

# **Linguistic Complexity in Firm Disclosures: Obfuscation or Information?**

Brian J. Bushee  
The Wharton School  
University of Pennsylvania  
[bushee@wharton.upenn.edu](mailto:bushee@wharton.upenn.edu)

Ian D. Gow  
Harvard Business School  
[igow@hbs.edu](mailto:igow@hbs.edu)

Daniel J. Taylor  
The Wharton School  
University of Pennsylvania  
[dtayl@wharton.upenn.edu](mailto:dtayl@wharton.upenn.edu)

November 2016

---

We thank Lauren Cohen, Mirko Heinle, Feng Li, Tim Loughran, Rabih Moussawi, Luis Palacios, Robert Verrecchia, Yuxing Yan, and seminar participants at Cubist, Florida International, George Washington, UC-Irvine, the University of Minnesota Empirical Accounting Conference, and University of Toronto for useful discussions, comments, and programming advice. We gratefully acknowledge the financial support of the Harvard Business School and of the Wharton School.

# Linguistic Complexity in Firm Disclosures: Obfuscation or Information?

## Abstract:

Prior research generally interprets complex language in firms' disclosures as indicative of managerial obfuscation. However, complex language can also reflect the provision of complex information, e.g., informative technical disclosure. As a consequence, linguistic complexity commingles two latent components—obfuscation and information—that are related to information asymmetry in opposite directions. We develop a novel empirical approach to estimate these two latent components within the context of quarterly earnings conference calls. We validate our estimates of these two latent components by examining their relation to information asymmetry following the call. Consistent with our predictions, we find that our estimate of the information component of managerial linguistic complexity is *negatively* associated with information asymmetry and our estimate of the obfuscation component is *positively* associated with information asymmetry. Our findings suggest that future research on linguistic complexity can construct more powerful tests by separately examining these two latent components of linguistic complexity.

*JEL Classification:* D82; G14; G18; M41; M45

*Keywords:* Disclosure; Linguistic complexity; Conference calls; Information asymmetry;

## 1. Introduction

An extensive prior literature examines the relation between the linguistic complexity of corporate disclosures and various firm outcomes. This literature generally interprets the use of complex language as an intentional choice by managers to obfuscate in order to increase information processing costs and delay the market reaction to news (e.g., Li [2008]). However, an alternative view is that complex language could be necessary to convey information about the firm's business transactions and operating strategy (e.g., Bloomfield [2008]). In reviewing the literature on linguistic complexity, Loughran and McDonald [2016] conclude, "researchers face the problem of separating the business and the document. These issues are intertwined because the document attempts to describe the economic reality of the business" (p. 1198).

In this paper, we show that linguistic complexity in firm disclosures commingles two latent components that have opposite relations to information asymmetry. One component represents the innate amount of information provided in the disclosure in the absence of obfuscation; e.g., linguistic complexity attributable to informative technical disclosure about the business. The other component represents obfuscation, which is intended to reduce the informativeness of the disclosure. As a result of this commingling, the relation between overall linguistic complexity and information asymmetry is theoretically and empirically ambiguous. However, we show that the relations between the two latent components of linguistic complexity and information asymmetry are theoretically and empirically unambiguous; i.e., the information (obfuscation) component of linguistic complexity is negatively (positively) associated with information asymmetry.

We develop a novel empirical approach to estimate the latent components of linguistic complexity within the context of quarterly earnings conference calls. We provide a simple

information structure to develop the intuition for our estimation strategy. The key to our approach is to identify the portion of the linguistic complexity that represents managers' appropriate disclosure response to business complexity in the absence of obfuscation. We exploit the fact that conference calls allow us to observe the linguistic complexity of an outside party—security analysts—in the same setting as the manager's linguistic complexity. Prior research finds that analysts have incentives to acquire and convey value-relevant information (Mayew [2008], Twedt and Rees [2012]), suggesting that they are unlikely to have obfuscation incentives during the call. As such, we use analysts' linguistic complexity as a benchmark for assessing the required level of linguistic complexity necessary to understand the firm in the absence of obfuscation. We also use a number of business complexity variables to incorporate any aspects of economic complexity not evident in analysts' language. Finally, because conference calls are a mix of both a scripted presentation and a more spontaneous response to questions, we examine linguistic complexity separately for the presentation and the response portions of the call.

Following an extensive prior literature on the complexity of both written and spoken language, we measure linguistic complexity using the Gunning [1952] Fog index. We decompose managerial linguistic complexity into the portion correlated with analyst linguistic complexity and business complexity (i.e. the information component) and the portion uncorrelated with analyst linguistic complexity and business complexity (i.e., the obfuscation component). We then validate our estimates of the latent components of linguistic complexity by testing whether the information (obfuscation) component of linguistic complexity is negatively (positively) associated with information asymmetry.

We first adopt the approach of prior work and examine the association between total linguistic complexity and information asymmetry, which we measure using illiquidity (Amihud

[1992]). We find a positive relation between linguistic complexity in the presentation part of the conference call and information asymmetry, consistent with prior work on 10-K disclosures that finds that higher linguistic complexity reflects obfuscation. However, in contrast to this work, we find a negative relation between linguistic complexity in the response part of the call and information asymmetry. While these results suggest the informativeness of complex language differs between scripted and unscripted language, an analysis of the latent components of linguistic complexity suggests a more nuanced conclusion.

When we decompose linguistic complexity into its two latent components, we find a negative (positive) relation between the estimated information (obfuscation) component of linguistic complexity and information asymmetry. These findings are similar for both the presentation and response. These results illustrate that the relation between overall linguistic complexity and information asymmetry is ambiguous because complex language is driven by two components that have opposite relations to information asymmetry. In contrast, the relation between the components themselves and information asymmetry is unambiguous.

These results are robust to a variety of sensitivity tests. First, the results continue to hold (and in some cases become stronger) when we control for additional language attributes on the conference call, such as number of words, industry jargon, forward-looking statements, and tone. Second, the results are robust to controls for information asymmetry prior to the conference call and to a estimating a within-industry, within-firm, and within-manager analysis, suggesting time-series variation in these components explains time-series variation in the firm's information asymmetry and that our results are not attributable to "manager style." Third, the results hold under three Monte Carlo placebo tests that use randomly selected non-event days, alternative conference call days, and language from another firm's conference call. Fourth, we further

validate our measures by showing that our estimate of the information (obfuscation) component of managers' linguistic complexity is positively (negatively) associated with the likelihood that managers provide quantitative earnings guidance on the call, as would be expected if managers act consistently in their voluntary disclosure decisions.

Finally, we examine our decomposition for loss firms, which prior work suggests use complex language to obfuscate (Li [2008]). However, Bloomfield [2008] speculates that this result could be driven by loss firms providing more information to explain the poor performance. We find that loss firms have significantly greater values of both information and obfuscation components during conference calls. While it might be more satisfying to attribute the complex language to either greater obfuscation or greater information, our results suggest both forces are at work. In the absence of obfuscation, the information provided by loss firms is higher. However, managers in loss firms also obfuscate to a greater extent to mask their performance.

Collectively, we find robust evidence that identifying the latent obfuscation and information components of linguistic complexity provides more powerful tests of the relation between linguistic complexity and capital market outcomes.

Our paper contributes to the large stream of literature that has examined the relation between the linguistic complexity of corporate disclosures and firm outcomes. Most prior literature focuses on 10-K filings and interprets complex language as obfuscation (see Loughran and McDonald [2016] for a review). Recent papers by Dyer, Lang, and Stice-Lawrence [2016] and Guay, Samuels, and Taylor [2016] suggest the linguistic complexity of the average 10-K might instead be driven by regulations and accounting standards, rather than intentional managerial actions. However, prior work does not examine the possibility that complex language could be informative. We extend this literature by adopting a measurement approach to show

that linguistic complexity has both latent information and obfuscation components that can be identified through the complex language of an external party (analysts). Our analysis illustrates that these latent components have countervailing effects on information asymmetry, such that the absence of an “on average” relation between linguistic complexity and a particular firm outcome should be interpreted with caution (e.g., Loughran and McDonald [2014]; Hoberg and Lewis [2015]). In this regard, we provide an empirical estimation strategy that future researchers can use to improve construct validity by identifying the latent obfuscation and information components of complex language.

Our paper also contributes to work that uses empirical models to identify discretionary reporting decisions. Leuz and Wysocki [2016] point out that many of these models, such as discretionary accruals, suffer from an inability to cleanly separate reporting decisions from economics. Our approach has a number of important differences over these prior models. First, our decomposition does not solely rely on firm characteristics that could proxy for both economics and reporting incentives. Rather, our identification strategy relies on using the linguistic complexity of analysts, who are outside of the firm’s financial reporting system and have little or no incentive to obfuscate in their language. Second, we have a clear setting for verifying that our model works because of the strong theoretical (and empirical) links between disclosure and information asymmetry in prior work. We expect to decompose a single variable—whose relation to information asymmetry is ambiguous—into two separate components with opposite sign predictions. This setting is in stark contrast to accrual models, where researchers rarely have opposite sign predictions for normal and abnormal accruals. Evidence of opposite directional effects for the two components of linguistic complexity should mitigate concerns that our estimates are biased in any one direction. For example, if our estimate of the information

component contains some obfuscation, it would bias the negative (positive) coefficient on the information (obfuscation) component upward (downward); and hence bias against finding opposite directional effects. While any empirical estimate of an unobserved latent variable will have some amount of noise and bias, our findings provide verification that the noise and bias is sufficiently mild that our estimates are empirical descriptive of the underlying latent components. We encourage future research using our approach to examine both components of linguistic complexity and to test for opposite sign predictions.

The remainder of the paper proceeds as follows. Section 2 reviews the relevant prior literature. Section 3 describes our research design. Section 4 presents the sample and descriptive statistics. Section 5 presents our findings, and Section 6 concludes.

## **2. Prior Literature**

Regulators have often expressed concern that the language in firms' disclosures has become increasingly complex (see Loughran and McDonald [2014]). To the extent that complex language increases information processing costs for investors, managers can use complex language to "obfuscate" the true nature of the firms' current and future performance (Li [2008]). Li [2008] uses the Gunning [1952] Fog Index to measure the "linguistic complexity" of firms' 10-K filings.<sup>1</sup> He finds that firms with higher 10-K Fog have lower current earnings performance and less persistent future earnings performance, consistent with managers using linguistic complexity to obfuscate poor performance.

Since Li [2008], a growing literature uses Fog to examine the relation between linguistic complexity of mandatory disclosures (e.g., 10-Qs or 10-Ks) and a variety of firm outcomes, such

---

<sup>1</sup> Li [2008] uses "readability" and "linguistic complexity" interchangeably to describe the construct that Fog measures. Throughout the paper, we use the latter term because it applies to both spoken and written words.



as investment efficiency, bid-ask spreads, delayed pricing of accounting information, voluntary disclosure, and short-sale constraints (see Loughran and McDonald [2016] for a review). Prior work also examines the relation between Fog and the decisions of external parties, such as trading of unsophisticated investors (Miller [2010]; Lawrence [2013]), analysts' outputs (Lehavy, Li, and Merkley [2012]; Bozanic and Thevenot [2015]), and disagreement among credit rating agencies (Bonsall and Miller [2016]).<sup>2</sup>

Despite the widespread use of the Fog index, there is an underlying concern about its construct validity because Fog potentially commingles informative text with uninformative text. Lang and Stice-Lawrence [2015] concisely summarize this concern:

*“Prior research provides evidence that Fog may be linked to managerial obfuscation (Li 2008) and higher processing costs (Miller 2010), but it is affected by multisyllabic words which can be associated with informative technical disclosure.”*

Loughran and McDonald [2014] find mixed evidence on the relation between the Fog of 10-Ks and the information environment, but stronger evidence using electronic file size of 10-Ks. Bonsall, Leone, and Miller [2015] suggest that file size results are driven by pictures and other graphical material that are unrelated to linguistic complexity.<sup>3</sup> In a survey piece, Loughran and McDonald [2016] offer a skeptical view on the construct validity of the Fog index.

We extend this literature in three ways. First, we show that the construct validity of linguistic complexity measures like Fog can be substantially improved. We recognize that linguistic complexity commingles two latent components that affect information in opposite directions. One component represents the innate amount of information provided in the

---

<sup>2</sup> While the vast majority of studies focus on the Fog of firms' disclosures, examples of studies examining Fog of disclosures by external parties include De Franco et al. [2012], who examine the Fog of sell-side analyst reports, and Dougal et al. [2012], who examine the Fog of a firm's media coverage.

<sup>3</sup> Bonsall, Leone, and Miller [2015] suggest an alternative proprietary measure developed by *StyleWriter*: the Bog Index. They note that the Bog Index takes grammatical features of written text into account that are omitted from the Fog index; e.g., passive voice. For this reason, it is less likely to be applicable to spoken language.

disclosure in the absence of obfuscation; e.g., Fog attributable to informative technical disclosure. The other component represents obfuscation, which is intended to reduce the informativeness of the disclosure. We develop an empirical approach for estimating these two latent components, and show that the information component is negatively associated with information asymmetry and the obfuscation component is positively associated with information asymmetry. Thus, rather than dismiss measures of linguistic complexity due their mixed relation with information quality on average, we show that the construct validity of Fog can be improved by decomposing the measure into its latent components.

Second, we focus on the use of complex language in conference calls, rather than in 10-Qs or 10-Ks.<sup>4</sup> Conference calls have become a prevalent disclosure medium that are informative to market participants and lead to reductions in information asymmetry (e.g., Frankel, Johnson, and Skinner [1999]; Bowen, Davis, and Matsumoto [2002]; Brown, Hillegeist, and Lo [2004]). Conference calls also provide a disclosure medium in which both a scripted presentation and a more spontaneous management response exist in the same setting, answering Bloomfield's [2008] call to examine complexity in more spontaneous disclosures (see also Larcker and Zakolyukina [2012]).<sup>5</sup> Recent papers by Brochet, Naranjo, and Yu [2014], Burgoon et al., [2014], Brochet, Loumioti, and Serafeim [2015], and Allee and DeAngelis [2015] also use Fog to measure the linguistic complexity of conference calls, but their focus is mainly on Fog as a mechanism to obfuscate information, through language barriers, short-termism, or incentives to reduce transparency.

---

<sup>4</sup> Early research suggests that “readability” measures, such as Fog and the Fleisch-Kinkaid index, also capture “listenability” (Allen [1952]; Harwood [1955], Fang [1967]). Notably, Harwood [1955] concludes that more “difficult” text was easier to comprehend when read than when heard.

<sup>5</sup> Lee [2015] shows that firms are more likely to “script” their responses to questions when performance is poor. Thus, a manager’s response to questions is not completely spontaneous, but should be relatively more so than the presentation portion or a 10-K. For example, Larcker and Zakolyukina [2012] find that some managers use swear words on conference calls, which would be very unlikely to appear in a vetted document.

Conference calls also arguably provide stronger managerial incentives to obfuscate than the 10-K setting. Li [2008] argues that the benefit of obfuscation is that “by increasing the processing cost of adverse information, managers hope that it is not reflected in stock prices or in prices with a delay” (p. 224). In a conference call, analysts and investors have 30-60 minutes of interactive discussion with managers on the earnings announcement date. The benefits of delaying the processing of adverse information in this setting are substantial: obfuscation could prevent follow-up questions on bad news and could delay market reaction to bad news until after the call, when there will be less focus on the company. Bushee, Matsumoto, and Miller [2003] report that there is high individual investor attention on the company during the call and that, prior to Reg FD, a commonly expressed reason for restricting individual investor access to calls was managers’ concerns that they would overreact to information released during the call. Thus, the higher degree of timeliness and visibility of the conference call compared to the 10-K release provide a powerful setting for examining obfuscation incentives.<sup>6</sup>

Third, we exploit the fact that conference calls include an outside party that lacks clear incentives to obfuscate—security analysts—to identify the extent to which managerial linguistic complexity during the same call reflects informative technical disclosure rather than obfuscation. Matsumoto, Pronk, and Roelofsen [2011] shows that the response portion of the call is associated with larger stock return reactions than the presentation portion, with greater analyst coverage enhancing the information content of the response portion. Mayew, Sharp, and Venkatachalam [2012] finds that analysts with superior private information are more likely to participate in

---

<sup>6</sup> The notion that obfuscation incentives are present in conference calls is not unique to our paper. For example, Larcker and Zakolyukina [2012, p. 499] state: “In contrast to prior studies on deception which use the text in written disclosures such as 10-Ks, we use arguably more spontaneous disclosures of conference calls. There are a number of limitations to using formal disclosures such as 10-Ks and 10-Qs in deception studies. First, formal disclosures are more scripted and prior research has shown that their content does not change much over time. Second, the different parts of the reports are written and edited by different individuals and these individuals are unlikely to be executives. Finally, these disclosures lack spontaneity that characterizes conference calls.”

conference calls. Thus, the evidence suggests that analyst participation in conference calls improves their information content. While it is possible that analysts also obfuscate (or are affected by managerial obfuscation), we show theoretically that this would work against our results by biasing the negative (positive) coefficient on the information (obfuscation) component upward (downward). We also empirically validate that any analyst obfuscation is sufficiently mild that our estimates are empirically descriptive of the underlying latent components.

### 3. Research Design

#### 3.1. Empirical estimation strategy

In this section, we provide a simple, formal framework for our estimation strategy and associated empirical tests.

##### 3.1.1. Representation of information and obfuscation

We conjecture that a manager’s linguistic complexity, which we represent as  $LC\_Manager$ , is determined by two countervailing forces. First, complex language is related to the intrinsic amount of informative technical disclosure that would be provided on the conference call by a benevolent manager; i.e., the information content of manager’s language in the *absence* of obfuscation (e.g., Bloomfield [2008]). We denote this component as  $Info^*$  (where the  $*$  indicates an unobservable variable). Second, complex language can also result from intentional obfuscation, or an intentional reduction in informative language (e.g., Li [2008]), which we denote  $Obfu^*$ . Thus, complex language can be thought of as an amalgamation of informative technical disclosure and potential obfuscation. Formally, we represent linguistic complexity of the managers as:

$$LC\_Manager = \varphi_0 + \varphi_1 Info^* + \varphi_2 Obfu^* + \varepsilon \quad (1)$$

where  $\varphi_1 > 0$  and  $\varphi_2 > 0$ .

The notion that managerial linguistic complexity is increasing in the intrinsic amount of technical information provided on a conference call ( $\varphi_1 > 0$ ) is supported by prior evidence that long sentences and complex words—two inputs to common measures of linguistic complexity (e.g., the Fog index)—do not necessarily indicate obfuscation. First, complex words are commonly used in financial contexts to convey informative technical details without any loss of comprehension. As Loughran and McDonald [2014] point out, “One of the longest words occurring with reasonable frequency in 10-Ks is ‘telecommunications’, a word not likely to force most readers to consult their dictionaries” (p. 1645). Second, longer sentences can convey more information because they better show the relationships among various topics than short sentences (Redish and Selzer [1985]).<sup>7</sup> Thus, in the absence of obfuscation incentives, it is reasonable to assume that longer sentences or complex words are necessary for better comprehension in certain firms or in certain settings.<sup>8</sup>

Prior work often asserts that complex language is unambiguously less informative than simple language. The important implication of equation (1) is that *LC\_Manager* commingles two latent (unobserved) components that affect the information content of manager’s language in opposite directions. Consequently, the theoretical relation between linguistic complexity and informativeness is ambiguous. Importantly, these issues apply to the construct of linguistic complexity generally—complex language increases both with informative technical disclosure and with obfuscation—and are not specific to any one empirical measure of the construct (e.g.,

---

<sup>7</sup> For example, Charrow and Charrow [1979] find that adding more words to highlight the relationships among information items in jury instructions leads to a decrease in readability scores (due to the longer sentences) but an increase in comprehension by jurors. Long sentences also often represent changes in “discourse plans,” where something the speaker says triggers related thoughts, causing the speaker to keep going with the sentence to add the additional information (Deese [1978]).

<sup>8</sup> This idea is similar to the reporting incentives literature, which finds that changes in firm circumstances or regulatory/legal institutions affect managerial incentives to provide more transparent disclosures (see Leuz and Wysocki [2016] for a summary of this literature).

the Fog index). This ambiguity could potentially explain the mixed empirical results that relate measures of linguistic complexity to measures of information content (e.g., Loughran and McDonald [2014]) and to fraudulent behavior (e.g., Hoberg and Lewis [2015]).

### 3.1.2. Identifying assumptions

We seek to recover empirical estimates of the latent variables,  $Info^*$  and  $Obfu^*$ . One unique feature of conference calls is that we can observe the linguistic complexity of an external party—security analysts—in the same disclosure setting as managers’ linguistic complexity. Prior research suggests analysts have incentives to acquire and convey value-relevant information during conference calls (Mayew [2008], Twedt and Rees [2012]). While managers may seek to obfuscate information provided on the call, analysts should not seek to do so. Accordingly, we assume that the analyst’s linguistic complexity on the call, which we represent as  $LC\_Analyst$ , does not reflect intentional obfuscation, but does reflect (at least to some extent) the intrinsic amount of informative technical disclosure on the call. Formally,

$$LC\_Analyst = \delta_0 + \delta_1 Info^* + v \tag{2}$$

where  $\delta_1 > 0$ . We expect  $\delta_1 > 0$  for the reasons cited earlier: complex words and longer sentences can increase comprehension and information content in a business context (Deese [1978], Charrow and Charrow [1979], Redish and Selzer [1985], Loughran and McDonald [2014]).

Under the identifying assumption that the linguistic complexity of analysts does not reflect intentional obfuscation, we can use the linguistic complexity of analysts on the call as a potential benchmark level of linguistic complexity that one would expect in the absence of obfuscation. While the assumption that the linguistic complexity of analysts is not affected by obfuscation may seem strong, in Appendix A we show that our empirical estimation strategy will remain valid as long as the effect of obfuscation on analysts is not too large (see Appendix A for

the precise condition). Ultimately, this is an empirical question, which we address in our validation tests.

### 3.1.3. Estimation Strategy

In addition to using linguistic complexity of analysts as a potential benchmark, our estimation strategy also uses a number of firm characteristics related to the economic complexity of the firm and its business transactions as a benchmark for the level of complexity necessary to communicate information (e.g., Barth, Kasznik, and McNichols [2001]; Chen, Cheng, and Lo [2010]). For example, prior work suggests that managers of complex firms provide more information on conference calls (e.g., Tasker [1998]) and more voluntary disclosure (e.g., Guay, et al., [2016]). Thus, while business complexity on its own should be positively associated with information asymmetry, the appropriate managerial disclosure response to such complexity (i.e., provision of additional information) should be negatively associated with information asymmetry. While adding business complexity variables allows the model to incorporate any aspects of business complexity that are not evident in analysts' language, in subsequent analyses, we find such additional characteristics are not necessary to obtain valid estimates of the two components.

We begin by regressing our measures of *LC\_Manager* on *LC\_Analyst* and the business complexity variables:

$$LC\_Manager = \beta_0 + \beta_1 LC\_Analyst + \sum \beta_i Business\ Complexity\ Variables + \eta \quad (3)$$

Using the estimated coefficients from this regression ( $b_0, b_1 \dots$ ), we decompose the linguistic complexity of managers into two components—a fitted value (i.e., the portion of managers' linguistic complexity correlated with analyst complexity and business complexity) and a residual value (i.e., the portion of managers' linguistic complexity uncorrelated with analyst complexity

and business complexity). The fitted value,  $Info$ , serves as our estimate of the latent information component ( $Info^*$ ) and the estimated residual value,  $Obfu$ , serves as our estimate of the latent obfuscation component ( $Obfu^*$ ).

#### 3.1.4. Validation

To validate our estimates of the two latent components, we rely on economic theory that suggests that obfuscatory (informative) financial disclosure is associated with greater (lower) information asymmetry between shareholders and managers,  $InfoAsym$  (e.g., Bloomfield [2002]). Formally,

$$InfoAsym = \gamma_0 + \gamma_1 Info^* + \gamma_2 Obfu^* + \eta \quad (4)$$

where  $\gamma_2 > 0$  and  $\gamma_1 < 0$ . Note that equation (4) expresses information asymmetry as a function of the *unobserved* latent components of linguistic complexity.

An important feature of our empirical validation is that, while  $LC\_Complexity$  is increasing in both latent components, the two components affect information asymmetry in opposite directions (see Figure 1). Hence, if our decomposition is empirically descriptive, we expect to decompose a single variable—whose relation to information asymmetry is ambiguous—into two separate components, one that is positively related to information asymmetry and one that is negatively related to information asymmetry. We leverage these differential predictions to validate our decomposition. Specifically, we replace the unobservable latent components in (4) with our empirical estimates of these components and estimate:

$$InfoAsym = \alpha_0 + \alpha_1 Info + \alpha_2 Obfu + \zeta \quad (5)$$

Appendix A solves for the coefficients  $\alpha_1$  and  $\alpha_2$  under the assumed information structure and shows: (i) that the coefficient on our estimate of the information component,  $\alpha_1$ , is predicted to be negative regardless of the magnitude of measurement error in our empirical measures, and



(ii) that the coefficient on our estimate of the obfuscation component,  $\alpha_2$ , is predicted to be positive so long as measurement error is not too large.

Appendix A also examines the consequences of a violation of our key identifying assumption; i.e., that the linguistic complexity of analysts does not reflect intentional obfuscation. We show that when the linguistic complexity of analysts is influenced by obfuscation (i.e., a setting where  $LC\_Analyst$  is also a function of  $Obfu^*$ ), the coefficient on  $Info$  ( $Obfu$ ) is biased upward (downward). The intuition for this result is that the predicted value,  $Info$ , would commingle both information and obfuscation, as would the residual value,  $Obfu$ . Because information and obfuscation have opposite effects on information asymmetry, the coefficients will be biased toward zero. Hence, this would bias against finding  $\alpha_1 < 0$  and  $\alpha_2 > 0$ . Our empirical results suggest that such effects are not large enough to invalidate our estimates.

Ultimately, it is an empirical question whether our simple formal framework and estimated decomposition are valid. If our estimates of the latent components are empirically descriptive, we expect a negative relation between the fitted value,  $Info$ , and information asymmetry ( $\alpha_1 < 0$ ) and a positive relation between the residual value,  $Obfu$ , and information asymmetry ( $\alpha_2 > 0$ ). Finding that the two latent components of  $LC\_Manager$  have opposite signs validates the notion that manager's linguistic complexity commingles both an information component and an obfuscation component, and validates our empirical estimation strategy.

### 3.2. Measures of key variables

#### 3.2.1. Information asymmetry

We measure of information asymmetry using the Amihud [2002] illiquidity construct (see, e.g., Lang and Maffett [2011]). Following Amihud [2002], *Illiquidity*, is defined as

$$Illiquidity_t = \frac{|R_t|}{DVOLUME_t} \quad (6)$$

where  $R_t$  is the daily return and  $DVolume_t$  is the daily dollar volume (in millions). The primary advantage of this variable is that it is easily calculated using CRSP data and is available for all firms in our sample.<sup>9</sup> We calculate this measure at the daily level, winsorizing at the 1st and 99th percentile, and then take the average value over the period starting the day of the call and ending twenty-five trading days subsequent to the call. To mitigate any effects of measurement error, we use the percentile ranks of *Illiquidity* in our tests (Balakrishnan, Core, and Verdi [2014]).

### 3.2.2. Linguistic complexity

Following an extensive body of literature discussed in Section 2, we measure linguistic complexity using the Gunning [1952] Fog index. The Fog index measures linguistic complexity as a function of the number of words per sentence and the percent of complex words, where complex words are those words with more than two syllables:

$$Fog(.) = 0.4 \times (\text{average number of words per sentence} + \text{percent of complex words}) \quad (7)$$

To calculate average words per sentence, we deviate from prior literature that follows Li [2008] and uses the *Lingua::EN::Fathom* Perl routine to calculate Fog. The *Fathom* routine can understate the amount of Fog in numerically intensive text because, among other issues, it mistakes decimal points for periods. We use the *Lingua::EN::Sentence* Perl routine that better identifies sentence breaks (see Appendix B for details).

Lower values of  $Fog(.)$  correspond to less complex text. Taken literally, the Fog index is intended to measure the hypothetical years of formal education necessary to comprehend the

---

<sup>9</sup> In Table IA-1 of the Internet Appendix, we estimate our tests with an alternative measure of information asymmetry: the information asymmetry component of the bid-ask spread, or  $\lambda$ . We compute  $\lambda$  using the estimation procedure of Madhavan, Richardson, and Roomans [1997] and Armstrong et al. [2011]. Because this measure requires intra-day data from TAQ, it is available for only approximately 75% of our sample. Our results hold when we use this alternative measure.

material. For example, a Fog index of 12 indicates formal education at the level of a high school senior (i.e., 12th grade), while a Fog index of 16 indicates college education.

We measure managerial Fog separately for the presentation and Q&A portions of the call. *Fog(Present)* represents the Fog index of managers' language during the presentation portion of the call, and *Fog(Response)* is derived from the managers' responses to questions on the call. Our primary variable to identify whether managerial Fog reflects information or obfuscation is *Fog(Analyst)*, which represents the Fog index of analysts' questions and statements during the Q&A portion of the call.

### 3.2.3. Business complexity variables

One challenge in measuring business complexity is that variables that proxy for business complexity could also reflect reporting incentives (Leuz and Wysocki [2016]). Thus, we attempt to choose variables that would explain cross-sectional variation in business complexity in the absence of obfuscation incentives. Moreover, any obfuscation incentives captured by this set of variables would likely be conditional on firm performance (e.g., earnings surprise), which we control for in our information asymmetry tests.<sup>10</sup> The complexity variables we use are firm size (*Size*); firm leverage (*Leverage*); book-to-market ratio (*BM*); historical stock performance (*Returns*); acquisitions (*Acquisitions*), capital intensity (*CapIntensity*), capital expenditures (*Capex*), research and development (*R&D*); debt and equity issuance (*Financing*); cash flow volatility ( $\sigma CFO$ ); goodwill impairments (*Goodwill*) and restructuring charges (*Restructuring*) (see, e.g., Barth, Kasznik, and McNichols [2001]; Chen, Cheng, and Lo [2010]). While adding

---

<sup>10</sup> For example, we use goodwill impairments as a measure of complexity, but they could also affect obfuscation incentives on the margin. Li and Sloan [2015] show that managers tend to delay goodwill impairments; hence, when an impairment is finally recognized, it represents a significant economic event for the company that will increase the complexity in understanding the business. It is likely that the economic complexity associated with the goodwill impairment is larger than the obfuscation incentives, especially since the decision to finally take a goodwill impairment indicates some degree of transparency.

business complexity variables allows the model to incorporate any aspects of business complexity that are not evident in analysts' language, in subsequent analyses, we find such additional characteristics are not necessary to obtain valid estimates of the two components.

To the extent that these variables reflect obfuscation incentives, it will weaken our identification of the latent components of managerial linguistic complexity. As we note above, no firm characteristic will perfectly capture only complexity or only reporting incentives. However, our information asymmetry validation tests, along with opposite sign predictions for the two latent components, will provide evidence on whether measurement error and bias is large enough to invalidate the decomposition.

#### 3.2.4. *Determinants of information asymmetry*

When we estimate our information asymmetry validation tests, we control for a number of determinants of information asymmetry. These controls, which we label as *Base Model Controls*, include: firm size (*Size*), book-to-market ratio (*BM*), returns over the quarter (*Returns*), idiosyncratic volatility over the quarter, (*IdioVol*), analyst coverage and dispersion (*Coverage* and *Dispersion*), an indicator for whether the firm provided an earnings forecast the day of the conference call (*MgmtForecast*), the firm's quarterly earnings surprise (*Surprise*), an indicator for whether firm had a loss that quarter (*Loss*), special items for the quarter (*SpecItems*), and an indicator for whether the firm meet or beat analyst forecasts by \$0.01 or less (*SmallBeat*).<sup>11</sup> Some of these variables are presumably also related to obfuscation incentives and outcomes (e.g., *Loss* and *SmallBeat*). Thus, by controlling for such variables in our information asymmetry

---

<sup>11</sup> We also considered controlling for share turnover, but it is highly correlated with other controls. When we include it as a control, its VIF is above 3 and it affects the significance of some of the other controls; e.g., analyst dispersion. Because analyst Fog is an important variable in our study, we decided to include analyst dispersion rather than turnover. In Table IA-6 of the Internet Appendix, we present results that include turnover, 1/price, earnings smoothness, and NYSE/NASDAQ listing indicators as additional controls. The coefficients on our estimates of the latent components remain significant at conventional levels after including these additional controls.

validation tests, we focus on estimating the incremental relation between the obfuscation (information) component and information asymmetry.

#### 4. Sample and Descriptive Statistics

We construct our sample using data on conference call transcripts from Thomson Reuters StreetEvents, analyst forecasts from I/B/E/S, and stock returns and accounting items from the CRSP/Compustat Industrial file. We begin the sample in 2002, when conference call transcripts are first available from StreetEvents, and conclude in 2011. The sample consists of 60,172 firm-quarters with conference call transcripts and the necessary CRSP, Compustat, and I/B/E/S data.

Table 1 presents descriptive statistics for variables used in our analysis; see Appendix C for variable definitions. Panel A shows that sample firms tend to be large (mean *Size* = 7.168) with high analyst coverage (mean = 9.240) and with PP&E (*CapIntensity*) accounting for 25% of the average firm's asset base. The average firm has *Acquisitions* of 0.6% of total assets, capital expenditures (*Capex*) of 1.2% of assets, R&D expense (*R&D*) of 1.1% of assets, and debt and equity issuances (*Financing*) that amounted to 3.6% of total assets during the quarter. The average (median) earnings surprise (*Surprise*) is 0.0% (0.1%) of market value at the beginning of the quarter, with 15% of the sample meeting or beating analyst forecasts by a penny or less (*SmallBeat*) and 16.7% of the sample reporting losses (*Loss*). Also, 2.8% of the sample took a goodwill impairment (*Goodwill*), 22.8% had a restructuring charge (*Restructuring*), and the average firm had income-reducing special items (mean *SpecItems* = -0.003).

Panel B reports descriptive statistics for the measures of linguistic complexity and information asymmetry. Analysts tend to use less complex language (mean *Fog(Analyst)* = 8.851) than managers, who use much more complex language in the presentation part of the call

(mean  $Fog(Present) = 15.861$ ) than in response (mean  $Fog(Response) = 11.956$ ). The differences in means and medians between  $Fog(Analyst)$ ,  $Fog(Present)$ , and  $Fog(Response)$  are statistically significant ( $p$ -values  $< 0.01$ , two-tailed). The lower complexity of the response compared to the presentation suggests that the vetting of the presentation by IR personnel and legal counsel creates more complexity than the spontaneous responses of managers. Also, Deese [1978] finds that simple sentences (i.e., fewer words) are more common in conversational exchanges than in monologues. Moreover, relative to prior work that focuses on 10-Ks, the language used on conference calls appears relatively less complex. For example, Li [2008] finds the MD&A section of the average 10-K from 1993 to 2003 has a Fog of 18.23.<sup>12</sup>

## 5. Results

### 5.1 Identifying the latent components in complex language

Table 2 reports results from regressions of managerial linguistic complexity on analysts' linguistic complexity and the business complexity variables (eq. (3)). We estimate regressions separately for the presentation and response portions of the call. We present results for three different models: a model based solely on linguistic complexity of analysts (columns (1) and (4)), a model based solely on business complexity variables (columns (2) and (5)), and a model based on both (columns (3) and (6)). We rank all of the variables into deciles and scale them to range from 0 to 1, which allows us to easily compare the relative economic significance of each variable as each coefficient represents the difference in the dependent variable between the top

---

<sup>12</sup> This difference is even more pronounced when one considers that prior work likely had artificially low Fog due to the technical error in the Perl routine discussed in Appendix B. For example, using a method without this error, Loughran and McDonald [2014] find the average 10-K from 2003 to 2011 has a Fog of 18.94

and bottom decile of the respective independent variable. This specification has the added advantage of being robust to outliers and nonlinearities.

Columns (1) and (4) report a positive and highly statistically significant coefficient on analysts' linguistic complexity ( $Fog(Analyst)$ ) for both the presentation and response portions of the call. Columns (3) and (6) indicate that the coefficient on  $Fog(Analyst)$  is generally unaffected by the inclusion of business complexity variables. In the  $Fog(Present)$  regression, the magnitude of the coefficient on  $Fog(Analyst)$  is comparable to the effect of Goodwill impairments, but much smaller than the effect of  $Size$ ,  $R\&D$ , and  $Leverage$ . This relatively low magnitude for  $Fog(Analyst)$  makes sense for the presentation part of the call, as it is prepared and delivered before the analysts ask questions. However, the coefficient on  $Fog(Analyst)$  is much larger in the  $Fog(Response)$  regression (0.22 vs. 0.09), and its magnitude is comparable to  $Size$ , book-to-market ( $BM$ ), and cash flow volatility ( $\sigma CFO$ ). Thus,  $Fog(Analyst)$  is an important determinant of managers' linguistic complexity in the response part of the call, with a coefficient that is comparable in magnitude to many of the well-established determinants of Fog.

For the business complexity variables, results in column (2) suggest managers use more complex language in the presentation when their firms are smaller, have greater leverage, lower historical stock returns, lower capital intensity, greater R&D, greater cash flow volatility, and take a goodwill impairment. Results in column (5) suggest managers use more complex language in the response when their firms are larger, have lower book-to-market, lower capital intensity, greater R&D, greater cash flow volatility, and take a goodwill impairment or restructuring charge. In the combined model for the presentation (column (3)), there is no change in the significant of the business complexity variables, whereas in the combined model for the response

(column (6)), the coefficient on prior stock returns becomes significant and the coefficient on *Goodwill* loses significance.<sup>13</sup>

Column (7) reports two-tailed *p*-values for a test of the difference in coefficients between columns (3) and (6). Firm size, leverage, and historical stock performance have a significantly stronger relation with the linguistic complexity of the presentation (*p*-values all  $\leq 0.10$ ), whereas book-to-market, capital intensity, restructuring charges, and the linguistic complexity of analysts have a significantly stronger relation with the linguistic complexity of the response (*p*-values all  $\leq 0.10$ ). The coefficients in columns (3) and (6) are generally of the same sign, with the exception of *Size* and *Restructuring*. These significant sign differences indicate that the model is identifying the managerial disclosure response to complexity in each part of the call, rather than just a consistent linear combination of business complexity variables.

Notably, the analyst-only model for the response (column (4)) has a substantially higher adjusted R-square (11.7%) than the analyst-only model for the presentation (2.2%) and both business-complexity-only models (5.0% and 3.7%). Moreover, the adjusted R-square of the combined model for the response (14.3%) is substantially larger than for the presentation (7.1%), consistent with analyst Fog better reflecting managerial Fog in the Q&A part of the call. While these adjusted R-squares are fairly low, they are comparable to other textual analysis work. For example, Li [2008] reports adjusted R-squares of 6% to 8% in explaining 10-K Fog after including year and industry fixed effects (p. 232). Thus, our R-squares are comparable to Li

---

<sup>13</sup> We also considered adding the number of segments as a measure of business complexity. However, Berger and Hann [2007] find that the number of segments a firm discloses is likely to be a strategic disclosure. In Table IA-3 of the Internet Appendix, we added number of business segments and number of geographic segments to the Table 2 model to see how they affect the results. The number of geographic segments is negatively related to Fog, which suggests that segments reflect a strategic disclosure decision. However, including number of segments in the decomposition model does not change our findings that the information (obfuscation) component is negatively (positively) associated with information asymmetry



[2008] for the *Fog(Present)* model and much larger for the *Fog(Response)* model, where analyst Fog provides the strongest identification of managerial Fog.

Consistent with our estimation framework described in Section 3, we use the regression models to provide the basis for our decomposition of linguistic complexity of the presentation and the response into latent information and obfuscation components. The fitted values of columns (3) and (6) are our estimates of the latent information components *Info(Present)* and *Info(Response)*, respectively. The residual values of columns (3) and (6) are our estimates of the latent obfuscation components, *Obfu(Present)* and *Obfu(Response)*, respectively.

Panel A of Table 3 reports descriptive statistics for our estimates of the latent components. Because the obfuscation components are regression residuals, the means are zero by construction; this does not suggest that the average level of obfuscation is zero. Similarly, the information component seems relatively large because we include the intercept in that component. Including the intercept from Table 2 in the calculation of the obfuscation components would increase (decrease) the mean of the obfuscation (information) components by 15.15 and 9.93, respectively, but would not affect any of our tests, which focus on the variation in the latent components.

Panel B of Table 3 reports correlations among *Fog(Present)*, *Fog(Response)*, and their respective latent components. Spearman (Pearson) correlations appear above (below) the diagonal. The correlation between *Obfu(.)* and *Fog(.)* is 0.96 for the presentation and 0.93 for the response; whereas the correlation between *Info(.)* and *Fog(.)* is 0.27 for the presentation and 0.38 for the response. These correlations indicate that the obfuscation component explains much of the variation in managerial Fog, as would be expected given the fairly low R-squares of the decomposition model. The correlation between *Obfu(Present)* and *Obfu(Response)* is 0.35,

suggesting that obfuscation manifests slightly differently in the presentation and the response. In contrast, the correlation between  $Info(Present)$  and  $Info(Response)$  of 0.70, which is consistent with the information components stemming from discussions of the firm’s business, which in the absence of obfuscation, should be similar across the two portions of the call.

## 5.2 Linguistic complexity and information asymmetry

We validate our decomposition of linguistic complexity into latent components by testing whether the information (obfuscation) component exhibits the expected negative (positive) association with information asymmetry. Table 4 reports results from estimating regressions of information asymmetry on the measures of linguistic complexity and controls. The regressions are based on eq. (5) and have the following form:

$$InfoAsym = \alpha_0 + \alpha_1 Info + \alpha_2 Obfu + \theta Base Model Controls + \eta \quad (8)$$

where  $InfoAsym$  is measured using *Illiquidity*, the *Info* and *Obfu* variables represent our latent components (as discussed above) and the *Base Model Controls* are the variables discussed in Section 3.2.3. We rank all of the variables into deciles and scale them to range from 0 to 1.

Before we assess the latent components, we estimate the regression using  $Fog(Present)$  and  $Fog(Response)$ , which is the approach followed by the prior literature that examines total Fog. Column (1) shows that the coefficient on  $Fog(Present)$  is positive and statistically significant, which is consistent with prior research that finds linguistic complexity is negatively associated with the information environment (e.g., Miller [2010]; Lang and Stice-Lawrence [2015]). In contrast, the coefficient on  $Fog(Response)$  is negative and statistically significant, suggesting linguistic complexity is conveying information in the Q&A. Taken together, these results could be interpreted as showing that managers use linguistic complexity to obfuscate in

the presentation and provide information in the response. However, our decomposition will show that this conclusion is biased by the commingling of the two latent components.

Column (2) present results after decomposing  $Fog(Present)$  and  $Fog(Response)$  into their respective latent components:  $Obfu(Present)$ ,  $Info(Present)$ ,  $Obfu(Response)$ , and  $Info(Response)$ . Recall from Table 3 that the correlation between  $Info(Present)$  and  $Info(Response)$  is 0.70, consistent with the two components being estimated as different linear combinations of the same set of variables (see Table 2). However, the correlation is sufficiently high that it induces multicollinearity in our regression when we include all four components together.<sup>14</sup> Thus, for the remainder of our analysis, we combine  $Info(Present)$  and  $Info(Response)$  into a single variable,  $Info(Both)$ , using the first principal component of the two variables.

In column (2), the coefficient on the information component,  $Info(Both)$ , is negative and highly significant. In contrast, the coefficients on the obfuscation components are both positive: the coefficient on  $Obfu(Present)$  is statistically significant at the 0.01 level and the coefficient on  $Obfu(Response)$  is insignificant (but with a  $p$ -value of 0.11). Thus, consistent with our predictions, the two components of the linguistic complexity are related to information asymmetry in opposite directions. The coefficients on the obfuscation components are similar in magnitude to the coefficient on meet-or-beat ( $SmallBeat$ ), which proxies for potential earnings management. The coefficient on the information component is similar in magnitude to another proxy for information complexity: analyst forecast dispersion ( $Dispersion$ ). Thus, the

---

<sup>14</sup> Whenever  $Info(Present)$  and  $Info(Response)$  are included together in a regression, the variance inflation factors are always above 7, whereas whenever  $Obfu(Present)$  and  $Obfu(Response)$  are included together in a regression, the variance inflation factors are always below 2.

coefficients on the latent components of linguistic complexity have an economic magnitude that is similar to other obfuscation and complexity proxies in explaining information asymmetry.<sup>15</sup>

In column (3), we control for other characteristics of the language in the presentation and response. For each part of the call, we include the number of words in the portion of the call (*Length*), the number of words that are industry jargon (*Jargon*), the proportion of sentences containing forward-looking statements (*ForwardLook*), the number of positive tone words (*PositiveTone*), and the number of negative tone words (*NegativeTone*).<sup>16</sup> Consistent with our earlier tests, each variable is measured separately for the presentation and response, ranked into deciles, and scaled to range between 0 and 1.<sup>17</sup> For these attributes, we find length and forward-looking sentences are negatively associated with information asymmetry and the number of negative (positive) tone words is positively (negatively) associated with information asymmetry. However, including these additional language attributes does not affect the signs or significance levels of the latent components of linguistic complexity, with the exception that the coefficient on *Obfu(Response)* is now significant at the 0.01 level. Thus, our results are not only robust to controls for additional measures of linguistic attributes, but become stronger.

---

<sup>15</sup> To examine whether our results are monotonic and not driven by extreme values, we report portfolio sorts in Table IA-2 of the Internet Appendix. We sort firms into quintiles based on the linguistic complexity measures and calculate the average abnormal information asymmetry for firms in each portfolio, where the abnormal information asymmetry is the residual from a regression of *Illiquidity* on the *Base Model Controls*. For both the presentation and response, we find that abnormal information asymmetry is monotonically increasing (decreasing) across quintiles of the obfuscation (information) component. Moreover, information asymmetry is significantly different between the top and bottom quintiles in each case.

<sup>16</sup> Industry jargon words are the 100 most common words with more than two syllables used by firms in the same two-digit SIC code; the measure of forward-looking sentences is from Li [2010] and Brochet, Loumioti, and Serafeim [2015]; and positive tone words and negative tone words are from Loughran and McDonald [2011]. In untabulated analysis, we find the number of words in the presentation and response is correlated 0.92 with the file size of the transcript. This should not be surprising given that conference call transcripts are exclusively comprised of text, unlike 10-K filings, which include pictures and advanced graphics. Thus, the inclusion of the number of words in the respective sections serves as a control for file size (e.g., Loughran and McDonald [2014]).

<sup>17</sup> In Table IA-4 of the Internet Appendix, we report correlations among these additional language attributes. *Length*, *Jargon*, *PositiveTone*, and *NegativeTone* are all highly correlated with each other (correlations in excess of 0.70). However, these measures have much lower correlations with Fog and its latent components (all correlations < 0.38). While multicollinearity among the additional language attributes does not affect the coefficients on the latent components, caution is needed when interpreting the coefficients on the additional language attributes.

Overall, the results from Table 4 show that our empirical decomposition of linguistic complexity into latent information and obfuscation components is sufficiently powerful to detect their opposite relations with information asymmetry. Moreover, our estimates of the latent components appear distinct from other attributes of language in the call. A research design using only overall linguistic complexity (column (1)) would have led to a conclusion that the relation between linguistic complexity and information asymmetry is ambiguous or differs based on the part of the call. However, our decomposition approach in columns (2) and (3) shows that the relation between the latent components of linguistic complexity and information asymmetry is unambiguous and does not differ between the presentation and response parts of the call.<sup>18</sup>

### 5.3 *Alternative specifications*

We provide a number of additional analyses in Table 5. All regressions specifications follow Table 4 and include both *Base Model Controls* and controls for additional linguistic attributes. For parsimony, coefficients on controls are not tabulated.

First, it is possible that more complex businesses have permanently higher information asymmetry, and that our latent components simply proxy for business complexity that is independent of the disclosure response to complexity. In column (1) of Table 5, we include a control for the average value of information asymmetry immediately prior to the call. Inclusion of the pre-call level of information asymmetry controls for any cross-sectional differences in the benchmark level of information asymmetry, as well as any time-varying economic complexity present prior to the call. Holding pre-call information asymmetry fixed, the obfuscation

---

<sup>18</sup> In Table IA-5 of the Internet Appendix, we repeat our analyses using the two components of Fog—average number of words per sentence and the percent of complex words—included separately and find generally similar results. However, the results are stronger when the two components are combined into the Fog index, supporting the finding of Gunning [1952] that both components together provide a more powerful measure of linguistic complexity. In Table IA-6, we also estimate the results in Table 4 with information asymmetry measured only on the day of the call and only on the day of the call and the day after. Our results for the Fog variables have the same sign and significance levels as the results for the 25-day window used in Table 4.

(information) component remains positively (negatively) associated with post-call information asymmetry.

Next, we examine whether our results are robust to including industry fixed effects (column (2)), firm fixed effects (column (3)), and both firm and manager fixed effects (column (4)). The inclusion of both firm and manager fixed effects controls for differences in managerial “style” with respect to the language of the conference call (e.g., Davis et al. [2015]).<sup>19</sup> The results for the information and obfuscation components are the same as in Table 4 with one exception. The coefficient on *Obfu(Present)* is positive but insignificant when firm fixed effects are included. This result suggest that within-firm variation in obfuscation in the presentation is likely limited due to the use of similar language, or scripting, across quarters (Lee [2015]). Despite this one exception, the results consistently indicate that within-industry, within-firm, and within-manager variation in the information (obfuscation) component of linguistic complexity is negatively (positively) associated with information asymmetry.

Finally, we assess the relative importance of analyst linguistic complexity and the business complexity variables to our empirical estimation strategy by estimating the latent components separately using only analyst complexity and only business complexity. In column (5) of Table 5, we report the results using latent components derived only from *Fog(Analyst)* (see columns (1) and (4) of Table 2). The results mirror Table 4 in showing that the latent components are significantly associated with information asymmetry in the predicted directions.

In column (6) of Table 5, we report the results using latent components derived only from the business complexity variables (see columns (2) and (5) of Table 2). The results hold for the

---

<sup>19</sup> We estimate manager fixed effects following Davis et al. [2015]. We obtain the identity of the CEO on Execucomp and require that the firm has changed CEOs during our sample period (2002–2011). These requirements allow us to distinguish a manager effect from a firm effect, but reduce our sample to 24,787 firm-quarters.

*Obfu(Present)* and *Info(Both)* components, but are the wrong sign and no longer significant for the *Obfu(Response)* component. Thus, not surprisingly, the business complexity variables do a poor job of explaining managerial Fog during the Q&A part of the call. Overall, these results show that analyst linguistic complexity is the main factor in driving a powerful identification of the latent components of managerial linguistic complexity on conference calls.<sup>20</sup>

#### 5.4 Monte Carlo placebo tests

In Table 6, we report the results using three Monte Carlo placebo tests. In the first two panels, we randomly select a different date for each conference call. In Panel A, we randomly select a non-conference call date within a 250-day window of the original conference call date, excluding dates within 25 trading days of the call. In Panel B, we randomly select another conference call date for the same firm between 2002 and 2011. We measure information asymmetry based on this random event date and re-estimate the regression in column (3) of Table 4. We repeat this step 1000 times, retaining coefficient estimates for each iteration. Then, rather than testing whether the coefficients in Table 4 are different from zero, we test whether they are different from the coefficients on the randomly selected day. This comparison allows us to rule out the possibility that we are documenting a general phenomenon that is not specific to the market's reaction to the conference call (e.g., Larcker, Ormazabal, and Taylor [2011]). In both placebo tests, we find that the relation between information asymmetry and the information

---

<sup>20</sup> This finding should also allay concerns that analyst Fog may reflect analyst incentives to obfuscate, thereby weakening the identification. As an additional test of this conjecture, we build on Mayew [2008], which finds that managers allow favorable analysts to ask questions early in the call. If favorable analysts help managers to obfuscate, then our results should be weaker for analyst Fog from the first half of the call. In Table IA-7 of the Internet Appendix, we estimate the model separately for *Fog(Analyst)* from the first half of the call ("early") and second half ("late"). We find that early and late *Fog(Analyst)* has similar explanatory power in the Table 2 model and that the latent components have almost identical relations with information asymmetry in the Table 4 regression. Thus, we do not find evidence consistent with favorable analysts facilitating managerial obfuscation.

(obfuscation) component of linguistic complexity is significantly more negative (positive) on the actual conference call date than on the random event dates, confirming our main results.

In Panel C, we randomly select a value of  $Fog(Present)$  and  $Fog(Response)$  from our sample and replace the actual value with the randomly selected value. Then, we estimate the decomposition model (Table 2) using the randomly selected value and test whether the resulting estimates of the two latent components are related to information asymmetry (Table 4). If our model is simply picking up a linear combination of the independent variables, then we should find the same results in the placebo test. However, if our model is capturing managers' specific disclosure responses to business complexity (i.e., the mapping between business complexity and managers' linguistic complexity rather than business complexity *per se*), the results will not hold in the placebo test. We find that the Table 4 results are significantly different than the placebo results, suggesting that our model captures the disclosure response to complexity, and not business complexity *per se*.

### 5.5 Validation using earnings guidance

We also validate our decomposition approach by examining the relation between the latent components of linguistic complexity and the decision to provide quantitative earnings guidance during the call. Guay et al., [2016] find a positive relation between linguistic complexity of the 10-K and voluntary disclosure, which they interpret as inconsistent with the linguistic complexity of the average 10-K resulting from an intentional choice to obfuscate. While Guay et al. [2016] make no attempt to separate and measure the two components of linguistic complexity, and study linguistic complexity in a mandatory disclosure setting that is heavily influenced by regulation and accounting rules (e.g., Dyer et al., [2016]), we can use a similar test in our setting to validate our estimates of the components of linguistic complexity.



In the conference call setting, managers are able to choose both the complexity of their language and whether to provide earnings guidance. If managers act consistently in their voluntary disclosure practices (i.e., managers who choose to be transparent, do so consistently), then managers that use complex language to provide information to shareholders are more likely to provide earnings guidance, whereas managers that use complex language to obfuscate are less likely to provide guidance. Thus, if our estimates of the two latent components of linguistic complexity are empirically descriptive, we expect that the information component is positively associated with the likelihood of providing earnings guidance during the call, whereas the obfuscation components are negatively associated with the likelihood of providing guidance.

In Table 7, we present results of the relation between the provision of quantitative earnings guidance and the latent components of linguistic complexity. The dependent variable is an indicator variable for whether the firm provides an earnings forecast during the conference call. In column (1), we report results using the total Fog measures, *Fog(Present)* and *Fog(Response)*. We find that *Fog(Present)* is negatively associated with guidance, which suggests that linguistic complexity in the presentation is associated with a lower probability of guidance. There is no significant relation between *Fog(Response)* and guidance.

In column (2), we include our estimates of the two latent components and the results provide a much clearer picture of managerial behavior. Both obfuscation components, *Obfu(Present)* and *Obfu(Response)*, are negatively associated with guidance, whereas the information component *Info(Both)* is positively associated with guidance. These results indicate that managers are acting consistently in their use of guidance and linguistic complexity. In column (3), we control for stickiness in guidance decisions by using guidance in the prior period (Billings, Jennings, and Lev [2014]). Our results hold for the *Obfu(Present)* and *Info(Both)*

components, but are weaker for the *Obfu(Response)* component ( $p$ -value = 0.11). Overall, these results further validate that our decomposition is empirically descriptive of two latent components of linguistic complexity.

### *5.6 Latent components of linguistic complexity and earnings*

Finally, as an application of our approach to prior work, we examine how the latent components of linguistic complexity vary with a firm's earnings. While Li [2008] reports that loss firms have higher 10-K Fog and attributes this finding to managers obfuscating poor performance, Bloomfield [2008] notes that this result could be driven by loss firms providing investors with more information to explain the poor performance.

Panel A of Table 8 presents mean and median values of the information and obfuscation components separately for profit and loss firms. First, consistent with Li [2008]'s findings for 10-Ks, we find loss firms have significantly higher Fog during both the presentation and response portions of the call. Second, we find that loss firms have significantly greater values of both information and obfuscation components during both parts of the conference call. While it might be more satisfying to attribute the higher Fog in loss firms to either greater obfuscation or greater information, our results suggest both forces are at work. In the absence of obfuscation, the information provided by managers of loss firms is higher. However, managers in loss firms also obfuscate to a greater extent to mask their poor performance.

Importantly, if our measures of the latent components of linguistic complexity are empirically descriptive, then an increase in obfuscation (or information) should be reflected in the *level* of the respective latent component, not the marginal effect of that component on information asymmetry. That is, the notion that managers in loss firms obfuscate more than

profit firms implies the *level* of obfuscation differs between profit and loss firms while the marginal effect of obfuscation on information asymmetry should be the same.

To investigate whether the marginal effects of obfuscation and information components vary between profit and loss firms, we re-estimate the regressions in Table 4 partitioning the sample into profit and loss firms. Panel B of Table 8 presents results. Note that estimating the regression within sample partitions leads to a reduction in power, especially for the loss firms, which constitute less than 1/6th of our sample (10,074 firms-quarters). Nevertheless, regardless of whether the firm is profitable or not, we continue to find a strong positive (negative) relation between the obfuscation (information) component and information asymmetry.

## **6. Conclusion**

Prior research generally interprets complex language in a firm's disclosures as indicative of managerial obfuscation. However, complex language can also reflect the provision of complex information, e.g., informative technical disclosure. As a consequence, linguistic complexity commingles two latent components—obfuscation and information—that are related to information asymmetry in opposite directions. In this paper, we develop a novel empirical approach to estimate these two latent components within the context of quarterly earnings conference calls. Specifically, we use the linguistic complexity of an external party—analysts—to identify the portion of managerial linguistic complexity related to obfuscation and the portion related to the provision of information. We then validate our estimates of the latent components of linguistic complexity by testing whether the information (obfuscation) component of linguistic complexity is negatively (positively) associated with information asymmetry.

Consistent with our estimates measuring the underlying latent components of linguistic complexity, we find a negative (positive) relation between the estimated information (obfuscation) component of linguistic complexity and information asymmetry. These relations hold across both the presentation and response parts of the conference call. The results are robust to controlling for other linguistic attributes of the call; to using within-industry, within-firm, and within-manager designs; and to placebo tests. We also validate our decomposition by showing that the information (obfuscation) component is positively (negatively) associated with the provision of quantitative earnings guidance on the call, as would be expected if managers act consistently in their disclosure decisions (i.e., the decision to be transparent). Finally, we show that our decomposition produces more nuanced conclusions for loss firms than existing work that suggests loss firms use Fog to obfuscate their performance. We show that some loss firms use complex language to provide more informative disclosures.

The conference call setting provides strong internal validity by allowing us to identify informative managerial linguistic complexity using analyst linguistic complexity in the same disclosure venue. Our findings suggest that researchers can improve the power of their tests by controlling for the level of linguistic complexity expected in the absence of obfuscation. One important caveat to our decomposition approach is that we cannot speak to the generalizability of the approach outside of the conference call setting; e.g., 10-K or 10-Q disclosures. We encourage future work to explore whether the decomposition needs to be modified or updated to be setting-specific. Regardless, our findings do suggest that regulators and researchers should exercise caution when assuming that complex language is necessarily less informative than simple language.

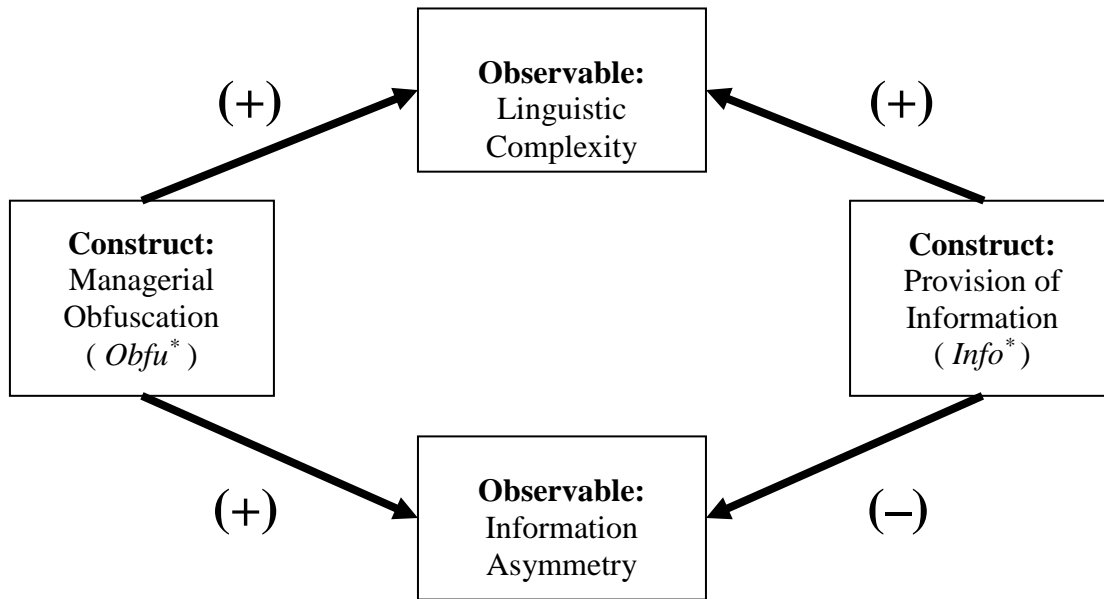
## References

- ALLEE, K., AND M. DEANGELIS. "The structure of voluntary disclosure narratives: Evidence from tone dispersion." *Journal of Accounting Research* 53 (2015): 241–274.
- ALLEN, W. "Readability of instructional film commentary." *Journal of Applied Psychology* 36 (1952): 164–168.
- AMIHUD, Y. "Illiquidity and stock returns: Cross-section and time-series effects." *Journal of Financial Markets* 5 (2002): 31–56.
- ARMSTRONG, C.; J. CORE; D. TAYLOR; AND R. VERRECCHIA. "When does information asymmetry affect the cost of capital?" *Journal of Accounting Research* 49 (2011): 1–40.
- BALAKRISHNAN, K; J. CORE; AND R. VERDI. "The relation between reporting quality and financing and investment: Evidence from changes in financing capacity." *Journal of Accounting Research* 52 (2014): 1–36.
- BARTH, M.; R. KASZNIK; AND M. MCNICHOLS. "Analyst coverage and intangible assets." *Journal of Accounting Research* 39 (2001): 1–34.
- BERGER, P., AND R. HANN. "Segment profitability and the proprietary and agency costs of disclosure." *The Accounting Review* 82 (2007): 869–906.
- BILLINGS, M.; R. JENNINGS, AND B. LEV. "On guidance and volatility." *Journal of Accounting and Economics* 60 (2015): 161–180.
- BLOOMFIELD, R. "Incomplete revelation hypothesis and financial reporting." *Accounting Horizons* 16 (2002): 233–243.
- BLOOMFIELD, R. "Discussion of 'Annual report readability, current earnings, and earnings persistence.'" *Journal of Accounting and Economics* 45 (2008): 248–252.
- BONSALL, S.; A. LEONE; AND B. MILLER. "A plain English measure of financial reporting readability." Working Paper, Ohio State University (2015).
- BONSALL, S., AND B. MILLER. "The impact of narrative disclosure readability on bond ratings and the cost of debt capital." *Review of Accounting Studies* (2016): forthcoming.
- BOWEN, R.; A. DAVIS; AND D. MATSUMOTO. "Do conference calls affect analysts' forecasts." *The Accounting Review* 77 (2002): 285–316.
- BOZANIC, Z. AND M. THEVENOT. "Qualitative disclosure and changes in sell-side financial analysts' information environment." *Contemporary Accounting Research* (2015): Forthcoming.
- BROCHET, F.; P. NARANJO; AND G. YU. "Causes and consequences of linguistic complexity in non-U.S. firm conference calls." Working paper, Harvard and MIT, 2014.
- BROCHET, F.; M. LOUMIOTI; AND G. SERAFEIM, "Speaking of the short-term: Disclosure horizon and managerial myopia." *Review of Accounting Studies* (2015): 1122–1163.
- BROWN, S.; S. HILLEGEIST; AND K. LO. "Conference calls and information asymmetry." *Journal of Accounting and Economics* 37 (2004): 343–366.
- BURGOON, J.; W. MAYEW; J. GIBONEY; A. ELKINS; K. MOFFIT; B. DORN; M. BYRD; AND L. SPITZLEY. "Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls." *Journal of Language and Social Psychology* (2015): 1–35.
- BUSHEE, B.; D. MATSUMOTO; AND G. MILLER. "Open versus closed conference calls: The determinants and effects of broadening access to disclosure" *Journal of Accounting and Economics* 34 (2003): 149–180.

- CHARROW, R., AND V. CHARROW. "A Psycholinguistic study of jury instructions." *Columbia Law Review* 79 (1979): 1306–1374.
- CHEN, X.; Q. CHENG; AND K. LO. "On the relationship between analyst reports and corporate disclosures: Exploring the roles of information discovery and interpretation." *Journal of Accounting and Economics* 49 (2010): 206–226.
- DAVIS, A.; W. GE; D. MATSUMOTO; AND J. ZHANG. "The effect of manager-specific optimism on the tone of earnings conference calls." *Review of Accounting Studies* 20 (2015): 639–673.
- DE FRANCO, G.; O. HOPE; D. VYAS; AND Y. ZHOU. "Analyst Report Readability." *Contemporary Accounting Research* 32 (2015): 76–104.
- DEESE, J. "Thought into speech: Linguistic rules and psychological limitations in processing information determine how we put our ideas into words." *American Scientist* 66 (1978): 314–321
- DOUGAL, C.; J. ENGELBERG; D. GARCIA; AND C. PARSONS. "Journalists and the stock market." *Review of Financial Studies* 25 (2012): 639–679.
- DYER, T.; M. LANG; AND L. STICE-LAWRENCE. "The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation" University of North Carolina Working Paper.
- FANG, I. "The 'easy listening' formula." *Journal of Broadcasting* 11 (1966–1967): 63–68.
- FRANKEL, R.; M. JOHNSON; AND D. SKINNER. "An empirical examination of conference calls as a voluntary disclosure medium." *Journal of Accounting Research* 37 (1999): 133–150.
- GLOSTEN, L., AND L. HARRIS. "Estimating the components of the bid/ask spread." *Journal of Financial Economics* 21 (1988): 123–142.
- GUAY, W.; D. SAMUELS; AND D. TAYLOR. "Guiding through the fog: Financial statement complexity and voluntary disclosure." *Journal of Accounting and Economics* (2016): forthcoming.
- GUNNING, R. 1952. "The technique of clear writing." New York, NY: McGraw-Hill International Book Co.
- HARWOOD, K. "Listenability and readability." *Speech Monographs* 22 (1955): 53–57.
- HOBERG, G., AND C. LEWIS. "Do fraudulent firms produce abnormal disclosure?" Working Paper, Vanderbilt Owen Graduate School of Management, 2015.
- LANG, M., AND M. MAFFETT. "Transparency and liquidity uncertainty in crisis periods." *Journal of Accounting and Economics* 52 (2011): 101–125.
- LANG, M., AND L. STICE-LAWRENCE. "Textual analysis and international financial reporting: Large sample evidence." *Journal of Accounting and Economics* 60 (2015): 110–135.
- LARCKER, D.; G. ORMAZABAL; AND D. TAYLOR. "The market reaction to corporate governance regulation." *Journal of Financial Economics* 101 (2011): 431–448.
- LARCKER, D., AND A. ZAKOLYUKINA. "Detecting deceptive discussions in conference calls." *Journal of Accounting Research* 50 (2012): 495–540.
- LAWRENCE, A. "Individual investors and financial disclosure." *Journal of Accounting and Economics* 56 (2013): 130–147.
- LEE, C., AND M. READY. "Inferring trade direction from intraday data." *Journal of Finance* 46 (1991): 733–746.
- LEE, J. "Can investors detect managers' lack of spontaneity? Adherence to pre-determined scripts during earnings conference calls." *The Accounting Review* (2015): 229–250.
- LEHAVY, R.; F. LI; AND K. MERKLEY. "The effect of annual report readability on analyst following and the properties of their earnings forecasts." *The Accounting Review* 86 (2012): 1087–1115.

- LEUZ, C., AND R. VERRECCHIA. "The economic consequences of increased disclosure." *Journal of Accounting Research* 38 (2000): 91–124.
- LI, F. "Annual report readability, current earnings, and earnings persistence." *Journal of Accounting and Economics* 45 (2008): 221–247.
- LI, F. "The information content of forward-looking statements in corporate filings: A naive Bayesian machine learning approach." *Journal of Accounting Research* 48 (2010): 1049–1102.
- LOUGHRAN, T., AND B. McDONALD. "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks." *Journal of Finance* 66 (2011): 35–65.
- LOUGHRAN, T., AND B. McDONALD. "Measuring readability in financial disclosures." *The Journal of Finance* 69 (2014): 1643–1671.
- MADHAVAN, A.; M. RICHARDSON; AND M. ROOMANS. "Why do security prices change? A transaction-level analysis of NYSE stocks." *Review of Financial Studies* 10 (1997): 1035–1064.
- MATSUMOTO, D.; M. PRONK; AND E. ROELOFSEN. "The information content of managers' presentations and analysts' discussion sessions." *The Accounting Review* 86 (2011): 1383–1414.
- MAYEW, W. "Evidence of management discrimination among analysts during earnings conference calls." *Journal of Accounting Research* 46 (2008): 627–659.
- MAYEW, W.; N. SHARP; AND M. VENKATACHALAM. "Using earnings conference calls to identify analysts with superior private information." *Review of Accounting Studies* 18 (2012): 386–413.
- MILLER, B. "The effects of reporting complexity on small and large investor trading." *The Accounting Review* 85 (2010): 2107–2143.
- NG, J.; T. RUSTICUS; AND R. VERDI. "Implications of transaction costs for the post-earnings announcement drift." *Journal of Accounting Research* 46 (2008): 661–696.
- REDISH, J. C., AND J. SELZER. "The place of readability formulas in technical communication." *Technical Communication* 32 (1985): 46–52.
- TASKER, S. "Bridging the information gap: Quarterly conference calls as a medium for voluntary disclosure." *Review of Accounting Studies* 3 (1998): 137–167
- TWEDT, B., AND L. REES. "Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports." *Journal of Accounting and Public Policy* 31 (2012): 1–21.
- VERRECCHIA, R. "Information quality and discretionary disclosure." *Journal of Accounting and Economics* 12 (1990): 365–380.

FIGURE 1. Graphical Representation of Latent Components



This figure presents a graphics representation of the latent components of managerial linguistic complexity. Predicted signs appear in parentheses.



APPENDIX A. Derivation of regression coefficients in eqn (5).

This appendix separately analyzes the effect of measurement error (A.1) and violation of the assumption that the linguistic complexity of analysts does not reflect obfuscation (A.2) on the estimated regression coefficients  $\alpha_1$  and  $\alpha_2$  in equation (5).

*A.1 Effect of Measurement Error*

To restate the assumptions given in Section 3.1.1 and 3.1.2:

$$InfoAsym = \gamma_0 + \gamma_1 Info^* + \gamma_2 Obfu^* + \eta$$

$$LC\_Manager = \varphi_0 + \varphi_1 Info^* + \varphi_2 Obfu^* + \varepsilon$$

$$LC\_Analyst = \delta_0 + \delta_1 Info^* + \nu$$

where  $*$  indicates unobserved variables;  $\eta$ ,  $\varepsilon$ , and  $\nu$ , represent measurement error; and  $\gamma_1 < 0$ ,  $\gamma_2 > 0$ ,  $\varphi_1 > 0$ ,  $\varphi_2 > 0$ , and  $\delta_1 > 0$ . For tractability, we additionally assume  $Info^*$ ,  $Obfu^*$ ,  $\eta$ ,  $\varepsilon$ , and  $\nu$  are all independently distributed with mean zero and variances  $\sigma_I^2$ ,  $\sigma_O^2$ ,  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ , and  $\sigma_\nu^2$  respectively, and set  $\varphi_0 = \delta_0 = 0$ ,  $\gamma_1 = -1$  and  $\gamma_2 = 1$ . We also assume that we have a sufficiently large sample, so that our estimated regression coefficients converge to their probability limits.

Our empirical estimate of  $Info^*$  ( $Obfu^*$ ), denoted  $Info$  ( $Obfu$ ), is the fitted (residual) value from a regression of  $LC\_Manager$  on  $LC\_Analyst$ :

$$Info = \hat{\beta} LC\_Analyst,$$

where  $\hat{\beta} = \frac{\varphi_1 \delta_1 \sigma_I^2}{\delta_1^2 \sigma_I^2 + \sigma_\nu^2}$ . Note that measurement error in  $LC\_Analyst$ ,  $\sigma_\nu^2$ , appears in the denominator and attenuates this coefficient. The fitted value,  $Info$ , is given by

$$\frac{\varphi_1 \delta_1^2 \sigma_I^2}{\delta_1^2 \sigma_I^2 + \sigma_\nu^2} Info^* + \frac{\varphi_1 \delta_1 \sigma_I^2}{\delta_1^2 \sigma_I^2 + \sigma_\nu^2} \nu^*.$$

We now solve for the coefficients in equation (5) in the presence of measurement error. The standard formulas for the slope coefficients from a regression of  $y$  on  $x_1$  and  $x_2$  are given by

$$\hat{\alpha}_1 = \frac{\text{var}(x_2) \text{cov}(x_1, y) - \text{cov}(x_1, x_2) \text{cov}(x_2, y)}{\text{var}(x_1) \text{var}(x_2) - \text{cov}(x_1, x_2)^2}$$

$$\hat{\alpha}_2 = \frac{\text{var}(x_1) \text{cov}(x_2, y) - \text{cov}(x_1, x_2) \text{cov}(x_1, y)}{\text{var}(x_1) \text{var}(x_2) - \text{cov}(x_1, x_2)^2}$$

From the above information structure, we have that

$$\begin{aligned}\text{var}(x_1) &= \hat{\beta}_1^2 (\delta_1^2 \sigma_I^2 + \sigma_v^2) \\ \text{var}(x_2) &= \sigma_o^2 \varphi_2^2 + (\varphi_1 - \hat{\beta}_1 \delta_1)^2 \sigma_I^2 + \sigma_\varepsilon^2 + \hat{\beta}_1^2 \sigma_v^2 \\ \text{cov}(x_1, x_2) &= \hat{\beta}_1 \delta_1 (\varphi_1 - \hat{\beta}_1 \delta_1) \sigma_I^2 - \hat{\beta}_1^2 \sigma_v^2 = 0 \\ \text{cov}(x_1, y) &= -\hat{\beta}_1 \delta_1 \sigma_I^2 \\ \text{cov}(x_2, y) &= \sigma_o^2 \varphi_2 - \sigma_I^2 (\varphi_1 - \hat{\beta}_1 \delta_1)\end{aligned}$$

Substituting in the respective variances and covariances and using some algebra yields

$$\hat{\alpha}_1 = -\frac{1}{\varphi_1} \quad \text{and} \quad \hat{\alpha}_2 = \frac{\varphi_2 \delta_1^2 \sigma_I^2 \sigma_o^2 + \sigma_v^2 (\varphi_2 \sigma_o^2 - \varphi_1 \sigma_I^2)}{\delta_1^2 \sigma_I^2 (\varphi_2^2 \sigma_o^2 + \sigma_\varepsilon^2) + \sigma_v^2 (\varphi_1^2 \sigma_I^2 + \varphi_2^2 \sigma_o^2 + \sigma_\varepsilon^2)}.$$

These results demonstrate that the sign of the coefficient on our estimate of the information component,  $\hat{\alpha}_1$ , is unambiguously negative and its magnitude is not affected by measurement error.

In contrast, the presence of measurement error does affect the sign of the coefficient on our estimate of the obfuscation component,  $\hat{\alpha}_2$ . The denominator of  $\hat{\alpha}_2$  is unambiguously positive, so we need only consider the sign of the numerator. The first term in the numerator is also unambiguously positive, so we need only consider the sign and magnitude of the second term.

Define  $\Delta_{Manager} = (\varphi_2 \sigma_o^2 - \varphi_1 \sigma_I^2)$ , where  $\Delta_{Manager}$  is the difference between the variation in *LC\_Manager* attributable to obfuscation and the variation in *LC\_Manager* attributable to information. For  $\Delta_{Manager} \geq 0$ , the variation in *LC\_Manager* attributable to obfuscation exceeds the variation attributable to information and  $\hat{\alpha}_2$  is unambiguously positive. For  $\Delta_{Manager} < 0$ , the variation in *LC\_Manager* attributable to obfuscation is less than the variation attributable to information and  $\hat{\alpha}_2$  is positive so long as measurement error in *LC\_Analyst* ( $\sigma_v^2$ ) is not too large.

To summarize, measurement error does not affect the sign of the coefficient on our estimate of the information component ( $\hat{\alpha}_1$ ), but does impart a negative bias on the coefficient on our estimate of the obfuscation component ( $\hat{\alpha}_2$ ). For sufficiently large values of measurement error ( $\sigma_v^2$ ),  $\hat{\alpha}_2$  will be *negative*. This is the opposite of our predicted sign, and illustrates that measurement error biases against finding  $\hat{\alpha}_2 > 0$ .

## A.2 Effect of Obfuscation in Linguistic Complexity of Analysts

We begin by reformulating the expression for *LC\_Analyst* to include an intentional obfuscation component:

$$LC\_Analyst = \delta_0 + \delta_1 Info^* + \delta_2 Obfu^*$$

where  $\delta_2 > 0$ . Note that expression is more general than the prior expression in which we assumed  $\delta_2 = 0$ . We employ the same assumptions as in A.1., except that for tractability, we do not additionally consider measurement error (i.e.,  $\sigma_v^2 = \sigma_\varepsilon^2 = 0$ ).

Our empirical estimate of  $Info^*$  ( $Obfu^*$ ), denoted  $Info$  ( $Obfu$ ), is the fitted (residual) value from a regression of  $LC\_Manager$  on  $LC\_Analyst$ :

$$Info = \hat{\beta} LC\_Analyst,$$

where  $\hat{\beta} = \frac{\varphi_1 \delta_1 \sigma_I^2 + \varphi_2 \delta_2 \sigma_O^2}{\delta_1^2 \sigma_I^2 + \delta_2^2 \sigma_O^2}$ . Note that as a result of  $\delta_2 > 0$ , the variance of the obfuscation component,  $\sigma_O^2$ , now appears in both the numerator and denominator. As a result, the fitted value,  $Info$ , is now given by

$$\frac{\varphi_1 \delta_1 \sigma_I^2 + \varphi_2 \delta_2 \sigma_O^2}{\delta_1^2 \sigma_I^2 + \delta_2^2 \sigma_O^2} Info^* + \frac{\varphi_1 \delta_1 \sigma_I^2 + \varphi_2 \delta_2 \sigma_O^2}{\delta_1^2 \sigma_I^2 + \delta_2^2 \sigma_O^2} v^*.$$

We now solve for the coefficients in equation (5) in the presence of  $\delta_2 > 0$ . Substituting in the respective variances and covariances into the standard formulas for the slope coefficients and using some algebra yields:

$$\hat{\alpha}_1 = \frac{\delta_2 \sigma_O^2 - \delta_1 \sigma_I^2}{\hat{\beta}(\delta_2^2 \sigma_O^2 + \delta_1^2 \sigma_I^2)}$$

$$\hat{\alpha}_2 = \frac{\sigma_O^2(\varphi_2 - \hat{\beta}\delta_2) - \sigma_I^2(\varphi_1 - \hat{\beta}\delta_1)}{\sigma_O^2(\varphi_2 - \hat{\beta}\delta_2)^2 + \sigma_I^2(\varphi_1 - \hat{\beta}\delta_1)^2}$$

First, consider the sign of  $\hat{\alpha}_1$ , the coefficient on our estimate of the information component. The denominator is unambiguously positive, and we need only consider the sign of the numerator. Define  $\Delta_{Analyst} = (\delta_2 \sigma_O^2 - \delta_1 \sigma_I^2)$ , where  $\Delta_{Analyst}$  is the difference between the variation in  $LC\_Analyst$  attributable to obfuscation and the variation in  $LC\_Analyst$  attributable to information. For  $\Delta_{Analyst} < 0$ , the variation in  $LC\_Analyst$  attributable to information exceeds the variation attributable to obfuscation and  $\hat{\alpha}_1 > 0$ , as predicted. However, for  $\Delta_{Analyst} > 0$ , the variation in  $LC\_Analyst$  attributable to obfuscation exceeds the variation attributable to information and  $\hat{\alpha}_1 < 0$ , counter to our prediction. This illustrates that our identification strategy relies on whether the linguistic complexity of analysts predominantly reflects the information component or predominantly reflects the obfuscation component, not that  $\delta_2 = 0$  *per se*. If the

analyst linguistic complexity predominantly reflects the information component, our prediction,  $\hat{\alpha}_1 < 0$ , holds.

Next, consider the sign of  $\hat{\alpha}_2$ , the coefficient on our estimate of the obfuscation component. The denominator is unambiguously positive, and we need only consider the sign of the numerator. Note that the numerator simplifies to  $\Delta_{Manager} - \hat{\beta}\Delta_{Analyst}$ . As such, we are left with a similar condition for  $\hat{\alpha}_2$  to be positive as in A.1, except that now the condition also depends on  $\Delta_{Analyst}$ . If both (1) the variation in  $LC_{Manager}$  attributable to obfuscation exceeds the variation attributable to information; i.e.,  $\Delta_{Manager} > 0$ , and (2) the variation in  $LC_{Analyst}$  attributable to information exceeds the variation attributable to obfuscation,  $\Delta_{Analyst} < 0$ , then  $\hat{\alpha}_2$  will be unambiguously positive, and our prediction holds. If either (1) or (2) is violated it is possible that  $\hat{\alpha}_2$  would be negative, counter to our predictions, depending on the precise values of  $\Delta_{Manager}$  and  $\Delta_{Analyst}$ .

To summarize, if  $LC_{Analyst}$  also contains an obfuscation component, it biases against finding the coefficient on the information (obfuscation) component is negative (positive).

## APPENDIX B. Technical Note on the Perl Routine used to Compute the Fog Index

Prior research has generally followed Li [2008] in using the Perl routine *Lingua::EN::Fathom* to calculate the Fog index.<sup>21</sup> However, this routine uses a simple regular expression (namely, `^b\s*[!?\s]*b/`) to identify sentence breaks; i.e., the routine simply looks for punctuation marks (“!?”) when calculating the number of sentences.<sup>22</sup> Consequently, the routine is confounded by punctuation used in numbers and abbreviations. For example, consider the following text:

*Thanks, Bob. Good morning everyone. Yesterday we reported second quarter revenues of 619 million. We reported net earnings of 24.3 million or 0.78 per diluted share for the quarter compared with net earnings of 15.5 million or 0.50 per diluted share for the same period last year. Our U.S. business has been particularly robust.*

The passage has 49 words, 11 complex words (i.e., more than two syllables), and five sentences. However, *Lingua::EN::Fathom* counts extra sentences due to “24.3,” “0.78,” “15.5,” “0.50,” “U.” and “S.” This issue affects the calculation of Fog, which is 10.76 with 11 sentences, compared to the correct value of 12.90 with five sentences.

Note that this measurement error is not random. Text passages with more numbers will tend to have erroneously low words per sentence and, thus, an erroneously low Fog index when using the *Lingua::EN::Fathom* routine. This technical issue could potentially affect prior literature that uses the Fog index. To correct for this issue, we use the *Lingua::EN::Sentence* routine, which correctly identifies five sentences in the example above, to calculate the number of sentences. Also, we note that the function *sent\_tokenize* in the Python *NLTK* (Natural Language Toolkit) package and the Fog index provided on the WRDS SEC Analytics Suite are not affected by this issue.

In Table IA-8 of the Internet Appendix, we estimate our results using the *Lingua::EN::Fathom* Perl routine (Li [2008]), which we call “FathomFog.” Panel A shows that one interesting difference from using *FathomFog* is that the coefficient on *Size* is positive and significant in explaining *FathomFog(Present)*, whereas as it is negative and significant for the correct *Fog(Present)*. In Panel B, we find that the results for *FathomFog* are similar to correct Fog for the *Obfu(Response)* and *Info(Both)* components, but do not hold for the *Obfu(Present)* component. Thus, *FathomFog* provides weaker identification of managerial Fog in the presentation part of the call, which is also the part of the call that is more likely to contain numerical test.

---

<sup>21</sup> Many of the papers cited in Section 2 either explicitly acknowledge their data comes from Li [2008] or state that they use the *Lingua::EN::Fathom* routine to calculate the Fog index.

<sup>22</sup> Source: <http://cpansearch.perl.org/src/KIMRYAN/Lingua-EN-Fathom-1.18/lib/Lingua/EN/Fathom.pm>

## APPENDIX C. Variable Definitions

### Measures of linguistic complexity and information asymmetry

---

<i>Fog(.)</i>	Fog index of the respective portion of the conference call. <i>Fog(Analyst)</i> refers to the Fog index of analysts during the call, <i>Fog(Present)</i> refers to the Fog index of managers during the presentation portion of the call, and <i>Fog(Response)</i> refers to the Fog index of managers' during the response portion of the call.
<i>Obfu(.)</i>	Estimated latent obfuscation component. <i>Obfu(Present)</i> refers to the presentation portion of the call, and <i>Obfu(Response)</i> refers to the response portion of the call. <i>Obfu(.)</i> is the residual regressions in columns (2) and (4) in Table 2.
<i>Info(.)</i>	Estimated latent information component. <i>Info(Present)</i> refers to the presentation portion of the call, and <i>Info(Response)</i> refers to the response portion of the call. <i>Info(.)</i> is the fitted value in columns (2) and (4) in Table 2.
<i>Info(Both)</i>	First principal component of <i>Info(Present)</i> and <i>Info(Response)</i>
<i>Illiquidity</i>	The average value of the Amihud [2002] measure of illiquidity over the period starting the day of the call and ending twenty-five trading days subsequent to the call. Firm days with zero volume are excluded. This restriction only affects 0.05% of the 2.1 million daily observations used to compute this measure.

### Firm characteristics

---

<i>Acquisitions</i>	Total acquisitions during the quarter, scaled by total assets at the beginning of the quarter.
<i>BM</i>	Book value of equity scaled by market value of equity at the beginning of the quarter.
<i>CapIntensity</i>	Net plant, property, and equipment scaled, scaled by total assets at the beginning of the quarter.
<i>Capex</i>	Amount of capital expenditures scaled by total assets at the beginning of the quarter.
<i>Coverage</i>	Number of analysts with earnings forecasts for the current quarter, measured prior to the conference call.
<i>Dispersion</i>	Standard deviation of analyst forecasts for the current quarter, measured prior to the conference call and scaled by price at the beginning of the quarter.
<i>Financing</i>	Amount raised from stock and debt issuances during the quarter scaled by total assets at the beginning of the quarter.
<i>Goodwill</i>	Indicator variable for whether the firm had a goodwill impairment charge that quarter
<i>IdioVol</i>	Standard deviation of residuals from a firm-specific three factor Fama-French model estimated using daily returns over the quarter.
<i>Leverage</i>	Long term debt plus short term debt, scaled by total assets at the beginning of the quarter.
<i>Loss</i>	Indicator variable for whether the firm reported a loss.
<i>MgmtForecast</i>	Indicator variable for whether the firm provides an earnings forecast during the conference call.
<i>R&amp;D</i>	Ratio of research and development expense to sales
<i>Restructuring</i>	Indicator variable for whether the firm had a restructuring charge that quarter
<i>Returns</i>	Buy-and-hold return over the quarter, in percent.
<i>Size</i>	Natural log of market value of equity at the beginning of the quarter.
<i>SmallBeat</i>	Indicator variable for whether earnings per share beat consensus forecasts by a penny or less
<i>SpecItems</i>	Special items scaled by market value of equity at the beginning of the quarter.
<i>Surprise</i>	Consensus forecast error scaled by market value of equity at the beginning of the quarter.
$\sigma$ CFO	Consensus forecast is the median analyst forecast on IBES measured prior to the conference call. Standard deviation of cash flows from operations scaled by total assets at over the prior five years.

TABLE 1. Descriptive Statistics

PANEL A. Firm Characteristics

Variable	Mean	Std. Dev.	P25	Median	P75
<i>Acquisitions</i>	0.006	0.025	0.000	0.000	0.000
<i>BM</i>	0.509	0.382	0.257	0.431	0.670
<i>CapIntensity</i>	0.247	0.242	0.059	0.159	0.367
<i>Capex</i>	0.012	0.016	0.003	0.007	0.015
<i>Coverage</i>	9.240	6.318	4.000	7.000	13.000
<i>Dispersion</i>	0.002	0.004	0.000	0.001	0.002
<i>Financing</i>	0.036	0.096	0.000	0.003	0.021
<i>Goodwill</i>	0.028	0.165	0.000	0.000	0.000
<i>Idiovol</i>	0.023	0.013	0.014	0.019	0.028
<i>Leverage</i>	0.211	0.200	0.028	0.176	0.328
<i>Loss</i>	0.167	0.373	0.000	0.000	0.000
<i>MgmtForecast</i>	0.514	0.500	0.000	1.000	1.000
<i>R&amp;D</i>	0.011	0.023	0.000	0.000	0.013
<i>Restructuring</i>	0.228	0.419	0.000	0.000	0.000
<i>Returns</i>	2.314	23.284	-11.012	1.826	14.209
<i>Size</i>	7.168	1.599	6.033	6.998	8.154
<i>SmallBeat</i>	0.150	0.357	0.000	0.000	0.000
<i>SpecItems</i>	-0.003	0.015	-0.001	0.000	0.000
<i>Surprise</i>	0.000	0.010	-0.001	0.001	0.002
<i>σCFO</i>	0.060	0.069	0.022	0.039	0.069

PANEL B. Measures of Linguistic Complexity and Information Asymmetry

Variable	Mean	Std. Dev.	P25	Median	P75
<i>Fog(Analyst)</i>	8.851	2.513	8.384	9.231	10.098
<i>Fog(Present)</i>	15.861	1.591	14.849	15.899	16.932
<i>Fog(Response)</i>	11.956	1.644	10.804	11.840	12.957
<i>Illiquidity</i>	0.076	2.297	0.000	0.002	0.008

This table presents descriptive statistics for firms in our sample. Our sample is constructed from the intersection of Thomson Reuters StreetEvents, I/B/E/S, and CRSP/Compustat. The sample spans the time period January 2002 to December 2011 and covers a total of 60,172 firm-quarter observations. Panel A reports the distribution of various firm characteristics used in our analysis, Panel B reports the distribution of measures of linguistic complexity and information asymmetry. All variables are winsorized at the 1<sup>st</sup> and 99<sup>th</sup> percentiles. See Appendix C for variable definitions.

TABLE 2. Estimating the Latent Components of Managers' Linguistic Complexity

Variable	Dependent Variable:						<i>p</i> -value: test of difference in coefficients (3) – (6)
	<i>Fog(Present)</i>			<i>Fog(Response)</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Fog(Analyst)</i>	<b>0.09***</b> (20.19)	.	<b>0.09***</b> (20.57)	<b>0.22***</b> (30.24)	.	<b>0.21***</b> (31.49)	[<0.001]
<i>Size</i>	.	<b>-0.56***</b> (-6.38)	<b>-0.58***</b> (-6.78)	.	<b>0.34***</b> (4.73)	<b>0.28***</b> (4.10)	[<0.001]
<i>Leverage</i>	.	<b>0.23***</b> (2.99)	<b>0.24***</b> (3.10)	.	0.08 (1.29)	0.09 (1.54)	[0.049]
<i>BM</i>	.	-0.08 (-1.09)	-0.07 (-0.98)	.	<b>-0.21***</b> (-3.39)	<b>-0.18***</b> (-3.20)	[0.10]
<i>Returns</i>	.	<b>-0.13***</b> (-4.54)	<b>-0.15***</b> (-5.93)	.	0.02 (0.64)	<b>-0.05**</b> (-2.11)	[<0.001]
<i>Acquisitions</i>	.	-0.08 (-1.26)	-0.09 (-1.40)	.	-0.03 (-0.57)	-0.05 (-1.00)	[0.488]
<i>CapIntensity</i>	.	<b>-0.47***</b> (-4.51)	<b>-0.39***</b> (-3.78)	.	<b>-0.78***</b> (-9.15)	<b>-0.59***</b> (-7.26)	[0.042]
<i>Capex</i>	.	-0.04 (-0.53)	-0.07 (-0.92)	.	0.07 (1.07)	0.00 (0.08)	[0.329]
<i>R&amp;D</i>	.	<b>0.58***</b> (5.54)	<b>0.57***</b> (5.57)	.	<b>0.44***</b> (4.99)	<b>0.42***</b> (5.05)	[0.128]
<i>Financing</i>	.	0.02 (0.46)	0.02 (0.45)	.	0.02 (0.54)	0.02 (0.51)	[0.975]
<i>σCFO</i>	.	<b>0.29***</b> (3.74)	<b>0.30***</b> (3.90)	.	<b>0.22***</b> (3.40)	<b>0.24***</b> (3.91)	[0.369]
<i>Goodwill</i>	.	<b>0.12**</b> (2.25)	<b>0.11**</b> (2.22)	.	<b>0.08*</b> (1.66)	0.07 (1.62)	[0.44]
<i>Restructuring</i>	.	0.00 (0.04)	-0.02 (-0.42)	.	<b>0.12***</b> (3.29)	<b>0.08**</b> (2.14)	[0.030]
<i>N-obs</i>	60,172	60,172	60,172	60,172	60,172	60,172	
<i>Adj R<sup>2</sup></i>	2.15	5.02	7.08	11.65	3.73	14.34	

This table presents results from estimating the linguistic complexity of managers during the respective portion of the conference call, *Fog(Present)* and *Fog(Response)*, as a function of linguistic complexity of analysts, *Fog(Analyst)*, and variables related to business complexity. We use the following variables to measure business complexity: firm size (*Size*); firm leverage (*Leverage*); book-to-market ratio (*BM*); historical stock performance (*Returns*); acquisitions (*Acquisitions*), capital intensity (*CapIntensity*), capital expenditures (*Capex*), research and development (*R&D*); debt and equity issuance (*Financing*); cash flow volatility (*σCFO*); goodwill impairments (*Goodwill*) and restructuring charges (*Restructuring*). See Appendix C for variable definitions. For ease of interpretation, each of these variables is ranked into deciles and scaled to range from 0 to 1. *t*-statistics appear in parentheses (two-tailed *p*-values appear in brackets) and are based on standard errors clustered by firm and date. \*\*\*, \*\*, and \* denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.



TABLE 3. Distribution of Estimated Latent Components of Linguistic Complexity

PANEL A. Descriptive Statistics

Variable	Mean	Std. Dev.	P25	Median	P75
<i>Obfu(Present)</i>	0.000	1.533	-0.978	0.028	1.026
<i>Info(Present)</i>	15.861	0.423	15.599	15.855	16.132
<i>Obfu(Response)</i>	0.000	1.521	-1.061	-0.122	0.913
<i>Info(Response)</i>	11.956	0.623	11.724	12.038	12.332

PANEL B. Correlation Matrix of Estimated Latent Components of Linguistic Complexity

Variable	(1)	(2)	(3)	(4)	(5)	(6)
<u>Presentation Portion</u>						
(1) <i>Fog(Present)</i>	1.00	<b>0.96</b>	0.25	0.37	0.32	0.19
(2) <i>Obfu(Present)</i>	<b>0.96</b>	1.00	0.00	0.31	0.34	0.02
(3) <i>Info(Present)</i>	0.27	0.00	1.00	0.25	0.02	<b>0.62</b>
<u>Response Portion</u>						
(4) <i>Fog(Response)</i>	0.38	0.32	0.26	1.00	<b>0.91</b>	0.42
(5) <i>Obfu(Response)</i>	0.34	0.35	0.00	<b>0.93</b>	1.00	0.09
(6) <i>Info(Response)</i>	0.19	0.00	<b>0.70</b>	0.38	0.00	1.00

This table presents descriptive statistics and correlations among the latent components of linguistic complexity. Panel A reports the distribution of our empirical estimates of the latent components of managers' linguistic complexity. Panel B reports the correlations among the latent components of linguistic complexity. *Fog(.)* is the Fog index of the respective portion of the conference call. *Obfu(.)* is the latent obfuscation component during the respective section of the call and *Info(.)* is the latent information component during the respective section of the call. Spearman (Pearson) correlations appear above (below) the diagonal and bold denotes correlations in excess of 0.5.

TABLE 4. Linguistic Complexity and Information Asymmetry

Variable	Linguistic Complexity		Linguistic Complexity Decomposed		Control for Additional Language Attributes	
	(1)	(2)	(3)	(4)	(5)	(6)
	coeff.	t-stat	coeff.	t-stat	coeff.	t-stat
<b>Linguistic Complexity</b>						
<i>Fog(Present)</i>	<b>0.01***</b>	<b>(3.41)</b>	.	.	.	.
<i>Fog(Response)</i>	<b>-0.01***</b>	<b>(-3.04)</b>	.	.	.	.
<b>Linguistic Complexity Decomposed into Latent Components</b>						
<i>Obfu(Present)</i>	.	.	<b>0.02***</b>	<b>(4.02)</b>	<b>0.01***</b>	<b>(2.95)</b>
<i>Obfu(Response)</i>	.	.	0.01	(1.58)	<b>0.01***</b>	<b>(4.07)</b>
<i>Info(Both)</i>	.	.	<b>-0.04***</b>	<b>(-8.56)</b>	<b>-0.03***</b>	<b>(-8.18)</b>
<b>Base Model Controls</b>						
<i>Size</i>	<b>-0.61***</b>	<b>(-73.75)</b>	<b>-0.62***</b>	<b>(-74.54)</b>	<b>-0.63***</b>	<b>(-78.98)</b>
<i>BM</i>	<b>0.03***</b>	<b>(6.27)</b>	<b>0.02***</b>	<b>(4.29)</b>	0.01	(1.19)
<i>Return</i>	<b>-0.10***</b>	<b>(-28.54)</b>	<b>-0.11***</b>	<b>(-29.59)</b>	<b>-0.10***</b>	<b>(-29.55)</b>
<i>IdioVol</i>	<b>0.10***</b>	<b>(13.92)</b>	<b>0.10***</b>	<b>(14.11)</b>	<b>0.10***</b>	<b>(14.51)</b>
<i>Coverage</i>	<b>-0.19***</b>	<b>(-29.08)</b>	<b>-0.19***</b>	<b>(-28.53)</b>	<b>-0.16***</b>	<b>(-25.82)</b>
<i>Dispersion</i>	<b>-0.04***</b>	<b>(-10.16)</b>	<b>-0.04***</b>	<b>(-10.06)</b>	<b>-0.05***</b>	<b>(-11.10)</b>
<i>MgmtForecast</i>	<b>-0.01***</b>	<b>(-6.19)</b>	<b>-0.01***</b>	<b>(-5.64)</b>	<b>-0.01***</b>	<b>(-3.56)</b>
<i>Surprise</i>	<b>-0.04***</b>	<b>(-16.96)</b>	<b>-0.04***</b>	<b>(-16.29)</b>	<b>-0.03***</b>	<b>(-14.16)</b>
<i>Loss</i>	<b>0.04***</b>	<b>(13.19)</b>	<b>0.05***</b>	<b>(14.06)</b>	<b>0.04***</b>	<b>(13.30)</b>
<i>SpecItems</i>	0.00	(-0.09)	0.00	(-1.06)	-0.00	(-0.47)
<i>SmallBeat</i>	<b>0.01***</b>	<b>(5.92)</b>	<b>0.01***</b>	<b>(6.15)</b>	<b>0.01***</b>	<b>(7.06)</b>
<b>Additional Controls for Characteristics of Presentation and Response Language</b>						
<i>Length(Present)</i>	.	.	.	.	<b>-0.06***</b>	<b>(7.45)</b>
<i>Jargon(Present)</i>	.	.	.	.	<b>0.02***</b>	<b>(2.91)</b>
<i>ForwardLook(Present)</i>	.	.	.	.	<b>-0.02***</b>	<b>(-4.64)</b>
<i>PostiveTone(Present)</i>	.	.	.	.	<b>-0.02***</b>	<b>(-4.22)</b>
<i>NegativeTone(Present)</i>	.	.	.	.	<b>0.06***</b>	<b>(14.06)</b>
<i>Length(Response)</i>	.	.	.	.	<b>-0.07***</b>	<b>(-9.07)</b>
<i>Jargon(Response)</i>	.	.	.	.	<b>0.01**</b>	<b>(2.09)</b>
<i>ForwardLook(Response)</i>	.	.	.	.	<b>-0.01**</b>	<b>(-2.33)</b>
<i>PostiveTone(Response)</i>	.	.	.	.	-0.01	(-1.21)
<i>NegativeTone(Response)</i>	.	.	.	.	<b>0.03***</b>	<b>(7.53)</b>
<i>N-obs / Adj R<sup>2</sup></i>	60,172 / 82.69		60,172 / 82.82		60,172 / 83.48	

This table presents results from estimating the relation between information asymmetry and the latent components of linguistic complexity. *Fog(.)* is the Fog index of the respective portion of the conference call, *Obfu(.)* is the latent obfuscation component, and *Info(.)* is the latent information component. *Info(Both)* is the first principal component of *Info(Present)* and *Info(Response)*. *Length(.)* is the number of words in the respective portion of the call. *Jargon(.)* is the number of words with more than two-syllables in the respective portion of the call that are industry jargon, where industry jargon is defined as the 100 most common words with more than two-syllables among firms in the same two-digit SIC code. *ForwardLook(.)* is the proportion of sentences containing forward-looking statements during the respective portion of the call (Li [2010]). *PositiveTone(.)* (*NegativeTone(.)*) is the number of positive (negative) tone words in the respective portion of the call (Loughran and McDonald [2011]). All other variables are as defined in Appendix C. For ease of interpretation, all independent variables are ranked into deciles and scaled to range from 0 to 1. *t*-statistics appear in parentheses and are based on standard errors clustered by firm and date. \*\*\*, \*\*, and \* denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.

TABLE 5. Linguistic Complexity and Information Asymmetry: Alternative Specifications

Variable	Controlling for Pre-existing Information Asymmetry	Industry Fixed Effects	Firm Fixed Effects	Firm + Manager Fixed Effects	Decomposition: Analyst Benchmark Only	Decomposition: Firm Characteristics Only
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Obfu(Present)</i>	<b>0.01</b> <sup>***</sup> (3.09)	<b>0.01</b> <sup>**</sup> (2.14)	0.003 (0.91)	0.001 (0.29)	<b>0.01</b> <sup>***</sup> (2.55)	<b>0.01</b> <sup>***</sup> (3.13)
<i>Obfu(Response)</i>	<b>0.01</b> <sup>***</sup> (4.07)	<b>0.01</b> <sup>***</sup> (3.02)	<b>0.02</b> <sup>***</sup> (7.18)	<b>0.02</b> <sup>***</sup> (4.92)	<b>0.01</b> <sup>***</sup> (3.85)	-0.001 (-0.25)
<i>Info(Both)</i>	<b>-0.01</b> <sup>***</sup> (-5.76)	<b>-0.05</b> <sup>***</sup> (-10.92)	<b>-0.04</b> <sup>***</sup> (-9.62)	<b>-0.04</b> <sup>***</sup> (-8.30)	<b>-0.02</b> <sup>***</sup> (-9.43)	<b>-0.02</b> <sup>***</sup> (-4.36)
<i>Illiquidity_Pre</i>	<b>0.59</b> <sup>***</sup> (94.96)					
Base Model Controls	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls for Characteristics of Presentation and Response Language	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	None	Industry	Firm	Firm + Manager	None	None
<i>N-obs / Adj R<sup>2</sup></i>	60,172 / 91.41	60,172 / 84.10	60,172 / 89.88	24,787 / 89.67	60,172 / 83.42	60,172 / 83.39

This table presents results from estimating the relation between the latent components of linguistic complexity and information asymmetry using four alternative regression specifications. Specification (1) presents results after control for the average value of the Amihud (2002) measure of illiquidity over the fiscal quarter ended immediately prior to the conference call (*Illiquidity\_Pre*). Specification (2) presents results from a within-industry design that includes industry fixed effects, where industries are defined based on two-digit SIC codes. Specification (3) presents results from a within-firm design that includes firm fixed effects. Specification (4) presents results from a within-manager design that includes both firm and manager fixed effects. This specification requires both the identity of the CEO (from Execucomp) and that the firm has multiple CEOs during our sample period. Specification (5) estimates the latent information and obfuscation components using only linguistic complexity of analysts as the benchmark, see specifications (1) and (4) of Table 2. Specification (6) estimates the latent information and obfuscation components using only firm characteristics as the benchmark, see specifications (2) and (5) of Table 2. All other variables are as defined in Appendix C. Regression specification follow column 3 of Table 4. For parsimony, coefficients on control variables are not tabulated. *t*-statistics appear in parentheses and are based on standard errors clustered by firm and date. \*\*\*, \*\*, and \* denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.

TABLE 6. Linguistic Complexity and Information Asymmetry: Monte Carlo Placebo Tests

PANEL A. Monte Carlo Placebo Test: Random Non-Conference Call Dates

Variable	Table 4	Avg. coeff	Diff
	Column 3	Random	(1) – (2)
	$\beta$	$E[\beta]$	$p$ -value
	(1)	(2)	(3)
<i>Obfu(Present)</i>	<b>0.011</b> ***	0.004	[<0.001]
<i>Obfu(Response)</i>	<b>0.013</b> ***	0.008	[<0.001]
<i>Info(Both)</i>	<b>-0.034</b> ***	-0.018	[<0.001]
Controls Included	Base Model Controls + Additional Controls for Characteristics of Presentation and Response Language		

PANEL B. Monte Carlo Placebo Test: Random Conference Call Dates

Variable	Table 4	Avg. coeff	Diff
	Column 3	Random	(1) – (2)
	$\beta$	$E[\beta]$	$p$ -value
	(1)	(2)	(3)
<i>Obfu(Present)</i>	<b>0.011</b> ***	0.010	[0.820]
<i>Obfu(Response)</i>	<b>0.013</b> ***	-0.002	[<0.001]
<i>Info(Both)</i>	<b>-0.034</b> ***	-0.015	[<0.001]
Controls Included	Base Model Controls + Additional Controls for Characteristics of Presentation and Response Language		

PANEL C. Monte Carlo Placebo Test: Random Decomposition

Variable	Table 4	Avg. coeff	Diff
	Column 3	Random	(1) – (2)
	$\beta$	Decomposition	$p$ -value
	(1)	$E[\beta]$	(3)
<i>Obfu(Present)</i>	<b>0.011</b> ***	0.000	[<0.001]
<i>Obfu(Response)</i>	<b>0.013</b> ***	0.000	[<0.001]
<i>Info(Both)</i>	<b>-0.034</b> ***	0.000	[<0.001]
Controls Included	Base Model Controls + Additional Controls for Characteristics of Presentation and Response Language		

This table presents results from three Monte Carlo placebo tests designed to compare our results against what one would expect under the null hypothesis. First, we estimate coefficients from a regression of information asymmetry on control variables and measures of linguistic complexity (e.g., column 3 of Table 4). Second, we compare test whether these estimated coefficients are different from those obtained from each of three Monte Carlo simulations ( $E[\beta]$ ). In Panel A, the simulation proceeds as follows. We randomly select a random date within a 250-day window

of the original conference call date, excluding dates within 25 trading days of the call. We then calculate each of our measures of information asymmetry assuming the random date is the conference call date, and re-estimate the regression. We repeat this step 1,000 times retaining coefficient estimates for each iteration. We then test whether estimated coefficients for the actual conference call date ( $\beta$ ) are significantly different from the average of the 1000 estimated coefficients for the randomly selected dates ( $E[\beta]$ ) using the empirical distribution of coefficients on random dates. In Panel B, the simulation proceeds similar to Panel A, except that we randomly select another conference call date for the same firm between 2002 and 2011. In Panel C the simulation proceeds as follows. For each call, we randomly select a value of  $Fog(Present)$  and  $Fog(Response)$  from the respective empirical distributions. We then estimate our decomposition for the randomly selected value of  $Fog(.)$  as in Table 2, and calculate each of  $Obfu(Present)$ ,  $Obfu(Response)$ , and  $Info(Both)$  on the randomly selected values. We then re-estimate the regression in column 3 of Table 4 for the random decomposition. We repeat this procedure 1,000 times retaining coefficient estimates for each iteration. We then test whether estimated coefficients for the actual decomposition of  $Fog(.)$  are significantly different from the average of the 1000 estimated coefficients from a decomposition of random  $Fog(.)$  using the empirical distribution of coefficients. For parsimony, coefficients on control variables are not tabulated.  $p$ -values (two-tailed) for the test  $\beta = E[\beta]$  appear in brackets, and \*\*\*, \*\*, and \* denote a statistical significant difference at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.

TABLE 7. Linguistic Complexity and Provision of Quantitative Guidance

Variable	Linguistic Complexity		Linguistic Complexity Decomposed		Controlling for Prior Quarter Guidance	
	coeff.	t-stat	coeff.	t-stat	coeff.	t-stat
Linguistic Complexity						
<i>Fog(Present)</i>	<b>-0.04**</b>	<b>(-1.97)</b>			.	.
<i>Fog(Response)</i>	0.00	(-0.04)			.	.
Linguistic Complexity Decomposed into Latent Components						
<i>Obfu(Present)</i>	.		<b>-0.06***</b>	<b>(-3.06)</b>	<b>-0.03***</b>	<b>(-3.58)</b>
<i>Obfu(Response)</i>	.		<b>-0.04***</b>	<b>(-2.79)</b>	-0.01	(-1.58)
<i>Info(Both)</i>	.		<b>0.08**</b>	<b>(4.34)</b>	<b>0.02**</b>	<b>(2.02)</b>
Base Model Controls						
<i>Size</i>	0.02	(0.70)	<b>0.05*</b>	<b>(1.91)</b>	0.01	(0.87)
<i>BM</i>	-0.01	(-0.59)	0.01	(0.50)	0.00	(-0.39)
<i>Return</i>	-0.01	(-1.58)	0.00	(-0.55)	0.00	(0.23)
<i>IdioVol</i>	<b>0.05***</b>	<b>(2.96)</b>	<b>0.05***</b>	<b>(2.81)</b>	<b>0.02**</b>	<b>(2.56)</b>
<i>Coverage</i>	<b>0.15***</b>	<b>(5.74)</b>	<b>0.14***</b>	<b>(5.24)</b>	<b>0.04***</b>	<b>(3.99)</b>
<i>Dispersion</i>	<b>-0.19***</b>	<b>(-11.39)</b>	<b>-0.20***</b>	<b>(-11.55)</b>	<b>-0.07***</b>	<b>(-10.08)</b>
<i>Surprise</i>	<b>0.09***</b>	<b>(9.74)</b>	<b>0.09***</b>	<b>(9.31)</b>	<b>0.03***</b>	<b>(6.18)</b>
<i>Loss</i>	<b>-0.08***</b>	<b>(-6.48)</b>	<b>-0.09***</b>	<b>(-6.94)</b>	<b>-0.03***</b>	<b>(-5.12)</b>
<i>SpecItems</i>	<b>-0.13***</b>	<b>(-11.27)</b>	<b>-0.12***</b>	<b>(-10.64)</b>	<b>-0.05***</b>	<b>(-8.58)</b>
<i>SmallBeat</i>	0.00	(0.46)	0.00	(0.37)	0.00	(0.96)
Additional Controls for Characteristics of Presentation and Response Language						
<i>Length(Present)</i>	<b>-0.13***</b>	<b>(-3.81)</b>	<b>-0.14***</b>	<b>(-4.08)</b>	<b>-0.04***</b>	<b>(-2.81)</b>
<i>Jargon(Present)</i>	<b>0.22***</b>	<b>(6.37)</b>	<b>0.23***</b>	<b>(6.74)</b>	<b>0.09***</b>	<b>(6.57)</b>
<i>ForwardLook(Present)</i>	<b>0.17***</b>	<b>(10.70)</b>	<b>0.18***</b>	<b>(11.23)</b>	<b>0.07***</b>	<b>(10.31)</b>
<i>PostiveTone(Present)</i>	<b>0.12***</b>	<b>(5.09)</b>	<b>0.12***</b>	<b>(4.95)</b>	<b>0.04***</b>	<b>(3.57)</b>
<i>NegativeTone(Present)</i>	<b>-0.05***</b>	<b>(-2.58)</b>	<b>-0.05**</b>	<b>(-2.50)</b>	<b>-0.03***</b>	<b>(-3.31)</b>
<i>Length(Response)</i>	0.02	(0.51)	0.01	(0.31)	0.01	(0.38)
<i>Jargon(Response)</i>	-0.04	(-1.51)	-0.03	(-1.08)	-0.01	(-0.39)
<i>ForwardLook(Response)</i>	0.01	(0.98)	0.02	(1.26)	0.00	(0.56)
<i>PostiveTone(Response)</i>	<b>0.15***</b>	<b>(6.87)</b>	<b>0.15***</b>	<b>(6.99)</b>	<b>0.05***</b>	<b>(4.23)</b>
<i>NegativeTone(Response)</i>	<b>-0.14***</b>	<b>(-7.61)</b>	<b>-0.13***</b>	<b>(-7.54)</b>	<b>-0.04***</b>	<b>(-4.91)</b>
Lagged Value of the Dependent Variable						
<i>MgmtForecast<sub>q-1</sub></i>					<b>0.66***</b>	<b>(105.74)</b>
<i>N-obs / Adj R<sup>2</sup></i>	60,172 / 9.07		60,172 / 10.09		57,513 / 48.99	

This table presents results from estimating the relation between the provision of quantitative guidance on the call and the latent components of linguistic complexity. The dependent variable is an indicator variable for whether the firm provides an earnings forecast during the conference call (*MgmtForecast<sub>q</sub>*). All other variables are as defined in Table 4. For ease of interpretation, all independent variables are ranked into deciles and scaled to range from 0 to 1. *t*-statistics appear in parentheses and are based on standard errors clustered by firm and date. \*\*\*, \*\*, and \* denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.

TABLE 8. Latent Components of Linguistic Complexity in Loss Firms

PANEL A. Loss Firms and the Latent Components of Linguistic Complexity

Sample partition: Variable	Profit Firms		Loss Firms		<i>p</i> -value: test of difference in means	<i>p</i> -value: test of difference in medians
	mean	median	mean	median		
<i>Fog(Present)</i>	15.754	15.794	16.396	16.426	[<0.001]	[<0.001]
<i>Obfu(Present)</i>	-0.049	-0.023	0.244	0.288	[<0.001]	[<0.001]
<i>Info(Present)</i>	15.803	15.812	16.152	16.184	[<0.001]	[<0.001]
<i>Fog(Response)</i>	11.906	11.788	12.201	12.093	[<0.001]	[<0.001]
<i>Obfu(Response)</i>	-0.038	-0.160	0.189	0.066	[<0.001]	[<0.001]
<i>Info(Response)</i>	11.944	12.023	12.012	12.109	[<0.001]	[<0.001]

PANEL B. Loss Firms and the Relation between Latent Components of Linguistic Complexity and Information Asymmetry

Dependent variable: <i>Illiquidity</i>			
Sample partition: Variable	Profit Firms	Loss Firms	<i>p</i> -value: test of difference in coefficients (1) – (2)
	(1)	(2)	(3)
<i>Obfu(Present)</i>	<b>0.012**</b> (2.83)	<b>0.015**</b> (2.13)	[0.68]
<i>Obfu(Response)</i>	<b>0.013***</b> (3.74)	<b>0.014***</b> (2.16)	[0.90]
<i>Info(Both)</i>	<b>-0.030***</b> (-6.93)	<b>-0.042***</b> (-6.02)	[0.10]
Base Model Controls	Yes	Yes	
Additional Controls for Characteristics of Presentation and Response Language	Yes	Yes	
<i>N-obs</i>	50,098	10,074	
<i>Adj R</i> <sup>2</sup>	82.69	73.30	

Panel A presents mean and median values of the latent components of linguistic complexity separately for profit and loss firms. Panel B presents results from estimating the relation between the latent components of linguistic complexity and information asymmetry separately for profit and loss firms. Regression specifications follow Table 4 and include all control variables. For parsimony, coefficients on control variables are not tabulated. *t*-statistics appear in parentheses (two-tailed *p*-values appear in brackets) and are based on standard errors clustered by firm and date. \*\*\*, \*\*, and \* denote statistical significance at the 0.01, 0.05, and 0.10 levels (two-tail), respectively.