |
| Size | -2.468\*\*\* | -2.470\*\*\* | -2.476\*\*\* |
|  | (-14.493) | (-14.505) | (-14.527) |
| Ln(Analyst Coverage) | 1.118\*\*\* | 1.119\*\*\* | 1.121\*\*\* |
|  | (5.414) | (5.419) | (5.423) |
| CAPM Beta | 0.009 | 0.009 | 0.007 |
|  | (0.077) | (0.072) | (0.059) |
| Institutional Ownership | -9.065\*\*\* | -9.060\*\*\* | -9.026\*\*\* |
|  | (-12.384) | (-12.379) | (-12.331) |
| Book to Market | 3.524\*\*\* | 3.520\*\*\* | 3.524\*\*\* |
|  | (7.315) | (7.303) | (7.303) |
| Leverage | 0.574\*\*\* | 0.574\*\*\* | 0.580\*\*\* |
|  | (6.537) | (6.532) | (6.601) |
| Dividend Payer | -0.067 | -0.069 | -0.065 |
|  | (-0.232) | (-0.237) | (-0.224) |
| Forecast Dispersion | -9.791\*\*\* | -9.800\*\*\* | -9.796\*\*\* |
|  | (-4.915) | (-4.920) | (-4.910) |
| Long-Term Growth | -8.720\*\*\* | -8.733\*\*\* | -8.739\*\*\* |
|  | (-4.693) | (-4.701) | (-4.697) |
|  |  |  |  |
| Observations | 13,604 | 13,604 | 13,604 |
| Year Fixed Effects | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes |
| Adjusted R-squared | 0.276 | 0.276 | 0.275 |

*Panel C: Inclusion of industry-year and firm fixed effects*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are alternative measures of 10-K readability. We include the control variables of the model (10). The regressions 1-3 include industry-year fixed effects. The regressions 4-6 include industry-year fixed effects and firm fixed effects. The variables are defined in Appendix A. We do not report the constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Industry-Year Fixed Effects | | | Industry-Year Fixed Effects  + Firm Fixed Effects | | |
| ICCA | (1) | (2) | (3) | (4) | (5) | (6) |
|  |  |  |  |  |  |  |
| **Fog Index** | **0.078\*\*\*** |  |  | **0.046\*\*\*** |  |  |
|  | **(4.117)** |  |  | **(2.837)** |  |  |
| **Grade Level** |  | **0.081\*\*\*** |  |  | **0.047\*\*\*** |  |
|  |  | **(4.165)** |  |  | **(2.735)** |  |
| **Reading Ease** |  |  | **-0.021\*\*** |  |  | **-0.016\*\*** |
|  |  |  | **(-2.520)** |  |  | **(-2.016)** |
| Size | -0.169\*\*\* | -0.171\*\*\* | -0.172\*\*\* | -0.574\*\*\* | -0.574\*\*\* | -0.575\*\*\* |
|  | (-3.677) | (-3.720) | (-3.729) | (-4.926) | (-4.927) | (-4.930) |
| Ln(Analyst Coverage) | -0.154\*\* | -0.153\*\* | -0.152\*\* | -0.060 | -0.060 | -0.060 |
|  | (-2.016) | (-2.003) | (-1.977) | (-0.788) | (-0.791) | (-0.792) |
| CAPM Beta | 0.145\*\*\* | 0.144\*\*\* | 0.145\*\*\* | 0.115\*\*\* | 0.115\*\*\* | 0.115\*\*\* |
|  | (3.556) | (3.544) | (3.552) | (2.695) | (2.690) | (2.684) |
| Institutional Ownership | -0.374 | -0.375 | -0.354 | -1.012\*\*\* | -1.012\*\*\* | -1.009\*\*\* |
|  | (-1.461) | (-1.466) | (-1.381) | (-3.265) | (-3.264) | (-3.255) |
| Book to Market | 2.527\*\*\* | 2.523\*\*\* | 2.528\*\*\* | 1.617\*\*\* | 1.617\*\*\* | 1.618\*\*\* |
|  | (13.698) | (13.677) | (13.679) | (7.147) | (7.146) | (7.146) |
| Leverage | 0.354\*\*\* | 0.353\*\*\* | 0.356\*\*\* | 0.183\*\*\* | 0.183\*\*\* | 0.183\*\*\* |
|  | (10.252) | (10.247) | (10.314) | (4.388) | (4.388) | (4.396) |
| Dividend Payer | 0.869\*\*\* | 0.870\*\*\* | 0.869\*\*\* | 0.951\*\*\* | 0.950\*\*\* | 0.950\*\*\* |
|  | (7.110) | (7.114) | (7.091) | (5.623) | (5.619) | (5.611) |
| Forecast Dispersion | 1.901\*\*\* | 1.900\*\*\* | 1.922\*\*\* | 0.384 | 0.386 | 0.395 |
|  | (2.780) | (2.779) | (2.805) | (0.540) | (0.541) | (0.553) |
| Long-Term Growth | 2.733\*\*\* | 2.729\*\*\* | 2.715\*\*\* | 3.968\*\*\* | 3.969\*\*\* | 3.954\*\*\* |
|  | (3.498) | (3.492) | (3.471) | (4.751) | (4.751) | (4.729) |
|  |  |  |  |  |  |  |
| Observations | 16,855 | 16,855 | 16,855 | 16,855 | 16,855 | 16,855 |
| Industry-Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | No | No | No | Yes | Yes | Yes |
| Number of firms | - | - | - | 2,288 | 2,288 | 2,288 |
| Firm Cluster | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.219 | 0.219 | 0.218 | 0.513 | 0.513 | 0.513 |

Table 4. Disentangling the effect of managerial obfuscation from intrinsic business complexity

*Panel A: Linguistic complexity purged from the effect of business complexity*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are alternative measure 10-K readability purged from the effect of business complexity. We obtain the purged measures of readability by using the residuals of a regression of the actual measures of readability on a constant, *Complexity – Business, Complexity – Geographic,* industry-year, and firm fixed effects. All regressions include industry-year and firm fixed effects. All the variables are defined in Appendix A. We do not report constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| ICCA | (1) | (2) | (3) |
|  |  |  |  |
| **Fog Index Orthogonal to Business Complexity** | **0.046\*\*\*** |  |  |
|  | **(2.836)** |  |  |
| **Grade Level Orthogonal to Business Complexity** |  | **0.047\*\*\*** |  |
|  |  | **(2.734)** |  |
| **Reading Ease Orthogonal to Business Complexity** |  |  | **-0.016\*\*** |
|  |  |  | **(-2.017)** |
| Size | -0.574\*\*\* | -0.574\*\*\* | -0.575\*\*\* |
|  | (-4.926) | (-4.927) | (-4.930) |
| Ln(Analyst Coverage) | -0.060 | -0.060 | -0.060 |
|  | (-0.788) | (-0.791) | (-0.792) |
| CAPM Beta | 0.115\*\*\* | 0.115\*\*\* | 0.115\*\*\* |
|  | (2.695) | (2.690) | (2.684) |
| Institutional Ownership | -1.012\*\*\* | -1.012\*\*\* | -1.009\*\*\* |
|  | (-3.265) | (-3.264) | (-3.255) |
| Book to Market | 1.617\*\*\* | 1.617\*\*\* | 1.618\*\*\* |
|  | (7.147) | (7.146) | (7.146) |
| Leverage | 0.183\*\*\* | 0.183\*\*\* | 0.183\*\*\* |
|  | (4.388) | (4.388) | (4.396) |
| Dividend Payer | 0.951\*\*\* | 0.950\*\*\* | 0.950\*\*\* |
|  | (5.623) | (5.619) | (5.611) |
| Forecast Dispersion | 0.384 | 0.386 | 0.395 |
|  | (0.540) | (0.541) | (0.553) |
| Long-Term Growth | 3.968\*\*\* | 3.969\*\*\* | 3.954\*\*\* |
|  | (4.751) | (4.751) | (4.729) |
|  |  |  |  |
| Observations | 16,855 | 16,855 | 16,855 |
| Industry-Year Fixed Effects | Yes | Yes | Yes |
| Firm Fixed Effects | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes |
| Adjusted R-squared | 0.513 | 0.513 | 0.513 |

*Panel B: Linguistic complexity purged from the effect of business complexity and ICC conditional on a firm meeting or just beating the earnings target*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are alternative measure 10-K readability purged from the effect of business complexity. We obtain the purged measures of readability by using the residuals of a regression of the actual measures of readability on a constant, *Complexity – Business, Complexity – Geographic,* industry-year, and firm fixed effects. For each proxy of readability, we run the regression in a sample of firms with an earnings surprise between 0 and 3 cents (meeting or just beating the earnings target = 1) and in a sample of firms with a negative earnings surprise or an earnings surprise greater than 3 cents (meeting or just beating the earnings target = 0). All regressions include industry-year and firm fixed effects. All the variables are defined in Appendix A. We do not report constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Earnings Per Share Surprise [$0,$0.03] | | | | | |
| ICCA | =0 | =1 | =0 | =1 | =0 | =1 |
|  |  |  |  |  |  |  |
| **Fog Index Orthogonal to Business Complexity** | **0.024** | **0.065\*\*** |  |  |  |  |
|  | **(0.971)** | **(2.517)** |  |  |  |  |
| **Grade Level Orthogonal to Business Complexity** |  |  | **0.023** | **0.066\*\*** |  |  |
|  |  |  | **(0.893)** | **(2.492)** |  |  |
| **Reading Ease Orthogonal to Business Complexity** |  |  |  |  | **-0.005** | **-0.027\*\*** |
|  |  |  |  |  | **(-0.444)** | **(-2.285)** |
| Size | -0.814\*\*\* | -0.153 | -0.814\*\*\* | -0.152 | -0.816\*\*\* | -0.153 |
|  | (-4.636) | (-0.812) | (-4.638) | (-0.811) | (-4.647) | (-0.813) |
| Ln(Analyst Coverage) | -0.181\* | 0.107 | -0.181\* | 0.107 | -0.182\* | 0.107 |
|  | (-1.711) | (0.664) | (-1.713) | (0.665) | (-1.719) | (0.662) |
| CAPM Beta | 0.210\*\*\* | 0.037 | 0.210\*\*\* | 0.037 | 0.210\*\*\* | 0.036 |
|  | (3.343) | (0.551) | (3.343) | (0.543) | (3.346) | (0.525) |
| Institutional Ownership | -0.389 | -1.204\*\* | -0.388 | -1.205\*\* | -0.379 | -1.199\*\* |
|  | (-0.853) | (-2.337) | (-0.849) | (-2.339) | (-0.832) | (-2.327) |
| Book to Market | 1.603\*\*\* | 1.638\*\*\* | 1.603\*\*\* | 1.638\*\*\* | 1.603\*\*\* | 1.644\*\*\* |
|  | (5.067) | (3.985) | (5.065) | (3.986) | (5.061) | (4.000) |
| Leverage | 0.227\*\*\* | 0.125\* | 0.227\*\*\* | 0.125\* | 0.227\*\*\* | 0.125\* |
|  | (3.643) | (1.817) | (3.642) | (1.815) | (3.640) | (1.820) |
| Dividend Payer | 0.654\*\*\* | 1.032\*\*\* | 0.654\*\*\* | 1.031\*\*\* | 0.652\*\*\* | 1.032\*\*\* |
|  | (2.713) | (3.596) | (2.713) | (3.591) | (2.707) | (3.595) |
| Forecast Dispersion | 0.515 | 1.046 | 0.517 | 1.047 | 0.524 | 1.073 |
|  | (0.588) | (0.538) | (0.590) | (0.538) | (0.597) | (0.552) |
| Long-Term Growth | 4.562\*\*\* | 3.210\*\* | 4.561\*\*\* | 3.212\*\* | 4.548\*\*\* | 3.201\*\* |
|  | (3.990) | (2.221) | (3.988) | (2.223) | (3.975) | (2.211) |
|  |  |  |  |  |  |  |
| Observations | 10,073 | 6,633 | 10,073 | 6,633 | 10,073 | 6,633 |
| Industry-Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted R-squared | 0.501 | 0.565 | 0.501 | 0.565 | 0.501 | 0.564 |

*Panel C: Effect of a change in linguistic complexity on a change in cost of equity conditional on a change in linguistic complexity increasing information asymmetry*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are changes in our proxies of 10-K readability. All the variables are first-differenced. We restrict our sample to firms that experience an increase in information asymmetry with respect to the previous fiscal year. We measure information asymmetry using the Amihud (2002)’s illiquidity construct. All regressions include industry-year and firm fixed effects. All the variables are defined in Appendix A. We do not report constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
|  | ∆ Information Asymmetry > 0 | | |
| ∆ICCA | (1) | (2) | (3) |
|  |  |  |  |
| **∆Fog Index** | **0.058\*\*** |  |  |
|  | **(2.01)** |  |  |
| **∆Grade Level** |  | **0.060\*\*** |  |
|  |  | **(1.979)** |  |
| **∆Reading Ease** |  |  | **-0.032\*\*** |
|  |  |  | **(-2.312)** |
| ∆Size | -2.638\*\*\* | -2.638\*\*\* | -2.634\*\*\* |
|  | (-7.915) | (-7.914) | (-7.907) |
| ∆Ln(Analyst Coverage) | -0.196\* | -0.196\* | -0.198\* |
|  | (-1.750) | (-1.753) | (-1.775) |
| ∆CAPM Beta | 0.187\*\* | 0.186\*\* | 0.188\*\* |
|  | (1.977) | (1.973) | (1.988) |
| ∆Institutional Ownership | 0.079 | 0.075 | 0.081 |
|  | (0.136) | (0.131) | (0.140) |
| ∆Book to Market | 0.691 | 0.690 | 0.695 |
|  | (1.445) | (1.444) | (1.455) |
| ∆Leverage | 0.113 | 0.112 | 0.111 |
|  | (1.523) | (1.517) | (1.508) |
| ∆Dividend Payer | 0.069 | 0.070 | 0.071 |
|  | (0.559) | (0.563) | (0.569) |
| ∆Forecast Dispersion | -1.552 | -1.553 | -1.541 |
|  | (-1.563) | (-1.564) | (-1.554) |
| ∆Long-Term Growth | 3.056\*\* | 3.054\*\* | 3.051\*\* |
|  | (2.012) | (2.012) | (2.013) |
|  |  |  |  |
| Observations | 4,157 | 4,157 | 4,157 |
| Year Fixed Effects | Yes | Yes | Yes |
| Adjusted R-squared | 0.118 | 0.118 | 0.119 |

*Panel D: Incremental complexity of the MD&A section and cost of equity capital*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are measures of MD&A readability or measures of MD&A readability orthogonal to the 10-K readability measures. All regressions include industry-year and firm fixed effects. All the variables are defined in Appendix A. We do not report constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Linguistic Complexity of the MD&A section of the 10-K annual report | | | Linguistic Complexity of the MD&A section orthogonal to the Linguistic Complexity of the 10-K annual report | | | |
| ICCA | (1) | | (2) | (3) | | (4) | (5) | (6) |
|  |  | |  |  | |  |  |  |
| **MD&A Fog Index** | **0.061\*\*\*** | |  |  | |  |  |  |
|  | **(2.616)** | |  |  | |  |  |  |
| **MD&A Grade Level** |  | | **0.053\*\*** |  | |  |  |  |
|  |  | | **(2.246)** |  | |  |  |  |
| **MD&A Reading Ease** |  | |  | **-0.018\*\*** | |  |  |  |
|  |  | |  | **(-2.338)** | |  |  |  |
| **MD&A Fog Index Orthogonal to 10-K Fog Index** |  | |  |  | | **0.053\*\*** |  |  |
|  |  | |  |  | | **(2.252)** |  |  |
| **MD&A Grade Level Orthogonal to 10-K Grade Level** |  | |  |  | |  | **0.045\*** |  |
|  |  | |  |  | |  | **(1.889)** |  |
| **MD&A Reading Ease Orthogonal to 10-K Reading Ease** |  | |  |  | |  |  | **-0.016\*** |
|  |  | |  |  | |  |  | **(-1.944)** |
| Size | -0.626\*\*\* | | -0.625\*\*\* | -0.625\*\*\* | | -0.626\*\*\* | -0.625\*\*\* | -0.625\*\*\* |
|  | (-4.778) | | (-4.768) | (-4.770) | | (-4.779) | (-4.770) | (-4.771) |
| Ln(Analyst Coverage) | -0.079 | | -0.078 | -0.078 | | -0.079 | -0.078 | -0.078 |
|  | (-0.851) | | (-0.846) | (-0.839) | | (-0.850) | (-0.844) | (-0.841) |
| CAPM Beta | 0.139\*\*\* | | 0.139\*\*\* | 0.139\*\*\* | | 0.139\*\*\* | 0.139\*\*\* | 0.138\*\*\* |
|  | (2.914) | | (2.910) | (2.906) | | (2.902) | (2.900) | (2.894) |
| Institutional Ownership | -0.573 | | -0.578 | -0.572 | | -0.574 | -0.578 | -0.573 |
|  | (-1.626) | | (-1.637) | (-1.622) | | (-1.627) | (-1.639) | (-1.625) |
| Book to Market | 1.472\*\*\* | | 1.474\*\*\* | 1.473\*\*\* | | 1.473\*\*\* | 1.476\*\*\* | 1.475\*\*\* |
|  | (5.711) | | (5.721) | (5.715) | | (5.717) | (5.726) | (5.725) |
| Leverage | 0.192\*\*\* | | 0.192\*\*\* | 0.193\*\*\* | | 0.193\*\*\* | 0.193\*\*\* | 0.194\*\*\* |
|  | (3.951) | | (3.962) | (3.973) | | (3.963) | (3.973) | (3.985) |
| Dividend Payer | 0.840\*\*\* | | 0.839\*\*\* | 0.838\*\*\* | | 0.839\*\*\* | 0.839\*\*\* | 0.837\*\*\* |
|  | (4.368) | | (4.363) | (4.359) | | (4.360) | (4.356) | (4.351) |
| Forecast Dispersion | 0.907 | | 0.906 | 0.906 | | 0.912 | 0.909 | 0.909 |
|  | (1.100) | | (1.099) | (1.099) | | (1.105) | (1.103) | (1.102) |
| Long-Term Growth | 3.845\*\*\* | | 3.844\*\*\* | 3.840\*\*\* | | 3.837\*\*\* | 3.837\*\*\* | 3.834\*\*\* |
|  | (4.375) | | (4.373) | (4.368) | | (4.365) | (4.364) | (4.361) |
|  |  | |  |  | |  |  |  |
| Observations | 13,301 | | 13,301 | 13,301 | | 13,301 | 13,301 | 13,301 |
| Industry-Year Fixed Effects | Yes | | Yes | Yes | | Yes | Yes | Yes |
| Firm Fixed Effects | Yes | | Yes | Yes | | Yes | Yes | Yes |
| Firm Cluster | Yes | | Yes | Yes | | Yes | Yes | Yes |
| Adjusted R-squared | 0.524 | | 0.523 | 0.523 | | 0.524 | 0.523 | 0.523 |

Table 5. Controlling for information quantity

*Panel A: Information readability and information quantity*

This panel presents the OLS regression results where the dependent variables are alternative measures of 10-K readability and the main dependent variable is the file size of the 10-K report. Regressions include industry-year and firm fixed effects. All the variables are defined in Appendix A. We do not report the constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
| 10-K Readability scores: | Fog Index | Grade Level | Reading Ease |
|  |  |  |  |
| **File Size** | **0.327\*\*\*** | **0.315\*\*\*** | **-0.710\*\*\*** |
|  | **(6.259)** | **(6.001)** | **(-4.618)** |
|  |  |  |  |
| Observations | 25,613 | 25,613 | 25,613 |
| Industry-Year Fixed Effects | Yes | Yes | Yes |
| Firm Fixed Effects | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes |
| Adjusted R-squared | 0.495 | 0.474 | 0.529 |

*Panel B: Linguistic complexity purged of the effect of information quantity and cost of equity capital*

This panel presents the results of OLS regressions where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are measures of 10-K readability orthogonal to the 10-K file size. We include industry-year and firm fixed effects in all the regressions. All the variables are defined in Appendix A. We do not report the constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| ICCA | (1) | (2) | (3) |
| **Fog Index Orthogonal to File Size** | **0.050\*\*** |  |  |
|  | **(2.255)** |  |  |
| **Grade Level Orthogonal to File Size** |  | **0.049\*\*** |  |
|  |  | **(2.110)** |  |
| **Reading Ease Orthogonal to File Size** |  |  | **-0.014\*** |
|  |  |  | **(-1.794)** |
| Size | -0.349\*\*\* | -0.349\*\*\* | -0.350\*\*\* |
|  | (-2.951) | (-2.951) | (-2.956) |
| Ln(Analyst Coverage) | -0.029 | -0.029 | -0.029 |
|  | (-0.380) | (-0.383) | (-0.391) |
| CAPM Beta | 0.083\*\* | 0.082\*\* | 0.082\*\* |
|  | (1.979) | (1.972) | (1.968) |
| Institutional Ownership | -1.241\*\*\* | -1.241\*\*\* | -1.238\*\*\* |
|  | (-4.219) | (-4.216) | (-4.209) |
| Book to Market | 1.688\*\*\* | 1.688\*\*\* | 1.686\*\*\* |
|  | (7.614) | (7.613) | (7.600) |
| Leverage | 0.194\*\*\* | 0.194\*\*\* | 0.194\*\*\* |
|  | (4.945) | (4.945) | (4.951) |
| Dividend Payer | 1.021\*\*\* | 1.021\*\*\* | 1.022\*\*\* |
|  | (5.880) | (5.876) | (5.876) |
| Forecast Dispersion | 0.287 | 0.288 | 0.292 |
|  | (0.398) | (0.399) | (0.405) |
| Long-term Growth | 4.369\*\*\* | 4.369\*\*\* | 4.363\*\*\* |
|  | (5.271) | (5.269) | (5.257) |
|  |  |  |  |
| Observations | 16,855 | 16,855 | 16,855 |
| Industry-Year Fixed Effects | Yes | Yes | Yes |
| Firm Fixed Effects | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes |
| Adjusted R-squared | 0.473 | 0.473 | 0.473 |

Appendix A. Variable definitions

|  |  |  |
| --- | --- | --- |
| Variable | Definition | Data Source |
| *Panel A: Readability Measures* | | |
| Fog | The fog index score computed based on the Gunning-Fox measure. | 10-K Annual Reports |
| Grade Level | The grade level score computed based on the Flesch-Kincaid measure. | 10-K Annual Reports |
| Reading Ease | The reading ease score computed based on the Flesch measure. | 10-K Annual Reports |
| *Panel B: Implied Cost of Equity Capital Measures* | | |
| ICCA | Implied Cost of Equity Capital measure based on analyst forecasts of future earnings per share. This is an average of the up to five individual methods we employ to estimate the cost of capital. | COMPUSTAT and I/B/E/S |
| ICCCSE | Implied Cost of Equity Capital measure based on cross-sectional earnings forecasts of future earnings per share (see Hou et al. (2012)).This is an average of the up to five individual methods we employ to estimate the cost of capital. | COMPUSTAT and I/B/E/S |
| *Panel C: Control Variables* | | |
| Size | The natural log of the market capitalization. | COMPUSTAT |
| Analyst Coverage | Number of analysts covering a firm. | I/B/E/S |
| Institutional Ownership | The percentage of shares owned by institutional investors. | 13-F Filings |
| CAPM Beta | Beta of the firm computed over the last five years using monthly returns. We use the value-weighted CRSP universe of stocks to define the benchmark market returns. | CRSP |
| Book to Market | Book-to-market ratio of equity. | COMPUSTAT |
| Forecast Dispersion | The standard deviation of the one-year-ahead EPS forecasts divided by the mean one-year-ahead EPS forecast. | I/B/E/S |
| Long-term Growth | The average of the I/B/E/S five-year earnings growth rate. | I/B/E/S |
| Leverage | The ratio of the total debt to total assets. | COMPUSTAT |
| Dividend Payer | A dummy variable that equals 1 if the firm pays dividends in that year and 0 otherwise. | COMPUSTAT |
| *Panel D: Other Variables* | | |
| Complexity - Business | The inverse of the concentration of sales based on the number of business units in the firm in a given year, where concentration is calculated using the Herfindahl-Hirshman Index. | COMPUSTAT |
| Complexity - Geographic | The inverse of the concentration of sales based on the geographic location of sales for the firm in a given year, where concentration is calculated using the Herfindahl-Hirshman Index. | COMPUSTAT |
| Earnings Surprises | EPS in excess of the last know I/B/E/S mean consensus EPS forecast. | COMPUSTAT and I/B/E/S |
|  |  |  |
| File Size | File size of the 10-K annual reports, computed following Loughran and McDonald (2014). We remove the pictures and the html coding, so the file size is a direct measure of the number of characters contained within the document. File Size is scaled by 1,000,000 to obtain megabytes. | 10-K Annual Reports |
|  |  |  |
| Illiquidity | We measure information asymmetry using Amihud (2002)’s illiquidity construct. Following Amihud (2002), illiquidity is defined as the daily return divided by the daily volume. We take the average value over the period starting the day of the publication of the annual report and ending 25 trading days subsequent to it (Bushee et al. 2018). We do not compute this measure for stock with a dollar value inferior to 5$. | CRSP |

Appendix B. Regressions of Table 3, Panel A, excluding utilities and financials

This table presents the results of OLS regression where the dependent variable is the implied cost of equity based on analyst earnings forecast (ICCA) and the main independent variables are alternative measures of 10-K readability. We include the control variables of the model (10). The regressions include industry and year fixed effects. We exclude financials (one-digit sic code = 6) and utilities (two-digit sic code =49). Variables are defined in Appendix A. We do not report the constant terms. Standard errors are robust to heteroscedasticity and clustered by firm. The t-statistics are reported in parentheses below each coefficient estimate. \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5% and 1% level, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| ICCA | (1) | (2) | (3) |
|  |  |  |  |
| **Fog Index** | **0.082\*\*\*** |  |  |
|  | **(4.487)** |  |  |
| **Grade Level** |  | **0.086\*\*\*** |  |
|  |  | **(4.512)** |  |
| **Reading Ease** |  |  | **-0.023\*\*\*** |
|  |  |  | **(-2.776)** |
| Size | -0.178\*\*\* | -0.180\*\*\* | -0.181\*\*\* |
|  | (-4.023) | (-4.070) | (-4.078) |
| Ln(Analyst Coverage) | -0.119 | -0.118 | -0.117 |
|  | (-1.609) | (-1.594) | (-1.575) |
| CAPM Beta | 0.129\*\*\* | 0.128\*\*\* | 0.129\*\*\* |
|  | (3.382) | (3.369) | (3.379) |
| Institutional Ownership | -0.515\*\* | -0.516\*\* | -0.493\*\* |
|  | (-2.170) | (-2.175) | (-2.075) |
| Book to Market | 2.402\*\*\* | 2.398\*\*\* | 2.404\*\*\* |
|  | (14.036) | (14.012) | (14.022) |
| Leverage | 0.367\*\*\* | 0.366\*\*\* | 0.369\*\*\* |
|  | (11.260) | (11.257) | (11.326) |
| Dividend Payer | 0.849\*\*\* | 0.849\*\*\* | 0.849\*\*\* |
|  | (7.155) | (7.157) | (7.136) |
| Forecast Dispersion | 1.692\*\* | 1.691\*\* | 1.717\*\* |
|  | (2.528) | (2.525) | (2.557) |
| Long-Term Growth | 3.213\*\*\* | 3.207\*\*\* | 3.196\*\*\* |
|  | (4.259) | (4.251) | (4.233) |
|  |  |  |  |
| Observations | 16,855 | 16,855 | 16,855 |
| Year Fixed Effects | Yes | Yes | Yes |
| Industry Fixed Effects | Yes | Yes | Yes |
| Firm Cluster | Yes | Yes | Yes |
| Adjusted R-squared | 0.190 | 0.190 | 0.189 |

1. We thank Thomas Bourveau, Olga Dodd, Robert Faff, Bart Frijns, Natalie Peng, Arthur Petit-Romec, Alireza Tourani-Rad, Jean-Philippe Weisskopf and participants at the FIRN UQ Women ECR Conference (2018), and AUT Research Seminar Series (2018) for their helpful comments. [↑](#footnote-ref-1)
2. Corresponding author: Auckland University of Technology, 55 Wellesley Street, Auckland 1010, New Zealand. Tel: +64 9 921 9999 ext 6076. E-mail: alexandre.garel@aut.ac.nz. [↑](#footnote-ref-2)
3. Auckland University of Technology, 55 Wellesley Street, Auckland 1010, New Zealand. Tel: +64 9 921 9999 ext 5713. E-mail: aaron.gilbert@aut.ac.nz. [↑](#footnote-ref-3)
4. Auckland University of Technology, 55 Wellesley Street, Auckland 1010, New Zealand. Tel: +64 9 921 9999 ext 5617. E-mail: ayesha.scott@aut.ac.nz. [↑](#footnote-ref-4)
5. See Leuz and Wysocki (2016) for a review of the empirical literature on the economic consequences of disclosure and financial reporting regulation. In particular, Chen et al. (2010) show that the Regulation Fair Disclosure, which changed how companies communicate with investors by increasing transparency and more frequent and timely communications, is associated with a decline in the cost of equity capital. In a similar vein, Li (2010b) and Daske et al. (2013) document that the adoption of IFRS is associated with a decline in the cost of capital. [↑](#footnote-ref-5)
6. The burgeoning literature on textual analysis is well summarized in Li (2010a) and Loughran and McDonald (2016). [↑](#footnote-ref-6)
7. Lo et al. (2017) show that, consistent with managerial obfuscation, firms that make disclosures more complex are more likely to have managed earnings. [↑](#footnote-ref-7)
8. “*…a section that is required by law but is also a medium in which managers have discretion over how to present an explanation of the company’s business, financial conditions, and results of operation. As opposed to conference calls and press releases, the structure and content of MD&As are fixed; consequently managers are “legally obliged to touch upon” (Warren Buffet) subjects they likely avoid in other disclosures*” in Lo et al. (2017). [↑](#footnote-ref-8)
9. For instance the adoption of SFAS 133 and SFAS 157. [↑](#footnote-ref-9)
10. Details on the cleaning process can be found at: https://sraf.nd.edu/data/stage-one-10-x-parse-data/ [↑](#footnote-ref-10)
11. Prior research shows that standard single-factor model, three-factor model and realized returns provide poor proxies for the cost of equity capital (Fama and French 1993, 1997; Elton 1999). It also show that the class of implied cost of capital models reasonably captures the time-variation in expected returns (Pástor et al. 2008). [↑](#footnote-ref-11)
12. In our key specifications, we use measures of ICC derived from analyst estimates because the inclusion of some of the key determinants of cost of equity identified by prior literature such as analyst coverage, long-term growth rate and forecast dispersion, require the firm to be covered by I/B/E/S analysts in the first place. As such, the coverage advantages that should come from the Hou et al. (2012)’s approach do not apply in our situation. [↑](#footnote-ref-12)
13. Due to differences in the available information required by the different models, whether firms pay dividends and whether iterative procedures can converge to a solution, in some instances, we cannot find an implied cost of equity for each model, and thus compute the average cost of equity based only on the models for which we can derive a cost of equity. When the implied cost of equity derived from one of the models we use (4 to 8) is negative or inappropriate (above 100%), we do not use this estimate in the computation of the average ICC. [↑](#footnote-ref-13)
14. Prior research shows that the effect of institutional ownership on ICC may depend on which type of institutional investor is prevalent in a firm’s ownership (e.g., Attig et al. 2008; Attig et al. 2013). [↑](#footnote-ref-14)
15. While prior literature excludes firms from the utilities and financial service industries because of their different operating and financial structure (e.g., Li 2008; Lo et al. 2017), the inclusion/exclusion of such firms does not qualitatively change our results (see Appendix B). We do include them in our main sample to maximize our sample size. Because our results are robust to the inclusion of industry, industry-year and firm fixed effects, it is highly unlikely that industry specificities drive our results. [↑](#footnote-ref-15)
16. The R-squared of the regressions including firm fixed effects is about twice the one of regressions not including them, which suggests that unobservable and observable persistent differences across firms explain an important share of the variation in ICC. To the extent that persistent differences across firms also explain differences in the linguistic complexity of the 10-K annual reports, not including them in our regressions would result in spurious coefficient estimates. In the rest of our analysis, we thus use regression specifications including industry-year and firm fixed effects. [↑](#footnote-ref-16)
17. Lo et al. (2017) show that, consistent with managerial obfuscation, firms that make disclosure more complex are more likely to have managed earnings. [↑](#footnote-ref-17)
18. Prior research indicates that meeting analyst forecasts is a more important benchmark than meeting prior year’s earnings or avoiding losses (Dechow et al. 2003), and that the market perceives missing analyst forecast as a more negative signal than a decrease in year-over-year earnings (Brown and Caylor 2005). [↑](#footnote-ref-18)
19. “*…a section that is required by law but is also a medium in which managers have discretion over how to present an explanation of the company’s business, financial conditions, and results of operation. As opposed to conference calls and press releases, the structure and content of MD&As are fixed; consequently managers are “legally obliged to touch upon” (Warren Buffet) subjects they likely avoid in other disclosures*” in Lo et al. (2017). [↑](#footnote-ref-19)