# The Plain English Mandate and Ex Ante Uncertainty:

# **Evidence from Initial Public Offerings**

Huu Phat Kevin Thai RMIT University kevin.thai@rmit.edu.au

Peter Clarkson UQ Business School, The University of Queensland <u>p.clarkson@business.uq.edu.au</u>

Kathleen Herbohn UQ Business School, The University of Queensland <u>k.herbohn@business.uq.edu.au</u>

June 2025

#### ABSTRACT

This study investigates the effect of Plain English Mandate (PEM) compliance on ex ante uncertainty about the valuation of initial public offerings (IPOs). Acknowledging that the PEM is a multifaceted notion, we conceptualise and operationalise its dimensions by employing a set of measures that are directly and faithfully linked to the spirit and language of the PEM. Using a sample of 1377 IPOs in the U.S. from 2002 to 2019, we find that complying with the PEM alleviates subsequent IPO return volatility. The benefit of adhering to plain English guidelines is evident in lengthy prospectuses and even short sub-sections such as the Use of Proceeds. Additional analyses reveal that the effect of complying with the PEM and its attributes on ex ante uncertainty varies across key sub-sections of the IPO initial prospectuses. Further, we conjecture and empirically document that there exists a turning point beyond which the benefits of increased compliance with the PEM diminish. Taken together, the findings demonstrate benefits of complying with the PEM, indicating that this reporting initiative is operationally sound.

Keywords: plain English, disclosure readability, initial public offerings, information processing cost.

#### JEL classification: M41

#### **1** Introduction

In October 1998, the U.S. Securities and Exchange Commission (SEC) issued the Plain English Mandate (PEM) with an aim of improving the clarity and utility of initial public offering (IPO) prospectuses. This initiative required prospective listed firms to comply with plain English principles by avoiding poor writing constructs (e.g., lengthy sentences, passive voice, legal jargon, abstract words, and numerous defined terms) and poor presentation characteristics (e.g., unreadable design and layout) in their prospectuses (SEC, 1997).<sup>1</sup> Lower frequencies or absence of such constructs are consistent with clear writing guidelines recommended by language experts (DuBay, 2004). This study investigates the effect of complying with the PEM in IPO prospectuses on ex ante uncertainty about IPO valuations.

The PEM is a multifaceted notion since it encompasses multiple factors that contribute to effective communication such as tone, sentence length, word complexity, and visual presentation. Building upon the extant IPO literature that highlights the effects of textual attributes on ex ante uncertainty, we contend that offering prospectuses failing to adhere to plain English principles hinder investors' ability to assimilate value relevant information. This in turn heightens the level of uncertainty about an IPO's prospects. An identical inference can be drawn from the lens of disclosure processing costs framework. That is, lower levels of compliance with the PEM can result in increased disclosure processing costs, especially in relation to integration costs (Blankespoor et al., 2020).

Although it is a prevailing assumption that greater compliance with plain English attributes is unambiguously beneficial (e.g., Li, 2008; Lehavy et al., 2011, Bonsall & Miller, 2017), a compelling argument suggests that an offering prospectus needs to be linguistically complex to an extent that it is deemed to be informative before it gets too difficult and costly to be processed. Echoing this standpoint, the SEC (1998a, p. 5) emphasised in its plain English handbook that plain English "does not mean deleting complex information to make the document easier to understand. For investors to make informed decisions, disclosure documents must impart complex information". In an experimental setting, Rennekamp (2012) documents supporting evidence that investors, particularly those who are the least

<sup>&</sup>lt;sup>1</sup> Plain English is described as an effective way of using language for communication purposes. Eagleson (2014) offers the following definition: "Plain English is clear, straightforward expression, using only as many words as are necessary. It is language that avoids obscurity, inflated vocabulary and convoluted sentence construction.... Writers of Plain English let their audience concentrate on the message instead of being distracted by complicated language." Given the term 'disclosure readability' is not clearly defined in most previous accounting and finance studies, it is important to note that this study specifically examines the compliance degree with the plain English guidelines outlined in the PEM, rather than addressing the broad notion of disclosure readability.

sophisticated, tend to overreact to disclosures that more closely follow the SEC's plain English guidelines. Combined, in this study, we also explore the possibility that the relation between the extent of adherence to the PEM and ex ante uncertainty is non-linear.

By considering the PEM as a *multifaceted* notion, we first conceptualise its various dimensions, namely conciseness, specificity, uncertainty, repetition, and bullets, and then evaluate the extent to which any of existing proxies fully captures these dimensions. An important inference from this review is that such proxies were not initially intended to directly measure the extent to which a business document complies with the SEC's guidelines on plain English writing and presentation, particularly in the context of initial offering prospectuses. To capture its multidimensionality, we then draw on prior literature to inform the measurement of the five dimensions and construct an aggregate measure of overall PEM compliance (*PEIndex*).

Using a sample of 1377 IPOs in the U.S. from 2002 to 2019, we document a negative association between *PEIndex* and return volatility within 60 days following the IPO date, indicating benefits of complying with the PEM in the context of IPO initial prospectuses. In terms of individual plain English feature, the proportion of uncertainty tone (specific details) is associated with increased (decreased) post-IPO return volatility. Contrary to what the SEC might expect, the finding shows that the use of bullet lists is associated with higher levels of ex ante uncertainty. Additional analyses show that the greater compliance to most plain English attributes and the overall PEM alleviates ex ante uncertainty, especially for longer initial prospectuses. We also document empirical evidence of a U-shaped relationship between overall PEM compliance and ex ante uncertainty, indicating the presence of a turning point beyond which adhering to the PEM is no longer beneficial. Nevertheless, it is revealed that the empirical switching point is near the extreme for the sample data (87<sup>th</sup> percentile), indicating that concerns around the unintended consequence of excessive PEM compliance are unnecessary.

The study then assesses the effect of adhering to plain English guidelines within four sections of the IPO initial prospectus on ex ante uncertainty. Prior research documents that different types of narratives included in prospectuses serve distinct communicative purposes and therefore are subject to varying levels of managerial discretion, regulatory influence, and auditor scrutiny (Li, 2010; Brown & Tucker, 2011). Following Hanley and Hoberg (2010), we focus on four key and distinct sub-sections within the initial prospectuses, namely the "Prospectus Summary" (PS), "Risk Factors" (RF), "Use of Proceeds" (UoP), and "Management's Discussion and Analysis" (MD&A). Our empirical findings reveal that the

impact of complying with specific plain English attributes on ex ante uncertainty differs significantly across sections, highlighting the varying relevance of such compliance in different sections of the offering prospectuses.

This study makes three contributions to the accounting literature. First, we provide empirical evidence on the economic benefits of complying with the PEM in the context of IPOs. Although a broad line of research documents favourable capital market consequences associated with enhanced disclosure (see Loughran & McDonald, 2016 for a review), these studies are typically one-dimensional, each focusing on a single textual attribute such as uncertainty tone (Loughran & McDonald, 2013; Ferris et al., 2013), specificity (Leone et al., 2007), and across-document textual similarity (Hanley & Hoberg, 2010). There has been minimal, if any, attention given to the plain English presentation features. Hence, we extend prior literature by acknowledging that the PEM is a multifaceted notion and employing a set of measures that are directly and faithfully linked to the spirit and language of the PEM. The dimensions represent both linguistics features (i.e., conciseness, uncertainty, specificity, and repetition) and presentation feature (i.e., bullet points). The findings collectively indicate that greater adherence to plain English guidelines aids investors in navigating information, especially within lengthy prospectuses.

Second, we extend a line of literature on disclosure readability by presenting novel empirical evidence indicating the presence of a turning point within our sample after which the benefits of complying with the PEM are less beneficial. The finding provides empirical support for experimental studies highlighting potential negative consequences from making disclosures too accessible to investors (Rennekamp, 2012; Asay et al., 2017; Asay et al., 2018). However, it is worth noting that the turning point in our data is situated very near the upper limit of the 17-year sample period. As such, concerns around the unintended consequence of excessive PEM compliance are unwarranted. Within the context of the Risk Factors section, we also demonstrate that the relationship between the proportion of uncertain words and subsequent return volatility exhibits an inverted U-shaped form, reconciling the two streams of literature on the positive and negative market reactions linked to the use of uncertain words (e.g., Arnold et al., 2010; Kravet & Muslu, 2013; Cazier et al., 2021).

Third, this study advances the understanding of the effects of various linguistic attributes in a context of a non-routine disclosure (i.e., IPO prospectuses) and its sub-sections. Since Li (2008), the literature has considered a broader set of textual attributes and measures such as similarity (i.e., stickiness), redundancy, repetition, salience, ambiguity, to name a few. Given most prior studies adopt a scattershot approach, usually focusing on a single textual feature without incorporating findings from earlier research, Dyer et al., (2017) questioned whether all of the constructs and their measures that have been identified are conceptually independent or incrementally important. Extending prior literature, our findings indicate varied informational roles of different plain English attributes in four key sub-sections of the initial offering prospectuses. In addition, exploratory factor analyses reveal that the proportion of uncertain and specific words as well as the number of bullet points capture the same dimension, highlighting the level of innate riskiness of an IPO. Corroborating Miller (2010) and the SEC's viewpoint, we document empirical evidence indicating that document length and PEM compliance are two distinct constructs.

The findings also have important implications for regulators, standard setters, and practitioners. For the last two decades, the SEC have expended efforts to implement the use of plain English principles on a broader scale with an aim to enhancing disclosure readability. For example, following a comprehensive review on current financial reporting requirements in 2013, the SEC proposed several amendments to eliminate excessive, unduly complex, and redundant disclosures (SEC, 2016). Collectively, several stakeholders including global and local regulators, standard-setters, and practitioners have turned their attention to disclosure effectiveness through enhanced disclosure readability in their enforcement practices. Given this stance, our study provides important evidence suggesting the SEC's PEM is operationally sound in alleviating ex ante uncertainty.

The remainder of this paper proceeds as follows. In section 2, we review the institutional background of the PEM and relevant literature, based on which two hypotheses are formulated. We conceptualise different dimensions of the PEM in section 3. Section 4 outlines steps in evaluating the existing proxies and then constructing the measures of the PEM compliance. Descriptive and multivariate results of the main model are displayed in Section 5. Section 6 examines the role of PEM compliance across four sub-sections. Section 7 presents a number of additional analyses. Section 8 concludes the study.

#### 2 Background and related research

#### 2.1 Institutional background

In an effort to address concerns that prospectuses were often impenetrable documents filled with legalese and jargons, the SEC issued the PEM, which required prospectuses used in registered public offerings of securities to be prepared in plain English (SEC, 1998a). As

depicted in Figure 1, the PEM includes a set of new plain English rules – Rule 421(d) and significant amendments to the existing Rule 421(b). Rule 421(d) requires firms to comply substantially with plain English principles (e.g., short sentences, concrete language, bullet lists) when preparing the front and back cover pages, the summary, and the risk factors section of the initial prospectuses. Rule 421(b) extends the implementation of Rule 421(d) to the remaining portions of the initial offering prospectuses with more specific guidance on how to make the entire prospectus clear, concise, and understandable. These additional requirements include avoiding boilerplate and repetitive explanations as well as using of legal highly technical business terminology. With an aim to assisting public companies to better prepare offering prospectuses in plain English, the SEC has continually devoted significant resources to assist and oversee the implementation of the PEM. These initiatives included the issuance of the Plain English handbook (SEC, 1998a), staff legal bulletins (SEC, 1998b, 1999) as well as staff training in filing review and comment letter issuance. Despite the PEM being mandatory only for IPO prospectuses, it is the intent of the SEC to encourage conformance with the PEM in all corporate filings.<sup>2</sup>

#### [INSERT FIGURE 1]

#### 2.2 Hypothesis development

A growing body of research has been examining the economic effects of financial disclosure readability, particularly focusing on financial reports. It is evident that investors are prone to limited attention, which provides explanations for many market anomalies (e.g., post-earnings announcement drift and accruals mispricing) (for a review, see Lu, 2022). When investors face disclosure processing costs, learning from a disclosure is an active economic choice similar to acquiring private information, and investors expect to be compensated for their costly processing activities (Grossman & Stiglitz, 1980). Supporting this view, the extant literature documents negative market outcomes associated with lengthy and boilerplate disclosures as well as positive market outcomes associated with improved disclosure consistency, specificity, and comparability (for reviews, see Libby & Emett, 2014 and Blankespoor et al., 2020).

<sup>&</sup>lt;sup>2</sup> The SEC states: "Our ultimate goal is to have all disclosure documents in plain English, and we have undertaken several initiatives to improve readability of these documents" (SEC, 1997, p. 14), and later in the document: "We also encourage you to use these techniques for drafting your other documents" (SEC, 1997, p. 39). The use of plain English became compulsory for disclosures relating to corporate governance, related-party transactions, executive and director compensation, and beneficial ownership (SEC, 2006) and documents from all federal agencies (PLAIN, n.d.).

In the context of offering prospectuses, an emerging line of research suggests that textual attributes of IPO prospectuses are useful in narrowing the estimate of the dispersion of the stocks in the secondary market, suggesting a negative association between textual features and ex ante uncertainty.<sup>3</sup> For instance, Leone et al. (2007) shows that IPO firms providing more specific information about the use of proceeds (measured by the dollar detail of an IPO issuer's intended use of proceeds) are associated with lower underpricing. Hanley and Hoberg (2010) find that offers with greater textual similarity (defined as the exposure to information in an IPO prospectus that is already contained in both recent and past industry IPOs) experience higher underpricing because there is limited firm-specific new information contained in the S-1 filings. In a similar vein, Loughran and McDonald (2013) document that shares of IPOs with a higher proportion of uncertain words (i.e., uncertainty tone) in their S-1 filings are associated with higher return volatility subsequent to the IPO date. Taken together, the findings support an inference that impairments of certain textual attributes are related to unfavourable capital market outcomes.

#### 2.2.1 The PEM compliance and ex ante uncertainty

We surmise that the PEM is multifaceted because it encompasses multiple factors that contribute to effective communication such as tone, sentence length, word complexity, and visual presentation. Building upon the extant IPO literature documenting the effects of textual attributes on ex ante uncertainty, we argue that offering prospectuses that violate several plain English principles (e.g., long sentences, complex words, vague and repetitive disclosures, poor layout) impede investors' ability to effectively assimilate value-relevant information. An identical inference can be drawn from the lens of disclosure processing costs framework. That is, lower levels of compliance with the PEM can result in increased disclosure processing costs, especially in relation to integration costs (Blankespoor et al., 2020). The release of a large volume of new information in a short time in the initial offering prospectuses exacerbates the issue of limited information processing capability, even for institutional investors (Cheynel & Levine, 2020).

Similarly, it is the SEC's longstanding belief that offering prospectuses with greater compliance with the PEM are more accessible and makes investors' valuation less challenging because "investors will be more likely to understand what they are buying" (SEC, 1998a, p. 3). As such, the SEC anticipated that greater conformance to the PEM allows "investors to make

<sup>&</sup>lt;sup>3</sup> Following prior literature, the notion of ex ante uncertainty in this study refers to the range of support for the value of an IPO (Beatty & Ritter, 1986; Leone et al., 2007; Lowry, 2010; Loughran & McDonald, 2013)

better-informed assessments of the risk and rewards of investment opportunities" and "reduce the likelihood that investors make investment mistakes because of incomprehensible disclosure documents".<sup>4</sup> This ultimately reduces the level of ex ante uncertainty about an IPO's prospects. Combined, this first hypothesis investigates whether the extent to which offering prospectuses comply with the PEM is *inversely* related to the level of ex ante uncertainty in the IPO pricing process. Formally:

# H1: the degree to which an IPO complies with the PEM is negatively associated with its level of ex ante uncertainty, ceteris paribus.

#### 2.2.2 Non-linear relation between the PEM compliance and ex ante uncertainty

As anticipated by the SEC and examined in the first hypothesis, it is a prevailing assumption that greater compliance with plain English attributes is unambiguously beneficial. Nevertheless, there is evidence indicating a possibility that the relation between PEM compliance and ex ante uncertainty is non-linear. Given financial disclosures primarily attempt to describe the economic reality of the business and meet regulatory reporting requirements (Bloomfield, 2008; Leuz & Wysocki, 2016; Loughran & McDonald, 2016), an alternative view is that complex language could be necessary to reflect the complex technical information. Echoing this standpoint, the SEC (1998a, p. 5) emphasised in its plain English handbook that plain English "does not mean deleting complex information to make the document easier to understand. For investors to make informed decisions, disclosure documents must impart complex information component is associated with reduced information asymmetry in the context of quarterly earnings conference calls. Combined, it is reasonable to expect that an offering prospectus needs to be linguistically complex to an extent that it is deemed to be informative before it gets too difficult and costly to be processed.

Contemporary experimental studies offer additional evidence that challenges the unequivocal benefits arising from greater plain English compliance. For instance, Rennekamp (2012) shows that investors, particularly those who are the least sophisticated, tend to overreact to disclosures that more closely follow the SEC's plain English guidelines. Extending this line of literature, Asay et al. (2017) document an unintended consequence of enhanced disclosure readability. That is, when firm-initiated disclosures are readable, investors are inclined to discount outside sources of information about firm, leading to an over-reliance on such

<sup>&</sup>lt;sup>4</sup> These benefits are from page 6375 of the final rules on the Plain English Disclosure released by the SEC on February 6, 1998, and available at https://www.govinfo.gov/content/pkg/FR-1998-02-06/html/98-2889.htm

information. Corroborating this finding, Asay et al. (2018) demonstrate that managers strategically enhance the readability of good news disclosure when they have a strong reporting goal to make the firm appear as favourable as possible.

Taken together, considering the SEC's clarification on the potential trade-off between PEM compliance and the indispensable role of complex information, along with contemporary experimental evidence, the benefits of increased adherence to the PEM appear to be less clearcut. As such, we postulate that there is a turning point after which benefits associated with greater compliance with the PEM are diminished. The second hypothesis is formally expressed as follows:

H2: The association between the PEM compliance and the level of ex ante uncertainty about IPO values exhibits a U-shaped form, implying a negative marginal association at low levels of PEM compliance and a positive marginal association at higher levels of PEM compliance.

#### 3 Methodology

#### 3.1 Dimensions of the PEM

Based on the premise that the PEM is multifaceted, we first conceptualise its various aspects by mapping the PEM requirements to five dimensions, namely conciseness, specificity, uncertainty, repetition, and bullets.<sup>5</sup> Table 1 outlines the mapping of plain English writing and presentation conventions to the five plain English dimensions.

#### [INSERT TABLE 1]

First, the SEC's plain English rules explicitly outline several writing conventions that supposedly enhance the clarity and understandability of disclosure documents (SEC, 1998a). The examples provided in Appendix B highlight that adhering to such recommendations can be linked to the use of shorter sentences, with most words being familiar to average investors. This in turn enhances the clarity of IPO prospectuses. Therefore, these writing conventions are grouped to the construct of *Conciseness*.<sup>6</sup> Based on the inference that concise disclosure

<sup>&</sup>lt;sup>5</sup> To ensure the accuracy regarding the interpretation and application of the plain English principles, we refer to the (1) Plain English handbook – the SEC's application guidance of the PEM (SEC, 1998a), (2) staff legal bulletins No.7 (SEC, 1998b) and No. 7A (SEC, 1999) for additional guidance and clarification. Appendix A presents evidence from the SEC's comment letters demonstrating its oversight and enforcement of issuer compliance with the plain English dimensions in prospectus drafting.

<sup>&</sup>lt;sup>6</sup> The English dictionary defines "conciseness" the extent to which a piece of writing communicates clear information in as few words as possible.

facilitates information processing, we predict a negative association between *Conciseness* and ex ante uncertainty.

Second, managers are recommended to avoid using vague explanations that are imprecise and readily subject to different interpretations when drafting the offering prospectus (SEC, 1997). In addition to the choice of words, as suggested in the plain English handbook, the quantification and/or inclusion of numbers with proper nouns can increase concreteness of a disclosure (e.g., "one share of IBM stock" is more concrete than "an asset") (SEC, 1998a, p. 24). Drawn from a review of prior literature, the most closely related textual attribute to this plain English requirement is the notion of disclosure specificity in Hope et al. (2016). This study defines a prospectus as more specific when there is a higher level of detail (e.g., information about locations of people or objects, or information about when the event happened or explicitly describes a sequence of events) provided in the offering prospectuses. In other words, the notion of specificity relates to the provision of concrete detail as opposed to vague and generic statement. The primary premise is that greater specificity (i.e., enhanced concreteness of the disclosed information) leads to higher level of information precision, which facilitates the process of information and lowers the associated processing costs. To this extent, we predict a negative association between *Specificity* and ex ante uncertainty.

Third, Rule 421(d) mandates IPO issuers to use *definite* language when drafting the IPO prospectuses (SEC, 1997). However, the SEC provides no further guidance regarding the definition and examples of definite words in any supplemental application materials. This study relates the lack of definite language to the notion of uncertainty that is drawn upon Loughran and McDonald (2011, 2013). Given *Uncertainty* is referred as a general notion of imprecision or definitiveness in characterising the company's narrative reporting, the lack of definite language points to the inability to assess the uncertainty relating to the probabilities of occurrence of a piece of disclosed information, which generates more uncertainties in future cash flow projections. On this premise, we predict a positive association between *Uncertainty* and ex ante uncertainty.

The fourth dimension of the PEM is disclosure repetition because Rule 421(b) notes that IPO issuers should avoid repeating disclosures in different sections of the document that increases the size of the document but does not enhance the quality of the information (SEC, 1997). Predicting the effect of disclosure repetition within IPO prospectuses on ex ante uncertainty is complex. On the one hand, the SEC explicitly advises issuers to avoid repeating disclosures across different parts of the documents in offering prospectuses, inferring that

repetition in the context of financial reporting is an undesirable textual feature (SEC, 1998a). On the other hand, from the lens of disclosure processing costs, repeating information helps to lower disclosure processing costs by reducing awareness costs (costs incurred to improve the probability of knowing a particular disclosure exists) (Blankespoor et al., 2020). The succession hypothesis from the communication literature and psychology literature also suggests that information repeated in multiple channels can be informative to investors, because it increases the possibility of that information to be identified and processed (Cacioppo & Petty, 1989; Stephens, 2007). Supporting this inference, Li (2019) shows that managers strategically use repetitive disclosures (measured by the extent to which MD&A content overlaps with audited financial statement notes) to enhance the salience of firm-specific events. Taken together, we argue that repeated disclosures are beneficial, particularly in the context of initial offering prospectuses where substantial new information of a company is revealed to the public all at once, as they enhance the likelihood that key information is identified and processed. Thus, we predict a negative association between *Repetition* and ex ante uncertainty.

The fifth dimension of the PEM is related to presentation requirements. In addition to writing guidelines, the SEC explicitly encourages the use of bullet points to list information whenever possible, which makes information easier to be absorbed in one quick glance (SEC, 1997).<sup>7</sup> From the lens of information processing costs, greater use of bullet points may facilitate more efficient information processing by investors, implying a negative association between bullet point usage and ex ante uncertainty. Supporting this inference, Miller (2010) constructs a proxy for report length by summing the number of words and table cells and finds a negative association between this measure and overall trading around the filing of the 10-Ks. In an experimental setting, Tan et al. (2015) show that firms with inconsistent benchmark performance strategically present positive performance in a more readable manner than negative performance. This improvement in readability is partly driven by the use of bullet points. However, this prediction should be interpreted with caution. Indeed, the market impact relating to the use of bullet points is contingent on other information characteristics, such as

<sup>&</sup>lt;sup>7</sup> We do not evaluate other presentation guidelines, such as tables and headings, due to significant measurement challenges. Following Miller (2010), we attempt to quantify the number of tables in S-1 filings using HTML tags. However, the reliability of these tags is limited. For example, we find that the tag '<TABLE' functions reasonably well in filings with minimal HTML formatting (pre-2005), while other tags often capture individual table cells or rows. A manual review of 100 randomly selected prospectuses further reveals substantial variation in table size and complexity. For instance, a single table may present either a single balance sheet item or a multi-year condensed income statement with numerous rows and columns. Overall, given the inconsistent tagging and flexible table construction across filings, it appears impractical to accurately identify the number of tables. Likewise, without a consensus in prior literature, evaluating the extent to which a heading or sub-heading is descriptive remains largely subjective.

the content (i.e., *what* is disclosed) and the density (i.e., *how much* information is disclosed) within each bullet point. For example, Beatty and Ritter (1986) find that IPOs providing more detailed intended uses of proceeds through a greater number of bullet points are associated with higher levels of underpricing, indicating higher level of ex ante uncertainty. Anecdotal evidence also suggests a common reporting practice in which companies use bullet points to enumerate various risk factors. Taken together, we posit that the use of bullet points facilitates investors' assessment and evaluation of the offering firm's riskiness level, thus indicating a positive association between *Bullets* and ex ante uncertainty.

#### 3.2 Research design

To test H<sub>1</sub>, a baseline model is used as follows:

$$ExanteUncertainty_i = \alpha + PEM \ compliance_i + Control \ Variables_i + \varepsilon; \tag{1}$$

*ExanteUncertainty* is the post-IPO total return volatility (Lowry et al., 2010; Loughran & McDonald, 2013). Total volatility is the market model root-mean square error for each IPO during days +5 to days +64 (a 60-day period) relative to the IPO day. The exclusion of the first four trading days after the IPO is to mitigate the effects of unusual trading activity (i.e., abnormal flipping shares) immediately following the offering. *PEM compliance* measures the extent to which an offering prospectus complies with the overall PEM and each of the five individual plain English dimensions.

To probe the non-linearity relation as proposed in H<sub>2</sub>, the baseline model also includes the squared values of PEM compliance, as expressed below:

#### *ExanteUncertainty*<sub>i</sub> = $\alpha$ + *PEM compliance*<sub>i</sub> + *PEM compliance*<sup>2</sup><sub>i</sub> + *Control Variables*<sub>i</sub> + $\varepsilon$ ;(2)

There are several control variables identified in prior research as determinants of ex ante uncertainty. Specifically, they include document length, firm age, length of registration days, share overhang, venture backing, pre-IPO market returns, audit quality, and high technology companies.<sup>8</sup>

<sup>&</sup>lt;sup>8</sup> Following Barth et al. (2017), both *HighTech* and industry effects are included because *HighTech* is not a proper subset of the Fama and French (1997) 12-industry classification. We collect two additional dummy variables, each equal to one if an IPO is classified as "Emerging Growth Companies" (ECG) under the JOBS Act or an equity spin-off (carve-out). There are 295 ECG firms and 216 spin-off firms in the sample. The main results remain qualitatively unchanged when these variables are included.

#### 4 **PEM measurement**

#### 4.1 Existing measures of disclosure readability

To assess the degree of compliance with the PEM, we evaluate the extent to which existing proxies capture the specific criteria outlined by the SEC in Sections 421(b) and 421(d). These include the Fog Index (Gunning, 1952); total word count (Li, 2008; You & Zhang, 2009; Lawrence, 2013); file size (Loughran & McDonald, 2014a); a composite measure of some plain English attributes (Loughran & McDonald, 2014b), and the Bog Index (Bonsall et al., 2017).

One of the most commonly used measures of readability in the extant accounting and finance literature is Robert Gunning's (1952) Fog Index. In essence, the Fog Index (and its variants) focuses on sentence length and polysyllabic words, estimating the number of years of formal education that is required to understand a piece of writing. Despite its simplicity and high adaptability, Loughran and McDonald (2014a) demonstrate that the calculation of average length sentence is prone to measurement errors in classifying financial reporting language. In addition, many prior studies in accounting and finance assess the degree of readability based on the quantity of textual disclosures, such as the number of words contained in a corporate filing (Li, 2008; You & Zhang, 2009; Lawrence, 2013) and net file size of 10-K filings (Loughran and McDonald, 2014a). Although these quantity-based measures address potential issues with parsing algorithms and thus, facilitate empirical replication, Bonsall et al. (2017) empirically show the variations in file size of financial disclosure across time can be driven by the non-textual components (e.g., HTML, XML, pdf, and jpeg file attachments), rather than the underlying text.

Loughran and McDonald (2014b) develop a plain English index (hereinafter, LM PE Index), which is an aggregated measure designed to capture some writing principles anchored in specific examples provided by the SEC's plain English handbook. While the LM PE Index represents an initial attempt in capturing some attributes of plain English writing recommended by the PEM, it has two main shortcomings. First, it measures word complexity by using the average word length, which is similar to the syllable count – the component that is directly criticised by Loughran and McDonald (2014a). Second, LM PE Index relies on a limited set of words/phrases to capture the use of technical jargon, legal jargon, and superfluous words (Bonsall et al., 2017).

Bonsall et al. (2017) proposes and validates Bog Index as a novel measure of disclosure readability. This is a multifaceted measure based on a linguistic software program, *StyleWriter* 

- *The Plain English Editor*. The Bog Index summarises writing attributes that tend to bog readers down (e.g., sentence length, word complexity, passive voice, weak verbs, overused verbs, financial jargon, and legalese). It can be computed as the sum of three components of writing clarity:

Bog Index = Sentence Bog + Word Bog - Pep, where the higher the Bog Index, the less readable the document is.

Bog Index overcomes several shortcomings inherent in other measures of readability in prior literature by employing proprietary wordlists to assess word difficulty and other plain English writing problems (Bonsall et al., 2017). Although the Bog Index is often viewed as a superior measure for disclosure readability, it is not designed to serve as a direct or comprehensive measure of PEM compliance. Notably, it overlooks key plain English dimensions outlined in section 3.1, including the use of concrete and definite words, the avoidance of repetitive disclosures and boilerplate explanations, and the use of bullet points.

#### 4.2 PEM measures

Given that existing measures of financial disclosure readability are, at best, only loosely aligned with the PEM's spirit and thus likely inadequate in capturing its five core dimensions, we draw on prior literature to inform the measurement of these dimensions and assess overall compliance with the PEM.

To measure disclosure conciseness (*Conciseness*), we begin by mapping the writing conventions associated with this construct (as identified in Appendix B) to compatible outputs that *StyleWriter* can reliably extract in the context of business writing. These problems include passive voice, hidden verbs, legalese (e.g., forthwith, in said agreement), wordy phrases (e.g., at a later date, due to the fact that), and personal pronouns. All but personal pronouns are generated from *StyleWriter*.<sup>9</sup> The SEC's plain English handbook specifically targets first-person plural and second-person singular personal pronouns to avoid the "he or she" dilemma (SEC, 1998b, p. 22). Thus, similar to Loughran and McDonald (2014b), we tabulate the frequency of the following personal pronouns "we", "us", "our", "ours", "you", "your",

<sup>&</sup>lt;sup>9</sup> *StyleWriter* appears to be the most advanced and appropriate computational tool for identifying and quantifying the frequency of Plain English writing issues, such as hidden verbs, passive voice, and superfluous words. We run a sample of randomly selected prospectuses to evaluate the software's technical precision, with a focus on minimising word misclassification in the context of corporate reporting. The analysis reveals that certain plain English style issues flagged by *StyleWriter* have limited or no relevance in this specific context. Examples include business cliches (e.g., please be advised, under date of), foreign words (e.g., inter alia, ipso facto), overwriting (e.g., completely, very, particularly) and jargon phrases (e.g., inventory management system, business management system). Therefore, they are excluded from the measure of *Conciseness*.

"yours". The search for those terms is case-insensitive. Since the expected proportion of each component may substantially vary, each of the five components are normalised (a mean of zero and a standard deviation of one). Given the SEC's explicit recommendation is to minimise the use of passive voice, hidden verbs, legalese, and wordy phrases and encourage the use of personal pronouns, the normalised values of all components except that of personal pronouns are negatively signed in the summation. For the ease of interpretation, we use the inverse of this summed amount that corresponds to a variable, *Conciseness*, where higher values indicate better compliance with the PEM (i.e., lower proportion of plain English style problems relating to linguistic conciseness).

There are three writing conventions that are identified in the plain English handbook but not explicitly accounted for in the measure of *Conciseness*, namely avoiding long sentences, multiple negatives, and unnecessary defined terms. First, measuring sentence length and multiple negatives is prone to significant measurement errors. Loughran and McDonald (2014a) show evidence highlighting that sentence ambiguation is particularly challenging in the presence of extensive lists, technical terminology, and other formatting complexities. We posit that adherence to other plain English principles that are already captured in the *Conciseness* measure (i.e., reducing legal jargon, avoiding hidden verbs, and eliminating unnecessary words) naturally contributes to keeping sentence lengths within a reasonable range. Likewise, Loughran and McDonald (2016, p. 1217) caution that identifying negation surrounding positive/negative words using automated linguistics techniques is "far more complex than what will be identified by looking for words like [no] and [not] preceding the target word" (e.g., not [...] until, not [...] except, not [...] unless). Many abbreviations and defined terms flagged by *StyleWriter* (e.g., SEC, EU, FDA, FASB) are likely well understood by the investing public.

To measure disclosure specificity (*Specificity*), following Hope et al. (2016), we count the number of specific words or phrases conveying specific information relevant to the disclosure firm, divided by the number of total words in an offer prospectus. The level of disclosure specificity includes specific words from the following categories: (1) names of persons, (2) names of locations, (3) names of organisations, (4) quantitative values in percentages, (5) money values in dollars, (6) times, and (7) dates. The Named Entity Recognition (NER) technique is used to identify and extract specific names from the above seven entity categories.

To measure *Uncertainty*, we use an aggregate wordlist, including 2612 words, based on three dictionaries of uncertain words, weak modal words, and negative words in Loughran and

McDonald (2013). They show that uncertain, weak modal, and negative wordlists are one dimensional in nature, providing a useful proxy for uncertainty in the context of the IPO offerings. In addition, these wordlists developed by Loughran and McDonald (2011) are widely used to measure textual sentiments in accounting and finance literature (see Loughran & McDonald [2016] for a review). For each offering prospectus, we count the number of uncertain words based on the aggregate wordlist, scaled by total word count of the document.

To measure the level of disclosure repetition, following Merkley (2014), we compute *within-document* cosine similarity for each prospectus. We first disaggregate S-1 filings into sentences and remove all articles and common prepositions from each sentence. Each word is then stemmed to its root to remove word tense and form. Next, we convert each sentence into a vector in an *n*-dimensional Euclidean space, where *n* is the number of unique words in a firm's prospectus. The value of each vector element is the frequency of a word in that document. We then compare the vectors based on their cosine similarity (bounded between 0 and 1) to determine how similar a sentence is relative to other sentences in the same document. Given the objective of this study is to measure the level of similarity within a prospectus, following Merkley (2014), we identify a sentence as repetitive if it is 90 percent, or more, similar to a previous sentence. The variable, *Repetition*, represents the total number of repetitive sentences in an offering prospectus.

To measure the use of bullet points, we employ a mixed approach that combines automated searches of commonly used HTML code for bullet point construction with extensive manual data collection.<sup>10</sup> Since nearly all firms use a consistent bullet point style throughout their prospectuses, the variable *Bullets* is the count of the most frequently used bullet point.

In addition to examining five individual plain English attributes in separate models, for the sake of parsimony, we develop an aggregate measure of plain English compliance, *PEIndex*. Because some of the variables are measured on different scales and their expected proportions vary, we normalise each of the five plain English components (mean zero, standard deviation of one) and sum these normalised values. Consistent with the directional predictions outlined in section 3.1 regarding the impact of individual plain English attributes on ex ante uncertainty, *Conciseness, Specificity*, and *Repetition* are positively summed while *Uncertainty* and *Bullets* 

<sup>&</sup>lt;sup>10</sup> We find that automated search method is not applicable for filings before 2002, the only way to count the number of bullet symbols is to scroll through such filings manually. Coupled with 1180 IPOs filed during the Internet bubble period (from 1997 to 2000), manual data collection of the *Bullets* variable before 2002 is extremely labour-intensive.

are negatively summed.<sup>11</sup> Higher values of *PEIndex* indicate higher compliance levels with the PEM, and thus are predicted to be associated with reduced ex ante uncertainty.

#### 4.3 Sample selection

The sample period is from 2002 to 2019 because the HTML tags used to locate the bullet points are only applicable for prospectuses with .html extension from 2002. Table 2 outlines details relating to the sample selection procedures. An initial list of firm-committed completed IPOs and relevant data are obtained from the SDC Platinum New Issues database. Following prior literature, this sample excludes American Depository Receipts, unit issues, Real Estate Investments Trusts (REITs), closed-end funds, financial firms, reverse leveraged buyouts, simultaneous international offerings, and firms with offer prices less than \$5 (Loughran & McDonald, 2013; Lowry et al., 2017). For each IPO, we use web-crawling algorithms in Python to identify its initial offering prospectuses based on the CIK and download from the EDGAR website. For each document, we follow the parsing procedures described in Loughran and McDonald (2011) and Bonsall et al. (2017). To ensure consistency in word counts, the LoughranMcDonald Master Dictionary (LM master dictionary), which was initially developed and validated in Loughran and McDonald (2011) is used. Financial variables are collected from CRSP. Data corrections regarding firm age are generously provided by Jay Ritter on his website. The final sample includes 1377 completed U.S. IPOs. All variables are defined in Appendix C.

#### [INSERT TABLE 2]

#### 5 Results

#### 5.1 Descriptive statistics

Table 3 provides summary statistics of the main variables. Given *Conciseness* is the inverse of the sum of normalised values of five writing problems, its summary statistics are somewhat more difficult to interpret. Of components of *Conciseness*, on average, an IPO prospectus contains 1730 wordy phrases (e.g., in relation to, with reference to, at a later date) and perhaps more surprisingly, only 102 words classified as legalese based on the proprietary wordlist of StyleWriter. In line with previous studies, the average proportion of uncertain words of a S-1 filing is 3.17% (Loughran & McDonald, 2013; Lowry et al., 2020). The mean for *Specificity* is

<sup>&</sup>lt;sup>11</sup> As a robustness test, we scale the number of bullet points (repetitive sentences) by the number of words (sentences) contained in the S-1 filings, which yields two new variables - *%Bullets* and *%Repetition*. The main results remain qualitatively unchanged.

3.73%, suggesting that, firms, on average, contain just less than four words describing specific entities per every 100 words in an initial offering prospectus. This finding indicates that the proportion of specific details in a S-1 filing is far less than the reported percentage of 19% in a 10-K filing in Hope et al. (2016). A novel insight revealed in Table 3 is that there are, on average, twenty-six clusters that include at least a pair of substantially repetitive sentences (cosine similarity being 0.9 or greater) and approximately 287 bullet points in an offering prospectus. Of control variables, the mean length of an offering prospectus is roughly 72,000 words with the longest one containing more than 253,000 words, which is consistent with Loughran and McDonald (2013). Nearly two-thirds of the IPO sample (61%) reports net operating losses in the year before going public. In addition, a majority of the offering prospectuses are audited by Big 4 firms while three-fifths of the IPO are traded on the Nasdaq stock exchange.

#### [INSERT TABLE 3]

Table 4 reports the correlations between variables. It is shown that all individual plain English attributes (except for *Conciseness*) and the *PEIndex* are highly linked to post-IPO return volatility. The univariate analysis provides initial supporting evidence for the first hypothesis. While *Bullets* are positively correlated with both *PEIndex* and *TotalVol*, there is a negative correlation between the overall compliance level of plain English rules and subsequent IPO volatility. In addition, high correlation coefficients among *Uncertainty*, *Specificity*, and *Bullets* point to a possibility that some individual plain English attributes capture the same dimensionality in nature, which is further examined in section 7.3.

#### [INSERT TABLE 4]

Untabulated analyses provide additional insights on variations of plain English attributes across IPO years and industries. Based on the Fama and French (1997) 12 industry classification, top three industries with the highest values of *PEIndex* are Chemicals and Allied Products, Consumer Durables, and Manufacturing, partly due to the high (low) score of *Uncertainty* (*Specificity*). Although offering prospectuses of the Healthcare and Business Equipment industries are more concise compared to those of peer industries, their low scores of *PEIndex* are due to high proportions of uncertain words, and low levels of specificity and repetition. It is important to note that the average level of plain English compliance for a year is dictated by a particular industry that dominates the total number of IPOs in that year. For instance, for the year periods 2014-15 and 2017-19, in which IPOs in the Healthcare industry significantly outnumbered other industries, there was a steep surge (decline) in terms of the

normalised values of *Uncertainty (Specificity*). As a result, *PEIndex* of these years dropped. The opposite prediction on *PEIndex* holds true for the years 2007-08, when there were far less IPOs from the Healthcare industry.

#### 5.2 Multivariate regressions

Table 5 reports the regression results from estimating Equation (1).<sup>12</sup> As revealed in column (1), the aggregate measure of plain English compliance, *PEIndex*, has a negative and statistically significant coefficient (t-statistics = -3.27), suggesting that IPOs with higher level of plain English compliance in their S-1 filings experience lower subsequent return volatility. Supporting the first hypothesis, our results indicate that the greater compliance with the plain English guidelines facilitates the processing of information of investors, particularly when they are required to decipher a large volume of new information disclosed in the initial offering prospectuses. As such, the findings support the benefits of adhering to the PEM as argued by the SEC (1998a) and corroborates a broader line of prior literature documenting favourable capital market consequences associated with improved disclosure readability (e.g., Li, 2008; Lehavy et al., 2011, Bonsall & Miller, 2017; Chen et al., 2023).

#### [INSERT TABLE 5]

As shown in columns (2) – (6), all five plain English attributes have the predicted signs, with three coefficients, namely *Uncertainty*, *Specificity*, and *Bullets*, being significant at varied level of statistical significance. In column (3), the slope coefficient on *Uncertainty* is significantly positive (0.417, t-statistic = 3.32), which is consistent with Loughran and McDonald (2013) and suggests that the frequency of uncertain words as an important factor explaining the post-IPO total return volatility. Column (4) reports a negative and statistically significant association between *Specificity* and subsequent IPO return volatility. Extending the finding of Hope et al. (2016) on the economic benefits of linguistic specificity within the RF section of annual reports, this result offers evidence that such benefits also extend to offering prospectuses.<sup>13</sup>

<sup>&</sup>lt;sup>12</sup> To provide additional insights on the determinants of PEM compliance, we rerun Equation (1) with the dependent variable being *PEIndex* and the control variables as independent variables, with industry and year fixed effects. Untabulated results reveal that IPOs with higher levels of PEM compliance are older, contain more words in their S-1 filings, and have longer filing period. Further, such firms are less likely to be audited by Big4, backed by venture capitalists, classified as HighTech firm, and listed on the Nasdaq exchange. All these results are statistically significant at the 1% level.

<sup>&</sup>lt;sup>13</sup> In untabulated analyses, we further examine which component of *Specificity* contributes most to the informational content of the S-1 filings. The number of specific details is disaggregated into quantitative (Money, Percentage, Date and Time) and qualitative groups (Organisation, Person, and Location). Both groups are then scaled by the number of words. Consistent with Hope et al. (2016), the results show that the specificity of quantitative information is highly significant while that of qualitative information is not statistically significant.

Column (6) shows marginal evidence on a positive association between *Bullets* and *TotalVol* (t-statistics = 1.82), indicating that the use of bullet lists is linked with increased ex ante uncertainty. Contrary to the SEC's view that bullet points help readers quickly absorb information (SEC, 1998a), this finding underscores that their impact in financial markets is contingent on factors such as the content and amount of information disclosed in each bullet point, as well as the section in which the bullet points appear.

Column (2) shows that the coefficient of *Conciseness* is negative, which is consistent with the prediction but not statistically related to ex ante uncertainty. Untabulated tests indicate that, out of five writing issues that compose the dimension of linguistic conciseness, only the coefficient of the percentage of legal words is statistically significant (t-statistics = -2.28), challenging the presumption that the use of legalese is always detrimental because impairs the readability of offering prospectuses. This result corroborates Cazier et al. (2021), who find that firms disclosing lengthy and boilerplate risk factor disclosures are less likely to be considered inadequate by judges in shareholder securities lawsuits and less likely to be targeted by a SEC comment letter. As revealed in column (5), while the coefficient on *Repetition* is negative as predicted, it is not statistically significant.<sup>14</sup> The lack of statistical significance highlights the empirical challenge in determining an optimal level of disclosure repetition that is beneficial to market participants, as argued in Cazier and Pfeiffer (2017).

Column (7) reports the results from estimating Equation (2), whereby the squared value of the overall PEM compliance (*PEIndex\_sqr*) is added. It is revealed that the coefficient of *PEIndex\_sqr* is positive and statistically significant at the 5% level (0.009, t-statistic =2.41), supporting the second hypothesis. In particular, the results indicate the presence of a turning point beyond which the benefits associated with greater compliance with the PEM are diminished because the information becomes too linguistically simplified and thus loses its informational content. To improve empirical rigor and result interpretation, additional analyses are conducted to provide more insights on the nature of the U-curve. As suggested by Lind and Mehlum (2010), obtaining a statistically significant coefficient of the squared value of the variable of interest alone appears insufficient to establish a quadratic relationship. Using the estimates of  $\beta_1$  and  $\beta_2$  of 0.071 and 0.009, the turning point (calculated as  $-\beta_1/(2\beta_2)$ ) is 3.95, which is positioned at the 87<sup>th</sup> percentile of the data. It is revealed that the slope at the upper bound of the data range (untabulated) is not sufficiently steep, suggesting that approximately

<sup>&</sup>lt;sup>14</sup> The results remain qualitatively unchanged when the threshold of *Repetition* is reduced from 0.9 to 0.6 as examined in Merkley (2014).

only half of a U-shape exists within the sample. The estimated 95-percent confidence interval using the Fieller method provides confirmatory evidence on the position of the turning point.

Taken together, we document empirical evidence on a U-shaped relationship between overall PEM compliance and ex ante uncertainty, suggesting the presence of a turning point beyond which adhering to the PEM is no longer beneficial. This result supports the experimental studies casting doubts on the unambiguous positive economic implications of improved disclosure readability (Rennekamp, 2012; Asay et al., 2017; Asay et al., 2018). Additional analysis shows that the turning point is located too close to the upper bound of the data range, which explains the observed negative association between *PEIndex* and post-IPO in the first hypothesis. More importantly, given the sample spans over a 17-year period, concerns around the unintended consequence of excessive PEM compliance are unnecessary. The achievement in ensuring an appropriate level of PEM compliance can be due to the significant SEC's oversight efforts in the context of IPOs (Li & Liu, 2017; Lowry et al., 2020). On balance, for the majority of IPOs, the SEC's PEM is operationally sound.

#### 5.3 The role of document length

Prior literature implies that disclosure length is equivalent to the complexity of the documents. For example, You and Zhang (2009) use a word count as a measure of 10-K complexity and find that investors' underreaction tends to be stronger for firms with lengthier annual reports. Relatedly, Miller (2010) documents that the effects of document length dominate those of disclosure readability (i.e., measured by Fog Index) when these two constructs are examined simultaneously. Nevertheless, the SEC explicitly notes the trade-off between complying with plain English guidelines and the length of disclosure, stating "writing disclosure in plain English can sometimes increase the length of particular sections of your document" (SEC, 1999, Question 11). We further address the nuance between disclosure length and the PEM compliance by investigating whether the length of initial prospectuses has a moderating effect on the association between the PEM compliance and ex ante uncertainty. Specifically, we split the sample is into quartiles by document length and exclude *Log(words)* from the Equation (1). As such, IPOs with the shortest prospectuses are placed in Quartile 1 while the longest prospectuses can be found in Quartile 4.

As shown in Table 6, the coefficients of all individual plain English attributes (except that of *Bullets*) as well as *PEIndex* are highly statistically significant at Quartile 4, suggesting that the compliance to most plain English dimensions and the overall PEM alleviates ex ante uncertainty, particularly for longer initial prospectuses. These results are consistent with

limited attention theory because heuristics tend to matter more when a greater amount of information is required to be analysed (Hirshleifer et al., 2011; Blankespoor et al., 2020). In the context of IPOs, our findings highlight the importance of substantial compliance with plain English guidelines, as advocated by the SEC, given investors are required to process a large amount of information presented in lengthy offering prospectuses.

#### [INSERT TABLE 6]

#### 6 The role of PEM compliance within IPO prospectus sub-sections

In addition to examining the effect of PEM compliance at the level of the whole prospectus, we are also interested in exploring whether adhering to plain English guidelines varies *across* sections of the IPO initial prospectus on ex ante uncertainty. Prior research shows that different types of narratives included in prospectuses serve distinct communicative purposes and therefore are subject to varying levels of managerial discretion, regulatory influence, and auditor scrutiny. For example, chairman's statements provide an overview of performance and key indicators of future prospects and are typically short (Clatworthy & Jones, 2001) and are thus found to be the most read (Courtis, 1982) and readable (Jones & Shoemaker, 1994). However, given this section is partly informational and partly promotional (Henry, 2008), its content is subject to impression management techniques because managers are likely (unlikely) to associate themselves with company financial results when the underlying financial performance is positive (negative) (Clatworthy & Jones, 2006).

To assess whether the effect of PEM compliance on ex ante uncertainty differs across the four main sections of IPO prospectuses, we follow Hanley and Hoberg (2010), focusing on the four key sub-sections within the initial prospectuses, namely the "Prospectus Summary" (PS), "Risk Factors" (RF), "Use of Proceeds," (UoP) and "Management's Discussion and Analysis" (MD&A). To parse the sub-sections, the beginning and end of each section of offering prospectuses are identified by the pagination implied by the Table of Contents.

#### 6.1 Predictions

Prior IPO studies typically examine a single textual attribute within a specific section of the prospectus, offering scant evidence on how that particular writing feature influences capital markets *across* multiple sections of the document (Leone et al., 2007; Beatty & Welch, 1996; Arnold et al., 2010). This poses significant challenges in predicting whether and the extent to which compliance with a given plain English attribute in a specific sub-section alleviates ex ante uncertainty. To predict the relationship between compliance with plain English attributes

across the four sub-sections and ex ante uncertainty, we draw on empirical insights from prior studies, where available, and take into account the thematic focus of the narrative content within each sub-section. A summary of predictions is displayed in Table 7.

#### [INSERT TABLE 7]

The PS section is intended to provide prospective investors with a concise overview of the key information to make an initial assessment of the investment opportunity and determine whether they should review the entire prospectus (Hanley & Hoberg, 2010). From the lens of impression management framework, it is probable that underwriters strategically over-enhance the presentation of favourable information while diverting the readers' attention away from unfavourable information (Merkl-Davies & Brennan, 2007). This is consistent with experimental evidence suggesting that managers strategically enhance the readability of good news disclosures when they have a strong reporting goal to make the firm appear as favourable as possible (Tan et al., 2015; Asay et al., 2018). We argue that firms can achieve this strategy by maximising the repetition of favourable information and the use of bullet points to facilitate such discussions in the PS section. Both repeating favourable information and using bullet lists aid investors to process such information more efficiently via reduced awareness costs (Tan et al., 2015; Blankespoor et al., 2020). Thus, we posit a negative relation between ex ante uncertainty and *Repetition* and *Bullets* within the PS section.

Within the RF section, we make three predictions. First, prior literature generally documents a positive relation between the frequency of uncertain words and the market's assessment of a firm's risk, suggesting that the uncertainty tone of this section is indicative of a firm's level of riskiness (e.g., Arnold et al., 2010; Campbell et al., 2014; Beatty et al., 2019). However, contemporary empirical evidence suggests the opposite - that a heightened use of uncertainty tone may function as a strategic tool to mitigate litigation risk (Cazier et al., 2021; Huang et al., 2022). Accordingly, the association between uncertainty tone and level of riskiness is positive only up to a certain threshold, beyond which it becomes negative, because extensive and detailed risk disclosures (reflected in a higher density of uncertainty tone) are perceived to reduce litigation risk. Combined, drawn from the preceding discussions, we conjecture that the relationship between Uncertainty and in the RF section and ex ante uncertainty exhibits an inverted U-shaped. Second, following Hope et al. (2016), documenting positive economic benefits associated with more specific risk disclosures, we predict a negative association between the level of specificity and post-IPO return volatility. Prior literature proposes and validates the number of risk factors listed and discussed in initial prospectuses as a proxy for ex ante uncertainty (Simunic & Stein, 1987; Clarkson, 1994). In line with the SEC's

recommendation that companies should use bullet points to present information (where relevant), anecdotal evidence suggests that firms are more likely to use bullet points to list out risk factors. The third prediction is a positive relation between the use of bullet points and post-IPO return volatility.

In the context of the UoP section, we predict that more specific disclosure relating to the intended use of proceeds is negatively related to ex ante uncertainty (Leone et al., 2007) while the opposite directional prediction is for the link between *Bullets* and ex ante uncertainty (Beatty & Ritter, 1986). Regarding the MD&A section, we make two predictions. Following Ferris et al. (2013), we predict a positive relation between the use of uncertainty tone and ex ante uncertainty. To an extent that repetitive disclosures in the MD&A section of annual reports are used to emphasize firm-specific events (Li, 2019), we predict a negative association between disclosure repetition and subsequent return volatility. Since the statistical significance of the relationship between *PEIndex* and ex ante uncertainty is dictated by compliance level of the five individual plain English dimensions, we form no prediction for the effect of overall plain English compliance on ex ante uncertainty in the sub-section analyses.

#### 6.2 Results

Table 8 displays descriptive statistics regarding PEIndex and its attributes for the four subsections. A number of observations emerge. First, the proportion of uncertain words in the RF section is approximately three times higher than that in the other three sub-sections, confirming the tone of the RF section being predominantly negative/uncertain due to extensive risk discussion (Kravet & Muslu, 2013; Ferris et al., 2013, Campbell et al., 2014). Second, the MD&A (RF) section contains the highest (lowest) number of specific words. Given the MD&A section is intended to discuss the management's assessment of the business and financial performance, such disclosure likely includes numbers, dates, percentages, and dollar amounts. On the other hand, the observed low level of Specificity in the RF section may result from companies not being required to quantify the impact of disclosed risks, pointing to a concern of this section being boilerplate and generic (Arnold et al., 2010; Beatty et al., 2019). Third, a significant proportion of repetitive sentences appears to be clustered in the MD&A section. Last, the RF section has the most bullet points, consistent with the inference that firms tend to use bullet points to outline the number of risk factors. Untabulated analyses further reveal a steady rise of linguistic conciseness for the PS and RF sections while no similar pattern is observed for Uncertainty across all four sub-sections. In contrast to Hope et al. (2016) and Beatty et al. (2019), who document a deterioration in the specificity of risk disclosures in the RF section of 10-K filings, *Specificity* in the RF section of S-1 filings remains plateaued from 2002 to 2019.

#### [INSERT TABLE 8]

Table 9 reports the results of plain English compliance for the four sub-sections. For the PS section, we document a negative and statistically significant coefficient on *Repetition* at the five percent level (t-statistic = -2.57), indicating that more repetitive sentences are related to reduced ex ante uncertainty. In addition, the coefficient on *Bullets* is negative and statistically significant at the five percent level, suggesting that the number of bullet points used in the PS section is associated with reduced post-IPO return volatility. This result lends support to the view that IPOs issuing firms have significant discretion over disclosure content and strong incentives to over-enhance favourable information by utilising bullet points in the PS section. In contrast to the documented positive association between the use of bullet points for the whole prospectus and post-IPO return volatility, this finding corroborates the inference that the economic implications related to the use of bullets points are complex and depends on other information characteristics.<sup>15</sup>

Within the RF section, the coefficient of *Uncertainty* is positive and its squared value (*Uncertainty\_sqr*) is negative, with both being statistically significant at the 5% level, supporting our prediction of a non-linear association between the proportion of uncertain words and subsequent return volatility. Using the estimates of  $\beta_1$  and  $\beta_2$  of 1.657 and -0.123, the turning point (calculated as  $-\beta_1/(2\beta_2)$ ) is 6.74%, which is positioned at the 44<sup>th</sup> percentile of *Uncertainty*. Further, the slopes at both ends of the curve are statistically significant (untabulated) and therefore, sufficiently steep. Combined, we document the relation between the use of uncertainty tone and ex ante uncertainty is inverted U-shaped. This finding corroborates a stream of literature suggesting that a greater frequency of uncertain words in the RF section points to a higher level of a firm's riskiness and thus is associated with increased ex ante uncertainty (Loughran & McDonald, 2013; Arnold et al., 2010; Kravet & Muslu, 2013; Campbell et al., 2014). However, we empirically show that this association is positive only up to a certain point after which the higher proportion of uncertainty tone is linked to reduced ex ante uncertainty. This reduction in post-IPO return volatility could be attributed to the perceived reduction in litigation risk, aligning with the advantages associated with boilerplate

<sup>&</sup>lt;sup>15</sup> The result remains unchanged qualitatively when we add a dummy variable set equal to a value of one if there is at least a sentence that is substantially similar to other sentences in this section (cosine similarity being at least 0.9), and zero otherwise. Also, the inference of the finding remains qualitatively indifferent when a control variable representing the use of bullet points in other sections is added.

risk reporting (Cazier et al., 2021) and the more frequent use of forward-looking statements (Huang et al., 2022) in the RF section.

In the same sub-section, we also document a positive association between *Bullets* and *TotalVol*, suggesting that the more bullet points are used, the higher the level of post-IPO return volatility is.<sup>16</sup> We find a positive but statistically insignificant coefficient on the interaction between *Specificity* and *TotalVol*, which is in contrast to Hope et al., (2016). It is worth noting two methodological differences between the two studies. First, Hope et al. (2016) examine risk-factor disclosures in 10-K filings from 2006-2011 while this study examines the risk-factor disclosures in S-1 filings from 2002-2019. As noted in the Table 9, there is a significant difference between the average length of the RF section in initial offering prospectuses and that in annual reports. Specifically, the average number of words in the RF section in this study is 14,296, nearly three times higher than the number reported in Hope et al. (2016). Second, Hope et al. (2016) includes financial firms that accounts for almost 22% of the sample while this study excludes financial firms because certain words have different connotations for financial firms as compared to that for non-financial firms (Jegadeesh & Wu, 2013).

For the UoP section, the coefficient of *Specificity* is -0.067 and statistically significant at the 1% level, indicating a reduction in ex ante uncertainty related to increased linguistic specificity. This finding supports our prediction and prior literature documenting the benefits related to more specific UoP within the offering prospectuses (Ljungqvist & Wilhelm, 2003; Leone et al., 2007). We also find marginal evidence on a positive coefficient of *Bullets* of 0.037 at the ten percent level, corroborating Beatty and Ritter (1986), who documents firms with a greater number of uses of proceeds tend to exhibit higher IPO underpricing. It is shown that the coefficient of the aggregate measure of all five plain English attributes (*PEIndex*) is -0.047 and statistically significant at the five percent level. This finding confirms the role of overall PEM compliance on alleviating ex ante uncertainty, even in a short document.

Within the MD&A section, we document no evidence that complying with plain English guidelines reduces ex ante uncertainty. Contrary to our prediction, the coefficient on *Uncertainty* is negative and statistically insignificant. A possible explanation for the lack of statistical significance is the varied focus of discussions within the MD&A section. For

<sup>&</sup>lt;sup>16</sup> In an additional analysis, we split *Log(words)* into two components, namely *Log(words)\_RF* as the logarithm of total words of the RF section and *Log(words)\_others* as the logarithm of total words of the remaining sections in the prospectus. Our results show that the coefficient on *Log(words)\_RF* is positive and statistically significant at least at 5% level across all models while *Log(words)\_others* is negative and statistically significant at the 1% level. This finding supports previous studies suggesting the length of the RF section is a proxy of a firm's riskiness (e.g., Kravet and Muslu, 2013; Campbell et al., 2014) while the greater length of other sections points to the notion of greater quantity of information and thus, reduced information asymmetry (Merkley, 2014; Bozanic et al., 2017).

instance, the market impact related to the use of uncertainty tone on subsequent return volatility is likely limited if such discussions are boilerplate, as highlighted in Brown and Tucker (2011). In addition, in contrast to our prediction, the coefficient on *Repetition* is positive but statistically insignificant, indicating no evidence of the economic benefits of repeating information *within* the MD&A section. Considering the scant guidance from prior literature regarding the optimal balance between useful repetition and burdensome redundancy, combined with the lack of statistical significance in the results, the implications remain limited.

#### [INSERT TABLE 9]

#### 7 Additional analyses

#### 7.1 Bog Index

As a robustness test to the use of *PEIndex*, we use the Bog Index as an alternative aggregate measure of plain English attributes (Bonsall et al., 2017). We rerun the Equation (1) and (2) with Bog Index being the variable of interest. Table 10 presents the results of these analyses.

As reported in Panel A, the mean of Bog Index is 92.91, which is designated by StyleWriter as "poor" readability. Noticeably, this finding indicates that the plain English readability of S-1 filings in this study, on average, is lower than those filed in 1996 and 1997 (103.79) and higher than those filed in the period 1999-2000 (86.27) as reported in Bonsall et al. (2017). In addition, the mean of Bog scores of 123,033 U.S. annual reports from 2002 to 2019 is 84.04, which indicates that IPO prospectuses on average are far less readable, relative to the 10-K filings.<sup>17</sup> The correlation matrix (untabulated) shows that *PEIndex* and *Bog* is negatively correlated but not statistically significant, reiterating that the two measures fundamentally capture different aspects of the PEM compliance. The correlation coefficients between five plain English dimensions and *Bog* are statistically significant, with the highest being 0.6 of *Bullets*.

Panel B reports the multivariate results. Column (1) shows that the coefficient of *Bog* is positive as predicted but not statistically insignificant. Upon a disaggregation of three components, there is no evidence suggesting average sentence length (as measured by *Sentence Bog*) and the plain English style problems and word difficulty (as measured by *Word Bog*) can alleviate ex ante uncertainty. Contrary to our prediction, column (4) reports a positive and significant relation between *Pep* and the dependent variable at 1% significance level (t-

<sup>&</sup>lt;sup>17</sup> We thank Brian Miller, who generously provides the Bog scores of 10-Ks from 1994 to 2020 on his website.

statistics = 3.87). Echoing the concern of the construct validity of *Pep* discussed in the section 4.1, it is not entirely clear how the greater frequency of lively verbs, interesting nouns, names, and conversational style is 'pep up' the writing of prospectuses but associated with higher subsequent volatility.

As shown in column (5), the coefficient of *Bog* remains negative and highly significant while that of *Bog\_sqr* is positive and statistically significant at the 1% level. It is shown that additional analyses indicate that the estimated lower and upper bounds of the data range using the Fieller method are within the sample, ascertaining that the turning point is well positioned within the data range. Consistent with the primary results, it is revealed that IPO prospectuses with higher *Bog* experience lower post-IPO return volatility. This reduction in disclosure readability score is perhaps due to the innate linguistic complexity deemed necessary to convey complex yet essential technical information (Bloomfield, 2008; Leuz & Wysocki, 2016; Loughran & McDonald, 2016). However, this negative association likely reaches to a point beyond which the relation between the *Bog* and post-IPO return volatility becomes positive. This result suggests that significantly low PEM compliance hinders investors' ability to assess IPO, leading to increased ex ante uncertainty. This finding supports existing literature documenting the negative economic consequences of lower compliance with plain English guidelines (e.g., Li, 2008; Bonsall et al., 2017; and Bonsall & Miller, 2017).<sup>18</sup>

#### [INSERT TABLE 10]

#### 7.2 Alternative measure of ex ante uncertainty – Underpricing

A well-established stream of literature examining IPO pricing documents that ex ante uncertainty is evident in the IPO pricing process and that higher ex ante uncertainty leads to higher underpricing (see Ljungqvist [2007] for a review). An IPO stock is underpriced when the closing price on the first trading day is greater than its final offer price. We revisit the first hypothesis by employing underpricing is an alternative proxy of ex ante uncertainty. Following prior literature, we measure underpricing over three intervals using the closing price at the IPO date, the first trading day, and 30 trading days after the IPO (Barth et al., 2017). In addition to a set of control variables used in primary analyses, following Loughran and McDonald (2013),

<sup>&</sup>lt;sup>18</sup> For completeness, we also use Bog Index as the independent variable in Equation (1) to assess the effect of plain English compliance on ex ante uncertainty in the four sub-sections. Untabulated results show that only the coefficient of *Bog* within the RF section is statistically significant at the 5% level, implying that less readable RF disclosures are linked to reduced subsequent return volatility. This finding contrasts with the non-significant result reported in section 6.2, where *PEIndex* is used.

we include *Up Revision*, which is percentage upward revision in the offer price from the midpoint of the filing range if the offer price is greater than mid-point.

Untabulated results show that the coefficients of underpricing, measured at three different intervals, are directionally consistent with prediction but not statistically significant, suggesting no role of plain English compliance in reducing the level of ex ante uncertainty. The absence of statistical significance on the relation between IPO underpricing and the compliance with the PEM and its attributes could be due to measurement noises arising from the abnormal share flipping activities within the first four trading days following the IPO date (Loughran & McDonald, 2011).<sup>19</sup>

#### 7.3 Principal component analysis

To probe whether some individual plain English attributes capture the same dimensionality in nature, the plain English dimensions are combined into factors using exploratory factor (principal component) analysis. In untabulated analyses, both models (without and with factor rotation, where the rotation method is the orthogonal varimax with Kaiser normalisation) indicate that three factors should be retained because their eigenvalues exceed 1.0 and explain 81.71% of the five plain English attributes. The Kaiser-Myer-Olkin score at 0.564 (above a critical threshold of 0.5) justifies the sampling adequacy for factor analysis whereas the Barlett's test of sphericity being statistically significant at 1% level confirms the retainment of three factors (Caglio et al., 2020). Using an eigenvector threshold of 0.5, Factor 1 is loaded with high values of Uncertainty and Bullets as well as lower value of Specificity while Factor 2 primarily includes values of Repetition. Factor 3 represents the values of Conciseness.

Consistent with predictions, Factor 1, which is loaded with high values of *Uncertainty* and *Bullets* as well as lower value of *Specificity*, is significantly positively associated with post-IPO return volatility (t-statistics = 3.62). With Factor 2 (Factor 3) including values of *Repetition* (*Conciseness*), their coefficients are directionally consistent with prediction but statistically insignificant. When all factors are examined in a regression, the statistical significance of Factor 1 remains qualitatively unchanged while those of Factors 2 and 3 are statistically non-significant. Collectively, supporting the results from the primary analysis, the findings from

<sup>&</sup>lt;sup>19</sup> Loughran and McDonald (2013) measure the level of uncertainty tone by an aggregate wordlist comprising the wordlists of negative words, uncertain words, and weak modal words. They find that shares of IPOs with a higher proportion of uncertain words (i.e., uncertainty tone) in their S-1 filings are underpriced to a greater extent. Since this study also employs the same proxy to measure *Uncertainty*, for reconciliation purpose, we replicate by examining the role of uncertainty tone in the sampling period from 1997 to 2010 as similar to Loughran and McDonald (2013). We also use identical control variables to their study. Untabulated results show that the effect of uncertain tone is diminished entirely once the period from 1997 to 2001 is excluded from the sample.

factor analysis suggest that the greater use of uncertain words and bullet points as well as the lower frequency of specific words capture the same dimension, likely indicating the level of innate riskiness of IPOs. This dimensionality is associated with higher post-IPO return volatility.

#### 7.4 Components of total volatility

Following Barth et al. (2017), we disaggregate *TotalVol* into idiosyncratic (*IdioVol*) and systematic (*Beta*) components. *IdioVol* (*Beta*) is the standard deviation of residuals (slope coefficient) from a firm-specific market model estimated over a window of 60 trading days from the 5th day after the IPO. All results (untabulated) remain qualitatively identical to those reported in Table 5 when the dependent variable is *IdioVol*, suggesting that the documented reduction in ex ante uncertainty is primarily driven by firm-specific information whose revelation is facilitated by greater compliance with plain English principles.

#### 8 Conclusion

This study investigates the effect of complying with the PEM on ex ante uncertainty in the context of IPO initial prospectuses. Acknowledging that the PEM is a multifaceted notion, we consider the full breadth of its aspects by using a set of measures that are directly and faithfully linked to the spirit and language of the PEM. Based on a sample of 1377 IPOs in the U.S. from 2002 to 2019, we find evidence indicating benefits of complying with the PEM in the form of reduced ex ante uncertainty. The benefit of adhering to plain English guidelines is evident in long prospectuses as well as short sub-section such as the Use of Proceeds. Overall, the findings support the notion of reduced information processing costs arising from better plain English compliance and the perceived benefits of the SEC since the inception of this reporting initiative.

We also find a U-shaped relation between the degree of plain English compliance and ex ante uncertainty, ascertaining the inference that adhering too closely to the PEM could lead to a loss of necessary latent informational complexity. Additional analysis reveals that the empirical switching point is near the upper bound of the sample, indicating that concerns around the unintended consequence of excessive PEM compliance are unnecessary. Combined, using the sample firms in this study, I show that on balance, higher compliance with the PEM is beneficial, suggesting that the SEC's mandate for plain English is operationally sound.

Future research can further explore the economic implications when issuers comply with a broader set of plain English guidelines (i.e., not necessarily prescribed by the SEC) by employing unsupervised machine-learning analysis. This enables the discovery of latent linguistic features that may not necessarily examined in this study. Another avenue for future studies is to examine the role of media coverage, on the basis that offering prospectuses and media are two primary and dominant information channels through which prospective investors can collect and process IPO information.

#### References

- Asay, H. S., Elliott, W. B., & Rennekamp, K. M. (2017). Disclosure Readability and the Sensitivity of Investors' Valuation Judgments to Outside Information. *The Accounting Review*, 92(4), 1-25.
- Asay, H. S., Libby, R., & Rennekamp, K. M. (2018). Firm Performance, Reporting Goals, and Language in Narrative Disclosures. *Journal of Accounting and Economics*, 65(2-3), 380-398.
- Arnold, T., Fishe, R. P., & North, D. (2010). The effects of ambiguous information on initial and subsequent IPO returns. *Financial Management*, 39(4), 1497-1519.
- Barth, M. E., Landsman, W. R., & Taylor, D. J. (2017). The JOBS Act and information uncertainty in IPO firms. *The Accounting Review*, 92(6), 25-47.
- Beatty, A., L. Cheng, and H. Zhang. (2019). Are risk factor disclosures still relevant? Evidence from market reactions to risk factor disclosures before and after the financial crisis. *Contemporary Accounting Research* 36(2), 805–838.
- Beatty, R. P., & Ritter, J. R. (1986). Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics*, 15(1-2), 213-232.
- Beatty, R. P., & Welch, I. (1996). Issuer expenses and legal liability in initial public offerings. *The Journal of Law and Economics*, 39(2), 545-602.
- Blankespoor, E., deHaan, E., & Marinovic, I. (2020). Disclosure Processing Costs and Investors' Information Choice, and Equity Market Outcomes: A Review. *Journal of Accounting and Economics*, 70(2-3), 101344.
- Bonsall, S. B., Leone, A. J., Miller, B. P., & Rennekamp, K. M. (2017). A plain English measure of financial reporting readability. *Journal of Accounting and Economics*, 63(2-3), 329-357.
- Bonsall, S. B., & Miller, B. P. (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies*, 22(2), 608-643.
- Bozanic, Z., Dietrich, J. R., & Johnson, B. A. (2017). SEC comment letters and firm disclosure. *Journal of Accounting and Public Policy*, *36*(5), 337-357.
- Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2), 309-346.
- Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information?. *Journal of Accounting Research*, *56*(1), 85-121.
- Cacioppo, J. T., & Petty, R. E. (1989). Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology*, 10(1), 3-12.
- Caglio, A., Melloni, G., & Perego, P. (2020). Informational content and assurance of textual disclosures: Evidence on integrated reporting. *European Accounting Review*, 29(1), 55-83.
- Campbell, J. L., Chen, H., Dhaliwal, D. S., Lu, H. M., & Steele, L. B. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Review of Accounting Studies*, 19(1), 396-455.
- Cazier, R. A., McMullin, J. L., & Treu, J. S. (2021). Are lengthy and boilerplate risk factor disclosures inadequate? An examination of judicial and regulatory assessments of risk factor language. *The Accounting Review*, 96(4), 131-155.
- Cazier, R. A., & Pfeiffer, R. J. (2017). 10-K disclosure repetition and managerial reporting incentives. *Journal of Financial Reporting*, 2(1), 107-131.
- Cheynel, E., & Levine, C. B. (2020). Public disclosures and information asymmetry: A theory of the mosaic. *The Accounting Review*, 95(1), 79-99.
- Clarkson, P. M. (1994). The underpricing of initial public offerings, ex ante uncertainty, and proxy selection. *Accounting & Finance*, 34(2), 67-78.

- Clatworthy, M., & Jones, M. J. (2001). The effect of thematic structure on the variability of annual report readability. *Accounting, Auditing & Accountability Journal, 14*(3), 311-326.
- Clatworthy, M. A., & Jones, M. J. (2006). Differential patterns of textual characteristics and company performance in the chairman's statement. *Accounting, Auditing & Accountability Journal*, 19(4), 493-511.
- Courtis, J. K. (1982). Private shareholder response to corporate annual reports. *Accounting & Finance*, 22(2), 53-72.
- DuBay, W. H. (2004). The Principles of Readability. Impact Information, Costa Mesa, CA.
- Dyer, T., Lang, M., & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics*, 64(2-3), 221-245.
- Eagleson, R. (2014). Short Definition of Plain Language. Retrieved from <u>https://plainlanguage.gov/about/definitions/short-definition/</u> (Accessed on April 12, 2018).
- Ferris, S. P., Hao, Q., & Liao, M. Y. (2013). The effect of issuer conservatism on IPO pricing and performance. *Review of Finance*, *17*(3), 993-1027.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3), 393-408.
- Gunning, R. (1952). The Technique of Clear Writing. McGraw-Hill International Book Co. New York, NY.
- Haans, R. F., Pieters, C., & He, Z. L. (2016). Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. *Strategic Management Journal*, 37(7), 1177-1195.
- Hanley, K. W., & Hoberg, G. (2010). The information content of IPO prospectuses. *The Review* of *Financial Studies*, 23(7), 2821-2864.
- Henry, E. (2008). Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication (1973)*, 45(4), 363-407.
- Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2011). Limited investor attention and stock market misreactions to accounting information. *The Review of Asset Pricing Studies*, 1(1), 35-73.
- Hope, O. K., Hu, D., & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Review* of Accounting Studies, 21(4), 1005-1045.
- Huang, A. H., Shen, J., & Zang, A. Y. (2022). The unintended benefit of the risk factor mandate of 2005. *Review of Accounting Studies*, 27(4), 1319-1355.
- Jegadeesh, N., & Wu, D. (2013). Word power: A new approach for content analysis. *Journal* of Financial Economics, 110(3), 712-729.
- Jones, M. J., & Shoemaker, P. A. (1994). Accounting narratives: A review of empirical studies of content and readability. *Journal of Accounting Literature*, 13, 142-165.
- Kravet, T., & Muslu, V. (2013). Textual risk disclosures and investors' risk perceptions. *Review of Accounting Studies*, 18(4), 1088-1122.
- Lang, M., & Stice-Lawrence, L. (2015). Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics*, 60(2-3), 110-135.
- Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics*, *56*(1), 130-147.
- Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 86(3), 1087-1115.
- Leone, A. J., Rock, S., & Willenborg, M. (2007). Disclosure of intended use of proceeds and underpricing in initial public offerings. *Journal of Accounting Research*, 45(1), 111-153.

- Leuz, C., & Wysocki, P. D. (2016). The economics of disclosure and financial reporting regulation: Evidence and suggestions for future research. *Journal of Accounting Research*, 54(2), 525-622.
- Li, B., & Liu, Z. (2017). The oversight role of regulators: evidence from SEC comment letters in the IPO process. *Review of Accounting Studies*, 22(3), 1229-1260.
- Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2), 221-247.
- Li, F. (2010). The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research*, 48(5), 1049-1102.
- Li, H. (2019). Repetitive disclosures in the MD&A. Journal of Business, Finance, and Accounting, 46(9-10), 1063-1096.
- Libby, R., & Emett, S. A. (2014). Earnings presentation effects on manager reporting choices and investor decisions. *Accounting and Business Research*, *44*(4), 410-438.
- Ljungqvist, A. (2007). IPO underpricing. In *Handbook of Empirical Corporate Finance* (pp. 375-422). Elsevier.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65.
- Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, 109(2), 307-326.
- Loughran, T., & McDonald, B. (2014a). Measuring readability in financial disclosures. *The Journal of Finance*, 69(4), 1643-1671.
- Loughran, T., & McDonald, B. (2014b). Regulation and financial disclosure: The impact of plain English. *Journal of Regulatory Economics*, 45(1), 94-113.
- Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.
- Lowry, M., Officer, M. S., & Schwert, G. W. (2010). The variability of IPO initial returns. *The Journal of Finance*, 65(2), 425-465.
- Lowry, M., Michaely, R., & Volkova, E. (2017). Initial public offerings: A synthesis of the literature and directions for future research. *Foundations and Trends in Finance*, 11(3-4), 154-320.
- Lowry, M., Michaely, R., & Volkova, E. (2020). Information Revealed through the Regulatory Process: Interactions between the SEC and Companies ahead of Their IPO. *The Review of Financial Studies*, *33*(12), 5510-5554.
- Lu, J. (2022). Limited attention: Implications for financial reporting. *Journal of Accounting Research*, 60(5), 1991-2027.
- Merkl-Davies, D. M., & Brennan, N. M. (2007). Discretionary disclosure strategies in corporate narratives: incremental information or impression management?. *Journal of Accounting Literature*, 27, 116-196.
- Merkley, K. J. (2014). Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review*, 89(2), 725-757.
- Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review*, 85(6), 2107-2143.
- Plain Language Action and Information Network (PLAIN). (n.d.). Retrieved from <u>https://www.plainlanguage.gov</u>
- Rennekamp, K. (2012). Processing fluency and investors' reactions to disclosure readability. *Journal of Accounting Research*, 50(5), 1319-1354.
- Securities and Exchange Commission (SEC). (1997). *Plain English Disclosure*. Retrieved from https://www.sec.gov/rules/proposed/34-38164.txt

- Securities and Exchange Commission (SEC). (August, 1998a). A Plain English Handbook: How to create clear SEC disclosure documents. SEC Office of Investor Education and Assistance. Retrieved from <u>http://www.sec.gov/pdf/handbook.pdf</u>.
- Securities and Exchange Commission (SEC). (September 4, 1998b). *Staff Legal Bulletin No.* 7 (*CF*). Retrieved from <a href="https://www.sec.gov/interps/legal/slbcf7.htm">https://www.sec.gov/interps/legal/slbcf7.htm</a>.
- Securities and Exchange Commission (SEC). (June 7, 1999). *Updated Staff Legal Bulletin No.* 7. Retrieved from <u>https://www.sec.gov/interps/legal/slbcf7a.htm</u>.
- Securities and Exchange Commission (SEC). (2006). *Executive compensation and related person disclosure*. Retrieved from https://www.sec.gov/rules/final/2006/33-8732a.pdf
- Securities and Exchange Commission (SEC). (July 13, 2016). SEC proposes Amendments to Update and Simplify Disclosure Requirements As Part of Overall Disclosure Effectiveness Review. Retrieved from https://www.sec.gov/news/press-release/2016-141
- Simunic, D. A., & Stein, M. T. (1987). Product differentiation in auditing: Auditor choice in the market for unseasoned new issues (Vol. 13). Canadian Certified General Accountants' Research Foundation.
- Stephens, K. K. (2007). The successive use of information and communication technologies at work. *Communication Theory*, *17*(4), 486-507.
- Stephens, K. K., & Rains, S. A. (2011). Information and communication technology sequences and message repetition in interpersonal interaction. *Communication Research*, *38*(1), 101-122.
- Tan, H. T., Ying Wang, E., & Zhou, B. O. (2014). When the use of positive language backfires: The joint effect of tone, readability, and investor sophistication on earnings judgments. *Journal of Accounting Research*, 52(1), 273-302.
- Tan, H. T., Wang, E. Y., & Zhou, B. (2015). How does readability influence investors' judgments? Consistency of benchmark performance matters. *The Accounting Review*, 90(1), 371-393.
- You, H., & Zhang, X. J. (2009). Financial reporting complexity and investor underreaction to 10-K information. *Review of Accounting studies*, 14(4), 559-586.

Figure 1: Summary of the Plain English Mandate



Plain English	Writing principle/convention	<b>Reference to the PEM</b>			
dimension		Rule 421(b)	Rule 421(d)	Plain English handbook	
1. Conciseness	Use active voice		$\checkmark$	pp. 19-20	
	Use strong verbs			pp. 20-21	
	Avoid legalese and highly technical business terms	✓	<b>√</b>	pp. 30-31	
	Omit superfluous words			pp. 25-26	
	Use personal pronouns			page 22	
	Use short sentences	✓	✓	pp. 28-29	
	No multiple negatives		✓	page 27	
	Avoid frequent reliance on glossaries and defined terms	~		page 13	
2. Specificity	Avoid vague explanations (by using more concrete terms) that are imprecise and readily subject to different interpretations	~	~	pp. 23-24	
3. Uncertainty	Use definite words		~		
4. Repetition	Avoid repeating disclosure in different sections	✓		page 12	
5. Bullets	Use bullets to list information wherever possible	~	~	page 48	

# Table 1 Dimensions capturing Plain English writing and presentation principles

# Table 2Sample selection procedures

Final sample size	1,377
SEC's EDGAR.	
less firms of which online initial prospectuses filings are not available on the	(11)
less firms of which online initial prospectuses filings not machine-readable.	(164)
and match with the SDC database.	
CIK is required to download the filings from the SEC's EDGAR website	
less firms that do not have CIK (matched by PERMNO and company names).	(6)
matched by PERMO and CUSIP (either 6 digits or 8 digits).	
maximize the sample size, IPO data and financial variables from CRSP are	
less firms that do not have sufficient information to match with CRSP. To	(284)
value of shares of these firms are combined.	
both types. These duplicates are identified and removed. Total number and	( )
<i>less</i> firms can offer shares that are either fully paid, partially paid, or a mix of	(186)
database.	
<i>less</i> firms with founding dates unable be found in Filed-Ritter's founding date	(21)
firms issue common shares.	
partnerships, stock units (Lowry et al., 2017). This means that the sampling	
less firms that offer the following types: units, natural resource limited	(68)
CRSP (Amex. NYSE, and NASDAO).	
offerings, penny stocks (offering price less than \$5), and stocks not listed on	
financial and insurance firms closed-end funds simultaneous international	2,117
Firm-commitment completed IPOs from 2002 to 2019, excluding REITs, ADRs,	2.117

Variables	Mean	Std. Dev.	P1	Q1	Median	Q3	P99
Dependent variables							
TotalVol	3.62	1.57	1.25	2.51	3.39	4.44	9.98
Independent variables							
Conciseness	-0.36	3.25	-24.3	-0.60	-0.29	0.45	8.84
Passive verbs	952.04	363.28	381	698	911	1,148	2,227
Hidden verbs	266.69	106.75	95	192	246	330	587
Legal words	102.02	61.44	24	63	88	123	369
Wordy phrases	1,730	604.72	712	1,283	1,698	2,082	3,626
Personal pronouns	4,310.6	1,316.8	1,909	3,264	4,313	5,183	7,684
Uncertainty (%)	3.17	0.46	2.19	2.83	3.13	3.49	4.24
Specificity (%)	3.73	0.74	2.51	3.17	3.59	4.19	5.81
Repetition	26.14	8.90	2.46	19.90	24.60	30.72	4.07
Bullets	286.67	114.63	73	202	276	366	570
PEIndex	0	2.88	-5.55	-1.94	-0.04	1.91	6.45
Control variables							
Firm characteristics							
Age	20.70	25.02	1	7	11	24	132
Share overhang	3.00	2.01	0	1.71	2.72	3.96	10.35
Positive EPS	0.39	0.49	0	0	0	1	1
Offer characteristics							
Words	71,852	23,235	30,195	53,605	70,948	86,620	142,234
Audit	0.83	0.38	0	1	1	1	1
VC	0.52	0.50	0	0	1	1	1
Days	121.06	134.24	24	38	85	127	832
Market characteristics							
Pre mkt ret	0.93	2.67	-6.82	-0.62	1.05	2.65	7.05
HighTech	0.31	0.46	0	0	0	1	1
Nasdaq	0.60	0.49	0	0	1	1	1

#### Table 3Summary Statistics

The table presents descriptive statistics for variables used in primary tests. The sample includes 1,377 completed U.S. IPOs from 2002 to 2019. **TotalVol** is the standard deviation of daily equity returns over the window of 60 trading days beginning the 5<sup>th</sup> day after the IPO. **PEIndex** is an aggregate measure of plain English compliance of which *Uncertainty* and *Bullets* (*Conciseness, Specificity* and *Repetition*) are negatively (positively) summed. All variables are defined in Appendix C. All continuous variables are winsorised at the 1st and 99th percentile.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1)TotalVol	1						
(2) PEIndex	-0.411***	1					
(3) Conciseness	0.009	$0.097^{***}$	1				
(4) Uncertainty	$0.407^{***}$	-0.784***	$0.072^{**}$	1			
(5) Specificity	-0.377***	$0.794^{***}$	-0.021	$-0.600^{***}$	1		
(6) Ln(Repetition)	-0.144***	$0.373^{***}$	0.033	-0.142***	$0.229^{***}$	1	
(7) Bullets	$0.238^{***}$	-0.583***	$0.090^{***}$	$0.525^{***}$	-0.447***	$0.273^{***}$	1
(8) Log(words)	$0.058^{*}$	-0.148***	$0.070^{**}$	$0.259^{***}$	-0.169***	$0.715^{***}$	$0.697^{***}$

Table 4Correlation Matrix

The sample includes 1,377 completed U.S. IPOs from 2002 to 2019. See Appendix C for detailed definitions of variables. \*, \*\*, and \*\*\* represent significance levels of 0.10, 0.05, and 0.01, respectively.

	Exp			Depende	nt variable =	TotalVol		
	Sign	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PEIndex	-	-0.067***						-0.071***
		(-3.27)						(-3.73)
Conciseness	-		-0.013					
			(-1.31)					
Uncertainty	+			0.417***				
				(3.32)				
Specificity	-				-0.172**			
					(-2.17)			
Ln(Repetition)	-					-0.208		
						(-1.49)		
Bullets	+					· · ·	0.071*	
							(1.82)	
PEIndex sqr	+							0.009**
_ 1								(2.41)
Log(words)	-	-0.931***	-1.229***	-1.039***	-1.003***	-0.836**	-1.537***	-1.036***
		(-3.01)	(-3.92)	(-3.28)	(-3.16)	(-2.30)	(-4.64)	(-3.45)
Ln(1+Age)	-	-0.201***	-0.217***	-0.212***	-0.192***	-0.212***	-0.210***	-0.204***
		(-4.01)	(-4.41)	(-4.26)	(-3.75)	(-4.28)	(-4.24)	(-4.06)
Share Overhang	+	0.045**	0.045**	0.042*	0.047**	0.044**	0.045**	0.045**
C		(2.20)	(2.20)	(2.03)	(2.44)	(2.20)	(2.22)	(2.23)
Positive EPS	-	-0.246**	-0.260**	-0.226**	-0.269**	-0.261**	-0.259**	-0.238**
		(-2.35)	(-2.44)	(-2.17)	(-2.53)	(-2.43)	(-2.45)	(-2.31)
Audit	-	-0.448***	-0.415***	-0.425***	-0.433***	-0.429***	-0.425***	-0.447***
		(-3.85)	(-3.41)	(-3.59)	(-3.67)	(-3.71)	(-3.58)	(-3.88)
VC	+	0.571***	0.652***	0.607***	0.596***	0.629***	0.629***	0.564***
		(7.49)	(7.91)	(6.72)	(8.28)	(8.56)	(7.59)	(7.72)
Ln(Days)	+	0.062	0.047	0.066	0.060	0.049	0.052	0.068
		(0.84)	(0.63)	(0.88)	(0.81)	(0.66)	(0.69)	(0.91)
Pre mkt ret	-	-0.034**	-0.033**	-0.034**	-0.034**	-0.033**	-0.033**	-0.034**
		(-2.36)	(-2.28)	(-2.29)	(-2.33)	(-2.29)	(-2.29)	(-2.30)
HighTech	-	-0.397**	-0.370**	-0.416**	-0.388**	-0.372**	-0.376**	-0.389**
		(-2.78)	(-2.66)	(-2.79)	(-2.81)	(-2.65)	(-2.67)	(-2.72)
Nasdaq	+	0.267**	0.286**	0.258**	0.279**	0.284**	0.279**	0.266**
		(2.70)	(2.80)	(2.56)	(2.78)	(2.77)	(2.73)	(2.67)
Constant		8.619***	9.861***	7.795***	9.525***	8.681***	11.186***	9.048***
		(5.39)	(6.07)	(4.25)	(5.97)	(5.28)	(6.97)	(5.99)
Adj R-squared		0.326	0.322	0.328	0.324	0.322	0.322	0.328

## Table 5PEM compliance and ex ante uncertainty

The sample includes 1,377 completed U.S. IPOs from 2002 to 2019. **TotalVol** is the standard deviation of daily equity returns over the window of 60 trading days beginning the 5<sup>th</sup> day after the IPO. **PEIndex** is an aggregate measure of plain English compliance of which *Uncertainty* and *Bullets* (*Conciseness, Specificity* and *Repetition*) are negatively (positively) summed. All variables are defined in Appendix C. All regressions include fixed effects associated with Fama and French (1997) 12-industry classification and IPO year. All continuous variables are winsorised at the 1<sup>st</sup> and 99<sup>th</sup> percentile. The t-statistics are in parentheses with the robust standard errors clustered by Form S-1 filing year. \*, \*\*, and \*\*\* represent significance levels of 0.10, 0.05, and 0.01, respectively.

Plain English				A11		
dimensions	_	Q1 (shortest)	Q2	Q3	Q4 (longest)	sample
PEIndex		-	+	**	*** -	*** -
Conciseness		+	-	-	** <b>-</b>	-
Uncertainty		+	-	+	+***	+***
Specificity		-	+	***	*** -	**
Ln(Repetition)		-	+	-	***	-
Bullets		-	-	+	+	+*
n	=	344	344	344	345	1,377

#### Table 6The role of document length

Dependent variable is **TotalVol**, defined as the standard deviation of daily equity returns over the window of 60 trading days beginning the 5<sup>th</sup> day after the IPO. For brevity, control variables are not reported but their statistical significance remains consistent with the main analyses. All regressions include fixed effects associated with Fama and French (1997) 12-industry classification and IPO year. All continuous variables are winsorised at the 1<sup>st</sup> and 99<sup>th</sup> percentile. The t-statistics are in parentheses with the robust standard errors clustered by Form S-1 filing year. \*, \*\*, and \*\*\* represent significance levels of 0.10, 0.05, and 0.01, respectively.

	Conciseness	Uncertainty	Specificity	Repetition	Bullets
A. Whole prospectus	- neg	+ pos	- neg	- neg	+ pos
B. Sub-section					
Prospectus summary				- neg	- neg
Risk Factors		inverted U-shaped	- neg		+ pos
Use of Proceeds			- neg		+ pos
MD&A		+ pos		- neg	

#### Table 7Predictions between PEM dimensions and ex ante uncertainty

This table summarises the predictions on the association between ex ante uncertainty and individual plain English compliance across the four sub-sections of offering prospectuses. + pos (- neg) denotes a predicted positive (negative) relation. The inverted U-shaped indicates a positive association up to a point, after which the relationship becomes negative.

Variables	Mean	Std. Dev.	Q1	Median	Q3
Conciseness					
Whole document	-0.36	3.25	-0.60	-0.29	0.45
Prospectus Summary	-0.46	4.62	-0.73	-0.38	0.41
Risk Factors	-0.14	2.19	-0.61	-0.34	0.33
Use of Proceeds	-0.41	4.15	-0.96	-0.51	0.57
MD&A	-0.21	3.65	-0.79	-0.36	0.52
Uncertainty					
Whole document	3.17	0.47	2.83	3.13	3.49
Prospectus Summary	2.17	0.82	1.59	2.06	2.62
<b>Risk Factors</b>	6.77	0.73	6.35	6.84	7.28
Use of Proceeds	2.72	0.94	2.15	2.74	3.31
MD&A	2.42	0.65	2.03	2.31	2.65
Specificity					
Whole document	3.73	0.75	3.17	3.59	4.19
Prospectus Summary	4.32	1.48	3.24	4.12	5.22
<b>Risk Factors</b>	1.55	0.55	1.17	1.46	1.81
Use of Proceeds	3.06	1.85	1.78	2.52	3.95
MD&A	5.69	1.64	4.54	5.52	6.69
Repetition					
Whole document	26.14	8.90	19.90	24.60	30.72
Prospectus Summary	0.75	1.19	0	0	1.41
<b>Risk Factors</b>	2.13	1.78	1	2	3.16
Use of Proceeds	0.10	0.38	0	0	0
MD&A	5.83	3.35	3.74	5.29	7.48
Bullets					
Whole document	286.67	114.63	202	276	366
Prospectus Summary	18.14	12.81	9	15	25
<b>Risk Factors</b>	91.47	56.92	50	77	116
Use of Proceeds	1.36	2.01	0	0	3
MD&A	21.94	18.02	8	19	32
PEIndex					
Whole document	0	2.88	-1.94	-0.04	1.91
Prospectus Summary	0	2.28	-1.39	0.13	1.44
<b>Risk Factors</b>	0	2.46	-1.68	-0.36	1.28
Use of Proceeds	0	2.39	-1.51	0.02	1.50
MD&A	0	2.66	-1.60	0.23	1.73

# Table 8 Summary Statistics – Sub-sections

The sample includes 1,377 completed U.S. IPOs from 2002 to 2019. All variables are defined in Appendix C. All continuous variables are winsorised at the 1st and 99th percentile.

	Dependent variable = TotalVol					
	Prospectus Summary	Ris	Risk Factors Use of Proceeds		MD&A	
Conciseness	-0.002	0.005		0.002	-0.010	
	(-0.27)	(0.38)		(0.33)	(-1.10)	
Uncertainty	0.033	0.032	1.657**	0.074	-0.043	
	(0.51)	(0.55)	(2.34)	(1.64)	(-0.81)	
Uncertainty_sqr			-0.123**			
			(-2.25)			
Specificity	-0.025	0.008		-0.067***	-0.056	
	(-1.02)	(0.12)		(-3.48)	(-1.69)	
Ln(Repetition)	-0.160**	-0.042		-0.143	0.097	
	(-2.57)	(-0.83)		(-1.23)	(0.86)	
Bullets	-0.008**	0.003***		0.037*	0.002	
	(-2.72)	(3.54)		(1.85)	(0.81)	
PEIndex	-0.010	-0.014		-0.047**	-0.010	
	(-0.52)	(-0.75)		(-2.73)	(-0.65)	
Control variables	Yes	Yes	Yes	Yes	Yes	

Table 9PEM compliance and ex ante uncertainty by sub-section

The sample includes 1,377 completed U.S. IPOs from 2002 to 2019. **TotalVol** is the standard deviation of daily equity returns over the window of 60 trading days beginning the 5<sup>th</sup> day after the IPO. Each independent variable is regressed in separate models with the coefficients being tabulated. For brevity, control variables are not reported but their statistical significance remains consistent with the main analyses. All variables are defined in Appendix C. All regressions include fixed effects associated with Fama and French (1997) 12-industry classification and IPO year. All continuous variables are winsorised at the 1<sup>st</sup> and 99<sup>th</sup> percentile. The t-statistics are in parentheses with the robust standard errors clustered by Form S-1 filing year. \*, \*\*, and \*\*\* represent significance levels of 0.10, 0.05, and 0.01, respectively.

#### Table 10Bog Index and ex ante uncertainty

Variables	Mean	Standard Deviation	P1	P25	Median	P75	P99
Bog	92.91	7.52	78	87	92	99	110
Sentence Bog	21.70	3.32	15	19	21.90	24	30.98
Word Bog	80.80	5.64	69	77	80	85	93.24
Pep	9.61	0.99	8	9	9	10	12

Panel A: Descriptive statistics

Bog is the Bog Index, which is reported by Editor Software's StyleWriter 4 and provides a measure of several plain English principles. Sentence Bog, Word Bog, and Pep are the three components of the Bog Index, and defined in section 4.1.

Panel B: Multivariate analysis

	Exp		<b>Dependent variable = TotalVol</b>				
	Sign	(1)	(2)	(3)	(4)	(5)	
Bog	+	0.003				-0.405***	
		(0.37)				(-3.92)	
Sentence Bog	+		0.022				
			(1.26)				
Word Bog	+			0.007			
				(0.68)			
Pep	-				0.133***		
					(3.87)		
Bog_sqr	-					0.002***	
						(3.91)	
Log(words)	-	-1.295***	-1.434***	-1.316***	-1.500***	-1.322***	
		(-3.49)	(-3.79)	(-3.77)	(-4.54)	(-3.59)	
Control variables		Yes	Yes	Yes	Yes	Yes	
Adjusted R-squared		0.321	0.322	0.322	0.327	0.327	
Turning point						93.065	
$-\beta_1/(2\beta_2)$						(59 <sup>th</sup> percentile)	
95% Fieller interval						[88.49; 97.95]	

The sample includes 1,377 completed U.S. IPOs from 2002 to 2019. **TotalVol** is the standard deviation of daily equity returns over the window of 60 trading days beginning the 5<sup>th</sup> day after the IPO. All variables are defined in Appendix C. For brevity, control variables are not reported but their statistical significance remains consistent with the main analyses. All regressions include fixed effects associated with Fama and French (1997) 12-industry classification and IPO year. All continuous variables are winsorised at the 1<sup>st</sup> and 99<sup>th</sup> percentile. The t-statistics are in parentheses with the robust standard errors clustered by Form S-1 filing year. \*, \*\*, and \*\*\* represent significance levels of 0.10, 0.05, and 0.01, respectively.

# Appendix A Evidence of the SEC's enforcement on plain English attributes

To provide confirmatory evidence on how the SEC oversees and enforces compliance with plain English constructs in IPO prospectuses, we conduct a keyword search of the plain English requirements in a random sample of the SEC's comment letters from 2005 (the first year these letters became available) to 2019. All the comment letters are available on the SEC's website.

# > Specificity

In the comment letter issued on November 7, 2017, the SEC required Dropbox Inc. to disclose the following information in its offering prospectus (*emphases added*):

- The *number* of active registered users, unique registered users, and unique paying users for each period in its "Prospectus Summary" section. (p. 1)
- The retention *rates* regarding the paying user base and the customer base for each period in its "Key Business Metrics" section. (p. 2)
- The *amount* of cash held overseas in order to provide an understanding of your sources of cash and the portion of total cash and cash equivalent that are not currently available to fund domestic operations without incurring taxes upon repatriation in its MD&A's section. (p. 5)
- The *name* of the court in which Synchronoss Technologies, Inc. has brought a claim against you for patent infringement in its MD&A's section. (p. 5)

## Uncertainty

in its Business and Risk factors sections of the initial prospectus, TriNet Group, Inc. claimed that its "in-house approach" presents a "significant opportunity...to penetrate and expand [the company's] presence in the SMB market" and later "[the company's] success *depends* on growth in market acceptance of the human resources outsourcing and related services we provide" (*emphasis added*). In the comment letter issued on December 19, 2013, the SEC required the company to further elaborate on reasonable opportunities for growth in its services and the associated challenges to such growth.

# Repetition

In the comment letter in December 2018, the SEC required Uber Technologies Inc. to revise its "Overview" section in the MD&A because "this section substantially repeats the overview you present in the Summary and Business sections" (SEC, p. 2).

## Appendix B Conciseness – examples from the plain English handbook

The following examples are extracted from the plain English handbook (SEC, 1998a) with *[emphasis added]* 

- IPO issuers are strongly recommended to use *short sentences* to communicate in their offering prospectuses because "lengthy and information-packaged sentences choke many prospectuses today" (p. 28).
- The use of *active voice* and *strong verbs* makes sentences "shorter and easier to understand" (p. 20).
- Legalese and highly technical business terminology (i.e., abstract words) should be replaced with concrete and everyday words (p. 20, pp. 30-31). Using concrete and everyday language is to avoid legalese such as "subject to the procedures described herein", "pursuant to the terms and conditions set forth" and "effected largely pursuant to transactions". In a similar vein, Rule 421(b) recommends IPO issuers avoid having complex information copied directly from legal documents without any clear and concise explanation of the provision(s). Although the SEC does not provide further guidance on this writing convention, it is reasonable to expect that financial disclosures that are buried in legal jargon and obtuse language would be comprised of *long sentences* and hard to understand.
- Good writing should avoid *superfluous words*. Superfluous words are those that can be replaced with fewer words that mean the same thing (p. 25). For example, "owing to the fact that" should be replaced with "because" or "since".
- "Sentences will be shorter and easier to understand" if a negative phrase (i.e., *multiple negatives*) is replaced with a single word that means the same thing (p. 27).
- The use *of personal pronouns* enables writers to "use more concrete and everyday language" and "keep the sentences short" (p. 22).
- In most cases, *defined terms* should be avoided. For example, "where acronyms, such as REIT, are widely understood to the investing public, they can safely be used without creating confusion" (p. 31). In rare instances in which defined terms are necessary, the choice of an intuitive and logical term "reduces the number of new words and phrases the reader needs to memorize to understand the document" (p. 31).

Variable	Definition
Dependent varia	ables
TotalVol	The standard deviation of daily equity returns over the window of 60 trading days beginning the 5 <sup>th</sup> day after the IPO.
Independent va	riables
Conciseness	The inverse of an aggregate measure of normalised values of five plain English style issues (i.e., passive verbs, hidden verbs, legal words, wordy phrases, personal pronouns). The normalised values of all components except that of personal pronouns are negatively signed in the summation.
Uncertainty	The number of uncertain words in the S-1 filings relative to total number of words where the number of uncertain words is from an aggregate wordlist combining the uncertain, weak modal, and negative word lists in Loughran and McDonald (2011).
Specificity	The provision of higher level of detail given firms decide to disclose a particular information (e.g., number of specific entity names, names of persons, locations, and organizations; quantitative values in percentages; money values in dollars; times; and dates, all scaled by the total number of words) (Hope et al., 2016).
Repetition	The number of repetitive sentences in an offering prospectus with a repetitive sentence identified as repetitive if its cosine similarity is equal or greater than 0.9 (Merkley, 2014).
Bullets	The number of bullets in an initial offering prospectus.
PEIndex	An aggregate measure of five plain English dimensions, being equal to the sum of standardised values of ( <i>Conciseness – Uncertainty +</i> Specificity + Repetition – Bullets)
Bog	A measure of disclosure readability created by Editor Software's plain English software, <i>StyleWriter</i> . The formula is based on several plain English writing problems identified in the SEC's plain English handbook (Bonsall et al., 2017).
Control variabl	es
Age	The difference between IPO year and the firm's founding date where founding dates are obtained from the Field-Ritter dataset (Loughran & Ritter, 2004).
Audit	Indicator variable that equals 1 if the firm's auditor was one of the "Big 4" accounting firms (Beatty, 1989).
Days	Number of days between filing date (S-1 filing) and issue date.
Dro mist rat	Value weighted of Nasdag for the 15 trading days prior to the IDO date

Appendix C V	ariable	Definiti	ons
--------------	---------	----------	-----

Pre mkt ret Value-weighted of Nasdaq for the 15-trading days prior to the IPO date, ending on day t-1 (Loughran & McDonald, 2013). This measure is expressed in percentage.

Variable	Definition
VC	A dummy variable set to one if an IPO is backed by venture capitalists, else zero (Megginson &Weiss, 1991).
Nasdaq	A dummy variable equal to one if an IPO is listed on the Nasdaq exchange, else zero
Share overhang	Defined as the number of shares retained by the pre-IPO shareholders divided by the number of shares issued in the initial offering.
HighTech	A dummy variable equal to 1 if a firm resides in a technology industry as identified in Loughran and Ritter (2004, p. 35).
Top-tier dummy	A dummy variable set to one if the lead underwriter (or joint underwriter) of the IPO has an updated Carter and Master (1990) rank of eight or more, else zero. For the period in which the underwriters' rankings are missing, an average of their rankings in prior period is used.
Positive EPS	A dummy variable set to one if the IPO has positive earnings per share (EPS) in the 12 months prior to going public, else zero.
Words	The number of words in an initial offering prospectus (S-1 filing). The wordlist used for word counts is the LoughranMcDonald Master Dictionary in Loughran and McDonald (2011).
HighTech	A dummy variable set to one if the IPO firm is a technology company based on the Loughran and Ritter (2004) classification.