



2019 - 100 pages - 2021.10.31

- \* → LIWC causal word list (Appendix B)
- ref. of management research on MD&A causal disclosure
- useful for motivation part
- useful steps / implementation details of preparing

(-) just count #  
of causal words  
not actually  
extract  
causal phrases

This item is the archived peer-reviewed author-version of:

- identification of MD&A section (begin + end)
- extraction of relevant sentences (sales, cost ...)

Causal language intensity in performance commentary and financial analyst behaviour

#### Reference:

Zhang Shuyu, Aerts Walter, Pan Huifeng.- Causal language intensity in performance commentary and financial analyst behaviour  
Journal of business finance & accounting - ISSN 0306-686X - 46:1-2(2019), p. 3-31  
Full text (Publisher's DOI): <https://doi.org/10.1111/JBFA.12351>  
To cite this reference: <https://hdl.handle.net/10067/1534050151162165141>



# Accepted Article

Causal language intensity in performance commentary and financial analyst behaviour<sup>1</sup>

SHUYU ZHANG<sup>2</sup>

University of International Business and Economics, School of Banking and Finance

Address: Chaoyang District Huixindong street 10, Beijing, China

WALTER AERTS

---

<sup>1</sup> We are grateful for valuable comments from the editor, the associated editor Steven Young and the anonymous referee.

<sup>2</sup> Correspondent author

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1111/jbfa.12351](https://doi.org/10.1111/jbfa.12351).

This article is protected by copyright. All rights reserved.

1. Antwerp University, Accountancy and Finance

Stadscampus Prinsstraat 13 S.B.327 2000 Antwerpen België

Tel. 032654110 walter.aerts@uantwerpen.be

2. University of Tilburg, School of Economics and Management Accountancy

Warandelaan 2 5037 AB Tilburg

HUIFENG PAN

University of International Business and Economics, School of Banking and Finance

Address: Chaoyang District Huixindong street 10, Beijing, China

**Abstract:** We use automated techniques to measure causal reasoning on earnings-related financial outcomes of a large sample of MD&A sections of US firms and examine the intensity of causal language in that context against extent of analyst following and against properties of analysts' earnings forecasts. We find a positive and significant association

This article is protected by copyright. All rights reserved.

between a firm's causal reasoning intensity and analyst following and analyst earnings forecast accuracy respectively. Correspondingly, analysts' earnings forecast dispersion is negatively and significantly associated with causal reasoning intensity. These results suggest that causal reasoning intensity provides incremental information about the relation between financial performance outcomes and its causes, thereby reducing financial analysts' information processing and interpreting costs and lowering overall analyst information uncertainty. Additionally, we find that decreases in analyst following are followed by more causal reasoning on performance disclosure. We also find that firms with a considerable increase of causal disclosure especially attract new analysts who already cover many firms. Overall, our evidence of the relation between causal reasoning intensity and properties of analyst behaviour is consistent with the proposition that causal reasoning is a generic narrative disclosure quality characteristic, able to provide incremental information to analysts and guide analysts' behaviour.

**Keywords:** causal reasoning, narrative disclosure, analyst following, dispersion, accuracy

**JEL Codes:** M41

## 1. Introduction

Explanation of performance outcomes is central to management commentary in annual reports. The Securities and Exchange Commission (SEC) and the International Accounting Standard Board (IASB) insist on providing meaningful causal explanation and related discussion of performance results in management commentary. The SEC has long argued that investor understanding should be fostered by providing meaningful context and interpretation of its performance, market position and progress within their Management Discussion and Analysis report (SEC, 1989, 2002). In its 'Practice Statement' on management commentary, the IASB (2010) voices similar arguments and contends that firms should use management commentary to include their own perspective on how their business is evolving with an analysis of how different factors are interacting to assist market participants in interpreting the financial statements and in comprehending their content in relation to management's objectives and strategies and action to achieve those objectives. Causal reasoning on performance is central in this respect (IASB, 2010). It should provide

incremental information by elaborating causes, reasons or motives, and other facilitating antecedents for results of operations which are not straightforwardly apparent from the financial statements themselves. Clarifying and complementing the numerical information and accompanying notes in the financial statements, these disclosures are generally seen as a useful extension of the financial reporting package (Baginski, Hassell, & Hillison, 2000; Clarkson, Kao, & Richardson, 1999; Cole & Jones, 2005).

Both the SEC and the IASB present causal reasoning as an overall quality of management commentary that underpins its usefulness to market participants. Their depiction of useful causal reasoning embraces a repertoire of performance explanation instances that may arise in management commentary<sup>3</sup>. We will investigate the assertion that

---

<sup>3</sup> Prior research has studied aspects of management's explanatory behaviour from an impression management perspective (Merkl-Davies et al., 2011). A basic assumption of our research is that impression management and information sharing are not necessarily mutually exclusive. Although, managerial impression management behaviour is often portrayed as deceptive self-presentation, impression management devices may also be used more innocently, as a natural tendency to resonate with an audience, or more proactively in order to signal management's

the intensity of causal language with regard to performance outcomes is relevant and useful to market participants by looking into the association of causal reasoning intensity and different characteristics of the behaviour of financial analysts as primary beneficiaries of firm-specific disclosures. Prior research documents that quality disclosure is able to attract more analysts and to positively affect the quality of their work. In concert with regulators' recommendations for more causal explanation, we argue that causal reasoning on performance has the capacity to significantly increase the precision of public information and decrease the costs of financial analysts' work, including their information processing, interpretation and private search costs. Both higher information precision and lower costs would enhance the analysts' ability to correctly impound all pertinent information in their recommendation and forecasting work. If public causal reasoning on performance is useful for analysts in these respects, we expect its intensity to be associated with more extensive

---

private information and, by doing so, reduce information asymmetries. It is our contention that impression management related devices in periodic accounting narratives need to subsume a substantial degree of truth and veracity to stay credible. Otherwise its persuasiveness would be limited.

analyst following, more consensus in their earnings forecasts, and more accurate earnings prediction.

We focus on the performance-related sections of the annual MD&A report of a large sample of US firms and use automated content analysis techniques to capture the extent and intensity of causal reasoning in these narratives. Our sample covers the period 1998 to 2012. We document a positive association between a firm's causal reasoning on performance and the number of analysts following the firm. We also find a positive relation between causal reasoning intensity and the consensus of analysts' earnings forecasts. The specific usefulness of more performance explanation for the quality of the analysts' work is further corroborated by a positive association between causal reasoning and the accuracy of earnings prediction.

In supplemental tests, we elaborate on the association between causal reasoning intensity and analyst following. We find that a decrease in analyst following is followed by more intense causal reasoning on performance. This relation seems to be more significant for thinly-covered firms for which the marginal value of analyst following would be larger. In

additional tests, we also estimate whether causal reasoning disclosure is more attractive to analysts who cover many firms. Our results suggest that firms with a considerable increase of causal disclosure especially attract new analysts who already cover many firms.

As these analysts can probably allocate less time to each individual firm, they may benefit more from causal reasoning intensity reducing information processing and interpretation time and related costs.

This study makes contributions along several dimensions. First, we extend the literature on causal reasoning in financial disclosure (Koonce, Seybert, & Smith, 2011) by measuring causal reasoning intensity as an overall characteristic of management commentary and investigating its relevance in large-scale archival research. Second, we add to the literature on voluntary disclosure and properties of analyst behaviour. Sedor (2002) and Kadous et al.

(2006) apply an experimental approach and study specific ways in which causal narratives affect analysts' decision making. The archival approach of our paper tends to complement and add to their work. Where Sedor (2002) and Kadous et al. (2006) demonstrate the relevance of causal narratives for analyst decision making in a highly controlled laboratory

setting, our archival approach is based on a large sample and helps to establish the empirical generalizability of the link between managerial explanation and analyst behaviour. It shows that the intensity of causal reasoning language in management commentary (including the types of causal scenario that Sedor (2002) and Kadous et al. (2006) use in their experimental setting) interacts with properties of analyst behaviour that are well-established in the empirical literature. We elaborate in what respects causal reasoning intensity is useful for analysts and how firms use causal reasoning intensity in response to analyst behaviour. Our investigation of the usefulness of causal reasoning in performance-related management commentary for market participants should be especially relevant for standard setters and regulators who have called for the provision of more causal context in their recommendations for more meaningful management commentary.

In our empirical analysis we are not able to disentangle incremental information provision and self-serving tendencies in causal disclosure. Indeed, the choice and selectivity of reported causal relations may introduce bias akin to impression management (self-serving attributional bias) or lead to purposeful obfuscation of reported performance. Yet, some recent research explicitly documents that impression management and information sharing

often coincide (Aerts & Yan, 2017; Solomon et al., 2013). Distinguishing these properties in discretionary disclosures is practically unfeasible. However, it is our contention that causal disclosures in periodic management commentary need to subsume a substantial degree of truth and veracity to be effective and stay credible, especially in the eyes of expert users such as financial analysts who process them in an intertemporal manner. Expected regulatory and legal costs are also likely to constrain manipulative behaviour in causal disclosure (Rogers, Van Buskirk, & Zechman, 2011).

In the remainder of this paper, we will first, in Section 2, review the relevant literature and develop hypotheses. In Section 3, we elaborate on our data and methods. Section 4 presents empirical results. Finally, we conclude in Section 5.

## 2. Literature review and hypothesis development

*Causal reasoning intensity as an overall disclosure quality characteristic*

**Causal disclosures**, linking corporate events and outcomes with internal and external causes, motives and facilitating factors, are generally seen as useful by policy makers (SEC, 1989, 2002; IASB, 2010) and market participants (Baginski *et al.*, 2000; Baginski, Hassell, & Kimbrough, 2004). The SEC has long recognized that financial statements as such may fall short in assisting investors “to judge the quality of earnings and likelihood that past performance is indicative of future performance” (SEC, 1989). Management commentary should provide the audience an opportunity to assess the firm “through the eyes of management” (SEC, 1989; IASB, 2010). In order to provide investors with more useful information, the SEC urges firms to focus on issues like liquidity, capital resources and performance outcomes (results of operations). Although the SEC and the IASB offer only broad recommendations on topics of interest and useful explanatory formats, they both tend to portray causal reasoning intensity as an overall quality of management commentary. Their description of useful causal reasoning can be conceived as a broad array of discursive

responses to the 'why' and 'how' questions that may arise with regard to the firm's financial position and performance outcomes as depicted in its financial statements. Content-wise, causal disclosure may elaborate on issues of strategic, moral, legal and practical conduct and refer to agency, strategy and plans, difficulties, responsibility, blame, and external constraints. In practice, managers frequently communicate using causal scenarios, linking past and current states, actions and plans and current or anticipated outcomes in a causal order (Sedor, 2002). To the extent that these causal disclosures are descriptive of actual causality, reflect management's private information and allow a view on management's business model reasoning, they enable refined inferences from financial statement figures and increase the precision of public information.

Potential performance explanations are plentiful, as they may relate to company-wide as well as to disaggregated segment and even product-level information. We will use the term 'causal reasoning' in this study for the entire spectrum of instances of performance explanation that arise in management commentary as a discursive context. As a general cognitive property of text, causal reasoning can use multiple linguistic and discursive formats, such as sequences of sentences, verbs, clausal structures, particles or indirectly

through descriptions of related events and outcomes (Antaki, 1994; Edwards & Potter, 1993).

Although firms are likely to have considerable leeway in constructing their management commentary, both content-wise and in terms of discursive formats, causal reasoning on performance is not costless (Clarkson *et al.*, 1999). Consequential costs of causality-based performance disclosures stem from the proprietary nature of the information, potential litigation risk and disclosure precedent avoidance (Aerts, Cheng, & Tarca, 2013). Causal disclosures may reveal proprietary information to competitors in the product markets. In addition, causal explanations may expose firms to regulatory response and legal action. Expected legal and regulatory costs constrain the use of causal explanations and promote more formal language in contextualizing performance (Aerts & Tarca, 2010). Their discretionary, selective and sometimes tentative and biased (e.g., by offering excuses for weak performance while taking credit for successes) nature could make causal disclosures vulnerable and litigable (Rogers *et al.*, 2011). Firms may also refrain from using causal explanations in order to avoid setting disclosure precedents that will be difficult to maintain in the future (Graham, Harvey, & Rajgopal, 2005).



Very useful referred point !!!

*Financial analysts as beneficiaries of causal disclosures*

Financial analysts are professional users of company information and act as intermediaries, receiving, processing and interpreting information for investors. The quality of their work is highly dependent on the quality of the information (both public and private) they use. As sophisticated processors of firm-specific information, they are expected to better understand the implications of narrative causal disclosure in management commentary, and, thus can be conceived as a primary target to examine the decision-usefulness of such disclosures. Prior research on analysts' decision inputs indicates that analysts rely extensively on qualitative and disaggregated information, with management communications being a primary source (Bowen, Davis, & Matsumoto, 2002; Lang & Lundholm, 1996; Previts, Briker, Robinson, & Young, 1994; Ramnath, Rock, & Shane, 2008; Rogers & Grant, 1997). More detailed discussion of operations and related financials is a key factor valued by analysts (Healy, Hutton, & Palepu, 1999). Sedor (2002) and Kadous, Krische and Sedor (2006) document that management's detailed causal narratives, elaborating causal

orderings between current states, planned actions and future outcomes (causal scenarios),  
are persuasive for analysts and affect their forecasts, suggesting that those causal  
narratives directly affect analysts' cognitive processing and attitudes.

#### *Causal reasoning intensity and properties of analyst behaviour*

Higher disclosure quality tends to have a positive impact on the accuracy and consensus  
of analysts' earnings forecasts. Analysts' earnings forecast accuracy and analysts' earnings  
forecast dispersion vary inversely with the quality of a large set of disclosures, including  
annual reports (Lang & Lundholm, 1996), press releases (Nichols & Wieland, 2009) and  
conference calls (Bowen *et al.*, 2002). The accuracy of analyst forecasts improves and the  
dispersion in their forecasts declines with higher SEC ratings of the quality of firms' MD&A  
disclosures (Barron, Kile, & O'keefe, 1999). Lehavy, Li, and Merkley (2011) find that firms  
with less readable annual reports have greater analysts' earnings forecast dispersion and  
less analysts' earnings forecast accuracy.

Research findings also document that higher disclosure quality tends to attract analysts and increase information intermediation (e.g., Lang & Lundholm, 1996; Lehavy *et al.*, 2011). Assuming that the extent of analyst following increases visibility, engenders more efficient information transmission, and eventually results in higher liquidity of the firm's stock and lower cost of capital, firms are expected to benefit from greater analyst following (Beyer, Cohen, Lys, & Walther, 2010). Volume and precision of information disclosure are likely to attract analysts' attention and increase their willingness to invest in following a firm (Botosan & Harris, 2000; Eccles, 2001; Nichols & Wieland, 2009; Yu, 2010).

It should be noted that we introduce causal reasoning intensity as an overall disclosure quality characteristic, and not as a readability characteristic. Causal reasoning intensity is related to disclosure understandability and reader comprehension, as causal frames are likely to create meaning for the reader, whereas readability is about reading ease with no relation to meaning creation. In that sense, the understandability and readability are two different issues with potentially quite different interpersonal consequences. Aerts and Yan (2017) show, for example, that rational explanation adds meaning and transparency by providing causal knowledge, but at the detriment of readability measures that have been

commonly used in prior narrative accounting research. Adding causal inferences and related explanations tends to lead to longer sentences and more complex textual relations and negatively affects readability as captured by the classic readability measures.

Moreover, higher causal reasoning intensity may increase the demand of analysts' work and reports. As causal reasoning intensity tends to increase the number and interrelatedness of informational cues in the disclosures (as well as decrease text readability), retail investors, lacking sophisticated earnings prediction models, may get confronted with a level of informational complexity that they can no longer cope with. Due to increasing informational complexity, and consistent with the findings of Lehavy et al. (2011), investors may rely more on analyst reports for firms with more causal reasoning disclosures. Consequently, the demand for analyst intermediation will increase for those firms.

### Hypotheses

We start our analyses by examining the relation between causal reasoning intensity and the number of analysts following a firm. To the extent causal reasoning intensity is a quality characteristic valued by analysts, management has incentives to offer more causal context

in management commentary to attract analysts and enjoy the benefits of increased information intermediation. Causal disclosures are argued to be attractive for analysts as these disclosures are likely to easily feed into analysts' information processing models, refine the precision of their inputs and projections and, ultimately, increase the efficiency of their work. As an aggregate disclosure quality proxy, more intense causal reasoning on performance outcomes will reduce the overall costs of analyst coverage, including information processing cost, interpreting cost and private search cost. Moreover, to the extent that more intense causal reasoning increases the complexity of management commentary and negatively affects the MD&A's readability for retail investors who lack appropriate information processing capacity, more causal disclosure increases the demand for analyst services and therefore analyst following (Lehavy *et al.*, 2011)<sup>4</sup>. Regardless of the firm's future prospects as determinants of a firm's attractiveness for an analyst, information

<sup>4</sup> Lehavy *et al.* (2011) show that less readable MD&A reports are associated with higher analyst following. They argue that more complex (less readable) reports will increase the demand for analyst services from retail investors. Lehavy *et al.* (2011) use the FOG index as a readability measure. This measure is essentially affected by the use of longer words and by the length of sentences used in textual discourse.

quality and increased demand for analyst services are likely to have a positive impact on the analyst's expected payoff from covering a firm. This brings us to hypothesize:

*Hypothesis 1 (H1): Causal reasoning intensity in the performance-related MD&A section of the annual report is positively associated with the number of analysts following the firm.*

Analysts have to take into account firms' operating uncertainties in forecasting future earnings. To the extent narrative causal disclosures mirror management's superior private information, they may be valuable to reduce analyst uncertainty and refine the precision of analyst projections. Baginski, Hassell, and Wieland (2011) find that explanations (especially external attributions) in management earnings forecasts enhance financial analysts' earnings forecast revisions, suggesting that those causal explanations are helpful in assessing the credibility of management's earnings claims.

Analyst forecast dispersion is generally used as a crude proxy of analyst confidence. It measures disagreement over common information and the effects of differences in private information among analysts and reflects the distribution of possible beliefs about expected earnings which an average analyst might hold (Baginski *et al.*, 2011). If causal reasoning on performance increases interpretation of firm-specific information and enhances information precision, it is to be expected that higher causal reasoning intensity results in a more level informational playing field among sophisticated users and tightens the distribution of analysts' beliefs. Moreover, if causal reasoning information reduces the analysts' costs of processing and interpreting firm disclosures, it will enhance their ability to correctly impound all pertinent information in their earnings assessments and lead to less equivocality and more agreement among analysts (Lehavy *et al.*, 2011). If causal reasoning intensity is instrumental in mitigating information asymmetry of analysts, then we hypothesize:

*Hypothesis 2 (H2): Causal reasoning intensity in the performance-related MD&A sections of the annual report is negatively associated with analysts' earnings forecast dispersion.*

If more causal reasoning on earnings-related issues makes it easier for analysts to forecast earnings, credible causal disclosures should also enhance the accuracy of the analyst consensus forecast as well. Although the effect of increased public disclosure can be expected to have a different impact on analyst forecast dispersion than on analyst forecast accuracy according to the level of analyst uncertainty (Barron *et al.*, 1999), additional credible public information signals are expected to affect both (Barron, Kim, Lim, & Stevens, 1998). We therefore hypothesize:

*Hypothesis 3 (H3): Causal reasoning intensity in the performance-related MD&A sections of the annual report is positively associated with analysts' earnings forecast accuracy.*

### 3. Data and content analysis

*Sample*

We investigate the MD&A section of the annual 10-K filing of US firms, covering a 15-year period from fiscal year 1998 to 2012. We collect our data from the following datasets: CRSP/COMPUSTAT Merge (CCM), CRSP monthly stock return file (annually updated), I/B/E/S database, and SEC Edgar. The MD&A sections of the 10K-filing are downloaded from SEC Edgar. Causal reasoning content in management commentary is identified and measured using automated text analysis procedures.

Our sample is based on all electronic 10-Ks filed from January 1, 1998 to December 31, 2013 and available in text format (99,874 in total). 10-K filings with less than 50 lines of text (7,317 in total) were dropped. Through Java programming we were able to extract 64,671 MD&A sections. 35,203 MD&A sections in 10-K filings could not be appropriately identified and extracted for the following reasons: (1) for 10,169 10-K filings, the program could only identify the ending of the MD&A section matching, but not the beginning of the MD&A section matching; (2) for 20,340 10-K filings, the program could identify the beginning of the

MD&A section matching, but not the ending of the MD&A section matching; and (3) the 10-K filing has a JPG (a picture file type) format or other format unreadable by our program (4,694)<sup>5</sup>. We use an identifier named 'CIK' to merge the 10-K filings with the Compustat/CRSP Merge database. This yielded 53,186 firm-year observations. Finally, we merged our database with the I/B/ES summary file<sup>6</sup>, yielding 34,264 firm-year observations for our final sample, including 4,938 observations for financial institutions<sup>7</sup>.

We use two methods to capture causal reasoning content: a word-based approach and a sentence-based approach. We measure 'causal reasoning word intensity' by counting the

---

<sup>5</sup> For further details on the identification and extraction process we refer to Appendix A which describes the automated text analysis procedures. Appendix A also elaborates how performance-related MD&A paragraphs are identified and how causal reasoning intensity is measured.

<sup>6</sup> We only use a one-year-ahead forecast horizon. As analyst forecast error tends to increase as the forecast horizon lengthens and given that we do not differentiate between different time horizons in causal explanation intensity in the main analyses, we expect causal explanation to be most beneficial for shorter time horizons.

<sup>7</sup> Dropping these observations did not materially affect our main empirical results.

relative frequency of causal reasoning words in the performance-related paragraphs of the MD&A. The identification of the causal reasoning words is based on a list of causal words used by Linguistic Inquiry and Word Count (LIWC), a widely-used text analysis program. The LIWC causal words list is included in Appendix B. Causal reasoning word intensity is measured as the number of causal reasoning words scaled by total number of words in the performance-related MD&A paragraphs, multiplied by 100. An alternative measure (referred to as 'causal reasoning sentence intensity') is based on a sentence as unit of analysis. A causal reasoning sentence is defined as a sentence that includes at least one of the causal words in the LIWC causal words list. Causal reasoning sentence intensity is measured as the number of causal reasoning sentences in the performance-related MD&A paragraphs scaled by the total number of sentences in the performance-related MD&A paragraphs, multiplied by 100 (for further details, see Appendix A). To test the validity of our causal reasoning sentence classification approach, two independent coders with an MBA level (unaware of the LIWC causal words list) were instructed to manually identify causal reasoning sentences among a sample of 4,000 randomly selected sentences. Coding instructions were broadly formulated, based on whether the meaning of a sentence related

to causality in terms of antecedent-consequence relations. Disagreements among coders were subject to a timed reconciliation with final arbitration by an independent coding supervisor. Following Manning and Schütze (1999), we use precision and recall measures to evaluate identification and classification performance and arrive at an overall retrieval accuracy score (measured by their F1 score) of 86.82%. *good enough!  
with experimental evidence!*

The word-based intensity measure will figure as the main proxy for causal reasoning on performance. We use the sentence-based causal reasoning intensity measure as a robustness check to take into account potential bias of the word-based measure. If a sentence contains several causal reasoning words, it would still be counted as only one incidence of causal reasoning in the sentence-based measure. Although the word intensity and the sentence intensity measure are highly correlated (near to 0.733), they may capture somewhat different aspects of how causal language is used in management commentary. The causal word intensity measure better captures the presence of multiple causal reasoning instances in one sentence or related to a single topic, whereas the causal sentence intensity measure gives more weight to the variety of explained performance

outcomes. Documenting both measures would corroborate the robustness of our empirical findings.

#### 4. Empirical models

We model the association between performance-related causal reasoning intensity and analyst following, analysts' earnings forecast dispersion, and analysts' earnings forecast accuracy as follows:

$$\begin{aligned} AnalystFollowing_{i,t}(Dispersion_{i,t}, Accuracy_{i,t}) = \\ \alpha + \gamma CausDisc_{i,t} + \beta_1 Fog_{i,t} + \beta_2 FirmSize_{i,t} + \beta_3 MD\&ASize_{i,t} + \beta_4 Market2Book_{i,t} + \\ \beta_5 LossFirm_{i,t} + \beta_6 ROE_{i,t} + \beta_7 IntangiblePercentage_{i,t} + \beta_8 Std(AdjustedStockReturn)_{i,t} + \\ \beta_9 Leverage_{i,t} + \beta_{10} AnalystFollowing_{i,t} + \sum \beta_i YearEffects_{i,t} + \sum \beta_j IndustryEffects_{i,t} + \varepsilon_{i,t} \end{aligned}$$

(Model 1: Baseline model)

The baseline model is examined using panel data on a sample ranging from 1998 to 2012<sup>8</sup>. We use causal reasoning word intensity as the main proxy for causal reasoning on performance. The number of analysts following a firm is used as a control variable in the models regressing analysts' earnings forecast dispersion, analysts' earnings forecast accuracy as dependent variables, but not in the model regressing extent of analyst following. Following prior work (Lehavy *et al.*, 2011; O'Brien, McNichols, & Lin, 2005), the extent of analyst following is the number of analysts following a firm in the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the 10-K filing. Analysts' earnings forecast dispersion is measured as the standard deviation of the individual analyst forecasts scaled by share price 90 days before the consensus forecast date. Analysts' earnings forecast accuracy is calculated as -1 times the squared difference between I/B/E/S reported earnings and the mean of analyst earnings forecasts scaled by the

---

<sup>8</sup> Adding the number of days between the 10-K filing date and the consensus forecast date as an additional control in supplementary tests did not materially affect our main findings.

share price 90 days before the consensus forecast date. Our key test variable, *CausDisc*,

refers to causal reasoning word intensity as previously defined.

Consistent with prior research on the relation between disclosure choices and analyst behaviour, we use the following variables as controls: text readability, loss firm, stock return volatility, firm size, MD&A size, leverage, market-to-book, ROE, intangible assets, and analyst following (Barth, Kasznik, & McNichols, 2001; Brown, 2001; Hope, 2003; Irani & Karamanou, 2003; Lang & Lundholm, 1996; Lehavy *et al.*, 2011)<sup>9</sup>. These variables proxy for the quality of the information environment, the inherent difficulty of earnings prediction and the demand and supply of management commentary.

We include the Fog index to control for the readability of the MD&A section (Li, 2008). The Fog index assumes that more syllables per word or more words per sentence make a text more difficult to read ( $\text{Fog index} = [\text{words per sentence} + \text{percent of complex words}] * 0.4$ , where complex words are defined as words with three syllables or more). A text is

---

<sup>9</sup> In supplementary tests (see later) we also seek to control for forward-looking content.

considered unreadable if the Fog index is equal or larger than 18 (Li, 2008). We control for loss firms as earnings of loss firms are generally more difficult to predict (Brown, 2001; Irani & Karamanou, 2003). Loss firm equals 1 if earnings before extraordinary items is less than 0, and zero otherwise. We include monthly return volatility as a measure of information uncertainty (Lang & Lundholm, 1996). Return volatility is measured as the standard deviation of the monthly adjusted stock returns in the prior year. We control for firm size as larger firms tend to have a better information environment, potentially more complex operations and greater demand for investment advice (Barth *et al.*, 2001; Bhushan, 1989; Lang & Lundholm, 1996; Lehay *et al.*, 2011; O'Brien & Bhushan, 1990). Firm size is measured as the natural logarithm of the firm's total assets. As our causal reasoning measure may be affected by other MD&A disclosures, we control for the MD&A size, defined as the natural logarithm of total number of words. We use leverage to control financial risk and measure it as long-term liabilities divided by total assets. We control for growth firms (and use market-to-book value as a proxy for growth opportunities), because growth firms may attract more analyst following (Barth *et al.*, 2001) while their future performance is more difficult to predict. Barth *et al.* (2001) find that firms with higher intangible assets are followed by more analysts.

Intangible percentage is measured as intangible assets scaled by total assets. All variables are winsorized at 1% and 99% to minimize the effect of extremely values. Moreover, all regression models include industry-fixed effects and year-fixed effects to capture variation in properties of analyst behaviour across industries and over time. We use the Fama-French industry classification (Lehavy *et al.*, 2011).

**Possible endogeneity: Reverse causality**

We expect that more causal reasoning leads to more analyst coverage, higher analyst earnings forecast accuracy, and less analyst earnings forecast dispersion. However, reverse causality, with analyst behaviour bringing firms to disclose more/less causal reasoning, may also apply (Muslu *et al.*, 2015). !

In the presence of possible endogeneity the use of traditional linear regression may not be appropriate as the OLS estimators become unreliable, inconsistent and inefficient (Hossain *et al.*, 2005). In order to control for this potential predicament and test the robustness of our results, we estimate a two-stage least-square (2SLS) specification in which causal reasoning intensity and properties of analyst behaviour are treated as jointly

determined endogenous variables. Next to the baseline mode, we estimate a causal reasoning whereby other variables on the right-hand side of the Baseline model are treated as exogenous variables.

We use CFO characteristics<sup>10</sup> as instrument variables, as these variables tend to be directly related with disclosure, but not with analyst behaviour. Prior research suggests that executives face considerable difficulties in assessing costs and benefits of disclosure (Clarkson et. al., 2008) and that they tend to rely on their own knowledge and experience in disclosure strategy decisions (Delmas and Toffel, 2008). Bamber *et. al.* (2010) find that top managers have an economically significant impact on their firms' voluntary disclosures. Personal background of executives plays an important role in disclosure style and affects disclosure quality (Plöckinger *et al.*, 2016).

We use CFO gender and CFO education as instruments. Prior research finds that gender relates to several aspects of accounting disclosure, such as tone, type, quality of

---

<sup>10</sup> Applying CEO characteristics as instruments do not materially affect the main results.

earnings, and openness to environmental issues. Gul *et al.* (2011) contend that firms with more gender-diverse boards make more price relevant disclosures to investors and, thus, create a richer information environment. Davis *et al.* (2015) find that female managers use less positive tone than their male counterparts during conference calls. Hollindale *et al.* (2017) find that firms with multiple female directors make more soft disclosures, such as information on vision and strategy, environmental profile and environmental initiatives. Female executives tend to report more conservatively and typically display more risk-averse accounting behaviour than their male counterparts. Krishnan and Parson (2008) find that gender diversity in senior management improves the quality of reported earnings. For example, firms with a female CFO have higher quality accruals (Barua *et al.*, 2010). Hollindale *et al.* (2017) find that firms with more than one female director make greenhouse gas emissions related disclosures that are of higher quality.

With regard to CFO education, prior research suggests that executives with a MBA degree tend to make different decisions than executives without such a degree (Finkelstein *et al.*, 2009). Lewis *et al.* (2014) find that CEOs with MBA degree are more likely to respond to the carbon

disclosure project. Graham and Harvey (2001) find that CFOs with MBA degree use more sophisticated valuation techniques than those without a MBA.

We first predict causal reasoning intensity in Model 2, and then we use the 'predicted causal reasoning intensity' to replace the 'causal reasoning intensity' in our Baseline model.

The causal reasoning model is specified as follows:

$$\begin{aligned} CausDisc_{i,t} = & \alpha + \gamma AnalystFollowing_{i,t}(Dispersion_{i,t}, Accuracy_{i,t}) + \beta_1 CFOGender_{i,t} + \\ & \beta_2 CFO_{MBAi,t} + \beta_3 Fog_{i,t} + \beta_4 FirmSize_{i,t} + \beta_5 MD\&ASize_{i,t} + \beta_6 Market2Book_{i,t} + \\ & \beta_7 LossFirm_{i,t} + \beta_8 ROE_{i,t} + \beta_9 IntangiblePercentage_{i,t} + \beta_{10} Std(AdjustedStockReturn)_{i,t} + \\ & \beta_{11} Leverage_{i,t} + \sum \beta_i YearEffects_{i,t} + \sum \beta_j IndustryEffects_{i,t} + \varepsilon_{i,t} \end{aligned}$$

(Model 2: Causal reasoning model)

Causal reasoning word intensity is the dependent variable in the causal reasoning models.

The variables for properties of analyst behaviour (analyst following, analysts' earnings forecast dispersion and analysts' earnings forecast accuracy) are used in separate models. CFO gender equals 1 if the CFO is female, otherwise 0. CFO MBA equals to 1 if the CFO holds a MBA degree or equivalent, otherwise 0. We add the following control variables: loss firm, standard deviation of monthly adjusted stock return in current year, firm size, MD&A size, leverage, market-to-book ratio, return on equity, intangible percentage.

Larger firm size is commonly associated with more extensive disclosure (Alsaeed, 2006; Wang & Claiborne, 2008), possibly because larger firms have lower information production costs (Lang & Lundholm, 1993). The market-to-book ratio is a proxy for growth potential and is measured as the market value of equity at the end of the fiscal year scaled by the book value of equity at that date. Brown and Tucker (2011) find that MD&A has become longer in the past decade. We use MD&A size to control this pattern. In addition, we use MD&A size to further control for the complexity of firm's operating environment (Li, 2010). We use an indicator variable for loss firms and financial leverage to control for financial risk. Prior research evidences a significant relation between disclosure and leverage (Singh & Zahn, 2008). Miller (2002) finds that managers' disclosure choice is related to a firm's performance. Therefore, we use ROE to measure profitability level. Adjusted stock return variability proxy for the demand for causal reasoning due to larger information uncertainty (Bhagat & Bolton, 2008). Additionally, all regression models include industry-fixed effects and year-fixed effects. Table 1 presents variable definitions of the dependent and independent variables used in the baseline and causal reasoning models.

<Insert Table 1 here>

## 5. Results

Table 2 presents descriptive statistics of the dependent and explanatory variables used in the empirical models. Average causal reasoning (word) intensity amounts to 2.17, meaning that performance-related management commentary contains, on average, 2.17% causal words from the LIWC causal words list. Average causal reasoning sentence intensity is 37.81%. Sample firms are, on average, followed by 6.85 analysts. The average analysts' earnings forecast dispersion is 0.04, and the average analysts' earnings forecast accuracy amounts to -0.03.

<Insert Table 2 here>

<Insert Table 3 here>

Table 3 presents a correlation matrix. Causal reasoning (word) intensity is significantly correlated with most variables. The correlation coefficient between causal reasoning intensity and analyst following is positive and significant. Causal reasoning word intensity is significantly and positively related to analysts' earnings forecast accuracy. The correlation coefficient of firm size and the causal reasoning measure is significantly negative. Causal reasoning is also negatively related to MD&A size.

<Insert Table 4 here>

Table 4 presents results (regression coefficients and corresponding t-value in parentheses) for the association of causal reasoning on performance (causal reasoning word intensity) and properties of analyst behaviour in a pooled regression. Overall, the results are consistent with hypotheses H1-H3. Model I (Table 4) shows that the regression coefficient for causal reasoning intensity on analyst coverage is positive and highly significant, suggesting that more intense causal reasoning on performance stimulates analyst coverage. This is consistent with H1. The coefficient amounts to 0.714, implying that an increase of one unit of standard deviation causal reasoning word intensity is associated with roughly a 0.54 increase in analyst following (an 8% increase)<sup>11</sup>. Analysts seem to value more causal performance disclosure, suggesting that firms may provide more causal performance explanations in order to attract and enjoy the benefits of more analyst following.

---

<sup>11</sup> Model I coefficient of causal reasoning word intensity (0.714) multiplied by 0.76 (standard deviation of causal reasoning word intensity).

Consistent with H2, Model II (Table 4) shows that causal reasoning intensity is negatively and significantly associated with analysts' earnings forecast dispersion, suggesting that more intense causal performance explanation is able to reduce analyst uncertainty and analysts' information search and processing costs. More intense causal reasoning on performance seems to enhance analysts' information base, create a more level playing field among them and result in less ambiguity and more agreement on future earnings among the analysts following a firm. The coefficient estimates suggest that an increase of one unit of standard deviation causal reasoning word intensity is associated with roughly a 5% decrease of analysts' earnings forecast dispersion. Finally, Model III (Table 4) offers evidence that causal reasoning intensity is positively and significantly related to analysts' earnings forecast accuracy, implying that credible causal reasoning enhances analysts' ability to correctly impound firm-specific information in their prediction models and, thus, increases the accuracy of their earnings forecast. The estimated coefficient is 0.006, suggesting that an increase of one unit of standard deviation causal reasoning word intensity is associated with roughly a 17% increase of analysts' earnings forecast accuracy. Taken together, these findings provide considerable evidence that causal reasoning intensity can be considered to

be a valued aggregate disclosure quality dimension and that it is incrementally informative and useful to analysts<sup>12</sup>.

The regression coefficient between causal reasoning sentence intensity and analyst following (Model IV, H1) is statistically positive and significant, as is the coefficient of forecast accuracy (Model VI, H3). Causal reasoning sentence intensity is negatively and significantly associated with analysts' earnings forecast dispersion (Model V, H2). The coefficients are 0.046, -0.0002, and 0.0004 in Model IV, V and VI respectively. This indicates that an increase of one unit of standard deviation causal reasoning sentence intensity is associated with a 0.46 increase in analyst following in Model IV (nearly 7% of the sample mean), a 0.002 decrease in analysts' earnings forecast dispersion in Model V (5% of the sample mean); and a 0.004 increase in analysts' earnings forecast accuracy (more than 13% of the sample median).

---

<sup>12</sup> When using fixed effect regression for the baseline model, results are not materially different.

In general, the coefficients on control variables are consistent with prior research.

Intangible percentage is positively and significantly associated with analyst following, and forecast accuracy, and negatively with forecast dispersion. Larger firms are more likely to have higher analyst following, less forecast accuracy, and less forecast dispersion. Monthly stock return volatility is positively and significantly associated with analyst following, indicating that firms with a less stable environment attract more analysts.

</Insert Table 5 here>

Table 5 shows the results of the second stage estimation where actual causal reasoning intensity is replaced by predicted causal reasoning intensity<sup>13</sup>. Predicted causal reasoning

---

<sup>13</sup> Firstly, in the first stage of Model I (analyst following), the F value is 111.31, partial R square is 0.065; the F value in the first stage of Model II (analysts' earnings forecast dispersion) is 82.08, partial R square is 0.048; the F value is 80.72, partial R square is 0.047 in the first stage of Model III (analysts' earnings forecast accuracy).

intensity is derived from the aforementioned causal reasoning model. In order to corroborate the validity of the instrument variable, we ran a number of tests (Harford *et al.*, 2014).

Table 5 reveals that correcting for potential simultaneity bias in this way does not affect our prior findings on the association between causal reasoning intensity and properties of analyst behaviour. Predicted causal reasoning on performance is still significantly and positively associated with analyst following and negatively with analysts' earnings forecast dispersion. The positive association between causal reasoning intensity and forecast accuracy also remains significant<sup>14</sup>.

---

These results indicate that the instrument is significant in explaining the endogenous variables. Secondly, we ran the weak identification test (Bendinelli *et al.*, 2016; Docquier *et al.*, 2008) - the Cragg-Donald Wald F statistic value in Model I, II, and III, are 111.31, 82.08, and 80.72. The null hypothesis is rejected for all three models. Therefore, we conclude that our instrument variable is not a weak instrument. Finally, we ran the Sargan test for all three Models and found that our instrument does not suffer from the overidentification problem (chi-square p-value < 0.05 in all Models).

<sup>14</sup> The magnitude of the coefficient increases substantially. As the first stage regression only captures a portion of the variation of the explanatory variable, its

## 6. Supplementary tests

### *Association between causal reasoning intensity and a decrease in analyst following*

Our findings corroborate that causal language intensity in performance commentary is positively related with extent of analyst following. This may imply that firms strategically use causal disclosures to manage the level of analyst monitoring (Arya & Mittendorf, 2007). Prior research shows that analysts' portfolios of followed firms are dynamic and that the composition of the analysts following a firm may change considerably over time (Hong & Kubik, 2003). Analysts routinely take decisions to initiate new coverage or to terminate prior coverage. These decisions may have considerable consequences at firm level. In this regard, Kecskés and Womack (2008) show that firms confronted with a decrease in analyst coverage experience negative returns, suggesting that a fallout in analyst following constitutes a significant performance predicament. Knowing that narrative behaviour, and

---

coefficient inflates accordingly in the second stage regression.

causal disclosures in particular, may be to a large extent defensive in nature and become more intense when accountability predicaments arise (Bloomfield, 2008; Firth, 1995; Merkl-Davies & Brennan, 2007; Sonenshein, Herzenstein, & Dholakia, 2011; Waring, 2007), firms may tend to provide more causal disclosures in a remedial mode in order to restore analyst interest and provide additional informational incentives to follow the firm. Relatedly, Anantharaman and Zhang (2011) find that managers increase the volume of disclosed guidance after a reduction in analyst following, consistent with the view that managers have incentives to increase disclosure level in order to preserve or restore prior analyst following.

To test this more dynamic scenario of the relation between causal disclosure and analyst following, we examine whether firms provide more causal reasoning on performance after a drop in analyst following. We also test whether the impact of a decrease in analyst following is different for a firm that is heavily covered by analysts, compared to a firm that is only thinly-covered. Leone (2004) provides anecdotal evidence that especially top managers in thinly-covered firms are under increasing pressure to keep their firms from losing analyst following. Moreover, heavily-covered firms are likely to experience a smaller positive stock price movement to the announcement of analyst following initiation, compared to thinly-covered

firms (Branson, Guffey, & Pagach, 1998). Anantharaman and Zhang (2011) find that managers use a higher amount of disclosure in response to declining analyst following, when the perceived value of analyst following to managers is greater, as in the case of firms that have become relatively thinly-covered. This evidence suggests that the marginal value of an additional analyst following the firm decreases in the extent of analyst following. In this regard, we explore whether the use of causal reasoning on performance as a response to a decline of analyst following differs according to the extent of analyst following. To do so, we introduce a dummy variable, ‘thinly covered’, which equals 1 if a firm’s analyst following is less than the median value and an interaction term between thinly-covered and causal disclosure. We expect that the association of the interaction term and a decline of analyst following is positive and significant. We use the following model to test the association:

$$\begin{aligned} \text{AnalystFollowingDecline}_{i,t} = & \alpha + \beta_1 \text{CausDisc}_{i,t} + \beta_2 \text{ThinlyCover}_{i,t} + \beta_3 (\text{Causal}_{i,t} \times \text{ThinlyCover}_{i,t}) \\ & + \beta_4 \text{MD\&ASize}_{i,t} + \beta_5 \text{FirmSize}_{i,t} + \beta_6 \text{Market-to-Book}_{i,t} + \beta_7 \text{Leverage}_{i,t} + \beta_8 \text{ROE}_{i,t} \\ & + \beta_9 \text{EarningsGrowth}_{i,t} + \beta_{10} \text{Profit2loss}_{i,t} + \beta_{11} \text{IntangiblePercentage}_{i,t} \\ & + \beta_{12} \text{Std(AdjustedStockReturn)}_{i,t} + \sum \beta_i \text{YearEffects}_{i,t} + \sum \beta_j \text{IndustryEffects}_{i,t} + \varepsilon_{i,t} \end{aligned}$$

(Model 4: Analyst following decline model)

Analyst following decline is calculated using the following procedure: first, we identify the analyst following record with a recording date in the I/B/E/S summary file which is closest to (but preceding) 90 days before the firm's annual report submit date ( $NUMESTi,t-1$ ); next, we pick the analyst following record with a recording date in the I/B/E/S summary file which is closest to (but after) the firm's annual report submit date in the prior year ( $NUMESTi,t-2$ ) and calculate the change in analyst following between these dates ( $NUMESTi,t-1$  minus  $NUMESTi,t-2$ ); if the difference is larger than or equal to 0, we set the variable equal to 0, otherwise it equals 1 (see also Figure 1). In this measurement, we use the window in order to (1) eliminate the impact of information coming out of the prior year annual report, and (2) to allow firms to be aware of the change in analyst following when elaborating the annual management commentary. A firm's performance deterioration may be another reason for an analyst following decline. In order to take this issue into account, we include two performance-related control variables: earnings growth rate and profit2loss. Earnings growth is measured as the earnings without extraordinary items growth rate. Profit2loss is an indicator variable measured as 1 if the firm made a profit in the previous year (year T-1), but suffers a loss in the current year (year T), and 0 otherwise. Profit or loss is measured as

earnings before extraordinary items. We also control for the year and industry fixed effects.

As additional controls, we include firm size, MD&A size, market-to-book, leverage, earnings growth, profit2loss, return on equity (ROE), intangible percentage and standard deviation of monthly adjusted stock return in current year.

*<Insert Table 6 here>*

Table 6 shows results of the logistic regression models testing the association between causal reasoning intensity and analyst following decline. Analyst following decline is positively and significantly associated with causal reasoning intensity, suggesting that managers consider the decline of analyst following as a performance predicament, bringing them to provide more causal disclosures in a remedial mode. The marginal effect of causal reasoning word intensity in Model I is 0.023. Moreover, the marginal effect of causal reasoning sentence intensity is 0.001 in Model II. Earnings growth is negatively related to analyst following decline, and Profit2loss is positively and significantly associated with

analyst following decline. The coefficient on the interaction term between causal reasoning intensity and thinly-covered firms is positive and significant. The marginal effect of interaction term is 0.018 in Model I, and 0.001 in Model II. Moreover, the magnitude and statistically significance of the interaction effect may vary by observation<sup>15</sup>. These results suggest that managers perceive the marginal value of analyst following to be lower when their firm is already followed by a larger number of analysts.

#### *Impact of analyst work load*

To further test whether analyst following in response to causal language intensity varies with analyst characteristics, we perform a first intuitive test to investigate whether an

---

<sup>15</sup> Ai and Norton (2003) argue that the magnitude of the interaction effect in a non-linear model is not the same as the marginal effect of the interaction term. The magnitude and statistical significance cannot be evaluated simply by sign, and the interaction effect varies by observation. Applying the Ai and Norton (2003) approach to investigate the interaction effect shows that 99.97% of the observations on the interaction term between causal reasoning word intensity and thinly-covered firm and 99.96% of the observations on the interaction term between causal reasoning sentence intensity and thinly-covered firm are positive.

analyst's work load in terms of the number of firms followed by the analyst makes a difference and, thus, examine whether causal reasoning disclosure is more attractive to analysts who already follow many firms. The dependent variable is a dummy variable measuring whether a firm increases causal reasoning from T-2 to T-1 by at least 10%. We require a 10% increase as we assume that analysts may not notice a minor change in narrative disclosure<sup>16</sup>. In Model I of Table 7, the dependent variable equals 1 if causal reasoning (word) intensity in T-2 is less than causal reasoning (word) intensity in T-1 by at least 10%, otherwise 0. In Model II of Table 7, the dependent variable equals 1 if causal reasoning (word) intensity in T-2 is less than causal reasoning (sentence) intensity in T-1 by at least 10%, otherwise 0. Next, we include 'New analyst status' as an extra independent variable. A new analyst is an analyst who did not cover the firm in prior years, but starts covering the firm in the current year. New analyst status is measured as the average number of firms that the analyst followed in the prior year. More analyst following equals 1 if a firm has more analyst following compared to the prior year, otherwise 0.

---

<sup>16</sup> We also applied 1%, 5%, 8%, 13%, 20% as a threshold and the significance level of our main results remained largely unchanged.

We use the Ai and Norton (2003) approach to assess the magnitude of the interaction effect in a non-linear model. The coefficient on the interaction term between “New analyst status” and “More analyst following” is positive and significant. The interaction effect of nearly all observations are positive and significant<sup>17</sup>. The marginal effect of interaction term between ‘new analyst status’ and ‘more analyst following’ in Model I is 0.001, while the marginal effect of interaction term in Model II is 0.002. These results suggest that causal reasoning intensity would be more beneficial (probably by reducing information processing and interpretation costs) to analysts who already follow many firms and, thus, can allocate less time to each individual firm.

*Forward-looking versus non-forward looking disclosures*

---

<sup>17</sup> More specifically, 96.39% of the observations on the interaction term between “New analyst status” and “More analyst following” with causal reasoning word intensity and 98.95% of the observations on the interaction term between “New analyst status” and “More analyst following” with causal reasoning sentence intensity are positive.

As a preliminary test to investigate whether forward-looking and non-forward looking causal disclosures equally affect analyst behaviour, we apply a machine learning technique to differentiate between both types of causal disclosure. A Naïve Bayesian model (Buehlmaier & Whited, 2016; Mitchell, 2006; Jiang, Wang & Zhang, 2016; Feng, Guo, Jing & Sun 2015; Huang, Zang & Zheng, 2014; Henry & Leone, 2015) is used to identify forward-looking causal reasoning and non-forwarding looking causal reasoning sentences<sup>18</sup>. Results (non-tabulated, but available from the authors on request) indicate that both forward-looking and

---

<sup>18</sup> Our preliminary Naïve Bayesian model includes three stages. First, we randomly pick 4,000 sentences from all causal sentences in our sample as a training set. We manually categorize these sentences into forward-looking and non-forward-looking causal reasoning sentences. Second, we estimate the probability of whether a causal sentence is forwarding-looking or not, by using the naïve Bayesian model in the training set. When using the training set, the Naïve Bayesian model can simulate the relation between the probability of forward-looking causal reasoning sentence and word appearances in order to predict the probability of forward-looking sentences in the full sample. Third, we use the fitted probability model to predict whether a causal reasoning sentence is forward-looking or not. In order to check the reliability of the model, we apply the training error validation method. Results of the reliability tests show that the model correctly classifies forward-looking causal reasoning sentences with a probability of 90.65%.

non-forward-looking causal reasoning disclosure are likely to provide incremental information and to facilitate analysts' earnings forecasting. These results are, however, preliminary and should be treated with caution due to the limits of the machine learning procedures applied (relatively low number of sentences of the training sample and absence of a validity test in a holdout sample).

## 7. Concluding remarks

Prominent regulators and standard setters have called for increased use of causal disclosures on business performance as an overall quality of management commentary to improve investor understanding and thereby increase the usefulness of financial reporting to market participants. There is, however, no strong empirical evidence to show that causal reasoning intensity in management commentary is really beneficial to investors and analysts in particular. This research partly fills this gap by measuring causal reasoning on performance as an overall disclosure quality characteristic of the MD&A section of a large

sample of US firms and by relating the intensity of causal reasoning language to properties of analyst behaviour (analyst following, analyst earnings forecast dispersion and analyst earnings forecast accuracy).

We use automated techniques to measure causal reasoning on earnings-related performance outcomes in the MD&A section of a large sample of 10-K filings of US firms and examine the intensity of causal language in that context against extent of analyst following and properties of analysts' earnings forecasts. We find a positive and significant association between a firm's causal reasoning intensity and analyst following and analyst earnings forecast accuracy respectively. Correspondingly, analysts' earnings forecast dispersion is negatively and significantly associated with causal reasoning intensity. These results suggest that causal reasoning intensity provides incremental information about the relation between performance outcomes and its causes, thereby reducing financial analysts' information processing and interpreting costs and lowering overall analyst information uncertainty. In addition, our results indicate that analysts not only perceive causal reasoning on performance as incrementally informative, but, considering the consistent relation between causal reasoning intensity and analyst forecast accuracy, also deem causal

reasoning on performance to be truthful incremental information that adds to real understanding of how internal and external factors affect current and future earnings performance. Although our empirical design does not enable us to directly distinguish deceptive causal disclosure (potentially leading to obfuscation of reported performance) and truthful causal reasoning on performance, the strength and consistency (among the three main properties of analyst behaviour) of our results lend support to an information sharing view, with causal disclosures effectively reducing information asymmetry.

We also investigate a more dynamic relation between the discretionary use of more causal reasoning and analyst following by examining whether a decline in analyst following would incline firms to provide more causal performance context. The results of these supplementary tests indicate that decreases in analyst following are indeed followed by more causal reasoning on performance disclosure and suggest that managers view decreases in analyst following as an information environment predicament to which they are likely to respond by providing more causal performance explanations. This relation shows to be more significant for thinly-covered firms for which the marginal value of analyst following is higher.

We also provide some first analyses to investigate whether analyst characteristics affect how analyst cope with causal disclosures. Specifically, we test whether causal reasoning disclosure is more attractive to analysts who already follow many firms. Our results suggest that analysts who already follow many firms seem to benefit more from more intense causal disclosures. These preliminary results on the effect of analyst characteristics could be further elaborated in future research.

Overall, the evidence from our various investigations of causal reasoning intensity and properties of analyst behaviour is consistent with the proposition that causal reasoning intensity is a generic narrative disclosure quality characteristic, able to provide incremental information to analysts and possibly guide analyst behaviour.

While we seek to make causal inferences in this research, the research design is such that these inferences should be interpreted with a degree of caution. Although we try to control for potential endogeneity bias in our empirical models by testing a simultaneous estimation system as a robustness check, the findings coming out of this research may still

be subject to the limitation that the documented relations only reflect association and not necessarily causality.

## Appendix A - Measuring causal reasoning on performance

We use the following procedures to identify and measure causal reasoning in firms' relevant management commentary sections. In this study, we use Java and PERL to (1) download 10-K filings from the SEC Edgar database, (2) extract relevant Management Discussion and Analysis (MD&A) sections from the filing, (3) extract performance-related paragraphs, and (4) measure causal reasoning on performance content. In the following, we describe the content analysis process by elaborating these four steps.

### Step 1 - Downloading the electronic 10-K filings

We use Java to download the 10-K filings (annual reports) of our sample firms for the period 1997-2009 from the SEC Edgar ftp website<sup>19</sup>. We use an external open source java

---

<sup>19</sup> <http://ftp.sec.gov>

package, the FTP<sup>20</sup> Client Apache package<sup>21</sup>, in this phase of programming. We first download the company index files from SEC Edgar. Each company index file contains at least firm name, central index key (*i.e.* CIK), report type and report URL. We filter out all information relating to the 10-K filing (*i.e.* company name, URL, and CIK). The CRSP/COMPUSTAT Merge (CCM) database and the SEC Edgar data both include the central index key (*i.e.* CIK code) as an identifier. In general, each firm has an unique CIK code. We retain the CIK code for subsequent data merge purposes. Finally, we download the 10-K filing using the company index file.

*Step 2 - Extracting the MD&A sections from the 10-K filings*

Consistent with Li (2008), we proceed as follows:

---

<sup>20</sup> FTP is an acronym for File Transfer Protocol. FTP is used to transfer files between computers on a network. You can use FTP to download files from remote computer accounts.

<sup>21</sup> <http://www.apache.org/licenses/LICENSE-2.0>

a) Transfer of html-language to plain language text - First, we extract 'central index key',

'conformed period of report' (refers to fiscal year-end date), and 'filed as of date' (refers to

10-K filing submitting date) from each 10-K filing. Second, all tables that begin with

<TABLE><sup>22</sup> and end with </TABLE> are deleted<sup>23</sup>; all paragraphs that contain <S> or <C>

are deleted as well. Third, we replace html-language format '&nbsp;' with blanks and remove

other html-language formatting<sup>24</sup>. Moreover, 10-K filings containing less than 50 lines are

dropped (7,317 in total). Finally, we reprocess each file by reading paragraph by paragraph

---

<sup>22</sup> All html tag language is matched on a case-insensitive basis.

<sup>23</sup> Li (2008) notes that <S> and <C> html tags are used by some firms to present tables

<sup>24</sup> Li (2008) replaces all html tags with blanks. We believe that our treatment will deliver similar results as Li (2008), since we already replaced '&nbsp;' with blanks. Moreover, it is possible that some files may use html tags to decorate the MD&A section header. If we replace these html tags with blanks, it may cause a defective matching pattern. For example, if we replace in the extract '<B>M</B>anagement's discussion and analysis...' the html tag (*i.e.* <B> and </B>) with blanks, the line would become: 'Management's discussion and analysis...', whereas if we remove the html tag, the line would be read as: 'Management's discussion and analysis...'.

(*i.e.* the second round), to make sure that all tables, tabulated text or financial statements are excluded. In this round, all paragraphs with more than 50% of non-alphabetic characters (*e.g.* white spaces or numbers) are deleted.

b) Extract the MD&A section - Within the remaining text, the program first removes the leading and trailing blanks in each line. The MD&A section is the content lying in between the starting matching content (*i.e.* the ‘starting point’) and the ending matching content (*i.e.* the ‘ending point’). The program identifies a line that satisfies one of the following criteria as the beginning of the MD&A section (*i.e.* starting point): (1) the line starts with ‘management’s discussion’; (2) the line contains both of ‘management’s discussion’ and (*‘item’ + one or more white space + ‘7’*) or (*‘item’ + one or more white space + ‘6’*)<sup>25</sup> and does not contain the word ‘see’, ‘refer to’, or ‘refers to’; (3) the line starts with ‘management’s discussion’; or (4) the line contains ‘managements discussion’ and (*‘item’ + one or more white space + ‘7’*) or (*‘item’ + one or more white space + ‘6’*) and does not contain the words or phrases: ‘see’,

---

<sup>25</sup> We add the ‘item 6’ into the beginning matching pattern, because some firms present their MD&A section in item 6.

'refer to', or 'refers to'. We save the matching content of the beginning MD&A section (*i.e.* starting point).

The program identifies a line that satisfies one of the following criteria as the ending of the MD&A section (*i.e.* ending point) (1) the line begins with 'Financial Statements'; (2) the line contains 'item' + one or more white spaces + '8' and the matching content of the beginning of the MD&A section (*i.e.* starting point) does not contain '6'; (3) the line contains 'Supplementary Data'; or (4) the line begins with 'SUMMARY OF SELECTED FINANCIAL DATA'; (5) the line contains 'item' + one or more white spaces + '7' and the line does not contain 'management' and the beginning of the matching content contains '6'. The selected MD&A section is the content lying in between the starting matching content and the ending matching content. This procedure failed to extract the MD&A sections of 35,203 10-K filings for the following reasons: (1) for 10,169 10-K filings, the program could only identify the ending of the MD&A section matching, and not the beginning of the MD&A section matching; (2) for 20,340 10-K filings, the program could identify the beginning of the MD&A section matching, but not the ending of the MD&A section matching; and (3) the 10-K filing has a JPG (a picture file type) format or other format unreadable by our program (4,694)

This article is protected by copyright. All rights reserved.

### Step 3 - Extracting financial performance related paragraphs from the MD&A section

First, we split each identified management commentary text into paragraphs. Second, we identify whether the paragraph contains performance-related content, based on a dictionary of 'financial performance' items. If a paragraph includes at least one of the words in the performance wordlist, it is retained for further analysis. The dictionary list contains expense-related words and income-related words. The expense-related wordlist includes the following words or phrases: amortization, cost, depreciation, disposition, expense, research and development, R&D, impairment, loss, write off. The income-related word list includes earnings, EBIT, EBITDA, income, sale, revenue, profit, margin, benefit, break-even, contribution, EPS, and return. Before identifying causal reasoning content, we correct for dot-signs such as: 'i.e.' to 'ie', 'U.S.' to 'US', 'No.' to 'number', 'Corp.' to 'corporation', 'et al.' to 'et al'.  
*preprocessing*

### Step 4 - Measuring causal reasoning content

For each performance-related paragraph identified in the previous step, we apply two methods to measure causal reasoning content: a word-based approach and a sentence-

based approach. We measure 'causal reasoning word intensity' by counting the relative frequency of causal reasoning words in the performance-related paragraphs of the MD&A.

The identification of the causal reasoning words is based on a list of causal words used by

LIWC (Linguistic Inquiry and Word Count). LIWC is a text analysis software program designed by James W. Pennebaker, Roger J. Booth, and Martha E. Francis<sup>26</sup>. It is an automated content analysis software for the purpose of analyzing linguistic features of text.

The software<sup>27</sup> processes text on a word-by-word basis and calculates the number of words that match pre-defined word categories (Merkl-Davies, Brennan, & McLeay, 2011; Pennebaker, Chung, Ireland, Gonzales, & Booth, 2007; Pennebaker & Francis, 1996).

According to LIWC, insertion of causal words in a sentence has a substantial impact on

---

<sup>26</sup> <http://www.liwc.net/>

<sup>27</sup> The LIWC Dictionary is composed of 2,290 words. Each word or word-stem defines one or more word categories or sub-dictionaries. For example, the word 'cried' is part of four word categories: sadness, negative emotion, overall affect, and a past tense verb. Hence, if it is found in the target text, each of these four sub-dictionary scale scores will be incremented. Many of the LIWC categories are arranged hierarchically.

comprehension and memory for text. This is consistent with prior research showing that the number of causal connections and causal chain membership is positively related with recall, summarization and judged importance of text messages. The LIWC causal words list is presented in Appendix B.

An alternative measure (referred to as 'causal reasoning sentence intensity') is based on a sentence<sup>28</sup> as unit of analysis. A causal reasoning sentence is defined as a sentence that includes at least one of the causal words in the LIWC causal words list.

---

<sup>28</sup> A sentence is identified on the occurrence of a dot '.', after the procedure explained in step 3 of the content analysis procedures.

## Appendix B - LIWC causal words list

<Insert Table Appendix B here>

## Appendix C - Examples of causal reasoning sentences, highlighting the causal reasoning words

*File name in SEC Edgar: 0000002178-09-000007.txt*

Absent the inventory items, crude oil earnings from operations were reduced in 2008 as a  
result of escalated prices for the diesel fuel consumed in the trucking function of this  
business.

Refined product driven operating earnings were reduced during 2008 because of an  
increased allowance for doubtful accounts receivable through a bad debt charge of \$.

Crude oil operating earnings improved in 2007 relative to 2006 because of the \$ million in inventory liquidation gains coupled with improved end-market pricing received from the companies customers relative to crude oil acquisition costs.

*File name in SEC Edgar: 0000849343-97-000004.txt*

Because of the strong positive acceptance of the Company's product by the marketplace, the Company was unable to meet the demand during the past hurricane season due to production capacity limits.

All of these areas are susceptible to hurricanes and therefore have a need for the Company's products.

## References

- Aerts, W., Cheng, P., & Tarca, A. (2013). Management's earnings justification and earnings management under different institutional regimes. *Corporate Governance: An International Review*, 21(1), 93-115.

- Aerts, W., & Tarca, A. (2010). Financial performance explanations and institutional setting. *Accounting and Business Research*, 40(5), 421-450.
- Aerts, W., & Yan, B. (2017). Rhetorical impression management in the letter to shareholders and institutional setting: A metadiscourse perspective. *Accounting, Auditing & Accountability Journal*, 30, 404-432.
- Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123-129.
- Alsaeed, K. (2006). The association between firm-specific characteristics and disclosure: The case of Saudi Arabia. *Managerial Auditing Journal*, 21(5), 476-496.
- Anantharaman, D., & Zhang, Y. (2011). Cover me: Managers' responses to changes in analyst coverage in the post-regulation FD period. *The Accounting Review*, 86(6), 1851-1885.
- Antaki, C. (1994). *Explaining and arguing: The social organization of accounts*. In Antaki, C. (eds). London: Sage Publications Ltd.

- Arya, A., & Mittendorf, B. (2007). The interaction among disclosure, competition between firms, and analyst following. *Journal of Accounting and Economics*, 43(2), 321-339.
- Baginski, S. P., Hassell, J. M., & Hillison, W. A. (2000). Voluntary causal disclosures: Tendencies and capital market reaction. *Review of Quantitative Finance and Accounting*, 15(4), 371-389.
- Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. (2004). Why do managers explain their earnings forecasts? *Journal of Accounting Research*, 42(1), 1-29.
- Baginski, S. P., Hassell, J. M., & Wieland, M. M. (2011). An examination of the effects of management earnings forecast form and explanations on financial analyst forecast revisions. *Advances in Accounting*, 27(1), 17-25.
- Bamber, L. S., Jiang, J., & Wang, I. Y. (2010). What's my style? the influence of top managers on voluntary corporate financial disclosure. *The Accounting Review*, 85(4), 1131-1162.

Barron, O. E., Kile, C. O., & O'keefe, T. B. (1999). MD&A Quality as Measured by the SEC

and Analysts' Earnings Forecasts. *Contemporary Accounting Research*, 16(1), 75-109.

Barron, O. E., Kim, O., Lim, S. C., & Stevens, D. E. (1998). Using analysts' forecasts to

measure properties of analysts' information environment. *The Accounting Review*, 73(4), 421-433.

Barth, M. E., Kasznik, R., & McNichols, M. F. (2001). Analyst coverage and intangible

assets. *Journal of Accounting Research*, 39(1), 1-34.

Barua, A., Davidson, L. F., Rama, D. V., & Thiruvadi, S. (2010). CFO gender and accruals

quality. *Accounting Horizons*, 24(1), 25-39.

Bendinelli, W. E., Bettini, H. F. A. J., & Oliveira, A. V. M. (2016). Airline delays, congestion

internalization and non-price spillover effects of low cost carrier entry. *Transportation Research Part A Policy & Practice*, 85, 39-52.

- Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, 50(2), 296-343.
- Bhagat, S., & Bolton, B. (2008). Corporate governance and firm performance. *Journal of Corporate Finance*, 14(3), 257-273.
- Bhushan, R. (1989). Firm characteristics and analyst following. *Journal of Accounting and Economics*, 11(2-3), 255-274.
- Bloomfield, R. (2008). Discussion of annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, 45(2-3), 248-252.
- Botosan, C. A., & Harris, M. S. (2000). Motivations for a change in disclosure frequency and its consequences: An examination of voluntary quarterly segment disclosures. *Journal of Accounting Research*, 38(2), 329-353.
- Bowen, R. M., Davis, A. K., & Matsumoto, D. A. (2002). Do conference calls affect analysts' forecasts? *The Accounting Review*, 77(2), 285-316.

- Branson, B. C., Guffey, D. M., & Pagach, D. P. (1998). Information conveyed in announcements of analyst coverage. *Contemporary Accounting Research*, 15(2), 119-143.
- Brown, L. D. (2001). A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research*, 39(2), 221-241.
- Brown, S. V., & Tucker, J. W. (2011). Large-sample evidence on firms' year-over-year MD&A modifications. *Journal of Accounting Research*, 49(2), 309-346.
- Buehlmaier, M. M., & Whited, T. M. (2016). Are financial constraints priced? Evidence from textual analysis. Working paper, Available at:  
<http://dx.doi.org/10.2139/ssrn.2435116>.
- Clarkson, P. M., Kao, J. L., & Richardson, G. D. (1999). Evidence that management discussion and analysis (MD&A) is a part of a firm's overall disclosure package. *Contemporary Accounting Research*, 16(1), 111-134.

- Clarkson, PM, Li, Y, Richardson, GD, Vasvari, FP (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, Organizations and Society* 33(4–5): 303–327.
- Cole, C. J., & Jones, C. L. (2005). Management discussion and analysis: A review and implications for future research. *Journal of Accounting Literature*, 24(ABI/INFORM Global), pg. 135.
- Davis, A. K., Ge, W., Matsumoto, D., & Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls. *Social Science Electronic Publishing*, 20(2), 639-673.
- Delmas, MA, Toffel, MW (2008). Organizational responses to environmental demands: opening the black box. *Strategic Management Journal* 29(10): 1027–1055.
- Docquier, F., Faye, O., & Pestieau, P. (2008). Is migration a good substitute for education subsidies? *Journal of Development Economics*, 86(2), 263-276.

- Eccles, R. G. (2001). *The value reporting revolution: Moving beyond the earnings game.* In (eds). John Wiley & Sons Inc.
- Edwards, D., & Potter, J. (1993). Language and causation: a discursive action model of description and attribution. *Psychological Review, 100*(1), 23-41.
- Feng, G., Guo, J., Jing, B. Y., & Sun, T. (2015). Feature subset selection using naive Bayes for text classification. *Pattern Recognition Letters, 65*, 109-115.
- Finkelstein, S, Hambrick, DC, Cannella, AA (2009). Strategic leadership: theory and research on executives, top management teams, and boards. Oxford University Press: New York, NY.
- Firth, A. (1995). 'Accounts' in negotiation discourse: A single-case analysis. *Journal of Pragmatics, 23*(2), 199-226.
- Graham, JR, Harvey, CR (2001). The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics 60*(2–3): 187–243.

- Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, 40(1-3), 3-73.
- Gul, F. A., Hutchinson, M., & Lai, K. M. Y. (2013). Gender-diverse boards and properties of analyst earnings forecasts. *Accounting Horizons*, 27(3), 511-538.
- Harford, J., Klasa, S., & W. F. M. (2014). Refinancing risk and cash holdings. *Journal of Finance*, 69(3), 975–1012.
- Healy, P. M., Hutton, A. P., & Palepu, K. G. (1999). Stock performance and intermediation changes surrounding sustained increases in disclosure. *Contemporary Accounting Research*, 16(3), 485-520.
- Henry, E., & Leone, A. J. (2015). Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. *The Accounting Review*, 91(1), 153-178.
- Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, 89(6), 2151-2180.

- Hollindale, J., Kent, P., Routledge, J., & Chapple, L. (2017). Women on boards and greenhouse gas emission disclosures. *Accounting & Finance*.
- Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance*, 58(1), 313-351.
- Hope, O. K. (2003). Disclosure practices, enforcement of accounting standards, and analysts' forecast accuracy: An international study. *Journal of Accounting Research*, 41(2), 235-272.
- Hossain, M., Ahmed, K., & Godfrey, J. (2005) Investment opportunity set and voluntary disclosure of prospective information: A simultaneous equations approach. *Journal of Business Finance and Accounting*, 32(5), 871-907.
- International Accounting Standards Board (IASB) (2010). *Management Commentary*. IFRS Practice Statement. London, IFRS Foundation.
- Irani, A. J., & Karamanou, I. (2003). Regulation Fair Disclosure, analyst following, and analyst forecast dispersion. *Accounting Horizons*, 17(1), 15-30.

- Jiang, L., Li, C., Wang, S., & Zhang, L. (2016). Deep feature weighting for naive Bayes and its application to text classification. *Engineering Applications of Artificial Intelligence*, 52, 26-39.
- Kadous, K., Krische, S. D., & Sedor, L. M. (2006). Using counter-explanation to limit analysts' forecast optimism. *The Accounting Review*, 81(2), 377-397.
- Kecskes, A., & Womack, K. (2008). Adds and drops of coverage by equity research analysts. *Working Paper, Available at:*  
[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=960501](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=960501).
- Krishnan, G. V., & Parsons, L. M. (2008). Getting to the bottom line: an exploration of gender and earnings quality. *Journal of Business Ethics*, 78(1-2), 65-76.
- Koonce, L., Seybert, N., & Smith, J. (2011). Causal reasoning in financial reporting and voluntary disclosure. *Accounting, Organizations and Society*, 36(4-5), 209-225.
- Lang, M., & Lundholm, R. (1993). Cross-sectional determinants of analyst ratings of corporate disclosures. *Journal of Accounting Research*, 31(2), 246-271.

Lang, M. H., & Lundholm, R. J. (1996). Corporate disclosure policy and analyst behavior.

*The Accounting Review*, 71(4), 467-492.

Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst

following and the properties of their earnings forecasts. *The Accounting Review*,  
86(3), 1087-1115.

Leone, M. (2004). The flight of the sell-side analysts. *CFO.com Magazine*, July 8, 2004.

Lewis, B. W., Walls, J. L., & Dowell, G. W. S. (2014). Difference in degrees: CEO  
characteristics and firm environmental disclosure. *Strategic Management Journal*,  
35(5), 712–722.

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

Li, F. (2010). The information content of forward-looking statements in corporate filings—A  
naïve Bayesian machine learning approach. *Journal of Accounting Research*, 48(5),  
1049-1102.

Manning, C., & Schütze, H.(1999). *Foundations of Statistical Natural Language Processing*,

MIT Press. Cambridge, MA

Matsumoto, D., Pronk, M. and Roelofsen, E., 2011. What makes conference calls useful?

The information content of managers' presentations and analysts' discussion sessions. *The Accounting Review*, 86, 1383-1414.

Merkl-Davies, D., & Brennan, N. (2007). Discretionary disclosure strategies in corporate

narratives: Incremental information or impression management? *Journal of Accounting Literature*, 26(1), 116-194.

Merkl-Davies, D. M., Brennan, N. M., & McLeay, S. J. (2011). Impression management and

retrospective sense-making in corporate narratives: A social psychology perspective. *Journal of Accounting, Auditing & Accountability*, 24(3), 315-344.

Mitchell, T. M. (2006). The discipline of machine learning (Vol. 3). Carnegie Mellon

University, School of Computer Science, Machine Learning Department.

- Miller, G. S. (2002). Earnings performance and discretionary disclosure. *Journal of Accounting Research*, 40(1), 173-204.
- Miller & Rodger Y.M. (2008) Economic importance and statistically significance: Guidelines for communicating empirical research. *Feminist Economics*, 14(2), 117-149.
- Muslu, V., Radhakrishnan, S., Subramanyam, K. R., & Lim, D. (2015). Forward-looking MD&A disclosures and the information environment. *Management Science*, 61(5), 931-948.
- Nichols, C., & Wieland, M. (2009). Do firms? Non-financial disclosures enhance the value of analyst services. *Working paper, Cornell University*.
- O'Brien, P. C., & Bhushan, R. (1990). Analyst following and institutional ownership. *Journal of Accounting Research*, 28(1), 55-76.
- O'Brien, P. C., McNichols, M. F., & Lin, H. W. (2005). Analyst impartiality and investment banking relationships. *Journal of Accounting Research*, 43(4), 623-650.

- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The development and psychometric properties of LIWC2007. *Austin LIWC Manual*.
- Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. *Cognition & Emotion*, 10(6), 601-626.
- Previts, G. J., Brioker, R. J., Robinson, T. R., & Young, S. J. (1994). A Content Analysis of Sell-Side Financial Analyst Company Reports. *Accounting Horizons*, 8(2), 55-70.
- Plöckinger, M., Aschauer, E., Hiebl, M. R. W., & Rohatschek, R. (2016). The influence of individual executives on corporate financial reporting: a review and outlook from the perspective of upper echelons theory. *Journal of Accounting Literature*, 37, 55-75.
- Ramnath, S., Rock, S., & Shane, P. (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting*, 24(1), 34-75.
- Rogers, J. L., Van Buskirk, A., & Zechman, S. L. C. (2011). Disclosure tone and shareholder litigation. *The Accounting Review*, 86(6), 2155-2183.

- Rogers, R. K., & Grant, J. (1997). Content analysis of information cited in reports of sell-side financial analysts. *Journal of Financial Statement Analysis*, 3(1), 17-31.
- Securities and Exchange Commission (SEC). (1989). Securities Act Release No. 6835 (May 18). Available at: <http://www.sec.gov>.
- Securities and Exchange Commission (SEC). (2002). SEC Release Nos. 33-8056; 34-45321; FR-61 (Jan 2002). Available at: <http://www.sec.gov>.
- Sedor, L. M. (2002). An explanation for unintentional optimism in analysts' earnings forecasts. *The Accounting Review*, 77(4), 731-753.
- Singh I., & JL W. Mitchell Van der Zahn. (2008). Determinants of intellectual capital disclosure in prospectuses of initial public offerings. *Accounting & Business Research*, 38(5), 409-431.
- Solomon, J. F., Solomon, A., Joseph, N. L., & Norton, S. D. (2013), "Impression management, myth creation and fabrication in private social and environmental

- reporting: Insights from Erving Goffman", *Accounting, Organizations and Society*, 38, 195-213.
- Sonenshein, S., Herzenstein, M., & Dholakia, U. M. (2011). How accounts shape lending decisions through fostering perceived trustworthiness. *Organizational Behavior & Human Decision Processes*, 115(1), 69-84.
- Waring, H. Z. (2007). The multi-functionality of accounts in advice giving. *Journal of Sociolinguistics*, 11(3), 367-391.
- Wang, K., & Claiborne, M. C. (2008). Determinants and consequences of voluntary disclosure in an emerging market: Evidence from China. *Journal of International Accounting, Auditing and Taxation*, 17(1), 14-30.
- Yu, M. (2010). Analyst forecast properties, analyst following and governance disclosures: A global perspective. *Journal of International Accounting, Auditing and Taxation*, 19(1), 1-15.

**Table 1 Variable definition**

<b>Causal reasoning (word) intensity</b>	Number of LIWC causal words in the performance-related MD&A sections, scaled by total number of words of those sections, multiplied by 100.
<b>Causal reasoning (sentence) intensity</b>	Number of LIWC causal reasoning sentences in the performance-related MD&A sections, scaled by total number of sentences in those sections, multiplied by 100.
Analyst following	Number of analysts following the firm is taken from the I/B/E/S summary database, with the recording date closest (but after) the announcement date.
Analysts' earnings forecast dispersion	Analysts' earnings forecast dispersion is measured as the standard deviation of the individual analyst forecasts, scaled by share price 90 days before the consensus forecast date. Standard deviation of the individual analyst earnings forecasts is taken from the I/B/E/S summary database, with the recording date closest (but after) the filing date of 10-K report.
Analysts' earnings forecast accuracy	Analysts' earnings forecast accuracy is measured as -1 times the square difference between I/B/E/S reported earnings and the mean of analyst forecast (I/B/E/S), scaled by share price 90 days before the consensus forecast date.
Analyst following decline	Analyst following decline is calculated using the following procedure: first, we identify the analyst following record with a recording date in the I/B/E/S summary file which is closest to (but preceding) 90 days before the firm's annual report submit date ( $NUMEST_{i,t}$ ); next, we pick the analyst following record with a recording date in I/B/E/S summary file which is closest to (but after) the firm's annual report submit date in the prior year and calculate the change in analyst

	following between these dates ( $NUMEST_{i,t1}$ minus $NUMEST_{i,t2}$ ); if the difference is larger than 0, otherwise it equals to 1.
Fog index (Readability)	Fog index = (words per sentence + percent of complex words) * 0.4; where complex words are defined as words with three syllables or more.
Std (Adjusted stock return)	The standard deviation of the monthly adjusted stock return in current year.
Market-to-book	The market value of equity at the end of the fiscal year divided by the book value of equity at that date.
Leverage	Long-term liabilities divided by total assets.
Firm size	The natural logarithm of total assets.
MD&A size (words)	The natural logarithm of total words in the performance related MD&A section.
Return on equity	Net income scaled by shareholders' equity.
Earnings growth	Earnings without extraordinary items growth rate.
Loss firm	Indicator variable which equals 1 if earnings before extra items is less than 0, and zero otherwise.
Profit2loss	Indicator variable which equals 1 if the firm made a profit in the previous year (year T-1), but suffers a loss in the current year (year T), and 0 otherwise. Profit or loss is measured as earnings before extraordinary items.
Intangible percentage	Intangible assets scaled by total assets.
CFO gender	Indicator variable which equals to 1 if the CFO is female, otherwise

	0.
CFO MBA	Indicator variable which equals to 1 if the CFO holds a MBA degree or equivalent, otherwise 0.

**Table 2 Descriptive statistics on dependent and independent variables (*N* = 34,264)**

	MEA N	MEDIA N	STD DEV	1st	25th	75th	99th
Causal reasoning (word) intensity	2.17	2.08	0.76	0.52	1.67	2.57	4.33
Causal reasoning (sentence) intensity	37.81	38.33	10.22	7.79	32.1 0	44.1 4	61.5 4
Analyst following	6.85	5.00	5.75	2.00	3.00	9.00	27.0 0
Analysts' earnings forecast dispersion	0.04	0.03	0.13	0.00	0.01	0.07	0.50
Analysts' earnings forecast accuracy	-0.03	0.00	0.17	-0.36	0.01	0.00	0.24
Analyst following decline	0.24	0.00	0.42	0.00	0.00	0.00	1.00
Fog Index (Readability)	18.00	17.41	3.24	13.14	15.7 4	19.5 7	25.7 1
Firm size	6.79	6.73	1.84	3.02	5.48	7.90	11.8

								4
MD&A size (words)	8.88	9.06	1.09	6.02	8.34	9.59	8	11.0
MD&A size (sentences)	5.57	5.73	1.09	2.40	5.10	6.20	8.00	
Market-to-book ratio	4.15	1.83	250.31	0.00	1.07	3.07	9	20.9
Leverage	0.18	0.10	0.22	0.00	0.01	0.28	0.87	
Loss firm	0.24	0.00	0.43	0.00	0.00	0.00	1.00	
Return on equity	-0.59	0.06	126.06	-3.39	0.00	0.14	2.16	
Intangible percentage	0.12	0.03	0.17	0.00	0.00	0.19	0.72	
Std(Adjusted stock return)	0.14	0.10	0.13	0.03	0.07	0.16	1.00	
Earnings growth	0.17	0.00	85.57	-	-	0.24	2	12.3
Profit2loss	0.07	0.00	0.26	0.00	0.00	0.00	1.00	
CFO gender	0.06	0.00	0.23	0.00	0.00	0.00	1.00	
CFO MBA	0.28	0.00	0.45	0.00	0.00	1.00	1.00	

**Table 3 Pairwise correlation of main dependent and independent variables (*N* = 34,264)**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R

Causal reasoning (word intensity, A)	1. 00 0																					
Causal reasoning (sentence intensity, B)	0. 73 3	1. 00 0																				
Analyst following, C	0. 02 6	0. 00 4	1. 00 0																			
Analysts' earnings forecast dispersion, D	- 0. 00 5	- 0. 03 0	0. 19 6	1. 00 0																		
Analysts' earnings forecast accuracy, E	0. 04 6	0. 02 2	0. 13 2	0. 28 4	1. 00 0																	
Analyst following decline, F	0. 05 3	0. 02 9	0. 01 3	0. 00 5	- 0. 00 4	1. 00 0																
Readability, G	0. 29 8	0. 22 9	0. 01 9	0. 04 9	- 0. 02 9	- 0. 03 3	1. 00 0															
Firm size, H	- 0. 20 0	- 0. 15 4	- 0. 50 8	- 0. 19 7	- 0. 00 5	- 0. 04 4	- 0. 12 3	1. 00 0														
MD&A size (words), I	- 0. 09 9	- 0. 06 2	- 0. 02 0	- 0. 02 6	- 0. 02 9	- 0. 00 6	- 0. 18 7	- 0. 10 6	1. 00 0													
MD&A size	- 0. 0.	- 0. 02	0. 01	0. 0.	- 0. 00	0. 0.	- 0. 0.	0. 09	0. 96	1. 00												

(sentences), J	05 0	03 0	5	2	02 3	0	03 0	1	7	0							
Market-to-book ratio, K	0. 00	0. 00	0. 00	0. 00	0. 00	0.	0. 00	0.	0. 00	0.	1. 00						
Leverage, L	- 0. <b>12</b> 7	- 0. <b>09</b> 4	- 0. <b>03</b> 7	- 0. <b>07</b> 0	- 0. <b>10</b> 0	0. 02	0. <b>04</b>	0. <b>15</b>	0. 03	0. <b>02</b>	0. <b>01</b>	1. 00					
Loss firm, M	0. 09	0. 10	0. 14	0. 33	0. 24	0. 07	<b>0.</b> <b>01</b>	0. 33	0. 00	0. 00	0. 00	0. <b>06</b>	1. 00				
ROE, N	- 0. <b>01</b> 5	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. <b>01</b>	0. <b>01</b>	0. 01	0. 01	0. 01	0. <b>01</b>	1. 00					
Intangible percentage, O	0. 08	0. 06	0. 09	0. 04	0. 03	0. 07	0. <b>04</b>	0. 01	0. 01	0. <b>01</b>	0. 00	0. <b>12</b>	0. <b>01</b>	0. 00	1. 00		
Std(Adjusted stock return), P	0. 05	0. 05	0. 13	0. 21	0. 16	<b>0.</b> <b>01</b>	<b>0.</b> <b>03</b>	<b>0.</b> 29	<b>0.</b> 01	<b>0.</b> 01	0. 00	0. <b>01</b>	0. <b>32</b>	0. 00	<b>0.</b> <b>01</b>	1. 00	
Earnings growth, Q	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	- 0. 00	1. 00		
Profit2loss, R	0. 03	0. 04	0. 04	0. 18	0. 11	0. 08	0. 02	0. 03	0. 02	0. <b>01</b>	0. 00	0. 04	0. <b>49</b>	0. 00	0. <b>02</b>	0. <b>05</b>	1. 00
<b>Bold</b> numbers imply significance at the 99% level, <b><i>bold+italic</i></b> significance at the 95% level, and <i>italic</i>																	

This article is protected by copyright. All rights reserved.

significance at the 90% level (two-tailed). The variables are defined in Table 1

**Table 4: Pooled regression of properties of analyst behaviour on causal reasoning intensity (1998-2012)**

	Model I	Model II	Model III	Model IV	Model V	Model VI
	Analyst following	Analysts' earnings forecast dispersion	Analysts' earnings forecast accuracy	Analyst following	Analysts' earnings forecast dispersion	Analysts' earnings forecast accuracy
Causal reasoning (word) intensity	0.714*** [8%]	-0.003*** [5%]	0.006*** [17%]			
	(19.425)	(-2.997)	(4.873)			
Causal reasoning (sentence) intensity				0.046*** [7%]	-0.000*** [5%]	0.000*** [13%]
				(17.239)	(-3.159)	(4.253)
Readability	- 0.083***	0.002***	0.000	- 0.164***	0.002***	0.000* (-1.695)
	(-9.771)	(8.430)	(-0.677)	(-19.389)	(10.044)	

<b>Firm size</b>	1.789***	-0.002***	-0.018***	1.790***	-0.002***	-0.018***
	(112.63 2)	(-4.864)	(-27.858)	(112.28 5)	(-4.912)	(-27.866)
<b>MD&amp;A size (words)</b>	- 0.094***	0.002***	-0.001	- 0.125***	0.003***	-0.002*
	(-3.857)	(4.085)	(-1.614)	(-5.121)	(4.298)	(-1.929)
<b>Market-to-book ratio</b>	0.000	0.000	0.000	0.000	0.000	0.000
	(0.233)	(-0.271)	(0.226)	(0.251)	(-0.272)	(0.231)
<b>Leverage</b>	- 3.333***	0.029***	-0.037***	- 3.392***	0.029***	-0.038***
	(- 26.768)	(9.618)	(-8.720)	(- 27.251)	(9.648)	(-8.839)
<b>Loss firm</b>	0.510***	0.081***	-0.098***	0.521***	0.081***	-0.098***
	(7.637)	(49.931)	(-43.283)	(7.791)	(49.940)	(-43.243)
<b>Return on equity</b>	0.000	0.000	0.000	0.000	0.000	0.000
	(0.497)	(0.683)	(-0.888)	(0.229)	(0.721)	(-0.957)
<b>Intangible percentage</b>	3.423***	-0.025***	0.024***	3.444***	-0.025***	0.024***
	(22.703)	(-6.881)	(4.586)	(22.809)	(-6.851)	(4.613)
<b>Std(Adjusted stock return)</b>	0.848***	0.097***	-0.155***	0.781***	0.097***	-0.156***
	(3.911)	(18.442)	(-21.024)	(3.597)	(18.495)	(-21.104)
<b>Analyst following</b>		-0.003***	0.005***		-0.003***	0.005***

		(-20.013)	(28.281)		(-20.057)	(28.425)
Constant	- 4.593*** (- 15.717)	0.029*** (4.057)	0.065*** (6.547)	- 3.051*** (- 11.497)	0.024*** (3.709)	0.079*** (8.720)
<i>N</i>	34,264	34,264	34,264	34,264	34,264	34,264
<i>Adjusted R</i> <sup>2</sup>	30.05%	15.09%	10.78%	29.89%	15.09%	10.77%
<p>Regressions are clustered at firm level. Causal reasoning word intensity is the main proxy for causal reasoning on performance. It is measured as the amount of causal reasoning words scaled by total number of words in the performance-related MD&amp;A paragraphs and then multiply by 100. The identification of the causal reasoning words is based on a list of causal words used by Linguistic Inquiry and Word Count (LIWC). Causal reasoning sentence intensity is measured as the amount of causal reasoning sentences in the performance-related MD&amp;A paragraphs scaled by the total number of sentences in the performance-related MD&amp;A paragraphs and then multiply by 100. A causal reasoning sentence is defined as a sentence that includes at least one of the causal words in the LIWC causal words list. Analyst following is the number of analysts following a firm in the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the 10-K filing. Analysts' earnings forecasts dispersion is measured as the standard deviation of the individual analyst forecasts scaled by share price 90 days before the consensus forecast date. Analysts' earnings forecast accuracy is calculated as -1 times the squared difference between I/B/E/S reported earnings and the mean of analyst earnings forecasts scaled by the share price 90 days before the consensus forecast date. Readability is measured by Fog index. Fog index = (words per sentence + percent of complex words) * 0.4, where complex words are defined as words with three syllables or more. Firm size is measured as the natural logarithm of the firm's total assets. MD&amp;A size is measured as the natural logarithm of total words. Leverage is measured as long-term liabilities divided by total assets. Loss firm</p>						

equals 1 if earnings before extra items is less than 0, and zero otherwise. Intangible percentage is measured as intangible assets scaled by total assets. Std(Adjusted stock return) is standard deviation of the monthly adjusted stock returns in the prior year. All variables are winsorized at 1% and 99% to minimize the effect of extremely values. \*\*\* significance at the 99% level, \*\* significance at the 95% level, and \* significance at the 90% level (two-tailed).

**Table 5<sup>29</sup>:** Second stage of two-stage least-square regression of properties of analyst behaviour on causal reasoning intensity (1998-2012)

	Model I	Model II	Model III	Model IV	Model V	Model VI
	Analyst followin g	Analysts' earnings forecast dispersio n	' earnings forecast accurac y	Analyst followin g	Analysts' earnings forecast dispersio n	' earnings forecast accurac y
Causal reasoning (word) intensity	9.938** *	-0.032**	0.173***			
	(12.899 )	(-2.422)	(7.669)			
Causal reasoning				0.898**	-0.003**	0.015***

<sup>29</sup> We do not report the first stage of the regressions to save the space of this paper.

(sentence) intensity				*		
				(11.379)	(-2.402)	(7.170)
Readability	- 0.494** * 0.000		- 0.010*** 0.000	- 0.394** * 0.000		- 0.009*** 0.000
	(- 9.852)	(0.035)	(-7.312)	(- 6.566)	(0.325)	(-6.122)
Firm size	2.274** * -0.005***		-0.004** -0.004**	2.410** * -0.005***		-0.002 -0.002
	(46.917) )	(-3.961)	(-2.266)	(37.149) )	(-3.750)	(-0.998)
MD&A size (words)	- 0.103** 0.002***		-0.002* -0.002*	- 17.155* ** 0.065***		- 0.306*** 0.306***
	(- 2.324)	(2.817)	(-1.761)	(- 10.567)	(2.677)	(-7.001)
Market-to-book ratio	0.000	0.000	0.000	0.000	0.000	0.000
	(0.310)	(-0.175)	(0.122)	(0.269)	(-0.175)	(0.110)
Leverage	-0.241	0.020***	-0.018**	-0.571	0.019***	-0.023**
	(- 0.663)	(3.667)	(-1.982)	(- 1.306)	(3.254)	(-2.240)
Loss firm	0.486** * 0.084***		- 0.115*** 0.115***	0.586** * 0.084***		- 0.116*** 0.116***

	(3.475)	(39.965)	(-32.164)	(3.577)	(38.484)	(-29.530)
<b>Return on equity</b>	0.001** *	0.000	0.000	0.001	0.000	0.000
	(3.152)	(0.030)	(-1.375)	(1.479)	(0.360)	(-0.318)
<b>Intangible percentage</b>	0.012	-0.016***	0.030***	0.525	-0.015**	0.038***
	(0.033)	(-2.910)	(3.083)	(1.126)	(-2.418)	(3.430)
<b>Std(Adjusted stock return)</b>	1.469** *	0.095***	-0.142***	0.990**	0.096***	-0.150***
	(3.983)	(17.399)	(-15.379)	(2.380)	(17.812)	(-15.505)
<b>Analyst following</b>		-0.002***	0.003***		-0.002***	0.003***
		(-9.043)	(6.549)		(-9.201)	(6.396)
<b>Constant</b>	-40.074* **	0.142***	0.580***	-9.661** *	0.034*	0.305***
	(-13.382)	(2.762)	(6.601)	(-6.577)	(1.649)	(8.255)
<b>N</b>	<b>34,264</b>	<b>34,264</b>	<b>34,264</b>	<b>34,264</b>	<b>34,264</b>	<b>34,264</b>
<b>Adjusted R<sup>2</sup></b>	<b>20.53%</b>	<b>12.50%</b>	<b>9.70%</b>	<b>19.10%</b>	<b>11.53%</b>	<b>7.90%</b>
Regressions are clustered at firm level. Causal reasoning word intensity is the main proxy for causal reasoning on performance. It is measured as the amount of causal reasoning words						

This article is protected by copyright. All rights reserved.

scaled by total number of words in the performance-related MD&A paragraphs, multiplied by 100. The identification of the causal reasoning words is based on a list of causal words used by Linguistic Inquiry and Word Count (LIWC). Causal reasoning sentence intensity is measured as the amount of causal reasoning sentences in the performance-related MD&A paragraphs scaled by the total number of sentences in the performance-related MD&A paragraphs, multiplied by 100. A causal reasoning sentence is defined as a sentence that includes at least one of the causal words in the LIWC causal words list. Analyst following is the number of analysts following a firm in the first analyst consensus annual earnings forecast issued after the 10-K filing for the fiscal period following the 10-K filing. Analysts' earnings forecasts dispersion is measured as the standard deviation of the individual analyst forecasts scaled by share price 90 days before the consensus forecast date. Analysts' earnings forecast accuracy is calculated as  $-1 \times (\text{standard deviation of forecasts} / \text{share price})^2$ . Readability is measured by Fog index. Fog index = (words per sentence + percent of complex words) \* 0.4, where complex words are defined as words with three syllables or more. Firm size is measured as the natural logarithm of the firm's total assets. MD&A size is measured as the natural logarithm of total words. Leverage is measured as long-term liabilities divided by total assets. Loss firm equals 1 if earnings before extra items is less than 0, and zero otherwise. Intangible percentage is measured as intangible assets scaled by total assets. Std(Adjusted stock return) is standard deviation of the monthly adjusted stock returns in the prior year. All variables are winsorized at 1% and 99% to minimize the effect of extremely values. \*\*\* significance at the 99% level, \*\* significance at the 95% level, and \* significance at the 90% level (two-tailed).

**Table 6: Logistic regression of analyst following decline on causal reasoning intensity**

	Model I	Model II

	Analyst following decline	Analyst following decline
Causal reasoning (word) intensity	0.129*** (5.220)	
Thinly covered × Causal reasoning (word) intensity	0.100*** (3.009)	
Causal reasoning (sentence) intensity		0.008*** (3.985)
Thinly covered × Causal reasoning (sentence) intensity		0.007*** (2.829)
Thinly covered	-0.097** (-2.229)	-0.153** (-2.335)
Firm size	0.090*** (10.804)	0.086*** (10.398)
MD&A size (words)	-0.014 (-1.209)	-0.391*** (-7.636)
Market-to-book ratio	0.000 (-1.321)	0.000 (-1.356)
Leverage	0.127**	0.110*

	(2.074)	(1.797)
Earnings growth	-0.001	-0.001
	(-1.368)	(-1.366)
Profit2loss	0.619***	0.627***
	(13.917)	(14.103)
Return on equity	0.000	0.000
	(0.033)	(0.025)
Intangible percentage	0.948**	0.959***
	(13.309)	(13.481)
Std(Adjusted stock return)	0.013	0.015
	(0.122)	(0.145)
Constant	-2.210***	-0.891***
	(-15.749)	(-4.655)
<i>N</i>	34,264	34,264
<i>Pseudo R</i> <sup>2</sup>	1.75%	1.68%
Regressions are clustered at firm level. Analyst following decline is calculated using the following procedure: first, we identify the analyst following record with a recording date in the I/B/E/S summary file which is closest to (but preceding) 90 days before the firm's annual report submit date (NUMEST <sub>i,t-1</sub> ); next, we pick the analyst following record with a recording date in the I/B/E/S summary file which is closest to (but after) the firm's annual report submit date in the prior year (NUMEST <sub>i,t-2</sub> ) and calculate the change in		

analyst following between these dates ( $\text{NUMEST}_{i,t-1}$  minus  $\text{NUMEST}_{i,t-2}$ ); if the difference is larger than or equal to 0, we set the variable equal to 0, otherwise it equals 1. Causal reasoning word intensity is measured as the amount of causal reasoning words scaled by total number of words in the performance-related MD&A paragraphs, multiplied by 100. The identification of the causal reasoning words is based on a list of causal words used by Linguistic Inquiry and Word Count (LIWC). Thinly covered is a dummy variable, which equals to 1 if a firm's analyst following is less than the median value of analyst following, 0 otherwise. Readability is measured by Fog index. Fog index = (words per sentence + percent of complex words) \* 0.4, where complex words are defined as words with three syllables or more. Firm size is measured as the natural logarithm of the firm's total assets. MD&A size is defined as the natural logarithm of total words. Leverage is measured as long-term liabilities divided by total assets. Earnings growth is measured as the earnings without extraordinary items growth rate. Profit2loss is an indicator variable measured as 1 if the firm made a profit in the previous year (year T-1), but suffers a loss in the current year (year T), and 0 otherwise. Profit or loss is measured as earnings before extraordinary items. Intangible percentage is defined as intangible assets scaled by total assets. Std(Adjusted stock return) is standard deviation of the monthly adjusted stock returns in the prior year. All variables are winsorized at 1% and 99% to minimize the effect of extremely values. \*\*\* significance at the 99% level, \*\* significance at the 95% level, and \* significance at the 90% level (two-tailed).

Table 7: Logistic regression of More causal reasoning on characteristics of analyst following

	Model I	Model II
	More causal reasoning	More causal reasoning

	(word) intensity	(sentence) intensity
<b>New analyst status × More analyst following</b>	0.006*	0.009***
	(1.859)	(2.932)
<b>More analyst following</b>	0.409***	0.414***
	(13.285)	(13.414)
<b>New analyst status</b>	-0.001	0.000
	(-0.509)	(-0.236)
<b>Firm size <sub>t-1</sub></b>	0.133***	0.107***
	(7.165)	(5.797)
<b>MD&amp;A size (words) <sub>t-1</sub></b>	-0.047	0.048
	(-0.815)	(0.844)
<b>MD&amp;A size (sentences) <sub>t-1</sub></b>	-0.080	-0.191**
	(-1.088)	(-2.590)
<b>Market-to-book ratio <sub>t-1</sub></b>	0.030*	0.012
	(1.825)	(0.849)
<b>Leverage <sub>t-1</sub></b>	0.064	0.018
	(0.365)	(0.102)
<b>Loss firm <sub>t-1</sub></b>	0.079	0.099
	(0.583)	(0.776)

<b>Return on equity <math>t_{-1}</math></b>	0.001	-0.001
	(0.096)	(-0.049)
<b>Intangible percentage <math>t_{-1}</math></b>	0.083	0.076
	(0.255)	(0.235)
<b>Std(Adjusted stock return) <math>t_{-1}</math></b>	-1.081**	0.250
	(-2.333)	(0.700)
<b>HHI <math>t_{-1}</math></b>	0.266	-0.338
	(0.820)	(-1.012)
<b>Constant</b>	-1.260***	-1.265***
	(-60.406)	(-60.570)
<b>N</b>	<b>34,264</b>	<b>34,264</b>
<b>Pseudo R<sup>2</sup></b>	<b>1.64%</b>	<b>1.47%</b>
Regression clustered at firm level. More causal reasoning words (sentences) intensity is a dummy variable which equals 1 if a firm increases causal reasoning words (sentences) intensity from T-2 to T-1 by at least 10%, otherwise 0. New analyst status is measured as the average number of firms that additional (new) analysts have been following before. An additional (new) analyst is identified as an analyst starting to cover the firm in the current year. More analyst following equals 1 if a firm has more analyst following compared to the prior year, otherwise 0. Firm size is measured as the natural logarithm of the firm's total assets. MD&A size is measured as the natural logarithm of total words. Leverage is measured as long-term liabilities divided by total assets. Loss firm equals 1 if earnings before extra items is less than 0, and zero otherwise. Intangible percentage is measured as intangible assets scaled by total assets. Std(Adjusted stock		

return) is standard deviation of the monthly adjusted stock returns in the prior year. All variables are winsorized at 1% and 99%. \*\*\* significance at the 99% level, \*\* significance at the 95% level, and \* significance at the 90% level (two-tailed).

Table: Appendix B – LIWC causal words list




activat*	changes	depended	how's	lead*	permit*	solves
affect	changing	depending	ignit*	led	pick	solving
affected	compel*	depends	implica*	made	produc*	source*
affecting	compliance	effect*	implied*	make	provoc*	stimul*
affects	complie*	elicit*	imply*	maker*	provok*	therefor*
aggravat*	comply*	experiment	inact*	makes	purpose*	thus
allow*	conclud*	force*	independ*	making	rational*	trigger*
attribut*	consequen *	foundation*	induc*	manipul*	react*	use
based	control*	founded	infer	misle*	reason*	used
bases	cos	founder*	inferr*	motiv*	respons e	uses
basis	coz	generate*	infers	obedien*	result*	using

because	create*	generating	influenc*	obey*	root*	why
boss*	creati*	generator*	intend*	origin	since	
caus*	cuz	hence	intent*	originat*	solution*	
change	deduc*	how	justif*	origins	solve	
changed	depend	hows	launch*	outcome*	solved	

Figure 1.

