Finding the Narrative in the Numbers: Long-Term Investors' Demand for Accounting Information

Federico Siano*
Boston University - Questrom School of Business
fsiano@bu.edu

May 2019

Abstract:

I study the MD&A section of 10-K filings and use disclosed numbers within the text of corporate narratives to proxy for the qualities of information that are likely demanded by long-term investors. I hypothesize and show that the prevalence of quantitative information in firms' written communication is positively associated with long-term institutional ownership. I show that numbers capture a dimension of business communication that is opposite in direction to a short-term disclosure horizon. Further tests suggest that current levels of disclosed numbers are relevant predictors of future changes in long-term ownership, but that present levels of long-term investors have no power in explaining future changes in numbers.

Key Words: *Disclosure; Long-term Investors; Quantitative Information; Textual Analysis; Short-termism; Analyst Following.*

Data Availability: Publicly-available data sets, Brian Bushee's classes of Institutional Investors.

[*]I wish to thank Peter Wysocki, Francois Brochet, Eddie Riedl and workshop participants at Boston University for helpful discussions and comments.

1. Introduction

This paper provides new evidence about the role of the voluntary disclosure system in reflecting *qualities* of information that are relevant for long-term investing. While much of prior literature in accounting and finance studies investment and disclosure horizon focusing on the drivers, attributes and implications of "short-termism", I offer insights on the association between long-term investors and the characteristics of firms' disclosures. I explore this link studying a fundamental, often neglected, set of accounting information: *numbers* within the narratives of corporate filings. First, I establish a strong relationship between long-term ownership and the prevalence of quantitative information in the text of the Management Discussion and Analysis (MD&A) section of firms' financial reports. I hypothesize and find a positive association, since I posit that numbers summarize and are correlated with the qualities of information demanded by longer-horizon investors. Second, I provide initial evidence on the direction of the link between numbers and long-term investors, by investigating the lead-lag association of current period numbers (long-term ownership) and future periods' changes in long-term ownership (numbers). Consistently with the signaling role of voluntary disclosure, I find that numbers within text are significant predictors of future changes in long-term investors, but not vice versa.

A characterization of investment horizon can be made around two main features. On the one hand, the trading behavior of agents, on the other hand, the set of information they employ. In this paper, I focus on the latter indicator and, in line with prior literature, I claim that while short-term investors focus on general news flows and the actions of other investors to predict stock price volatility, long-term investors are interested in the drivers of firms' value – such as cash-flow generation over time and future investment opportunities. In more details, they are likely to target information that improve forecasts about the magnitude and riskiness of expected cash-flows thus allowing more reliable business planning. Examples of relevant information could be trends in demand and the competitive environment, data about present or forthcoming investments, and details about the current and prospective financial or risk position. Ideal qualities of information sought by long-term investors are precision and objectivity.

Among the several outputs of a firm's information production function, voluntary disclosures represent the appropriate investigation set. In fact, the traditional accounting system (e.g., income statement, balance sheet, statement of cash-flows) hinges on financial reporting standards that are designed to meet the information needs of the maximum number of users within a trade-off setting of conflicting objectives (Statement of Financial Accounting Concepts No. 8). Therefore, given that financial statements cannot provide different stakeholders with a complete set of information, firms are allowed to exploit the voluntary disclosure system to convey incremental information. One important characteristic of voluntary disclosures, when compared to the traditional accounting system, is their pronounced forward-looking orientation. A clear example in this sense is represented by the Management Discussion and Analysis (MD&A) section of financial reports that provides managers with an opportunity to convey their competitive strategy, plans and expectations. The Title 17 of the Code of Federal Regulation (17 CFR 229.303) reports that MD&A requires the discussion of "known trends or any known demands, commitments, events or uncertainties" that likely result in material changes in the firm's (i) liquidity position; (ii) capital resources; (iii) net sales and revenues. In other words, even when managers do not make explicit predictions or projections about the future, and therefore do not directly anticipate a trend, the MD&A has to include *prospective* information. Moreover, the federal legislative corpus indicates that discussions included in the MD&A sections should not “merely repeat numerical data contained in the consolidated financial statements”. I thus focus, within voluntary disclosures, on the MD&A sections of 10-K filings.

The textual analysis literature in accounting and finance has traditionally examined, within financial reports, the text of corporate narratives. However, managers could exploit language to obfuscate firm’s performance (see Li, 2008), or employ “cheap talk” communication that does not reflect real actions. More recent studies highlight the role of numbers within text (see Siano and Wysocki, 2018). Quantitative information is precise, verifiable, objective, concise and likely refers to tangible business events or economic transactions. Therefore, numbers seem to possess or to be correlated with the qualities of information that long-term investors would likely find helpful. For the reasons stated above, I choose to investigate numbers within the narratives of the MD&A sections of 10-K filings.

My work offers insights into the active research area, in accounting and finance, that studies determinants, attributes and outcomes of corporate disclosures (see Leuz and Wysocki, 2016). In particular, this paper furthers the understanding of relevant *qualities* of accounting information that can be of particular interest to the users of financial reports. I add to the literature that examines the textual channel of disclosure, by extending the evidence on the role of numbers within text (see Siano and Wysocki, 2018) and suggesting a novel link with non-textual, capital market variables. I also contribute to the literature on institutional ownership that separates investors’ horizon, in a reveled preferences framework, through heterogeneity in their trading behavior (see Bushee, 1998). My study suggests the opportunity to characterize long-term investors more completely, taking into account the information set they likely exploit.

The structure of the remainder of the paper is as follows. In Section 2, I review previous literature. In Section 3, I develop my hypotheses. In Section 4, I discuss data and the research design. In Section 5, I summarize the results. In Section 6, I outline relevant conclusions and discuss future work.

2. Prior Literature

This work is related to and has implications for academic research on the textual attributes of disclosure and their connection with several financial, accounting, disclosure and economic outcomes (see, for instance, Li, 2011 and Loughran and McDonald, 2016). This paper also furthers the literature on institutional ownership by showing the qualities of information that are particularly relevant as the investment horizon lengthens and therefore indicating the possibility to complement the characterization of investment horizon through the information set employed by traders.

Recent empirical work in accounting and finance shows how the qualitative properties of firms' disclosures can reveal information about investment decisions, executives' actions and their incentives. For example, Merkely (2014) finds that narratives are distinctly informative about companies' R&D investments. Brochet et al. (2015) establish a significant relationship between language in management's conference calls and symptoms of managerial myopia. They validate a word-based measure of short-termism and test its association with accruals and real earnings management. In addition, Etruglu et al. (2017) investigate the link between annual report readability and corporate borrowing costs, while Lo et al. (2017) explore the connection between language complexity and earnings management. A different strand of the textual analysis literature has focused, indirectly or directly, on quantitative information within narratives. Lundholm et al. (2014) compare readability and "number of numbers" of foreign firms and domestic U.S. companies. The authors show that foreign firms' disclosures are characterized by both more numbers and a lower Fog index as compared to domestic companies. Siano and Wysocki (2018), more recently, find a strong correlation between linguistic complexity and the prevalence of numbers found in the narratives of business documents and discuss implications for past and future disclosure and textual analysis works.

Prior literature on the link between institutional ownership and disclosure has traditionally examined the *amounts* of firms' communication and classified investment horizon based on trading preferences. Bushee (1998) employs a principal component analysis to distinguish investment horizon through revealed trading behaviors and finds that high-turnover and momentum investors foster firms' myopic investment. Bushee and Noe (2001) investigate corporate disclosure practices and provide evidence about the association between higher AIMR scores and increased levels of institutional ownership – both long-term and short-term. Boone and White (2015) link higher institutional ownership to greater firms' propensity to provide voluntary disclosure through management forecasts and 8-K filings. Cadman et al. (2019) develop and test a structural framework in which short-term investors induce companies to commit to more frequent earnings-related voluntary disclosures. Warren (2014) provides a discussion about the multi-faceted nature of investment horizon and the major role that both trading behavior and the information set employed for decision-making play.

Finally, recent works on "short-termism" document (i) capital market and (ii) incentive-based pressured on the short-term, thus providing relevant insights into potential determinants of disclosure horizon in addition to firms' characteristics. On the one hand, sell-side analysts can induce managers to concentrate on short-term targets to meet or beat earnings forecasts (He and Tian, 2013). On the other hand, managerial compensation tied to stock performance likely induces executives to maximize their near-term benefits at the expense of longer term value creation (Edmans et al., 2014 and Gopalan, 2014).

My study complements and extends the line of research attempting to gauge relevant qualities of corporate disclosure, using the narrative channel. This paper provides new evidence about disclosure horizon and its relationship with investment horizon and supports the significance and role that numeric information plays in qualitative disclosures.

3. Hypothesis Development

It is safe to affirm that numbers are the primary element of interest within corporate disclosures. It is therefore reasonable to expect that stakeholders are willing to seek out quantitative information not only in the financial statements but also within the narratives that accompany and explain financial tables. Numbers synthetize relevant *qualities* of accounting information; they are precise, objective, factual and concise. Numeric data are very well suited to describe, precisely and effectively, the outcomes of past and present value creation strategies that, in general, are long-term in nature. Disclosed numbers are clearly elements of substantial interest to longer-horizon investors, who are less concerned with interim changes in asset prices and focus on income growth. Numbers' inclusion in corporate documents is influenced by stakeholders' demand, managers' evaluation of the associated disclosure costs and the incentives they face. Given the suitability of numbers, found within voluntary disclosures that describe *prospective* information (i.e. the MD&A sections of 10-K filings), to be the inputs of long-term value models, I hypothesize that their disclosure reflects the demand of longer-horizon institutional ownership. This leads to my first hypothesis that I state in the null form, given the absence of a clear theory linking the interactions between number, text and potential outcomes[1]. I will thus conduct a two-sided hypothesis test.

[1] As noted in Siano and Wysocki (2018), the academic linguistics literature does not provide any direct insights on the relationship between numbers and language and "has generally ignored hard numbers within corpuses".

H1: The amount of numbers reported within voluntary corporate narratives is unrelated to long-term institutional ownership and to capital market and incentive based pressures on the short-run.

It is worth noticing here, that this first hypothesis does not imply that numbers, to inform about long-term value creation, should necessarily be part of sentences that predict future events. In fact, it is plausible to expect that numbers are mostly referred either to the present or to the near past, since the future underlies uncertainties that are incoherent with the factual nature of quantitative data. Nonetheless, even if quantitative data are not used to make projections about the future, the description of current trends through numbers is *prospective* in nature.

If numbers summarize the traits of information that would be especially important for long-term investment decision making, as I posit, I expect them to be *positively* associated with long-term ownership and *negatively* related to capital market and incentive-based pressures on the short-run – after controlling for determinants of firms' disclosure horizon.

Proving an association, would still leave many interesting questions unanswered. In particular, it would be relevant to shed some light on the directional link between numbers and longer-horizon investors. An institutional ownership oriented to the long-term could induce changes in corporate disclosure practices that would be reflected in different amounts of disclosed numbers. On the other hand, firms could signal their long-term value creation type, to institutional investors, by using more quantitative information in their voluntary narratives. A third possibility is that long-term investors and disclosed numbers are jointly driven by an unobserved cause. By the logic of causality, a determinant precedes chronologically its effects. I therefore test my second hypothesis, again stated in the null form, in a lead-lag framework.

H2: Current period numbers (long-term investors) are unrelated to future periods' changes in long-term investors (numbers).

I do not formulate any expectations, but reason that if numbers (long-term investors) are a proper determinant of long-term investors (numbers) they should have power in predicting future changes in the outcome variable, whereas the opposite should not be true.

4. Data and Research Design

My empirical strategy is divided into two main parts. I first validate *Numbers/Words* by testing their cross-sectional association with *Long-Term Investors*, controlling for a wide array of financial and textual determinants of disclosure horizon. As part of this analysis, I replicate the main results in Brochet et al. (2015) and show how numbers and short-term words capture opposite dimensions of qualitative disclosures. I then examine the directional link between numbers and long-term investing by studying their lead-lag structure. The objective of the latter set of tests is to provide initial evidence about the causal relationship between numbers and long-term investors.

In the following subsections, I summarize the data samples used in the empirical analyses; describe the text analysis methodology; outline the modeling choices; and report the financial and textual variables used in the main tests.

4.1. Data

My primary text data is represented by EDGAR 10-K filings between 1994 and 2017. In more details, I download the plain text version of 10-K reports from the Bill McDonald's "Stage One 10-X" dataset available on the Notre Dame *Software Repository for Accounting and Finance*. I choose to focus on 10-K filings for three main reasons. To begin with, 10-K reports are among the most widely used corporate documents. Secondly, their vast time-series allows for large-scale analyses. Lastly, compared to filings issued more frequently (i.e. 10-Q), they exhibit higher cross-sectional and over time variation in textual attributes. Within 10-K filings, I study the Management Discussion and Analysis (MD&A) section. This decision not only fulfills the objective of analyzing a relevant component of the voluntary disclosure system, but also mitigates measurement error by limiting the processing of cautionary, forward-looking or other types of statements containing short-term and long-term words that are not directly related to firms' economic transactions.

To construct the variables included in the regression analyses, I download and match data from the following repositories over the period 1994-2017 (fiscal years): (i) the Annual Fundamental table and the Segments table of *Capital IQ* containing companies' fundamentals, (ii) the *Execucomp* Annual File that includes stock-based executives compensation, (iii) the *I/B/E/S* Summary and Surprise tables with information about analysts following, (iv) the Bushee (2001) classification of institutional investors and, (v) *Thomson Reuters 13F* that classify and list data on institutional holdings; (vi) the Bog Index Data available on Brian P. Miller's website; (vii) the Accounting Reporting Complexity (Arc) measure available on Udi and Rani Hoitash's website.

4.2. Textual Analysis Methodology

Once downloaded, I process the 10-K filings in two steps. First, I extract the type (e.g., 10-K or 10-K Amendment), the CIK, the filing date, the report date and the MD&A section from each filing. Next, I analyze the MD&A sections and extract the relevant textual variables.

I begin by excluding 10-K Amendments, 10-KSB and 10-KSB Amendments from the sample. I then take advantage of the Python *Glob* and *Regular Expression* libraries to parse the MD&A section of 10-K reports. I code starting signals (e.g., "ITEM 7", "Item 7") and ending signals (e.g., “ITEM 7A”, “ITEM 8”) to delimit the Management Discussion and Analysis section and develop tailored conditional statements to handle cases of multiple starting and/or ending signals. Whenever the MD&A section of a 10-K cannot be identified with sufficient reliability, I exclude the entire document from the sample. The number of excluded documents varies from year to year and is, on average, in the range 10%-20% of the total filings. I manually check 50 MD&A sections of 50 different companies in different years to assess the reliability of the outlined coding strategy.

I exploit the *Natural Language Toolkit Library* (NLTK) in Python to analyze text and extract relevant information. I start by sentence-tokenizing (i.e. dividing into sentences) each MD&A section and then word-tokenize (i.e. divide into words) each sentence. Sentences are identified through periods but the Library functions also allow to control for common textual features, within 10-K reports, that could lead to an improper sentence representation such as: (i) the possibility of abbreviations (e.g. U.S.A.) or (ii) the presence of decimal numbers which digits are separated by a period (e.g. “increased by 21.5%.”). Tables are excluded from the relevant text of MD&A sections. I mark a sentence, or a set of sentences, as a table whenever (i) the number of white spaces is at least 200 and the ratio of numbers to words is greater than 0.25 or (ii) the ratio of numbers to words is greater than 0.50[2]. For each document I count (i) the total number of relevant sentences; (ii) the total number of words; (iii) the total number of complex words (i.e. words with more than two syllables); (iv) the total number of numbers; (v) the total number of dates; and (vi) the number of words contained in pre-defined dictionaries to set up textual variables.

[2] These thresholds balance the trade-off between including too many tables (more likely to happen for high numbers of white spaces and ratios of numbers to words) and excluding too many relevant numbers (more likely to happen for low numbers of white spaces and ratios of numbers to words).

To count words and complex words I use the *lexicon_count*[3] and *difficult_words*[4] functions in the *Textstat* package. To count numbers, I first identify what a relevant number is. The MD&A section of corporate filings includes a wide variety of numbers. They can take the form of monetary amounts, percentage changes, ratios, dates or even numbers expressed in words. For the purpose of my analysis, I select the numbers most likely to convey quantitative information and only read those. Specifically, my parsing algorithm identifies and counts a number in the following cases: (i) the number is preceded by a dollar sign ("$"); (ii) the number is followed by the words million/billion/trillion; (iii) the number is followed by a percentage sign ("%") or by the words "percent" / "pct". In addition, the software identifies numbers in parentheses (negative sign) and/or for which the previous markers (i) are preceded by one or two white spaces; (ii) are not preceded by any white spaces; (iii) are capitalized, fully or in part (applies to words). With regards to dates, I first identify years, months, days and then count them as one or multiple dates depending on whether or not they are located in proximity one to the other[5].

4.3. Cross-Sectional Models and Variables

4.3.1. Validation of Numbers/Words

To infer a relevant association between numbers and long-term investors, I use an ordinary least square (OLS) regression in which the dependent variable is *Numbers/Words*. The baseline model includes *Long-Term Investors* as an explanatory variable, together with a wide set of financial determinants and textual attributes that prior studies have found to be correlated with

[3] The default *lexicon_count* function excludes a list of "easy words" from the count. I modify the function so that all words are counted.
[4] The *difficult_words* function uses the Pyphen library for words hyphenation. I manually test the function on 30 MD&A sections and find an accuracy of 85%.
[5] I include an extensive set of conditional statements to properly identify the elements of a date and avoid double counting. I manually check 100 dates in different formats in 50 documents and find an accuracy of 99%.

horizon and numbers within qualitative disclosures. Unobserved heterogeneity at the industry and fiscal year level is accounted for through fixed effects. The potential bias in the estimation of coefficients' variance, induced by serial correlation in the errors structure, is limited using a cluster correction at the firm level. Augmented versions of the model, used to mitigate concerns about correlated omitted variable bias, include (i) a first-order autoregressive term; (ii) the Stylewriter Bog Index (Bonsall et al. 2017) as an alternative to the Gunning Fog Index; (iii) the Arc (Hoitash and Hoitash 2018) measure for financial reporting complexity[6]. I report the baseline model thereafter (Model 1). Subscript *i* is used for firms and subscript *t* for fiscal years.

$$
\begin{aligned}
Numbers/Words_{i,t} &\\
= \ & \alpha + \beta_1 \ Long\ Term\ Investors_{i,t} + \beta_2 \ Analyst\ Following_{i,t} \\
& + \beta_3 \ Stock\ Compensation_{i,t} + \beta_4 \ CFO\ Volatility_{i,t} + \beta_5 \ Operating\ Cycle_{i,t} \\
& + \beta_6 \ Leverage_{i,t} + \beta_7 \ Liquidity_{i,t} + \beta_8 ROE_{i,t} + \beta_9 \ MTB_{i,t} + \beta_{10} Size_{i,t} \\
& + \beta_{11} \ Special\ Items_{i,t} + \beta_{12} Segments_{i,t} + \beta_{13} \ Restructure_{i,t} \\
& + \beta_{14} \ Short\ Term_{i,t} + \beta_{15} \ Fog_{i,t} + \beta_{16} \ 1/Words_{i,t} + Industry\ FE_i + Year\ FE_t \\
& + \varepsilon_{i,t}
\end{aligned}
$$

(Model 1)

Numbers/Words is the ratio of the relevant "number of numbers" divided by the total words count in the MD&A section of a corporate 10-K. *Long-term Investors* is the ratio of the shares held by *Dedicated* and *Quasi-Indexer* investors minus those held by *Transient* stockholders (Bushee, 2001) to the total number of shares outstanding. I conjecture numbers to capture dimensions of

[6] I scale the natural logarithm of Arc by the natural logarithm of total assets.

disclosure that are of particular interest to long-term investors and I therefore expect a positive association between the two. *Analyst Following* is one plus the maximum number of I/B/E/S analysts following a firm during a given fiscal period, transformed in natural logarithm and scaled by the natural logarithm of total assets[7]. Previous studies document how analysts increase managerial focus on the short-run, thus I foresee a negative association with numbers. *Stock Compensation* is the residual obtained from regressing the average top-5 executives' stock and option-based compensation on firm's market capitalization, market-to-book ratio, industry and year fixed effects following the Cheng et al. (2015) procedure. I hypothesize a negative correlation with the dependent variable since stock-based incentives have been shown to foster short-termism. *CFO Volatility* is the 5-year standard deviation of the company operating cash flow from operations, scaled by total assets. *Operating Cycle* is the length of the operating cycle of a firm calculated as the natural logarithm of *[(Inventory/COGS) * 360 + (Accounts Receivable/Sales) * 360]*. I employ *CFO Volatility* and *Operating Cycle* as proxies for operating risk. I posit that firms with higher cash flow volatility are more concentrated on the short-run, while companies with longer operating cycle are more oriented to the long-run. Therefore, I expect a negative and a positive coefficient, respectively. *Leverage* is the ratio of short-term and long-term liabilities to total assets, while *Liquidity* is calculated as current assets deflated by current liabilities. I include the latter two variables to control for financial distress. I foresee higher availability of financial resources to ease pressures on the short-term, thus I predict a positive coefficient on *Liquidity*. I do not formulate expectations on *Leverage* since healthy firms could choose high indebtedness to

[7] I substitute a zero whenever the data is missing to rescue observations. Results are statistically and qualitatively similar with no substitution. In untabulated regressions, I also study the association between numbers and the presence/absence of analysts (i.e. I code the Analyst Following variable as 1 if at least one analyst follows the company and 0 otherwise). I obtain similar results.

exploit the tax shield on financial interests. *ROE*, calculated as net income scaled by shareholders' equity, is used to control for profitability and I expect it to be positively associated with numbers. I further control for growth opportunities including the *MTB* ratio, the firm's stock price at the end of the fiscal year divided by the book value of an equity stake[8]. Growth firms are expected to have longer investment horizons; thus, I foresee a positive relation with numbers. I include *Size* and the number of business *Segments* to account for the complexity of business operations. *Size* is defined as the natural logarithm of the number of common shares outstanding multiplied by the firm's stock price at the end of the fiscal year. *Segments* is the natural logarithm of the number of business and operating segments. I expect higher business complexity to be positively related to longer disclosure horizon. However, I do not formulate any predictions on *Size*. In fact, on the one hand, bigger companies may benefit from higher reputation and therefore be less concerned by the short-run. On the other hand, there appears to be a non-linear relationship between the dimension of a firm and the numbers of words it reports, thus, a negative correlation (-16%) between *Size* and *Numbers/Words*. In addition, I control for extraordinary business transactions undertaken by firms through *Special Items*, the absolute value of reported special items scaled by total assets, and *Restructuring*, an indicator variable that equals 1 if reported pre-tax restructuring costs are different from zero and 0 otherwise. I expect a negative association with numbers as those business activities create a natural pressure on recovering the incurred costs. The baseline model includes also relevant textual variables expected to be associated with numbers. *Short-Term* is the Brochet et al. (2015) measure, defined as the ratio of short-term words to long-term words, used to capture short-term attitude in qualitative disclosures[9]. I posit that numbers capture qualities of disclosure

[8] In untabulated tests I use the beginning of fiscal period MTB and obtain comparable results.
[9] The short-term and long-term words are defined following the dictionary provided by Brochet et al. (2015).

that are opposite to those associated with short-term words, and I therefore expect a negative relationship between them. I further include the Gunning Fog Index, a measure of linguistic complexity introduced by Gunning (1952) that is calculated as *0.4*(Average number of words per sentence + % Complex words)*. Siano and Wysocki (2018) give evidence to the fundamental link between numbers and language complexity in business documents, by showing a strongly negative association between them. Finally, I include *1/Words* that is the scaling factor used to calculate the dependent variable. I do so to control for the total length of the analyzed documents and to alleviate concerns about spurious associations between numbers and the model's variables.

I conclude the validation process with a different specification of Model 1, in which the dependent variable is *Short-Term*, while *Numbers/Words* is used as a predictor and the additional variables are unchanged. The objective is to replicate the Brochet et al. (2015) analysis.

4.3.2. Directional Link between Numbers and Long-Term Investors

The modeling choices made so far underlie an implicit causal link, between long-term investors and numbers that has not been established. Therefore, I use *Long-Term Investors* as the dependent variable in Model 1 and test its association with the full set of determinants previously identified. I expect the association between long-term investors and the explanatory variables to be similar – with regard to sign and statistical significance – to that involving numbers as the dependent variable. Two relevant exceptions are represented by *Size* and *Analyst Following*. With respect to size, it is plausible to foresee a positive association with long-term investors due to the role that "quasi-indexers" (i.e. traders investing in a diversified low-turnover portfolio of corporate securities for dividend income and/or capital appreciation) play in this category of long-term oriented market agents. Therefore, moving on to analyst following, I do not make any signed prediction since higher following could both reflect short-term pressures or firms' attributes, correlated with size, that quasi-indexers target. Overall, I expect this analysis to yield results coherent with those of previous tests.

In order to explore the directional association between numbers and long-term investors, and shed some light on the potential causal link between the two, I model their lead-lag behavior. In other words, I use the set of current period determinants to predict the one-year ahead and two-year ahead change in both long-term investors and numbers. By the logic of causality, a determinant should possess explanatory power in the prediction of future periods' outcomes. At the same time, the present period outcome should not be relevant in explaining future periods' determinants. I report the model thereafter (Model 2). Subscript *i* is used for firms, subscript *t* for fiscal years and *n* goes from 1 to 2 periods forward.

$$
\begin{aligned}
&Long\ Term\ Investors_{i,t+n} \\
&\qquad = \alpha + \beta_1\ Long\ Term\ Investors_{i,t} + \beta_2\ Numbers/Words_{i,t} \\
&\qquad + \beta_3\ Analyst\ Following_{i,t} + \beta_4\ Stock\ Compensation_{i,t} + \beta_5\ CFO\ Volatility_{i,t} \\
&\qquad + \beta_6\ Operating\ Cycle_{i,t} + \beta_7\ Leverage_{i,t} + \beta_8\ Liquidity_{i,t} + \beta_9 ROE_{i,t} \\
&\qquad + \beta_{10}\ MTB_{i,t} + \beta_{11} Size_{i,t} + \beta_{12}\ Special\ Items_{i,t} + \beta_{13} Segments_{i,t} \\
&\qquad + \beta_{14}\ Restructure_{i,t} + \beta_{15}\ Short\ Term_{i,t} + \beta_{16}\ Fog_{i,t} + \beta_{17}\ Bog_{i,t} \\
&\qquad + \beta_{18}\ 1/Words_{i,t} + Industry\ FE_i + Year\ FE_t + \varphi_{i,t+n}
\end{aligned}
$$

$$
\begin{aligned}
Numbers/Words_{i,t+n} &= \alpha + \beta_1 \ Numbers/Words_{i,t} + \beta_2 \ Long\ Term\ Investors_{i,t} \\
&+ \beta_3 \ Analyst\ Following_{i,t} + \beta_4\ Stock\ Compensation_{i,t} \\
&+ \beta_5\ CFO\ Volatility_{i,t} + \beta_6\ Operating\ Cycle_{i,t} + \beta_7\ Leverage_{i,t} \\
&+ \beta_8\ Liquidity_{i,t} + \beta_9 ROE_{i,t} + \beta_{10}\ MTB_{i,t} + \beta_{11} Size_{i,t} \\
&+ \beta_{12}\ Special\ Items_{i,t} + \beta_{13} Segments_{i,t} + \beta_{14}\ Restructure_{i,t} \\
&+ \beta_{15}\ Short\ Term_{i,t} + \beta_{16}\ Fog_{i,t} + \beta_{17}\ Bog_{i,t} + \beta_{18}\ 1/Words_{i,t} \\
&+ Industry\ FE_i + Year\ FE_t + \theta_{i,t+n}
\end{aligned}
$$

(Model 2)

5. Results

5.1. Descriptive Evidence

The final sample comprises 14,253 firms over the 1994-2017 (fiscal years) time period. Table 1 reports descriptive statistics about the main variables – continuous metrics are all winsorized at the 1st and 99th percentiles. The average firm has a market capitalization of about $ 220 MM and is followed by 9 analysts. The proportion of shares held by “dedicated” and “quasi-indexers” investors (i.e. long-term) exceeds “transient” (i.e. short-term) ownership. Numbers represent 3% of total words within the MD&A section and every other sentence contains quantitative information. The Fog index is about 19 and short-term words are almost 50% of long-term ones. Interestingly, past and present verb tenses are concentrated in quantitative sentences (i.e. on average 2.88 per numeric sentence *versus* 2.62 per non-numeric sentence) while future verbs are more frequent in sentences that do not contain numbers (i.e. on average 0.03 per quantitative sentence *versus* 0.08 per non-quantitative sentence). The latter evidence indicates that numbers are more likely included in sentences that describe past and present business and less likely referred to future events. In fact, firms could find problematic referring to future uncertainties using verifiable and objective information.

5.2. Cross-sectional Validation of Numbers/Words

I test the association between the prevalence of numbers within text and determinants of disclosure horizon. Table 2 reports the regression results. Columns (1) and (2) show the baseline version of Model 1 without and with textual controls respectively. Consistently with my expectations, numbers are positively associated with long-term investors (i.e. a statistically-significant coefficient on *Long-Term Investors* of 0.002). *Numbers/Words* is negatively associated with capital-market and incentive-based pressures on the short-term (i.e. a negative and statistically-significant coefficient of -0.004 on *Analyst Following* and a negative and statistically-significant coefficient of -0.0001 on *Stock Compensation*). The relationship with other firm-level determinants of disclosure horizon is in line with the stated predictions. Interestingly, numbers are negatively correlated with the proportion of short-term words within MD&A sections, a finding that confirms how quantitative information capture qualities of disclosure that are opposite to those of short-term narratives. The regressions also give evidence to the appropriateness of controlling for textual attributes, especially *Fog*, that are highly correlated with numbers and increase remarkably the OLS explained variation (i.e. the adjusted R-squared increases from 18% to 31.5%). In Column (3) I substitute the Gunning *Fog* Index with the Stylewriter *Bog* Index, which prior literature introduces as a more accurate measure of readability, and I find comparable results. In Column (4) I include a first-order autoregressive term (i.e. *Lag Numbers/Words*) to mitigate concerns of potential correlated omitted variable bias. While the association between numbers and several determinants is subsumed by lagged numbers, the relationship with *Long-Term Investors* remains positive and statistically significant. Again, at this point, the modeling choice should only be interpreted as a robustness check and does not imply the intention to look at a *change* in the dependent variable. Finally, specification (5) controls for *Arc* that is meant to capture accounting complexity and is computed as the natural logarithm of the number of XBRL tags in a firm's financial statements scaled by the natural logarithm of a firm's total assets[10]. The positive and significant relationship between numbers and long-term investors is confirmed. In untabulated analyses, I use the natural logarithm of *Numbers/Words* as the dependent variable, to correct for skewness, and find qualitatively and statistically similar results.

In Table 3 I replicate the Brochet et al. (2015) determinants regression, using *Short-Term* as the dependent variable in Model 1. Columns (1) and (2), coherently with my hypothesis and with the authors' results, show reversed sign associations with respect to those highlighted in Table 2. In other words, *Short-Term* is negatively and significantly associated with *Long-Term Investors* and positively and significantly related to *Analyst Following* and *Stock Compensation*. Moreover, as previously outlined, numbers and *Short-Term* are strongly and negatively correlated. Using *Bog* in specification (3) does not change inferences. On the other hand, the inclusion of a first-order autoregressive variable (i.e. *Lag Short-Term*) in Column (4) makes the association between *Short-Term* and *Long-Term Investors* statistically insignificant. I notice here two things. First, the relationship between number and long-term investors was robust to lagged numbers. Second, we should read Column (4) results with caution as I am modeling a *change* in the dependent variable that could be prone to misspecification errors. Controlling for *Arc*, in Column (5), I again find an insignificant relationship with long-term investors.

[10] I download this measure, available for fiscal years 2008-2017, from http://www.xbrlresearch.com/.

Legitimately, the reader could be concerned that the determinants model, with *Numbers/Words* as the dependent variable, is not robust to the inclusion of *dates* as an additional independent variable. In this work, I posit that the inclusion of dates on the right-hand side of the regression specification is not appropriate from both a conceptual and an empirical perspective. Conceptually, it is worth noticing that – in the vast majority of cases – dates provide context in time to numbers, within business documents. Therefore, even without leveraging on a particular theory, it is reasonable to expect that numbers are not determined by dates. Therefore, when dates are employed to explain numbers, the regression is prone to misspecification errors. Empirically, the high correlation between numbers and dates (i.e. 0.88) is not surprising. Thus, multicollinearity is a relevant problem especially when numbers and dates are both accounted for as explanatory variables. I nonetheless include dates (in untabulated analyses) both as a dependent and independent variable and find consistent results[11].

5.3. The Link between Numbers and Long-Term Investors

I provide additional evidence about the potential link between numbers and long-term investors. To begin with, given the correlational nature of this study, In Table 4 I replicate previous analyses (in Table 2) using *Long-Term Investors* as the dependent variable. I find that numbers are strongly associated with investors oriented to the long run even when a more restrictive specification, that employs a first-order autocorrelated term, is used. The additional regressors follow the predictions that I previously formulated. It is interesting to notice that *Analyst Following* is positively associated to *Long-Term Investors*, consistently with "quasi-indexers" targeting firms' characteristics correlated with *Size* and therefore the presence of

[11] I use Dates/Words, (Numbers-Dates)/Words, Dates/Sentences.

analysts. Additional textual variables load in a statistically significant way but do not subsume the role of numbers within text. I then run further tests, reported in Table 5, representing a first attempt towards identification. In fact, the contemporaneous association between numbers and long-term investors leaves many interesting questions unanswered. For instance, do numbers reflect long-term investors' ownership and their disclosure preferences or, on the other hand, do companies tailor their communication to attract prospective value investors? In order to shed some light on this question, I examine the lead-lag relationship between numbers and long-term investors. In Columns (1) and (2) the dependent variable is *Long-Term Investors*, while in Columns (2) and (3) it is *Numbers/Words*. Columns (1) and (3) report the one-period forward value of each dependent variable, while Columns (2) and (4) the two-period forward one. Each specification includes current period explanatory variables identical to those used in previous tests, together with a first-order autoregressive term. In practice, the presented OLS regressions are used to model the one-period and the two-period ahead *change* in the dependent variable as a function of present period's regressors. The results show that while numbers within text are a relevant predictor of future periods long-term investors (i.e. a positive and statistically-significant coefficient of 0.198 for the one-year change and 0.443 for the two-year change), the opposite is not true. In other words, current period levels of long-term investors do not explain numbers in the next two fiscal years. The reported effects are robust to the additional inclusion of firms fixed effects (untabulated analyses). In summary, Table 5 provides preliminary evidence consistent with numbers being provided in qualitative disclosures to attract investments oriented to the long-term.

6. Conclusions and Future Work

A recent strand of the literature has started to characterize the *qualitative* attributes of corporate disclosure, with a special focus on voluntary narratives. However, in the vast majority of cases, researchers have removed or ignored *quantitative* information. In this paper, I offer an empirical test about how numbers within the text of the voluntary disclosure system synthetize *qualities* of information that are especially relevant to long-term institutional investors.

Consistent with the quantitative focus of traditional literature in accounting and finance, I argue that numbers found in the narratives accompanying the financial statements convey relevant information to company stakeholders. In particular, given their precision, objectivity and factual nature, numbers are well suited to describe trends that are likely to impact the fundamental value of business transactions. I use the universe of EDGAR 10-K filings and their MD&A section from 1994 to 2017 and show that quantitative information within the text of corporate reports are positively associated with the level of long-term institutional ownership and negatively associated with the level of capital market and incentive-based pressures on the short-term (i.e. *Analyst Following* and *Stock Compensation*, respectively). In this regard, it is important to notice two facts. First, numbers provide information oriented to the long-term regardless of their natural proximity to present and past verb tenses. In other words, the prevalence of numbers in sentences describing the present or the past does not impair the fundamental ability of numeric data to convey information that is relevant to long-term value. Second, I show that numbers capture a dimension of business narratives that is opposite in direction with respect to the word-based measure of disclosure horizon that Brochet et al. (2015) find to be associated with "short-termism" and managerial myopia. Overall, these results are consistent with the view that numbers disclosed in the body of qualitative narratives represent key covariates linked to the firms' disclosure policies and their related outcomes.

In this work, I provide insights into the directional link between numbers and long-term investors. In a set of lead-lag tests, I show that present period's numbers are a relevant predictor of the one-year and the two-year ahead changes in long-term investors. However, present period long-term ownership does not seem to affect changes in the amounts of disclosed numbers. This evidence represents a first attempt toward causality and is consistent with the signaling role of voluntary disclosure.

In addition, my study offers with new opportunities for the empirical characterization of long-term investors, not only based on their revealed trading behavior, but also related to an important set of information employed for investment decisions.

Future relevant contributions could result from the attribution of numbers within qualitative disclosure to the financial statement prototype (e.g., net income, general and administrative expenses) they belong to. In addition to that, disclosed numbers in text could be linked to recognized numbers in financial statements to assess more precisely their incremental informativeness. Finally, structural modeling might represent a promising avenue to parse out causality in the presented relationships.

References

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

Bolton, P., Scheinkman, J., and Xiong, W., 2006. Executive compensation and short-termist behavior in speculative markets. *Review of Economic Studies* 73(3), 577–610.

Bonsall, S., Leone A., Miller, B. P., and Rennekamp, K., 2017. A Plain English Measure of Financial Reporting Readability. *Journal of Accounting and Economics* 63 (2-3): 329-357.

Boone, A. L., and White, J. T., 2015. The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics* 117(3), 508–533.

Bozanic, Z., Roulstone, D. T., and Van Buskirk, A., 2018. Management earnings forecasts and other forward-looking statements. *Journal of Accounting and Economics* 65(1), 1-20.

Brochet, F., Loumioti, M., and Serafeim, G., 2015. Speaking of the short-term: disclosure horizon and managerial myopia. *Review of Accounting Studies* 20, 1122–1163.

Bushee, B., 1998. The influence of institutional investors on myopic R&D investment behavior. *The Accounting Review* 73(3), 305–333.

Bushee, B., 2001. Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research* 18(2), 207–246.

Bushee, B., and Noe, C. F., 2001. Corporate disclosure practices, institutional investors and stock-return volatility. *Journal of Accounting Research* 38(supplement), 171–202.

Cadman, B. D., Heinle, M. S., and Macciocchi, D., 2019. Investor horizon and disclosure policies. *Working paper*: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3175477.

Cheng, I. H., Hong, H., and Scheinkman, J. A., 2015. Yesterday's heroes: Compensation and creative risk-taking. *Journal of Finance* 70(2), 839–879.

Dyer, T., M. Lang, and L. Stice-Lawrence. 2017a. The evolution of 10-K textual disclosure: evidence from latent dirichlet allocation. *Journal of Accounting and Economics*.

Edmans, A., Fang, V., and Lewellen, K., 2017. Equity vesting and managerial myopia. *The Review of Financial Studies* 30, pp.2229-2271.

Ertugrul, M., Lei, J., Qiu, J., and Wan, C., 2017. Annual report readability, tone ambiguity, and the cost of borrowing. *Journal of Financial and Quantitative Analysis* 52, 811-836.

Frankel, R., Mayew, W. J., and Sun, Y., 2009. Do pennies matter? Investor relations consequences of small negative earnings surprises. *Review of Accounting Studies* 15, 220–242.

Gopalan, R., Milbourn, T., Song, F., and Thakor, A., 2014. Duration of executive compensation. *Journal of Finance* 69(6), 2777–2817.

Gunning, R., 1952. The technique of clear writing. New York, NY: *McGraw-Hill International Book Co*.

He, J., and Tian, X., 2013. The dark side of analyst coverage: The case of innovation. *Journal of Financial Economics* 109(3), 856–878.

Healy, P. and Palepu, K., 2001. Information asymmetry, corporate disclosure and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics* 31.

Hoitash, R., and Hoitash, U., 2018. Measuring Accounting Reporting Complexity with XBRL. *The Accounting Review* (93)1, 259-287.

Korbaenkova, A., Siano, F., and Wysocki, P. D., 2019. Recognition versus disclosure of numbers in 10-K filings: Measures, determinants and outcomes. *Boston University working paper*.

Lang, M., and Stice-Lawrence L., 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60, 110–35.

Lawrence, A., 2013. Individual investors and financial disclosure. *Journal of Accounting and Economics* 56, 130-147.

Laptev, N., Lawrence, A., Ryans, J., and Sun, E., 2018. Earnings announcement promotions: a Yahoo Finance field experiment. *Journal of Accounting and Economics* 66(2-3), 399-414.

Lehavy, R., F. Li, and K. Merkley, 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review* 86, 1087–115.

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

Levitt, A., 2000. *Speech at the Center of Law and Business.* New York University, May 10. https://www.sec.gov/news/speech/spch370.htm.

Li, F., 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45, 221–47.

Li, F., 2011. Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature* 29.

Lo, K., Ramos, F., and Rogo, R., 2017. Earnings management and annual report readability. *Journal of Accounting and Economics* 63.

Loughran, T., and McDonald, B., 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54, 187–230.

Lundholm, R., Rogo, R., and, Zhang, J., 2014. Restoring the tower of Babel: how foreign firms communicate with US investors. *The Accounting Review* 89, 1453-1485.

Merkley, K., 2014. Narrative disclosure and earnings performance: Evidence from R&D disclosures. *The Accounting Review* 89(2), 725–757.

Siano, F., and Wysocki, P. D., 2018. The primacy of numbers in financial and accounting disclosures: Implications for textual analysis research. *Boston University working paper*.

Warren, G., 2014. Long-term investing: what determines investment horizon? *CIFR working paper*: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2513088.

Variables Definitions

Variable	Description	Source
Long-Term Investors	The number of shares held by *dedicated* investors and *quasi-indexers* minus the shares held by *transient* investors scaled by total shares outstanding	Bushee (1998) classification and Reuters 13F
Analyst Following	ln(max num. estimates) / ln(total assets)	I/B/E/S
Stock Compensation	Residual from regressing average top-5 executives stock and option-based compensation on size, MTB and industry and year fixed effects	Execucomp
CFO Volatility	5-year standard deviation of operating cash-flows scaled by total assets	Compustat
Operating Cycle	Ln([(inventory/COGS)*360 + (receivables/sales)*360])	Compustat
Leverage	(ST-liabilities + LT-liabilities)/total assets	Compustat
Liquidity	Current assets/current liabilities	Compustat
MTB	Fiscal closing price / book value per share	Compustat
Special Items	Abs(special items)/total assets	Compustat
Restructure	Dummy = 1 if reported pre-tax restructuring > 0, and 0 otherwise	Compustat
ROE	Net income/book value of equity	Compustat
Size	Ln(market capitalization)	Compustat
Segments	Ln(business and operating segments)	Compustat
Numbers/Words	Number of numbers read in a firm's 10-K MD&A section scaled by the total number of words found in that section	EDGAR
Short-Term	Short-term oriented to long-term oriented words (Brochet et al.,2015) first introduced in the context of earnings conference calls	EDGAR
Fog	0.4*(average words per sentence + % complex words)	EDGAR
Bog	Stylewriter readability index introduced by Bonsall et al. (2017)	Brian P. Miller website
Arc	Ln(number of XBRL tags)/ln(total assets) introduced by Hoitash and Hoitash (2018)	Rani and Udi Hoitash website

Table 1
Sample Statistics for Key Variables in 10-K MD&A Sample

Firms: 14,253	# Obs.	Mean	SD	P25	Median	P75
Long Term Investors	55,810	0.23	0.22	0.05	0.21	0.41
Analyst Following	105,217	0.16	0.16	0.00	0.15	0.31
Stock Compensation	105,217	-0.03	0.76	-0.42	-0.00	0.38
CFO Volatility	93,836	0.11	0.29	0.02	0.04	0.08
Operating Cycle	96,551	4.91	1.39	4.17	4.71	5.32
Leverage	103,412	0.31	0.46	0.03	0.22	0.41
Liquidity	85,092	2.84	3.34	1.12	1.84	3.15
Roe	103,726	-0.01	1.31	-0.08	0.07	0.15
MTB	95,839	2.55	7.89	0.94	1.73	3.23
Size	96,447	5.36	2.27	3.84	5.41	6.94
Special Items	102,467	0.04	0.13	0.00	0.00	0.02
Segments	95,563	0.51	0.65	0.00	0.00	1.1
Restructuring	105,217	0.17	0.38	0.00	0.00	0.00
Numbers/Words	105,217	0.03	0.01	0.02	0.02	0.03
Short Term	104,873	0.55	0.50	0.22	0.41	0.07
Fog	105,217	18.96	1.54	17.96	18.91	19.92
Bog	101,322	83.75	7.38	79	84	89
Arc	26,153	1.10	1.02	0.71	0.84	1.06

This table presents summary sample statistics related to relevant variables used within regressions. The sample consists of 14,253 unique firms and data spans the 1994-2017 (fiscal years) time period. Variables are described in the *Variable Definitions* section. All continuous variables are winsorized at the 1st and 99th percentiles

Table 2
Validation of the Prevalence of Numbers within the Narratives of 10-K MD&A

Variables	Pred. Sign	Dependent Variable: *Numbers/Words*				
		(1)	(2)	(3)	(4)	(5)
Long-Term Investors	**(+)**	**0.002*** [3.4]**	**0.002*** [3.6]**	**0.001*** [3.1]**	**0.000** [2.0]**	**0.001** [2.1]**
Analyst Following	(-)	-0.004*** [-5.6]	-0.004*** [-6.7]	-0.003*** [-5.2]	-0.001*** [-4.8]	-0.002** [-2.2]
Stock Compensation	(-)	-0.000*** [-5.6]	-0.000*** [-4.4]	-0.000*** [-5.0]	0.000 [0.1]	-0.000** [-2.3]
CFO Volatility	(-)	-0.003*** [-8.5]	-0.003*** [-7.4]	-0.003*** [-8.2]	0.000 [0.1]	-0.004*** [-4.0]
Operating Cycle	(+)	0.000*** [3.1]	0.000* [1.7]	0.000*** [2.9]	0.000 [0.8]	0.000*** [3.1]
Leverage	(?)	0.001*** [3.0]	0.001*** [4.7]	0.001*** [3.1]	0.000*** [3.3]	0.001 [1.5]
Liquidity	(+)	-0.000* [-1.7]	-0.000 [-0.7]	0.000 [0.2]	0.000** [2.5]	0.000*** [2.7]
Roe	(+)	0.000*** [4.4]	0.000*** [3.5]	0.000*** [3.5]	-0.000 [-0.1]	-0.000 [-0.6]
MTB	(+)	-0.000 [-0.3]	0.000 [0.0]	-0.000 [-0.3]	-0.000 [-1.1]	-0.000 [-0.4]
Size	(?)	-0.000*** [-7.2]	-0.000*** [-4.4]	-0.000*** [-6.5]	-0.000 [-1.5]	-0.000 [-1.6]
Special Items	(-)	-0.004*** [-9.0]	-0.004*** [-8.3]	-0.004*** [-8.6]	-0.001*** [-4.0]	-0.005*** [-4.8]
Business Segments	(+)	0.001*** [7.0]	0.001*** [6.4]	0.001*** [8.1]	0.000*** [2.9]	0.001*** [3.5]
Restructuring	(-)	-0.000* [-1.8]	0.000 [0.3]	-0.000 [-0.7]	0.000* [1.9]	0.000 [1.4]
Short-Term	(-)		-0.002*** [-8.2]	-0.002*** [-8.2]	-0.001*** [-7.2]	-0.001*** [-2.7]
Fog	(-)		-0.002*** [-31.1]		-0.001*** [-26.1]	-0.002*** [-17.3]
1/Words	(-)		-2.480*** [-20.8]	-2.733*** [-24.1]	-1.475*** [-16.8]	-3.135*** [-13.4]
Bog	(-)			-0.000*** [-21.7]		
Lag Numbers/Words	(+)				0.729*** [131.7]	
Arc	(?)					0.000* [1.7]
Observations		40,561	40,314	39,469	36,820	12,190
Adjusted R-squared		0.180	0.315	0.273	0.727	0.248
SE Cluster		Firm	Firm	Firm	Firm	Firm
Industry Fixed Effects		YES	YES	YES	YES	YES
Year Fixed Effects		YES	YES	YES	YES	YES

This table shows the regression results (OLS) using as dependent variable the Numbers/Words measure, calculated for 1994-2017 (fiscal years) 10-K MD&A narratives, and as independent variables a contemporaneous set of disclosure horizon determinants. Continuous explanatory variables are winsorized at the 1st and 99th percentiles. Industry fixed effects are based on Fama and French 48-industry definitions. All regressions are estimated with an intercept (not reported). Cluster-robust-to-heteroskedasticity t-statistics are reported in [] parentheses. ***, **, * indicate significance at <0.01, <0.05, <0.10 respectively (two-tailed tests). Coefficients of special interest are in boldface type.

Table 3
Replication of the Brochet et al. (2015) Results on the Association between "Short-Term" and Determinants of Disclosure Horizon

Variables	Pred. Sign	Dependent Variable: *Short-Term*				
		(1)	(2)	(3)	(4)	(5)
Long-Term Investors	**(-)**	**-0.060***** **[-2.7]**	**-0.055**** **[-2.5]**	**-0.056**** **[-2.5]**	**-0.003** **[-0.4]**	**-0.033** **[-0.9]**
Analyst Following	(+)	0.276*** [7.2]	0.261*** [6.8]	0.267*** [6.9]	0.065*** [4.1]	0.074 [1.2]
Stock Compensation	(+)	0.019*** [4.4]	0.018*** [4.1]	0.017*** [3.9]	0.004** [2.0]	0.014** [2.1]
CFO Volatility	(+)	0.060** [2.4]	0.048** [2.0]	0.050** [2.0]	-0.010 [-0.9]	0.027 [0.5]
Operating Cycle	(-)	-0.001 [-0.2]	-0.001 [-0.1]	0.001 [0.1]	0.001 [0.2]	-0.006 [-0.7]
Leverage	(?)	-0.047*** [-3.2]	-0.040*** [-2.7]	-0.043*** [-2.9]	-0.010 [-1.6]	-0.041* [-1.8]
Liquidity	(-)	-0.001 [-0.4]	-0.001 [-0.4]	-0.001 [-0.4]	-0.000 [-0.3]	-0.003 [-1.0]
Roe	(-)	-0.003 [-1.2]	-0.003 [-1.1]	-0.003 [-1.2]	-0.002 [-1.1]	-0.007 [-1.6]
MTB	(-)	0.000 [0.3]	0.000 [0.2]	0.000 [0.2]	0.000 [0.4]	0.000 [0.1]
Size	(?)	-0.002 [-0.7]	-0.003 [-0.9]	-0.003 [-1.0]	-0.001 [-0.8]	-0.001 [-0.3]
Special Items	(+)	0.166*** [5.7]	0.154*** [5.3]	0.159*** [5.3]	0.089*** [3.7]	0.099* [1.7]
Business Segments	(-)	-0.050*** [-6.3]	-0.045*** [-5.7]	-0.046*** [-5.7]	-0.013*** [-4.7]	-0.032*** [-2.9]
Restructuring	(+)	0.060*** [6.5]	0.062*** [6.7]	0.060*** [6.4]	0.014*** [3.5]	0.010 [0.8]
Numbers/Words	(-)		-3.870*** [-7.2]	-3.549*** [-6.8]	-1.301*** [-5.6]	-2.863*** [-2.9]
Fog	(?)		-0.011*** [-2.7]		-0.003* [-1.9]	-0.020*** [-3.2]
Bog	(?)			-0.001 [-1.4]		
Lag Short Term	(+)				0.746*** [97.5]	
Arc	(?)					-0.029** [-2.0]
Observations		40,488	40,488	39,640	36,940	12,221
Adjusted R-squared		0.059	0.063	0.064	0.576	0.038
SE Cluster		Firm	Firm	Firm	Firm	Firm
Industry Fixed Effects		YES	YES	YES	YES	YES
Year Fixed Effects		YES	YES	YES	YES	YES

This table shows the replication results (OLS) of Short-Term, the Brochet et al. (2015) word-based measure for disclosure horizon calculated as #Short-Term Words / #Long-Term Words. Short and Long-Term words in this paper follow the Brochet et al. (2015) dictionary. The analysis is performed on 1994-2017 (fiscal years) MD&A sections of 10-K filings. Continuous explanatory variables are winsorized at the 1st and 99th percentiles. Industry fixed effects are based on Fama and French 48-industry definitions. All regressions are estimated with an intercept (not reported). Cluster-robust-to-heteroskedasticity t-statistics are reported in [] parentheses. ***, **, * indicate significance at <0.01, <0.05, <0.10 respectively (two-tailed tests). Coefficients of special interest are in boldface type.

Table 4
The Association between Long-Term Investors and Numbers using Long-Term Investors as the Dependent Variable

Variables	Pred. Sign	Dependent Variable: *Long-Term Investors*				
Numbers/Words	**(+)**	**0.619*** [3.4]**	**0.731*** [3.6]**	**0.615*** [3.1]**	**0.148*** [3.7]**	**0.561*** [2.8]**
Analyst Following	(?)	0.128*** [8.3]	0.133*** [8.6]	0.133*** [8.5]	0.092*** [15.8]	0.136*** [5.3]
Stock Compensation	(-)	-0.012*** [-7.6]	-0.012*** [-7.4]	-0.011*** [-7.1]	-0.002*** [-2.9]	-0.011*** [-4.7]
CFO Volatility	(-)	-0.050*** [-7.1]	-0.050*** [-7.1]	-0.051*** [-7.3]	-0.017*** [-5.7]	-0.072*** [-3.5]
Operating Cycle	(+)	0.011*** [4.0]	0.011*** [4.0]	0.011*** [4.1]	0.002** [2.1]	0.015*** [3.9]
Leverage	(?)	-0.006 [-1.0]	-0.005 [-1.0]	-0.004 [-0.7]	-0.003 [-1.5]	-0.011 [-1.1]
Liquidity	(+)	-0.003*** [-3.8]	-0.003*** [-3.9]	-0.002*** [-3.6]	0.001** [2.4]	-0.001 [-0.8]
Roe	(+)	-0.002** [-2.2]	-0.002** [-2.3]	-0.002** [-2.4]	-0.000 [-0.5]	0.000 [0.1]
MTB	(+)	-0.002*** [-10.0]	-0.002*** [-9.9]	-0.002*** [-9.9]	-0.001*** [-5.9]	-0.001* [-1.8]
Size	(+)	0.038*** [31.9]	0.038*** [31.9]	0.038*** [31.7]	0.007*** [16.0]	0.039*** [21.3]
Special Items	(-)	-0.008 [-0.9]	-0.005 [-0.5]	-0.004 [-0.4]	-0.007 [-1.1]	0.035* [1.7]
Business Segments	(+)	0.022*** [6.5]	0.022*** [6.5]	0.023*** [6.8]	0.003*** [2.6]	0.013*** [2.8]
Restructuring	(-)	0.036*** [9.5]	0.038*** [9.9]	0.038*** [10.0]	-0.002 [-1.0]	0.027*** [4.8]
Short-Term	(-)		-0.008** [-2.3]	-0.008** [-2.3]	-0.003** [-2.0]	-0.006 [-0.8]
Fog	(?)		-0.000 [-0.2]		0.000 [0.5]	-0.002 [-0.6]
1/Words	(?)		5.540*** [3.4]	4.705*** [2.9]	1.769** [2.2]	2.752 [0.9]
Bog	(?)			-0.001*** [-2.7]		
Lag Long-Term Investors	(+)				0.807*** [228.4]	
Arc	(?)					-0.012*** [-2.9]
Observations		40,561	40,314	39,469	33,284	12,190
Adjusted R-squared		0.278	0.278	0.278	0.761	0.300
SE Cluster		Firm	Firm	Firm	Firm	Firm
Industry Fixed Effects		YES	YES	YES	YES	YES
Year Fixed Effects		YES	YES	YES	YES	YES

This table shows the regression results (OLS) using as dependent variable Long-Term Investors, calculated from Bushee's (1998) classification and Reuters 13F, and as independent variables the determinants of disclosure horizon previously investigated. Data samples refer to 1994-2017 (fiscal years) 10-K MD&A narratives. Continuous explanatory variables are winsorized at the 1st and 99th percentiles. Industry fixed effects are based on Fama and French 48-industry definitions. All regressions are estimated with an intercept (not reported). Cluster-robust-to-heteroskedasticity t-statistics are reported in [] parentheses. ***, **, * indicate significance at <0.01, <0.05, <0.10 respectively (two-tailed tests). Coefficients of special interest are in boldface type.

Table 5
The Lead-Lag Association between Current Period Numbers/Words (Long-Term Investors) and Future Changes in Long-Term Investors (Numbers)

Variables	*Long-Term Investors*		*Number/Words*	
	(1)	(2)	(3)	(4)
	(T+1)	(T+2)	(T+1)	(T+2)
Numbers/Words (T)	**0.198*****	**0.443*****	**0.778*****	**0.630*****
	[2.8]	**[3.5]**	**[156.1]**	**[79.7]**
Long-Term Investors (T)	**0.795*****	**0.612*****	**0.000**	**0.000**
	[215.7]	**[89.0]**	**[1.3]**	**[1.1]**
Analyst Following	0.046***	0.059***	-0.000	-0.000
	[7.8]	[5.7]	[-1.6]	[-0.2]
Stock Compensation	-0.002**	-0.005***	-0.000**	-0.000
	[-2.3]	[-5.4]	[-2.5]	[-1.6]
CFO Volatility	-0.014***	-0.035***	-0.001***	-0.001***
	[-3.6]	[-4.1]	[-2.6]	[-2.8]
Operating Cycle	0.001	0.001	0.000***	0.000***
	[0.8]	[0.3]	[3.0]	[2.7]
Leverage	-0.006**	-0.012**	0.000*	0.000*
	[-2.5]	[-2.5]	[1.8]	[1.8]
Liquidity	0.001***	0.001*	-0.000	0.000
	[3.4]	[1.9]	[-0.1]	[0.6]
Roe	0.002***	0.004***	-0.000	-0.000**
	[3.1]	[4.1]	[-0.4]	[-2.3]
MTB	-0.000**	-0.000	-0.000	-0.000
	[-2.2]	[-1.1]	[-1.1]	[-1.5]
Size	0.012***	0.024***	-0.000**	-0.000***
	[25.7]	[27.1]	[-2.5]	[-3.7]
Special Items	-0.046***	-0.064***	0.001**	0.000
	[-7.3]	[-6.2]	[2.2]	[0.7]
Business Segments	0.000	-0.000	0.000***	0.000***
	[0.5]	[-0.2]	[3.8]	[4.3]
Restructuring	-0.010***	-0.011***	0.000	0.000
	[-6.5]	[-4.4]	[1.4]	[0.8]
Short-Term	-0.003**	-0.007***	-0.000**	-0.000
	[-2.5]	[-3.1]	[-2.1]	[-1.6]
Fog	-0.000	-0.002*	-0.000	-0.000
	[-0.4]	[-1.7]	[-0.7]	[-0.4]
Bog	-0.000***	-0.000**	-0.000***	-0.000***
	[-2.9]	[-2.1]	[-3.0]	[-4.1]
1/Words	1.325*	2.872**	0.379***	0.337***
	[1.7]	[2.3]	[4.4]	[3.3]
Observations	32,174	27,070	34,561	29,831
Adjusted R-squared	0.766	0.597	0.704	0.533

SE Cluster	Firm	Firm	Firm	Firm
Industry Fixed Effects	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES

This table shows the lead-lag regression results (OLS) using as dependent variables T+1 and T+2 Long-Term Investors and Number/Words. Data samples refer to 1994-2017 (fiscal years) 10-K MD&A narratives. All regression include a first-order autoregressive term. Continuous explanatory variables are winsorized at the 1^{st} and 99^{th} percentiles. Industry fixed effects are based on Fama and French 48-industry definitions. All regressions are estimated with an intercept (not reported). Cluster-robust-to-heteroskedasticity t-statistics are reported in [] parentheses. ***, **, * indicate significance at <0.01, <0.05, <0.10 respectively (two-tailed tests). Coefficients of special interest are in boldface type.