

Information Supporting Investor Valuations: Evidence from a Comparative Content Analysis of Analyst Reports and Form 10-K

Mary E. Barth*

Graduate School of Business
Stanford University

Ken Li

DeGroot School of Business
McMaster University

Charles G. McClure

Booth School of Business
University of Chicago

Preliminary and Incomplete. Do not quote without authors' permission.

March 2023

* Corresponding author. Graduate School of Business, Stanford University, 655 Knight Way, Stanford, CA 94305. Email: mbarth@stanford.edu

We thank seminar participants at the University of Melbourne for helpful comments and suggestions. We also thank the Stanford University Graduate School of Business Data, Analytics, and Research Computing for research support and Boyie Chen for excellent research assistance. We appreciate the support provided by the Stanford University Graduate School of Business, the Social Sciences and Humanities Research Council of Canada, and the Liew Faculty Fellowship at the University of Chicago Booth School of Business.

Information Supporting Investor Valuations: Evidence from a Comparative Content Analysis of Analyst Reports and Form 10-K

Abstract

We address whether financial reports include information supporting investor valuations. We view analyst reports (AR) as reflecting this information and employ topic modeling to compare the contents of AR and Form 10-K. Our main findings follow. (i) Regarding topic categories, Form 10-K focuses heavily on financial reporting, whereas AR focuses most on performance, followed by analysis and business. However, AR discusses financial reporting almost as much as business, which suggests financial reporting is a crucial component of AR. (ii) Regarding performance topics, AR and Form 10-K both focus most on revenues and margins. As expected, Form 10-K focuses more than AR on earnings and expenses, whereas AR focuses more on ratios, target prices, recommendations, and adjusted earnings. Surprisingly, AR focuses more on earnings than adjusted earnings or cash flows. (iii) Strikingly, Form 10-K's MD&A section focuses more than AR on performance, analysis, and business, which suggests MD&A discussion resembles AR discussion. (iv) Differences in AR content for economically dissimilar firms depend on firms' circumstances but are small. Together, our findings reveal financial reports include information supporting investors' valuations, which is inconsistent with financial reports lacking relevance.

JEL classification: G10, G18, M40, M41

Keywords: Capital Markets; Financial Reporting; Textual Analysis; Analyst Reports; Value Relevance

Information Supporting Investor Valuations: Evidence from a Comparative Content Analysis of Analyst Reports and Form 10-K

1. Introduction

The question we address is whether financial reports include information supporting investor valuations of firms. This question is fundamental to financial reporting because financial reports are designed to provide investors with information to help them make their decisions about providing capital to a firm. Yet, Form 10-K, a firm's primary financial report, has been criticized for lacking information that is relevant to investors. We view AR as containing information supporting investor valuations because AR describe the fundamental analysis that justifies the analyst's forecasts of the firm's performance and recommendations to investors regarding whether to buy, hold, or sell the firm's equity securities. Thus, we address our research question by analyzing and comparing the contents of equity analysts' reports (AR) and Form 10-K. Taken together, our findings reveal many similarities between the contents of Form 10-K and AR, which calls into question the criticism that financial reports lack relevance.

We compare the contents of Form 10-K and AR for three reasons. First, equity analysts comprise a large class of users of financial reports who have an equity investor perspective. Although analysts are not investors per se, they provide investors with analysis supporting their valuations and investment advice, and their recommendations are often a major input for institutional investors. Second, whereas the information in Form 10-K must comply with U.S. Securities and Exchange Commission (SEC) regulations and U.S. GAAP, analysts can choose the content of their report for each firm. Thus, the content of AR reveals what information supports the analyst's valuation analysis. Third, related prior research finds that AR content reflects information investors find relevant to valuing a firm's equity.

To address our research question, we assemble a sample of 26,757 equity analyst reports that provide fundamental analysis and Forms 10-K for 4,335 firms from 1996 to 2018. We then use a neural network-based word embedding topic modeling approach, Word2vec, to identify topics discussed in the combined AR and Form 10-K corpus. Word2vec groups together words with similar meanings and includes each word in the corpus into one of the groups. These groups represent topics discussed in AR and Form 10-K. We organize these topics into six broad categories: Performance, Analysis, Business, Financial Reporting, Regulatory, and Other. We then calculate the proportion of words in AR and Form 10-K in each topic. We use these proportions to analyze and compare the contents of AR and Form 10-K.

We also compare the contents of AR and two prominent sections of Form 10-K, Management's Discussion and Analysis (MD&A) and Financial Statement (FS). These two sections are designed to provide investors with different types of information. MD&A provides management's analysis of the firm's resources and operations, whereas FS reports on the firm's financial condition and performance. As a result, we expect AR to bear more resemblance to MD&A than to FS. We also compare AR content for several pairs of economically dissimilar types of firms to provide insights into how the content of AR varies depending on a firm's circumstances. These pairs are profit and loss, large and small, non-technology and technology, and non-financial and financial firms.

Regarding topic categories, we find that AR discusses Performance the most, followed by Analysis and Business, which is consistent with analysts' interpretation and analysis role. Perhaps surprisingly, analysts devote almost as much of their reports to Financial Reporting as to each of the three top categories. This finding reveals that, contrary to claims of lack of relevance, financial reporting is a crucial component of AR. As expected, we find that analysts

devote considerably less of their reports to Regulatory topics. We also find that the extent of Form 10-K discussion of all categories differs from that of AR. The largest difference relates to Financial Reporting. Perhaps surprisingly, Form 10-K discusses Financial Reporting almost twice as much as AR.

Regarding topics, we find that topics discussed in AR and Form 10-K differ most for Performance. In particular, AR has considerable discussion of Revenues/Margins, Ratios, Target Prices, and Recommendations. These findings, again, are consistent with the analysis role of analysts. However, we find that Earnings is the third most-discussed topic in AR, whereas Adjusted Earnings is the seventh and Cash Flows is the least. These findings are inconsistent with analysts viewing earnings as irrelevant and instead focusing on alternative performance measures, including cash flows. Although Revenues/Margins also is the top topic for Form 10-K, the second is Earnings, and the third is Expenses. As expected, Ratios, Target Prices, and Recommendations are the three least-discussed topics in Form 10-K.

Regarding Analysis and Business, we find that AR and Form 10-K generally discuss similar topics and discuss each topic to a similar extent. Most notably, for both AR and Form 10-K, the most-discussed Analysis topic is Markets and Industries followed by Trends and Forecasts, and the most-discussed Business topic is Products and Markets.

Regarding Financial Reporting, we find that AR focuses more than Form 10-K on Investments, Equity, Interest Rates, Issuances, Mergers & Acquisitions (M&A), Loans, and Debt and Notes. Form 10-K focuses more than AR on the other 11 topics in this category, except Fair Value, which is discussed in similar proportions. Regarding Regulatory, we find that AR focuses more on Responsibilities, Risks, Audit Report, and Disclaimers, whereas Form 10-K focuses more on Forms and Exhibits, Legal Language, Uncertainties, and Internal Controls.

Finding that AR contains considerable discussion of risk suggests the assertion that risk disclosures are uninformative is overstated.

Focusing our AR content comparisons on MD&A and FS rather than the entire Form 10-K reveals additional insights. Most notably, unlike the entire Form 10-K, we find that MD&A and FS both contain more discussion of Performance than AR. MD&A also contains more discussion of Analysis than AR, and the more extensive discussion of Business in Form 10-K is attributable to MD&A. That MD&A contains more discussion of these three topic categories is notable because they typify the role of analysts and are the top three categories discussed in AR. We also find that MD&A and FS both contain more discussion of Financial Reporting than AR, although—as one would expect—the difference is more pronounced for FS. In fact, these differences are the largest among the topic categories. Thus, MD&A and, particularly, FS discussions about Financial Reporting are least consistent with AR discussion. Perhaps surprisingly, both MD&A and FS contain less discussion of Regulatory than AR, not more, as is the case for the entire Form 10-K.

Regarding topics, we find that the discussions of Performance in MD&A and FS mirror those of the entire Form 10-K in that Revenues/Margins, Earnings, and Expenses are the most-discussed topics in all three. However, we find that MD&A (FS) contains more discussion of Revenues/Margins (Earnings and Expenses) than the entire Form 10-K, which is consistent with MD&A discussion being more similar to AR discussion than to FS discussion. Strikingly, almost half of MD&A discussion of Performance is devoted to Revenues/Margins, whereas only a quarter of AR and FS discussions are. The largest Performance differences between MD&A and FS relate to Revenues/Margins and Earnings.

Regarding the remaining topic categories, Form 10-K, MD&A, and FS discussions largely are similar. The most notable difference is that in the Analysis discussion MD&A contains more discussion of Increases and Decreases than FS and the entire Form 10-K. This difference likely reflects MD&A requirements for firms to explain large changes in financial statement amounts. The only notable finding for Business is that the greater discussion of Insurance and Health Care in Form 10-K largely is attributable to FS rather than MD&A, which likely reflects the extensive post-employment benefits and insurance footnotes. The two most notable findings for Financial Reporting are that FS contains more discussion of Equity than MD&A, and MD&A contains more discussion of Credit and Financing than FS.

Our comparisons of AR content for economically dissimilar firms reveal that differences depend on firms' circumstances. In particular, AR focuses more on Performance for Profit, Non-Technology, and Financial firms, which may arise because these firms' values more likely are attributable to existing operations than to growth options. AR focuses more on Business for Loss, Small, Technology, and Non-Financial firms, which may arise because these firms' businesses likely are less well-understood. AR focuses more on Financial Reporting for Profit, Large, Non-Technology, and Financial firms, which may arise because the values of these firms likely are more reflected in financial reports. However, these differences in AR content are small relative to differences between the contents of AR and Form 10-K. In addition, the contents of AR and Form 10-K differ more across industries than across profitability status or size, which could be attributable to analysts' tendencies to focus their coverage within industries.

This study contributes to the literature by addressing the question of whether financial reports contain information supporting investor valuations using a large-scale comparative content analysis of AR and Form 10-K. To our knowledge, this is the first study to do so.

Despite some notable differences, we find many similarities in the contents of Form 10-K and AR. Thus, taken together, our findings are inconsistent with financial reports lacking relevance.

The paper proceeds as follows. Section 2 relates our study to prior research and describes our predictions. Section 3 develops the research design, Section 4 describes the sample, and Section 5 reports the findings. Section 6 presents additional analyses, and Section 7 concludes.

2. Related literature and predictions

Financial analysts are sophisticated market participants who provide new information about firms' values and interpret previously released information (Asquith, Mikhail, and Au 2005). Prior research finds that information provided by sell-side analysts is a major input into investors' valuations and the market reacts to analyst forecasts, price targets, and recommendations (Schipper 1991; Womack 1996; Howe, Unlu, and Yan 2009). These findings imply that analyst-provided information is correlated with investors' valuation assessments. Importantly for our study, analysts do not provide forecasts and recommendations in isolation but instead issue reports that explain the analyst's analysis and the details behind it. Because there are no standards or requirements for the content of analyst reports, each report reveals what information supports the analyst's valuation analysis.¹

Early research examining the content of analyst reports, often based on a small sample, finds the reports tend to focus on performance measures, especially accounting-based measures that often deviate from US GAAP (Govindarajan 1980; Previts, Bricker, Robinson, and Young 1994). However, subsequent research finds that analysts do not justify their analysis solely on

¹ Analysts might explain the basis for their forecasts and recommendations using discussion topics similar to those in financial reports because they believe readers will find the reports more persuasive, not because the analyst developed the forecasts and recommendations using those topics. However, finding that analysts explain their forecasts and recommendations using topics typically found in financial reports is consistent with financial reports including information supporting investor valuations.

accounting-based measures. Breton and Teffler (2001) finds that analysts discuss soft information, such as business strategy and management, more frequently than accounting-based information. Bradshaw (2002) observes that this soft information is especially relevant when analysts have negative views of the firm.

A more recent literature uses text-based methods to analyze large samples of analyst reports to infer the role of analysts and tests whether the text can predict firm outcomes. Huang, Lehavy, Zang, and Zheng (2018) estimates that nearly one-third of analyst reports is devoted to the discovery of new information, whereas the remaining two-thirds interprets conference call data. However, Martineau and Zoican (2021) notes that the information content of analyst reports varies as a function of the number of analysts following the firm. Relatedly, Huang, Zang, and Zheng (2014) finds that the tone of analyst reports predicts returns and earnings, which suggests the reports' exposition offers insights into a firm's valuation. This research motivates our analysis of the textual content of analyst reports.

An extensive literature examines the value relevance of accounting information.² Several studies conclude that accounting information has lost its relevance (Core, Guay, and Van Buskirk 2003; Lev and Gu 2016). This evidence primarily is based on the low and declining associations between firms' equity market values and accounting amounts, particularly earnings. However, this conclusion is subject to three qualifications. First, this literature typically constrains assessment of value relevance to amounts presented in financial statements. Second, the approach measures correlations, which are not designed to determine whether investors use a particular amount in their decision-making. Third, these studies assume a valuation model that links prices and accounting amounts, such as Ohlson (1995), whereas investors use a variety of

² See Barth, Beaver, and Landsman (2001) and Holthausen and Watts (2001) for summaries.

valuation models (Joos and Plesko 2005). Barth, Li, and McClure (2023) uses machine learning to mitigate some of these issues and concludes that accounting information has not yet lost its relevance. Our comparative content analysis of AR and Form 10-K offers a new approach to assessing relevance to investors valuations that also mitigates some of these issues and reaches a similar conclusion regarding the relevance of financial reports.

Form 10-K includes not only firms' financial statements, but also discussion of firms' business and financial condition.³ As a result, Form 10-K provides equity investors with substantial information on which to base their investment decisions. However, the SEC, practitioners, and the academic literature observe that Form 10-K has become longer, more repetitive, and more boilerplate, not necessarily more informative (Dyer, Lang, and Stice-Lawrence 2017). To address these issues, the SEC has taken steps to simplify disclosures, including the Plain English Initiative and changes to Regulation S-K.⁴ Despite these changes, concerns about disclosure overload remain. Prior research identifies undesirable characteristics of mandatory disclosures, such as low readability, that obfuscate low earnings and increased crash risk (Loughran and McDonald 2014; Kim, Wang, and Zhang 2018). Furthermore, firms whose annual reports have more boilerplate have lower liquidity, analyst following, and institutional ownership (Lang and Stice-Lawrence 2015). This evidence suggests the content and exposition of Form 10-K affect investors' decisions.

Our comparison of the contents of AR and Form 10-K does not imply that we expect the topics discussed in Form 10-K and AR to be the same. Whereas Form 10-K focuses on presenting the firm's current financial conditional and performance, AR focuses on analyzing the valuation implications of that performance (Previts et al. 1994). Thus, we expect AR to discuss

³ <https://www.investor.gov/introduction-investing/investing-basics/glossary/form-10-k>.

⁴ <https://www.sec.gov/plainwriting>.

performance and analysis more than Form 10-K. However, because firms also discuss performance in Form 10-K, it is an empirical question whether AR or Form 10-K discusses performance more. Analysts also devote considerable attention to discussing a firm's business (Breton and Teffler 2001), which leads us to expect AR contains more discussion of business than Form 10-K. However, MD&A also gives a firm the opportunity to discuss its business, particularly as it relates to changes in the firm's performance. Thus, it also is an empirical question whether AR or Form 10-K discusses a firm's business more. The criticism that Form 10-K contains excess information that is of little use to investors (Dyer et al. 2017) suggests that Form 10-K discusses some topics more than AR. Because some of this criticism relates to requirements of accounting standards and regulation, we expect that Form 10-K contains more discussion of financial reporting and regulation topics than AR.

3. Research design

3.1 Overview

Addressing our research question requires us to analyze the contents of AR and Form 10-K and construct metrics to compare them. We begin by identifying the topics discussed in AR and Form 10-K and organizing these topics into broad categories. We then calculate the proportions of AR and Form 10-K devoted to discussing each topic and topic category. By using these proportions to analyze the content of AR, we provide insights into the information supporting investor valuations. By comparing the proportions of Form 10-K devoted to each category and topic to those of AR, we provide insights into how much the information in regulated annual reports differs from the information supporting investor valuations.

We also compare the contents of AR and the Management's Discussion and Analysis (MD&A) and Financial Statement (FS) sections of Form 10-K and compare the contents of MD&A and FS to each other. We do so because these two sections are designed to provide

investors with different types of information. MD&A is intended to provide management’s analysis of the firm’s resources and operations (Li 2010). As such, it bears some resemblance to AR. Also, MD&A is perhaps the most important and read section of Form 10-K (Tavcar 1998). We examine the FS section because financial statements, including the accompanying footnotes, provide many of the inputs to investor valuation models. We also compare AR content for several groups of economically dissimilar firms to provide insights into the extent to which the content of AR varies depending on a firm’s circumstances.

3.2 Identifying topics and topic categories discussed in AR and Form 10-K

A distinctive feature of our research design is that we identify topics discussed in AR and Form 10-K by implementing a machine learning neural network-based word embedding topic model, Word2vec. See the Appendix for details.⁵ In brief, we identify topics in six steps. First, we obtain AR from Thomson Reuters Investext, which collects millions of AR from over 1,600 investment banks and sell-side research firms. Our focus is on the content of equity research reports that provide fundamental analysis. Thus, we eliminate AR from brokers that primarily provide robo reports, event transcripts, proxy and governance reports, credit rating reports, and company descriptions. After eliminating these reports, we randomly select one AR for each firm-year to ensure our topic modeling approach is computationally feasible.⁶ We then convert AR from PDF to text for analysis. Second, we obtain Form 10-K from the EDGAR database. We remove HTML from these Form 10-K based on the approach in Loughran and McDonald

⁵ Word2vec is based on the idea that words that co-occur with similar neighboring words have similar meanings. Intuitively, the two words “earnings” and “net_income” would be identified as similar if they occur frequently in the context “We had record _____ this quarter.” In our sample, we find that the model produces coherent topics in that it includes words such as “properties” and “leasing” in the same topic. Word2vec has been used in a variety of fields, including speech recognition, machine translation, and question answering, as well as finance. For example, in finance, Li, Mai, Shen, and Yan (2021) uses Word2vec to develop a dictionary to measure corporate culture.

⁶ Chen, Cheng, and Lo (2010) finds that AR content can vary in terms of interpretation and discovery throughout the fiscal year. To construct a sample of AR that reflects this variation, we do not constrain the random selection procedure to particular points in the firm’s fiscal year.

(2016). We also identify the MD&A and FS sections of each Form 10-K based on the approach in Dyer et al. (2017).

Third, to focus our analysis on meaningful words and phrases, we eliminate uninteresting words such as punctuation symbols and single-character words. We also remove the top 50 most common words such as “the,” “of,” and “and,” unless the words could relate to a topic of potential substantive interest, such as “income,” “price,” and “cash.” In addition, we include multi-word phrases so that our topic modeling approach can better reflect the content of the documents. Specifically, we treat as single words sequences of up to 16 words that appear, on average, more than once per AR, by joining the words with underscore symbol. For example, we treat the word sequence “earnings per share” as the single word “earnings_per_share.” Hereafter, when we refer to “words” in AR and Form 10-K we interpret each of these sequences as one word. Because AR is shorter, on average, than Form 10-K, we scale the AR corpus to be the same length as the Form 10-K corpus. The result of these steps is a combined AR and Form 10-K corpus that underlies our analyses.

Fourth, we train the topic model on the combined corpus in that Word2vec converts each word to a vector in a 100-dimensional vector space in which words with similar (dissimilar) meanings are close together (far apart) in the vector space. Fifth, we cluster these word vectors into 100 clusters that minimize the within-cluster sums of mean squared vector differences. Each of the clusters represents a topic discussed in AR and Form 10-K. Together, these topics include all the words in the combined AR and Form 10-K corpus. Sixth, we label topics and, following Dyer et al. (2017), organize them into interpretable categories to ease interpretation.

We focus our analyses on the five most interpretable categories—Performance, Analysis, Business, Financial Reporting, and Regulatory.⁷

3.3 Quantifying the extent of topic and category discussion

We determine the amount a document discusses a particular topic by scaling the number of words in that topic by the total number of words. Thus, the proportion that document i in year t discusses topic j is

$$Proportion_{j,i,t} = \frac{\# \text{ of Words in } j_{i,t}}{\# \text{ of Words in } i_t} \quad (1)$$

To determine the amount a document discusses a topic category, we sum the proportion of topics in that category. Thus, the proportion that document i in year t discusses category c is

$$Category \ Proportion_{c,i,t} = \sum_{j \in c} Proportion_{j,i,t} \quad (2)$$

To determine the topic or category proportions across documents, such as within a document type (e.g., AR), we average the proportions across firm-year observations.

The proportions from Equations (1) and (2) provide the basis for us to analyze the content of AR and to compare it (i) to that of Form 10-K, (ii) to those of the MD&A and FS sections of Form 10-K and to compare the contents of MD&A and FS to each other, and (iii) for several groups of economically dissimilar firms. We test whether the discussion of a particular topic or category differs between two sets of documents (e.g., AR versus Form 10-K) using a t-test of the difference in proportion means for each document set.

3.4 Differences in AR content for economically dissimilar firms

We compare AR content for four sets of economically dissimilar firms to provide insights into the extent to which AR content varies across firms depending on a firm's circumstances.

⁷ These five categories include 59 of the 100 topics Word2vec identifies. We categorize the remaining 41 topics as "Other" but do not analyze them because they have no discernable topic focus. See Section 5.1.

Two comparisons reflect differences potentially applicable to all firms—profit versus loss and large versus small—and two reflect differences associated with industry membership—non-technology versus technology and non-financial versus financial. We compare AR content for profit and loss firms because loss firms are more difficult to value (Joos and Plesko 2005; Darrrough and Ye 2007). Loss firms are firms that report negative earnings. We compare AR content for large and small firms because the sources of their firm values likely differ. For example, large (small) firms are more likely to derive value from operations (growth options). Large (Small) firms are firms with equity market value above (below) the median for our sample.

We compare AR content for non-technology and technology firms because technology firms are emblematic of the New Economy. These firms typically have substantial intangible assets, whose value is not reflected in financial statements. Technology firms have three-digit SIC codes 283, 357, 360-368, 481, 737, and 873 (Francis and Schipper 1999; Core et al. 2003; Barth et al. 2023). We compare AR content for non-financial and financial firms because most financial firms' assets are financial, which are recognized and valued differently from other assets. This is why financial firms often are excluded from, or examined separately, in most accounting and finance research (e.g., Fama and French, 2006). Financial firms are members of the Fama-French 48 industries of banking (44), insurance (45), and trading (47).

4. Sample

Our sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018. Our sample period begins in 1996 when the SEC required Form 10-K to be submitted electronically on EDGAR and ends in 2018 when access to AR on Thomson Investext became limited. We require sample firms to have, in a given firm-year, data in the Compustat and CRSP databases, at least one fundamental analysis AR available from Investext, and a Form 10-K for

which we can identify the MD&A and FS sections using the approach in Dyer et al. (2017), as explained in the Appendix. Untabulated statistics reveal that the average AR contains 3,418 words, which is substantially fewer than the average Form 10-K, which contains 65,951 words.⁸ The average numbers of words in MD&A and FS are 10,162 and 15,345.

5. Findings

5.1 Topics discussed analyst reports and Forms 10-K

As Section 3.2 explains, we organize the topics Word2vec identifies in AR and Form 10-K into categories—Performance, Analysis, Business, Financial Reporting, Regulatory, and Other. Table A.1, Panels A through F, in the Appendix presents the topics discussed in each category, listed in the order of the most to the least discussed in AR. Each row of the table presents the 15 most common words in each topic, listed in order of the highest to the lowest frequency in the combined AR and Form 10-K corpus.

Panel A of Table A.1 reveals that Performance comprises eight topics: Revenues/Margins, Ratios, Earnings, Target Prices, Recommendations, Adjusted Earnings, Expenses, and Cash Flows. Regarding words in each topic, Revenues/Margins, for example, comprises words such as sales, revenues, and margins, whereas Expenses comprises words such as costs, expenses, and fees. Earnings comprises words such as earnings and net income. By

⁸ The average number of words in Form 10-K for our sample is larger than the 45,349 words in Dyer et al. (2017) for two reasons. First, our sample is more recent. Dyer et al. (2017) finds Form 10-K has become longer during that study's sample period. Untabulated statistics reveal both AR and Form 10-K also become longer during our sample period. The average annual increase in the average number of words in AR (Form 10-K) is 126 (1,310), which is 3.7% (2.0%) of the mean number of words during our sample period. The increase in Form 10-K words is similar to the median increase reported in Dyer et al. (2017) of 1,588 words ($= (50,000 - 23,000)/17$). Thus, the increase in number of words applies to AR as well. Untabulated statistics also reveal that for both AR and Form 10-K, all topic categories have become longer. Taken together, these statistics are consistent with the amount of information required to value firms increasing over time (Barth et al. 2023). Second, our sample firms are followed by analysts, which means that, on average, they are larger than the Dyer et al. (2017) sample firms. AR and Form 10-K typically are longer for larger firms than for smaller firms. Untabulated statistics for our sample reveal the mean number of words in AR and Form 10-K for Large (Small) firms are 4,011 and 74,383 (2,824 and 57,518).

contrast, Adjusted Earnings and Cash Flows comprise words such as ebitda, pro forma, cash_flow, and free_cash_flows.

Panel B reveals that Analysis comprises nine topics: Markets and Industries, Trends and Forecasts, Positives and Negatives, Estimates and Assessments, Increases and Decreases, Successes and Challenges, Views and Opinions, Economic Environment, and Affect and Effect. These topics comprise words that relate to interpretations for changes, such as strong, low, overall, increase, lower, decline, estimate, analysis, assumptions, believe, expect, opinion, impact, affect, and influenced. Panel C reveals that Business comprises 16 topics. These topics comprise words referring to (i) general business and operations such as services, technology, solutions, products, operations, customers, business, management, and strategy; (ii) industry-specific business descriptions such as clinical, patients, treatment, production, oil, gas, insurance, benefits, and coverage; and (iii) specific business details such as new_york, delaware, located, project, expansion, and area.

Panel D reveals that Financial Reporting comprises 18 topics. These topics comprise words referring to (i) accounting standards such as accounting and standards; (ii) various assets and liabilities such as assets, impairment, goodwill, property, equipment, lease, liabilities, payable, and reserves; and (iii) financial statements such as financial, statement, and balance_sheet. Panel E reveals that Regulatory comprises eight topics. The topics in this category relate to risk and uncertainty and regulatory or legal efforts to mitigate their effect on a firm. Words in this category include words such as risk, adverse, liability, rights, obligation, responsible, damages, and internal_control_over_financial_reporting.

The Other topics category comprises 41 topics; Panel E present only the top 15 for parsimony. We do not label or analyze the topics in this category because they appear to be

aggregations of generic words such as it, only, part, when, and however, or boilerplate language such as company, corporation, group, and llc.

5.2 Discussion of topic categories

Table 1 presents the findings relating to AR and Form 10-K discussions of the topic categories.⁹ Figure 1 displays the average proportion of the document devoted to each category. Table 1 reveals that, on average, analysts devote 14.11% of AR to discussing Performance, which is the most-discussed topic category. Analysts devote 12.90% and 11.71% of AR to discussing Analysis and Business. These proportions indicate that, as expected, AR contains considerable analysis of a firm's performance and business activities, which reflects analysts' interpretation and analysis role (Huang et al. 2018). Table 1 also reveals that, on average, analysts devote almost as much of AR, 11.41%, to discussing Financial Reporting as they do to these other three categories. This finding suggests financial reporting is a crucial component of AR. By contrast, analysts devote only 6.75% of AR to Regulatory, which is not surprising because regulatory topics are less likely to be relevant to analysts' valuation analysis. Table 1 reveals that 43.11% of AR comprises language with no discernible topic focus. Table 1 also reveals that, in aggregate, the extent to which AR and Form 10-K discuss the 100 topics are 23.2% different.¹⁰

More importantly for our research question, the findings reveal differences between AR and Form 10-K content for all topic categories, some expected and some unexpected. Table 1 reveals that, as expected, AR discusses Performance and Analysis more than Form 10-K (14.11% and 12.90% versus 10.08% and 10.03%), and Form 10-K discusses Regulatory more

⁹ Table 1, and Tables 2 through 6 that follow, also report results of statistical tests of significance of the reported differences. Unless otherwise noted, all differences we discuss in Sections 5 and 6 are significant.

¹⁰ This amount is the sum of the absolute differences in topic proportions, divided by two.

than AR (10.26% versus 6.75%). Although these differences are significant, they are not as large economically as perhaps some would expect. Unexpectedly, Form 10-K also discusses Business more than AR (12.84% versus 11.71%). The topic with the largest difference between AR and Form 10-K is Financial Reporting. Although we expect Form 10-K to discuss Financial Reporting more than AR, AR discusses this category only approximately half as much (11.41% versus 22.20%). The large Financial Reporting and Regulatory differences suggest that the concern of information overload in Form 10-K arises primarily from these two categories.

5.3 Discussion of topics within each topic category

Table 2, Panels A through E, presents the findings relating to discussion topics within each category. Figure 2, Panels A through E, displays these findings. Each panel presents the average proportion of each topic relative to the discussion of the topic category in that document type. These statistics provide insight into the topics on which AR and Form 10-K focus when discussing a particular category.

Regarding Performance, Panel A reveals that AR contains considerable discussion of all topics, except for Expenses and Cash Flows. Specifically, the AR proportions are only 4.58% and 2.30% for Expenses and Cash Flows, whereas the proportions range from 26.33% to 9.58% for the other Performance topics. In contrast, Form 10-K focuses on three topics, Revenues/Margins, Earnings, and Expenses. The Form 10-K proportions are 35.22%, 29.69%, and 18.17% for these three topics, whereas the proportions range from 5.85% to 0.90% for other topics. These findings reveal that within their discussions of Performance, AR and Form 10-K focus on different topics. In fact, Panel A reveals that, in aggregate, the extent to which AR and Form 10-K discuss Performance topics are 39.53% different, which is the largest difference among the five categories.

Regarding Analysis, Panel B reveals that the top five topics discussed in AR are Markets and Estimates, Trends and Forecasts, Positives and Negatives, Estimates and Assessments, and Increases and Decreases (proportions range from 27.15% to 10.92%). AR contains less discussion of Successes and Challenges, Views and Opinions, Economic Environment, and Affect and Effect (proportions range from 7.49% to 1.51%). Panel B also reveals that Form 10-K focuses on the same topics, except for Positives and Negatives. In fact, in aggregate, the discussions of Analysis topics in AR and Form 10-K are only 10.44% different, which is the second smallest difference among the five categories. This small difference is unexpected, given that analysis is a key analyst role.

Regarding Business, Panel C reveals that the topics discussed in AR and Form 10-K are similar. For both documents, the top three topics are Products and Markets, Business and Operations, and Services and Technology (proportions range from 16.08% to 12.40% for AR and from 16.27% to 11.86% for Form 10-K). Perhaps surprisingly, Form 10-K does not contain less discussion of these topics and, in fact, it contains slightly more. The four largest differences are that AR has more discussion of Cities and International Locations and Energy (Diff. = 3.05% and 2.10%) and Form 10-K has more discussion of Insurance and Healthcare and Business and Operations (Diff. = -3.99% and -2.17%). In addition, in aggregate, the discussions of Business topics in AR and Form 10-K are only 8.71% different, which is the smallest difference among the five categories.

Regarding Financial Reporting, Panel D reveals that AR focuses more than Form 10-K on Investments, Equity, Interest Rates, Issuances, M&A, Loans, and Debt and Notes, although the differences generally are not economically large (proportions range from 13.26% to 2.77% for AR and from 10.55% to 2.08% for Form 10-K). Regarding Regulatory, Panel E reveals that

AR focuses more on Responsibilities, Risks, Audit Report, and Disclaimers (Diff. ranges from 9.96% to 2.69%), and Form 10-K focuses more on Forms and Exhibits, Legal Language, Uncertainties, and Internal Controls (Diff. ranges from 12.33% to 1.18%). However, Panel D (Panel E) also reveals that, in aggregate, the discussions of Financial Reporting (Regulatory) topics in AR and Form 10-K are 20.95% (21.93%) different, which is the third (second) largest difference among the five categories. These findings echo those in Table 1 and suggest that Form 10-K might over-emphasize some Financial Reporting and Regulatory topics.

5.5 Differences in topic discussion in MD&A and financial statements

Table 3 presents the findings for the topic categories in AR and the MD&A and FS sections of Form 10-K. Figure 3 displays the findings. Table 3 reveals that, unlike the entire Form 10-K, MD&A and FS both contain more discussion of Performance and Analysis than AR (17.76% and 16.03% versus 14.11%). MD&A also has more discussion of Analysis and Business than AR (14.16% and 13.28% versus 12.90% and 11.71%), but FS contains less (8.61% and 8.65%).¹¹ That MD&A contains more discussion of Performance, Analysis, and Business is surprising because these three categories typify the role of analysts and Table 1 reveals these are the top three categories discussed in AR.

Table 3 also reveals, as expected, that MD&A and FS have more discussion of Financial Reporting than AR (20.04% and 30.59% versus 11.41%). Also as one might expect, the difference is most pronounced for FS. In fact, the Financial Reporting differences between AR and the MD&A and FS sections of Forms 10-K are the largest among the five categories (8.63% and 19.18%). Thus, again, discussions about Financial Reporting, particularly in FS, appear over-emphasized, relative to the coverage in AR. In addition, both MD&A and FS contain less

¹¹ Table 3 also reveals that the more extensive discussion of Business in Form 10-K revealed by Table 1 is attributable to MD&A, not FS (13.28% and 8.65% versus 11.71%).

discussion of Regulatory than AR (5.93% and 6.37% versus 6.75%). Despite this difference and, perhaps unexpectedly, these findings reveal that the discussion in MD&A is more similar to that in AR than that in FS.

Table 4, Panels A through E, presents the findings relating to discussion topics for each topic category, which we display in Figure 4. Untabulated statistics reveal that, in aggregate, the AR and the MD&A and FS sections of Form 10-K discussions of the 100 topics are 28.5% and 33.0% different, and the MD&A and FS discussions are only 19.9% different. These statistics reveal MD&A and FS are more similar to each other than AR is to either MD&A or FS. However, AR is more similar to MD&A than to FS.

Regarding Performance, Panel A reveals that the pattern of each topic's discussion in both in MD&A and FS mirrors that of entire Form 10-K in Table 2, Panel A. For example, the top three topics for both sections and the entire Form 10-K are Revenues/Margins, Earnings, and Expenses. However, Panel A reveals that MD&A (FS) contains more discussion of Revenues/Margins (Earnings and Expenses) than the entire Form 10-K, which is consistent with MD&A discussion being more similar to AR discussion than to FS discussion. Strikingly, almost half (48.35%) of MD&A is devoted to discussing Revenues/Margins, whereas AR and FS devote only 26.33% and 25.92%. The largest differences between MD&A and FS relate to discussion of Revenues/Margins and Earnings (22.43% and 13.54%). Panel A reveals that, taken together, the discussions of Performance in MD&A and FS are 44.99% and 39.50% different from the discussion in AR. Although the difference for FS is similar to that for the entire Form 10-K (39.53%), the difference for MD&A is considerably larger.

Regarding Analysis, Panel B reveals that the discussion in MD&A and FS largely are similar to that in the entire Form 10-K as presented in Table 2, Panel B. The most notable

difference is that MD&A contains more discussion of Increases and Decreases than FS and the entire Form 10-K contains (24.13% versus 6.04% and 11.14%). Again, this likely reflects the requirement for firms to explain in MD&A large changes in financial statement amounts. The second largest difference is that FS contains more discussion of Estimates and Assessments than MD&A (20.89% versus 12.21%). As with the corresponding difference in Table 2, Panel B, this likely reflects requirements for firms to disclose information about estimates included in financial statements.

Regarding Business, Panel C reveals patterns for MD&A and FS that are similar to those in Table 2 for the entire Form 10-K. The only notable finding is that the greater discussion of Insurance and Health Care for Form 10-K in Table 2 largely is attributable to FS rather than MD&A (15.25% and 8.47% versus 9.09%), which likely reflects firms needing to report information on their pension plans.

Regarding Financial Reporting, Panel D also reveals patterns for MD&A and FS that are similar to those in Table 2 for the entire Form 10-K. The two most notable findings are that FS contains more discussion of Equity than MD&A (11.98% versus 7.47%) and MD&A contains more discussion of Credit and Financing than FS (11.07% versus 6.14%). Regarding Regulatory, Panel E reveals that the top two topics for MD&A are Risks and Uncertainties (28.80% and 20.07%), whereas for FS they are Forms and Exhibits and Responsibilities (25.06% and 20.71%). For AR, the top two topics are Risks and Responsibilities (23.80% and 23.19%). These statistics again reveal that MD&A resembles AR more than FS does.

5.5 Differences in topic discussion for economically dissimilar firms

Table 5 presents mean differences in the discussion in AR of each topic category between the paired subsamples of firms. Table 5 reveals that AR for Profit firms contains more (less)

discussion of Performance, Analysis, and Financial Reporting (Business and Regulatory) than for Loss firms; the differences are 0.50%, 0.70%, and 0.50% (−0.79% and −0.26%). AR also contains more discussion of Financial Reporting for Large firms than Small firms (Diff. = 0.23%). AR of Large firms contain less (more) discussion of Performance and Analysis (Regulatory); the differences are −1.43%, −0.21%, and −0.22% (0.51%).

Table 5 also reveals that for Non-Technology firms, AR contains more (less) discussion of Financial Reporting (Business) than for Technology firms; the difference is 1.14% (−0.81). For Non-Financial firms, AR contains more (less) discussion of Analysis, Business, and Regulatory (Performance and Financial Reporting) than for Financial firms; the differences are 0.42%, 2.09%, and 0.24% (−0.44%, and −3.12%). The Financial Reporting difference of 3.12% is the largest of all categories among all four comparisons. Moreover, Table 5 reveals that this difference represents more than a quarter, 27.36%, of the average level of discussion of Financial Reporting in the full sample (= 3.12/14.11 in Table 1).

Several of these findings are intuitive, in that AR is more focused on Performance for Profit, Non-Technology, and Financial firms, whose value is more likely attributable to existing operations than growth options. They also reveal that AR is more focused on Business for Loss, Small, Technology, and Non-Financial firms, which may be because these firms' businesses likely are less well-understood. AR is more focused on Financial Reporting for Profit, Large, Non-Technology, and Financial firms, whose equity value is more likely to be reflected in financial reports.

Table 6, Panels A through E, presents differences between the paired subsamples relating to discussion within each topic category. Regarding Performance, Panel A reveals that for Profit firms AR contains more discussion of Earnings and Target Prices and less discussion of Ratios,

Expenses, and Cash Flows than for Loss firms. For Large firms, AR contains more discussion of Target Prices, Recommendations, and Cash Flows and less discussion of Revenues/Margins, Ratios, Earnings, Adjusted Earnings, and Expenses than for Small firms. For Non-Technology firms, AR contains more discussion of Earnings, Target Prices, and Expenses and less discussion of Revenues/Margins and Cash Flows than for Non-Technology firms. For Non-Financial firms, AR contains more discussion of Revenues/Margins and Cash Flows and less discussion of Ratios, Earnings, Target Prices, and Expenses than for Financial firms.

Taken together, these findings are consistent with earnings differing in relevance for different types of firms. Earnings is more relevant for Profit firms (Collins, Maydew, and Weiss 1997) and less relevant for Technology and Financial firms. Our results are consistent with AR of small, technology firms discussing Revenues/Margins (Damodaran 2009). In contrast, Financial firms discuss Earnings more and discuss Revenues/Margins less.

Regarding Analysis (Business), Panel B (Panel C) reveals that AR for Profit firms have less discussion of most Analysis topics than Loss firms, which could reflect analysts focusing on trend analysis for Loss firms. The exceptions are Positives and Negatives, Increases and Decreases, and Economic Environment. AR for Loss and Technology firms have more discussion of Services and Technology and Drugs and Clinical Trials than Profit and Non-Technology firms. These findings are consistent with the business activities in which Loss and Technology firms tend to engage (Collins et al. 1997; Srivastava 2014). Similarly, AR for Financial firms have more discussion of Banking. Overall, the Business topics discussed in AR reflect the nature of business activities in which different types of firms engage.

Regarding Financial Reporting, Panel D reveals that there is more variation in topics discussed between Non-Technology and Technology firms and between Non-Financial and

Financial firms, than between Profit and Loss firms and between Large and Small firms. This is consistent with several Financial Reporting topics being industry-specific. For example, Non-Technology firms discuss Interest Rates, Loans, and Credit and Financing more than Technology firms because the former are less likely to have loans. In contrast, Financial firms discuss each of these topics more than Non-Financial firms because of their lending activities. Combined with the finding in Table 3 that Technology (Financial) firms discuss Financial Reporting less (more) overall, these findings are consistent with Financial Reporting being more relevant for particular types of firms.

Regarding Regulatory, Panel E reveals that the most notable differences are that AR for Loss and Small firms contain more discussion of Uncertainties than Profit and Large firms, which reflects the greater uncertainty about the prospects of Loss and Small firms. AR for Large firms contain more discussion of Forms and Exhibits and use Legal Language than those for Small firms.

We also examine the extent to which AR and Form 10-K discussions devoted to the 100 topics differ across the paired subsamples. Table 7, Panels A through D, presents the findings. The table reveals fewer differences for AR and Form 10-K across the subsamples than for the subsamples across AR and Form 10-K. For example, Panel A reveals that the aggregate difference between topics discussed in AR for Profit and Loss firms is 3.4%, whereas the difference in Form 10-K is 5.9%. In contrast, the difference between AR and Form 10-K for Profit firms is 23.4% and for Loss firms is 23.1%. These findings suggest that although there are differences in AR between Profit and Loss firms, they are small relative to differences between AR and Form 10-K.

Perhaps more strikingly, the difference in topics discussed in AR for Profit and Loss firms, 3.4%, is smaller than the difference in topics discussed in Form 10-K for Profit and Loss firms, 5.9%. This is striking because AR largely is unregulated, whereas Form 10-K largely requires the same information for all firms. The same pattern holds for Technology and Non-Technology firms, Large and Small firms, and Financial and Non-Financial firms. Moreover, Table 7 reveals that AR and Form 10-K have larger differences between Technology and Non-Technology firms and between Financial and Non-Financial firms than between Profit and Loss firms and between Large and Small firms. These findings suggest that AR and Form 10-K are more different across industries than across profitability status or size, which could be attributable to analysts' tendencies to focus their coverage within industries.

6. Additional analyses: Analysis portion of analyst reports

Because AR contains standardized brokerage disclosures that do not relate to analysis (Huang et al. 2014), we examine separately the portion of AR that is unlikely to include these disclosures. We observe that when short (long) brokerage disclosures exist, they typically appear at the beginning (end) of the document. Thus, we identify and remove the first 10% and last 75% of the sentences in each AR and analyze the remaining content. We refer to this truncated portion of the AR as the analysis portion.

We first assess whether the analysis portion of AR is more likely to contain fundamental analysis and less likely to contain standardized brokerage disclosures. Specifically, we identify three topics in the Other category that appear unrelated to analysis and determine whether these topics are discussed more in the full AR than in the analysis portion. These topics are (i) brokerage names, with words such as Morningstar, ubs, morgan_stanley, jp_morgan, and investment_banking; (ii) subsidiaries and affiliates, with words such as subsidiaries, affiliates,

persons, person, and subsidiary; and (iii) broker regulations, with words such as regulated, authority, financial_services, uk, and united. Untabulated statistics reveal that each of these topics is discussed more in the full AR than in the analysis portion (2.62%, 1.11%, and 0.85% versus 1.03%, 0.38%, and 0.11%).

More importantly for our research question, untabulated findings from our content analysis based on the analysis portion of AR reveal inferences similar to those revealed by our tabulated findings. In particular, the analysis portion has more discussion of Performance than the full AR (17.76% versus 14.11%). Regarding Performance topics, the analysis portion has more discussion of Revenues/Margins, Ratios, and Adjusted Earnings and a similar amount of discussion of Earnings and Expenses (5.87%, 4.67%, 2.21%, 2.08%, and 0.71% versus 3.89%, 3.41%, 1.43%, 2.22%, and 0.70%). These findings are consistent with the inferences we draw from Table 1 that Performance is the most-discussed category in AR and from Table 2, Panel A, that among Performance topics, AR focuses more on Revenues/Margins and less on Expenses than Form 10-K.

In addition, the untabulated findings reveal that the analysis portion contains less discussion of Financial Reporting and Regulatory than the full AR (9.93% and 4.10% versus 11.41% and 6.75%). As with the full AR, the analysis portion discusses virtually every Financial Reporting topic less than Form 10-K. Regarding Regulatory topics, although the analysis portion contains somewhat less discussion of Risks and somewhat more discussion of Uncertainties than the full AR (1.19% and 0.77% versus 1.54% and 0.70%), Risks is the most-discussed Regulatory topic in the analysis portion of AR. These findings are consistent with analyst reports focusing less on Financial Reporting and Regulatory, except for Risks.

7. Conclusion

We address whether financial reports include information supporting investor valuations. Addressing this question is fundamental because financial reports are designed to provide investors with information to help them make their decisions about providing capital to a firm, but financial reports have been criticized for lack of relevance to investors. To address our research question, we analyze and compare the contents of Form 10-K, which is the firm's primary financial report, and equity analysts' reports (AR), which reveal information supporting investor valuations.

Using a word-embedding topic model, we identify the topics discussed in AR and Form 10-K and categorize the topics into broad categories on which we focus our analysis: Performance, Analysis, Business, Financial Reporting, and Regulatory. Most notably, we find that Performance, Analysis, and Business are the three most-discussed topic categories in AR. Surprisingly, we find that Financial Reporting topics are discussed almost as much as Business, which reveals that financial reporting is a crucial component of AR and, thus, supports investor valuation. However, our evidence also is consistent with Form 10-K over-emphasizing Financial Reporting in that this topic category is discussed almost twice as much in Form 10-K as in analyst reports.

Regarding Performance, we find that revenues and margins is the most-discussed topic in both AR and Form 10-K. However, AR and Form 10-K discussions of the other Performance topics differ considerably. In particular, we find Form 10-K focuses more on Earnings and Expenses, whereas analyst reports focus more on Ratios, Target Prices, Recommendations, and Adjusted Earnings. Surprisingly, analyst reports focus more on Earnings than on Adjusted Earnings or Cash Flows. These findings are inconsistent with analysts viewing earnings as

irrelevant and focusing on alternative performance measures, including cash flows. We also find that discussion in AR is more similar to Form 10-K's MD&A section than to discussion in Form 10-K's financial statements section.

Our comparisons of differences in AR content for economically dissimilar firms reveal that the differences depend on firms' circumstances. However, we find that differences in AR content between economically dissimilar firms are small relative to differences between AR and Form 10-K contents. We also find that the contents of AR and Form 10-K differ more across industries than across profitability status or size, which could be attributable to analysts' tendencies to focus their coverage within industries.

Despite some notable differences, we find many similarities in the contents of Form 10-K and AR. Taken together, our findings are inconsistent with financial reports lacking relevance to investor valuations.

References

- Asquith, P., M.B. Mikhail, and A.S. Au. 2005. Information content of equity analyst reports. *Journal of Financial Economics* 75(2): 245-282.
- Barth, M.E., W.H. Beaver, and W.R. Landsman. 2001. The relevance of the value relevance literature for financial accounting standard setting: Another view. *Journal of Accounting and Economics* 31: 77-104.
- Barth, M.E., K. Li, and C.G. McClure. 2023. Evolution in value relevance of accounting information. *The Accounting Review* 98(1): 1-28.
- Bradshaw, M.T. 2002. The use of target prices to justify sell - side analysts' stock recommendations. *Accounting Horizons* 16(1): 27-41.
- Breton, G., and R.J. Taffler. 2001. Accounting information and analyst stock recommendation decisions: A content analysis approach. *Accounting and Business Research* 31(2): 91-101.
- Chen, X., Q. Cheng, and K. Lo. 2010. On the relationship between analyst reports and corporate disclosures: Exploring the roles of information discovery and interpretation. *Journal of Accounting and Economics* 49(3): 206-226.
- Collins, D.W., E.L. Maydew, and I.S. Weiss. 1997. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* 24: 39-67.
- Core, J.E., W.R. Guay, and A. Van Buskirk. 2003. Market valuations in the New Economy: An investigation of what has changed. *Journal of Accounting and Economics* 34: 43-67.
- Damodaran, A. (2009). Valuing young, start-up and growth companies: estimation issues and valuation challenges. *Unpublished Working Paper*.
- Darrough, M., and J. Ye. 2007. Valuation of loss firms in a knowledge-based economy. *Review of Accounting Studies* 12(1): 61-93.
- Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of Form 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64: 221-245.
- Fama, E.F., and K.R. French. 2006. Profitability, investment and average returns. *Journal of Financial Economics* 82(3): 491-518.
- Francis, J., and K. Schipper. 1999. Have financial statements lost their value relevance? *Journal of Accounting Research* 37(2): 319-352.

- Govindarajan, V. 1980. The objectives of financial statements: An empirical study of the use of cash flow and earnings by security analysts. *Accounting, Organizations and Society* 5(4): 383-392.
- Holthausen, R.W., and R.L. Watts. 2001. The relevance of the value relevance literature for financial accounting standard setting. *Journal of Accounting and Economics* 31: 3-75.
- Howe, J.S., E. Unlu, and X. Yan. 2009. The predictive content of aggregate analyst recommendations. *Journal of Accounting Research* 47(3): 799-821.
- Huang, A.H., R. Lehavy, A.Y. Zang, and R. Zheng. 2018. Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science* 64(6): 2833-2855.
- Huang, A.H., A.Y. Zang, and R. Zheng. 2014. Evidence on the information content of text in analyst reports. *The Accounting Review* 89(6): 2151-2180.
- Investor Responsibility Research Center Institute (IRRC). 2016. *The corporate risk factor disclosure landscape*.
- Joos, P., and G.A. Plesko. 2005. Valuing loss firms. *The Accounting Review* 80(3): 847-870.
- Kim, C., K. Wang, and L. Zhang. 2019. Readability of 10 - K reports and stock price crash risk. *Contemporary Accounting Research* 36(2): 1184-1216.
- Lang, M., and L. Stice-Lawrence. 2015. Textual analysis and international financial reporting: Large sample evidence. *Journal of Accounting and Economics* 60(2-3): 110-135.
- Lev, B., and F. Gu. 2016. *The End of Accounting and the Path Forward for Investors and Managers*. John Wiley & Sons, Inc. Hoboken, NJ, USA.
- Li, F. 2010. The information content of forward - looking statements in corporate filings—A naïve Bayesian machine learning approach. *Journal of Accounting Research* 48(5): 1049-1102.
- Li, K., F. Mai, R. Shen, and X. Yan. 2021. Measuring corporate culture using machine learning. *Review of Financial Studies*, 34(7): 3265-3315.
- Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *Journal of Finance* 69(4): 1643-1671.
- Loughran, T., and B. McDonald. 2016. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research* 54(4): 1187-1230.
- Martineau, C., and M. Zoican. 2021. Measuring information in analyst reports: A machine learning approach. *Unpublished working paper*.

- Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv: 1301-3781*.
- Ohlson, J.A. 1995. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research* 11(2): 661-687.
- Previts, G.J., R.J. Bricker, T.R. Robinson, and S.J. Young. 1994. A content analysis of sell-side financial analyst company reports. *Accounting Horizons* 8(2): 55-70.
- Rehurek, R., and P. Sojka. 2010. Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*: 45-50.
- Schipper, K. 1991. Analyst forecasts. *Accounting Horizons* 5(4): 105-121.
- Tavcar, L.R. 1998. Make the MD&A more readable: Certified Public Accountant. *CPA Journal* 68(1): 10.
- Womack, K.L. 1996. Do brokerage analysts' recommendations have investment value? *Journal of Finance* 51(1): 137-167.

Appendix

To identify the information in analyst reports (AR) and Form 10-K, we use a neural network-based word embedding topic modeling approach, Word2vec. Word2vec groups together words with similar meanings. This grouping approach is appropriate for our setting because in accounting different words often are used when discussing the same topic. For example, pairwise “costs” and “expenses,” “repurchase” and “buyback,” and “eps” and “earnings per share” likely relate to the same topic. Word2vec identifies words as having similar meanings if the words co-occur with similar neighboring words. For example, Word2vec would identify “earnings” and “net income” as having similar meanings if they both occur frequently in the context of “We had record _____ this quarter.”¹²

We implement the Word2vec approach in six steps. First, we obtain AR from Thomson Reuters Investext. We eliminate AR from brokers that primarily provide robo reports, event transcripts, proxy and governance reports, credit rating reports, and company descriptions. To do this, we focus on the top 100 brokers by number of reports in our sample and review a random sample of five reports per broker. We keep reports from only brokers that primarily provide fundamental analysis. We randomly select one AR for each firm-year to ensure our topic modeling approach is computationally feasible. We convert AR from PDF to text for analysis. For PDF reports that are not machine readable, we use optical character recognition to convert the PDF to text.

¹² Word2vec differs from Latent Dirichlet Allocation (LDA), another topic analysis method. In LDA, words are probabilistically associated with latent topics, and words can appear in multiple topics. For example, in Dyer et al. (2017, Internet Appendix), the word “revenues” appears in the topics labeled subsidiaries, restaurants (franchises), software and systems, clients and revenues, semiconductor industry, and offshore oil drilling, among others. In contrast, Word2vec groups together words with similar meanings into topics, and each word appears in only one topic. Dyer et al. (2017) notes that having commonly used words, such as sales and income, in multiple topics can prevent LDA estimators from converging. Word2vec is less subject to this limitation because it ensures each word appears in only one topic (Mikolov, Chen, Corrado, and Dean 2013). This enables us to calculate how much discussion relates to each topic, even topics that include the words most commonly used in AR and Form 10-K.

Second, we obtain Form 10-K from the EDGAR database, and parse the text to remove HTML. This is necessary because in the EDGAR database Form 10-K contains HTML, which is not subject to our analyses. We base our approach for removing HTML on Loughran and McDonald (2016). In particular, we begin by removing all document segments that contain GRAPHIC, ZIP, EXCEL, JSON, PDF, and XML. We then parse HTML code—for example, we replace `\&NBSP` and `\ ` with a blank space—and remove SEC headers and footers and HTML tags. We also remove XBRL, but we retain text that appears in tables. Because AR generally does not contain HTML, this step is not necessary before we analyze AR.

In this step, we also extract two sections of Form 10-K, namely Management’s Discussion & Analysis (MD&A) and financial statements (FS). We base our approach for identifying the MD&A and FS sections on Dyer et al. (2017). In particular, we identify all instances of logical references to the MD&A and FS sections and impose a minimum section length of 2,000 characters. If multiple instances exist, we choose the longest.

Third, we process the text for analysis. In particular, we convert uppercase letters to lowercase and eliminate punctuation symbols, single-character words, and words that are not all alphabetical characters. We then remove the top 50 most common words such as “the,” “of,” and “and,” unless the words could relate to a topic of potential substantive interest, such as “income,” “price,” and “cash.” We concatenate together as a single word, using the underscore symbol, sequences of up to 16 words (i.e., 16-grams) that appear, on average, more than once per AR. For example, because the word sequences “net income” and “earnings per share” appear, on average, more than once per AR, we treat them as the single words “net_income” and “earnings_per_share.” To manage vocabulary size, we exclude words that appear, on average, less than once per 25 firm-year observations. Because AR is shorter, on average, than Form 10-

K, we scale the AR corpus to be the same length as the Form 10-K corpus. That is, we duplicate the AR corpus R times, where R is the ratio of Form 10-K total word count to AR total word count, truncated to an integer.

Fourth, we train the Word2vec model on the combined AR and Form 10-K corpus. We use the gensim library, which is an open-source Python package, to implement Word2vec (Rehurek and Sojka 2010). We follow the standard implementation and use a word window size of five and a vector size of 100. The word window specifies how many words before and after a given word Word2vec considers when determining the word's meaning. The vector size specifies how many dimensions of meaning Word2vec considers. See Mikolov et al. (2013) for details about Word2vec. This step effectively converts each word to a vector in a 100-dimensional vector space, in which words with similar (dissimilar) meanings are close together (far apart).

Fifth, we use K-means clustering to cluster the word vectors into clusters that minimize the within-cluster sums of mean squared vector differences. We instruct the model to create 100 clusters because untabulated statistics reveal that the rate of reduction in the sums-of-squares levels off at close to 100 clusters. Each cluster represents a topic. Together, these topics include all the words in the combined AR and Form 10-K corpus.

Sixth, for ease of discussion, we label and, following Dyer et al. (2017), organize the topics into major topic categories. We identify five categories—Performance, Analysis, Business, Financial Reporting, and Regulatory—on which we focus our analysis—and Other. For each of these categories, Table A.1, Panels A through F, presents the topics listed in the order of the most to least discussion in AR. It also presents the 15 most common words in each topic, listed in order of the frequency with which they occur.

Table A.1
Topic Categories, Topics, and Words

Panel A: Performance Topic Category

Topic	Words
Revenues/Margins	million was sales revenue were due revenues margin fiscal respectively segment compared year_ended december million_million margins
Ratios	ratio pe na dec fy cap source company mm book usd mil avg leverage figure nm
Earnings	net operating income loss tax earnings net_income losses foreign currency income_taxes taxes income tax gains gain
Target prices	price investors return stocks analysts sp total_return target index coverage_universe stock_price peer stars reit ranking
Recommendations	rating ratings sector outperform underperform rated past_months none usd_usd nr stewardship weight next_months ib medium
Adjusted earnings	eps per_share ebitda diluted adjusted flow shares_outstanding gaap basic earnings_per_share weighted_average pro forma calculated calculation
Expenses	costs cost expense expenses fees deferred recorded amortization recognized charges depreciation interest_expense incurred charge operating_expenses
Cash flows	cash cash_flow cash_flows net_cash free_cash cash_cash investing cash_equivalents cashflow undiscounted

Table A.1 (continued)
Topic Categories, Topics, and Words

Panel B: Analysis Category

Topic	Words
Markets and industries	market these new certain more industry number well those high some customer most specific generally
Trends and forecasts	current expected future performance results rates conditions level past relative levels actual periods forecast volatility
Positives and negatives	strong continued demand low overall activity positive negative upside recovery solid despite incremental pressure relatively
Estimates and assessments	estimates estimate valuation analysis estimated model using assumptions reporting analyses moat determine assessment methodology quantitative
Increases and decreases	growth increase increased primarily higher lower decrease down decline increases reduction reduced decreased increasing improvement
Successes and challenges	than one likely recent competitive important favorable recently consistent stable attractive difficult volatile successful critical
Views and opinions	if believe expect view opinion believes think assurance anticipate believed evidence uncertain confidence ultimate probable
Economic environment	economic environment caused economy damage weakness political weather delays lost challenges pressures problems downturn deterioration
Affect and effect	effect impact affect affected impacted affecting influenced impacting affects

Table A.1 (continued)
Categories, Topics, and Words

Panel C: Business Category

Topic	Words
Products and markets	us products companies operations customers markets clients facilities sources businesses segments relationships components firms arrangements
Servicies and technology	services systems technology system line software solutions technologies enterprise network communications delivery applications provider storage
Business and operations	business management development corporate key marketing program support resources strategy programs opportunities research development focus personnel
Banking	bank state federal trade local national commission banking banks department deposit canadian agency residents
Goods and retail	product retail consumer selling category distributor specialty brand goods food wholesale brands channel circulation merchandise
Energy	capital energy investment_banking_services lp sipc bluematrix wholly dominion noninvestment midstream petroleum gp resource incs qualifying
Prices and production	prices production contract oil gas volume capacity natural_gas supply volumes fuel commodity coal raw crude
Insurance and healthcare	benefit insurance benefits coverage plans health policy pension professional care retirement medical healthcare savings severance
States and regions	new_york delaware located american south america california north texas central west city pacific district florida
Manufacturing and materials	manufacturing materials industrial produced industries water generation steel parts paper semiconductor automotive waste natural manufacture
Cities and international locations	branch office street london floor offices square east box road tel avenue tower po tokyo
Transportation and utilities	service power orders electric transportation utility utilities air aircraft replacement vehicle star fleet terminal vehicles
Drugs and clinical trials	clinical patients treatment phase drug trial stage study trials response candidates patient cancer studies launch
Hotels and entertainment	country home stores store trademarks owners franchise centers entertainment family names gaming restaurants hotel territory
Projects and pipelines	project expansion area projects pipeline operation drilling region feet exploration location wells field proved plants
Partnerships and ventures	license partners partnership ii joint iii venture pharmaceuticals licensing arrangement exclusive partnerships royalty pharmaceutical ventures

Table A.1 (continued)
Categories, Topics, and Words

Panel D: Financial Reporting Category

Topic	Words
Investments	securities investment investments trading transactions exchange contracts agreements funds private fund financial_instruments shortterm instruments derivative
Equity	stock shares equity share common_stock common outstanding dividend dividends class shareholders stockholders offering issuance proceeds
Financial statements	financial statement see consolidated table_contents included reported statements financial_statements results_operations table thousands millions discussed balance sheet
Employees	employees executive you member who employee director committee termination employment directors board your participant board_directors
Interest rates	interest rate term longterm yield fixed interest rate life discount forward premium deposits interest_rates maturity contractual
Issuances	issued purchase distributed acquired sold held owned purchased offered accounted treasury valued traded closed trades
Credit and financing	credit loan principal obligations payment payments senior financing facility liquidity default commitments credit_facility collateral borrowings
M&A	acquisition sale control transaction acquisitions closing restructuring merger combination consolidation completion integration ma disposal disposition
Compensation	plan compensation option units options exercise incentive granted grant awards award restricted restricted_stock stock_options vesting
Loans	loans commercial portfolio trust mortgage real_estate finance residential bonds lending underwriting servicing lien housing portfolios
Regulations	requirements regulatory regulations laws compliance government environmental safety states protection restrictions jurisdictions fda practices governmental
Litigation	legal claims against action decision settlement judgment court litigation claim actions patent final proceedings patents
PPE	property equipment lease properties construction leases maintenance tenant building plant improvements land space rent intellectual
Intangibles/impairment	assets asset impairment goodwill inventory allowance carrying deferred_tax intangible_assets useful intangibles impaired lives longlived allowances
Liabilities	liabilities amounts accounts balance payable accrued reserves receivable reserve equivalents receivables premiums balances inventories contingent
Debt and notes	debt note notes indebtedness debentures
Fair value	value fair_value method fair values blackscholes splits basecase optionpricing
Accounting standards	accounting standard guidance standards sfas adoption principles poors recognition fasb presentation asu sharebased asc financial_reporting

Table A.1 (continued)
Categories, Topics, and Words

Panel E: Regulatory Category

Topic	Words
Responsibilities	but upon and/or respect except opinions otherwise without liability rights event relevant extent right obligation
Risks	change risk changes potential factors risks recommendations circumstances views events
Forms and exhibits	under form exhibit filed defined accordance pursuant covered dated connection amended intended among mentioned described
Legal language	agreement section act herein applicable terms item law rules regulation code securities exchange commission solicitation schedule rule
Uncertainties	material result significant adverse adversely financial_condition cause materially reduce significantly substantial lead occur limit aware
Audit report	registered firm public independent subject_change_without_notice preparation audit manager counsel llp sarbanesoxley redistribution author audited oversight
Disclaimers	responsible personal nor damages directly_indirectly arising conflict conflicts breach acts whatsoever liable fraud unauthorized violation
Internal controls	internal_control_over_financial_reporting holdneutral internal_control sponsoring treadway ratingsib coso controlintegrated unqualified

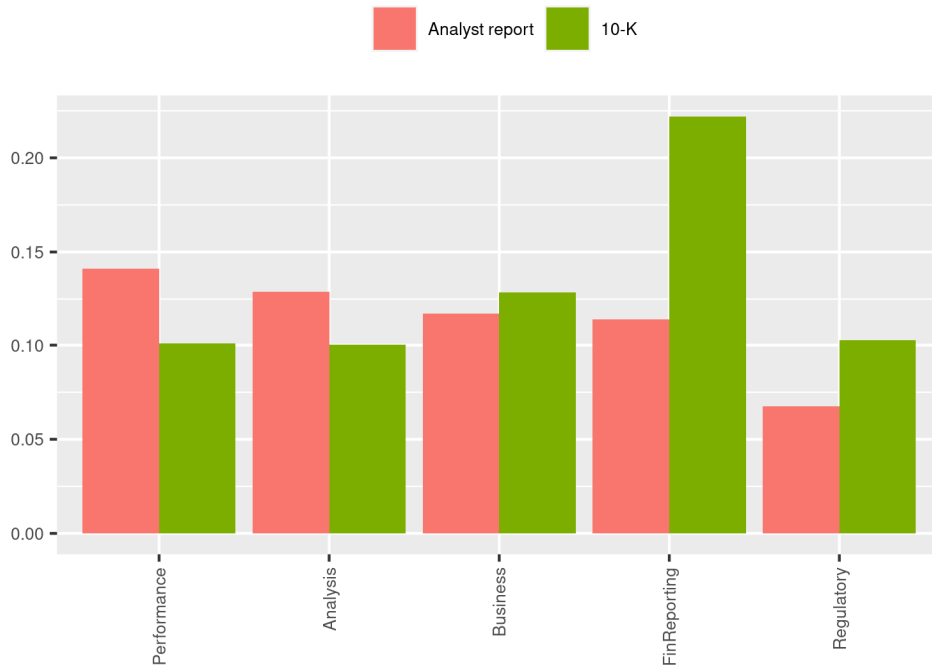
Table A.1 (continued)
Categories, Topics, and Words

Panel F: Other Category

Topic	Words
1	no it been also only part additional addition above when while however they currently further
2	year period during quarter years within time following prior first after end months last three
3	may will could available should required would do can does ability order reasonable must necessary
4	morningstar ubs morgan_stanley jp_morgan investment_banking jefferies deutsche_bank credit_suisse rbc_capital_markets morgan_stanley_research capital_markets sp_capital wells_fargo_securities_llc iq piper
5	total each over amount average basis annual up per approximately portion base percent unit less
6	companys com buy equity_research price_target neutral target_price recommendation update history mar overweight sep jun underweight
7	company corporation group llc between co corp holdings ag rights_reserved usa bancorp chase associates dean
8	limited general international united_states global third individual core primary separate single institutional major whole world
9	information page below disclosure disclosures please important_disclosures certification additional_information please_see information_data discussion list description definitions
10	including based related used about associated source regarding includes provides represents based_upon makes generated derived
11	sell provide make continue hold receive perform pay complete maintain remain become seek obtain meet
12	analyst cfa views_expressed registrant vice_president research_analyst mark certify president john michael james david former associate
13	provided made there effective paid received given determined prepared approved deemed established completed obtained indicated
14	their through use into offer existing own access providing improve develop drive operate ensure manage
15	had since making record remains maintaining known appears transition maintained revised created experienced assumes added

This table lists the topic categories, topics within each category, and the 15 top words for each topic estimated by Word2vec. For the category labeled “Other,” Panel F lists the top 15 of the category’s 41 topics. For each topic, words are listed in order of their frequency of occurrence in AR and Form 10-K.

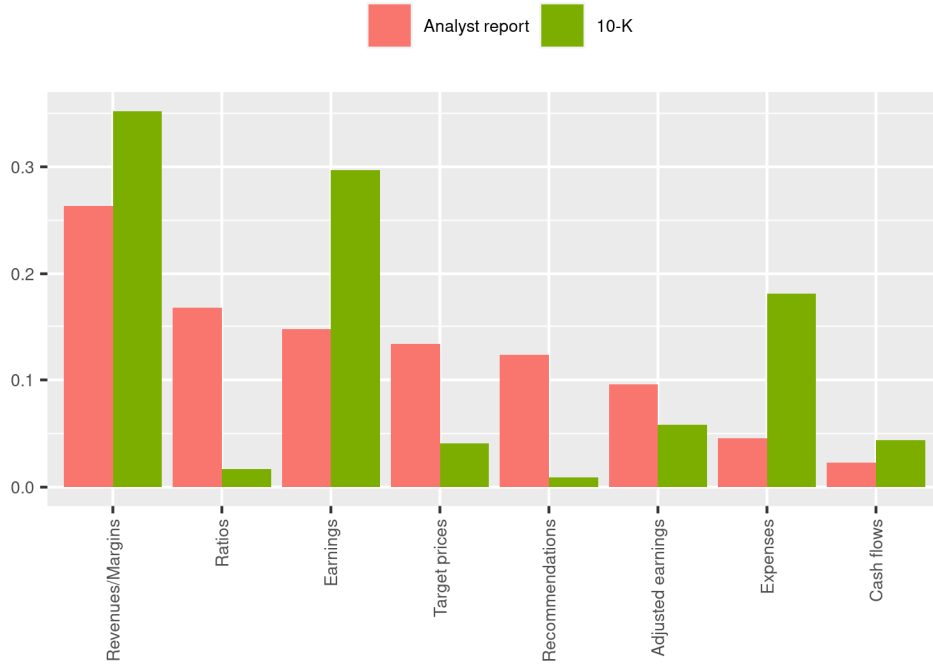
Figure 1
Analyst Report and Form 10-K Discussion by Topic Category



This figure presents the average discussion of each topic category in analyst reports and Form 10-K, as a proportion of the total discussion in each document. See Table A.1, Panels A to E, in the Appendix for the discussion topics in each category and the most frequent words in each topic. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Figure 2
Analyst Report and Form 10-K Discussion by Topic

Panel A: Performance Category



Panel B: Analysis Category

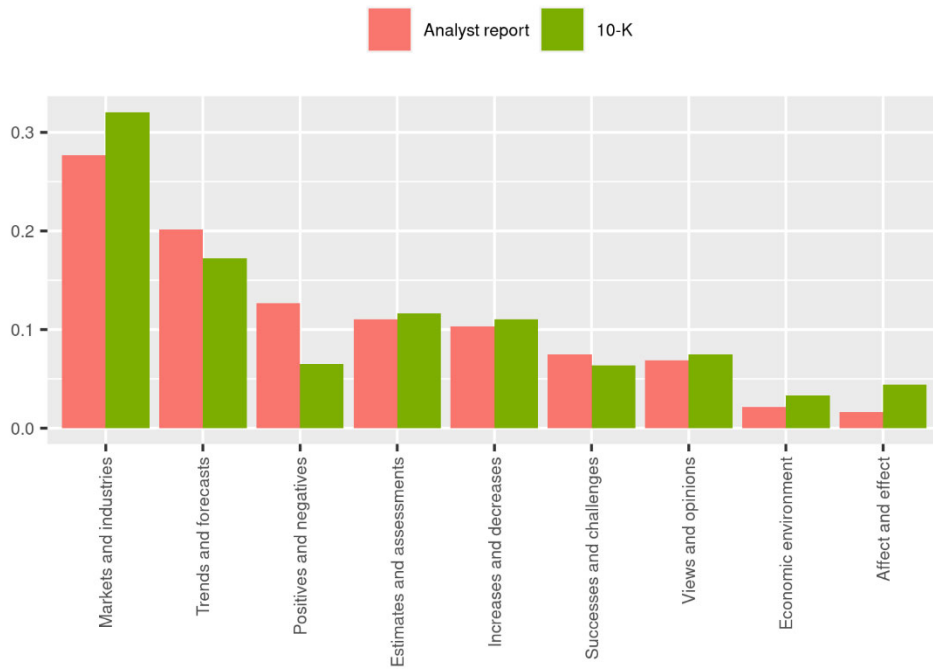
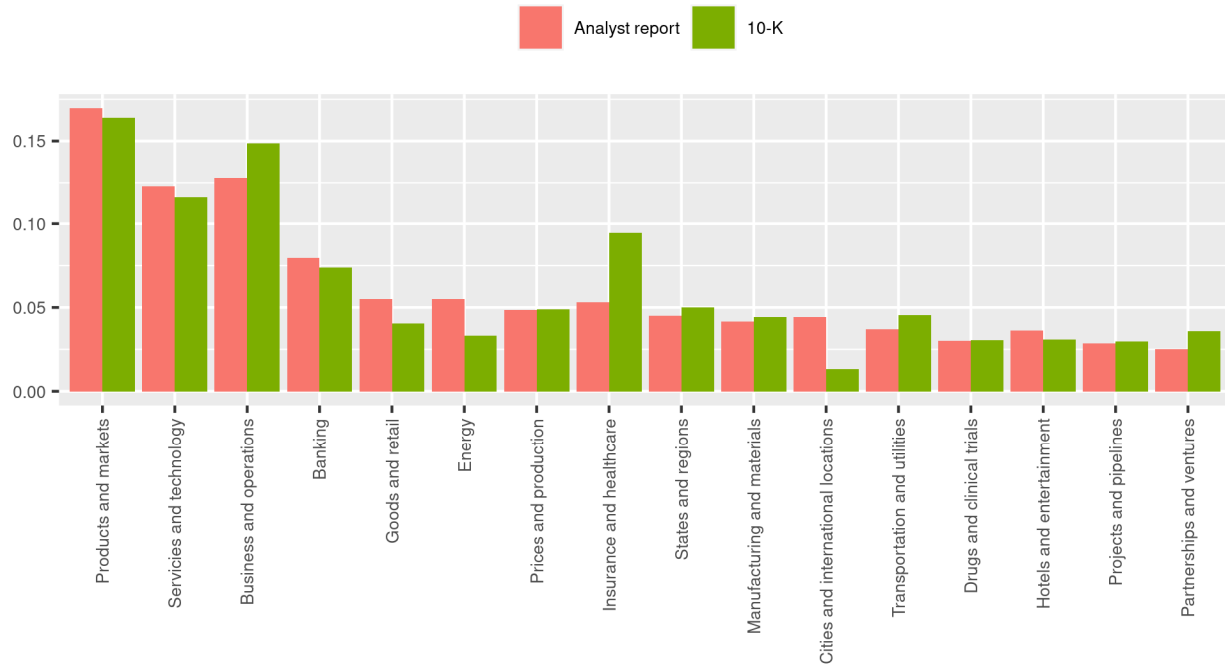


Figure 2 (continued)
Analyst Report and Form 10-K Discussion by Topic

Panel C: Business Category



Panel D: Financial Reporting Category

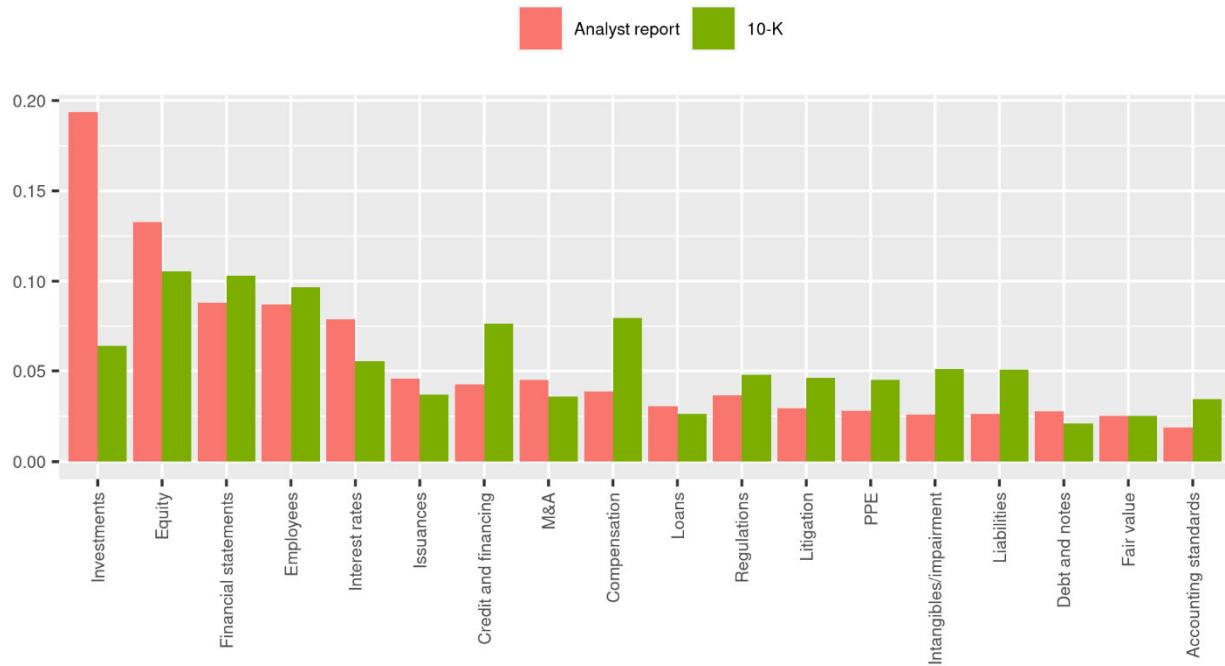
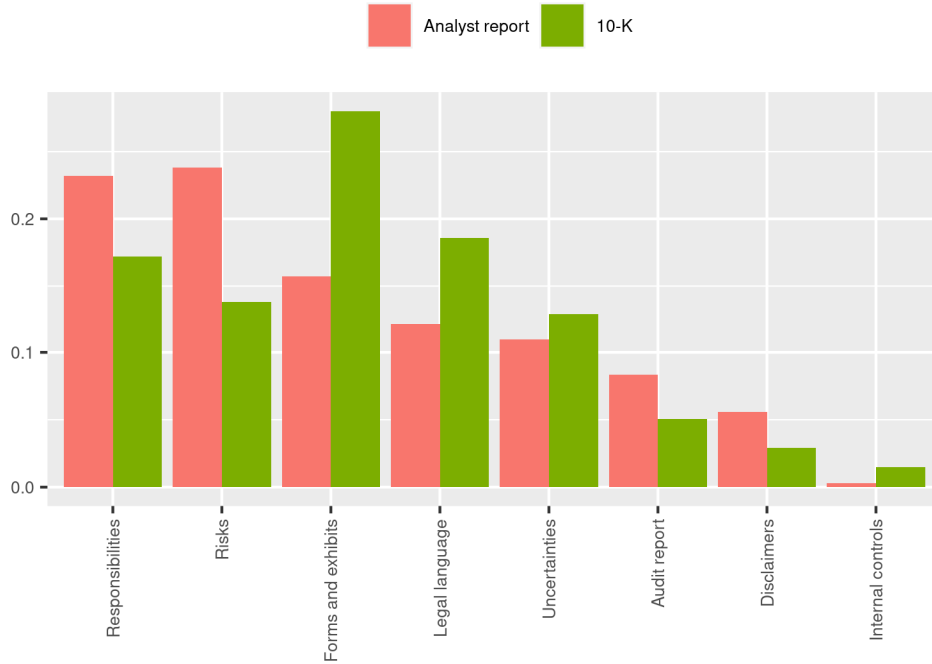


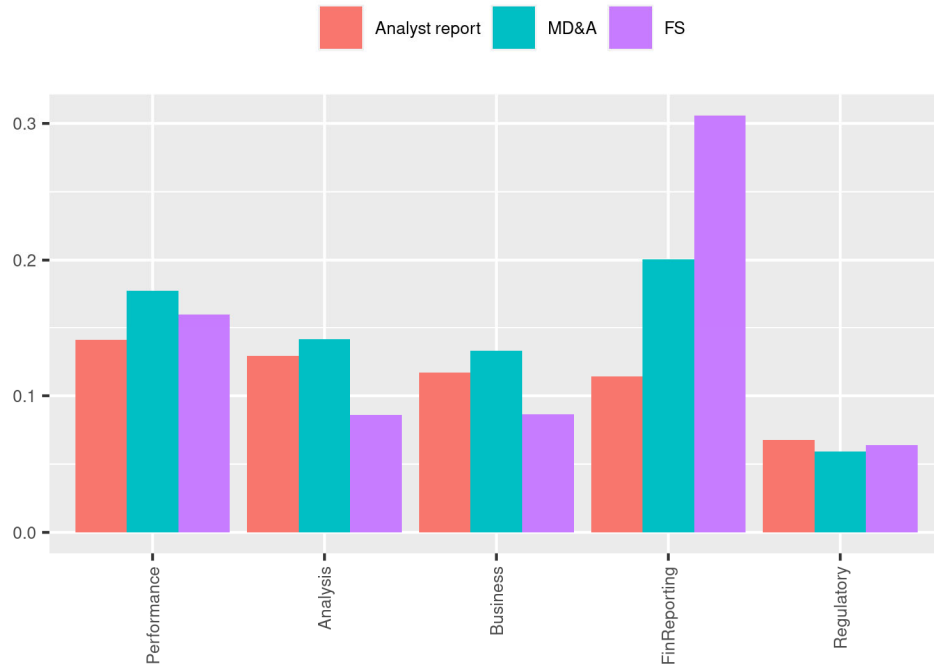
Figure 2 (continued)
Analyst Report and Form 10-K Discussion by Topic

Panel E: Regulatory Category



This figure presents the average discussion of each topic within each topic category in AR and Form 10-K, as a proportion of the total discussion in each document. Panels A through E present topics for the Performance, Analysis, Business, Financial Reporting, and Regulatory categories. See Table A.1, Panels A through E, for the most frequent words in each topic for each category. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

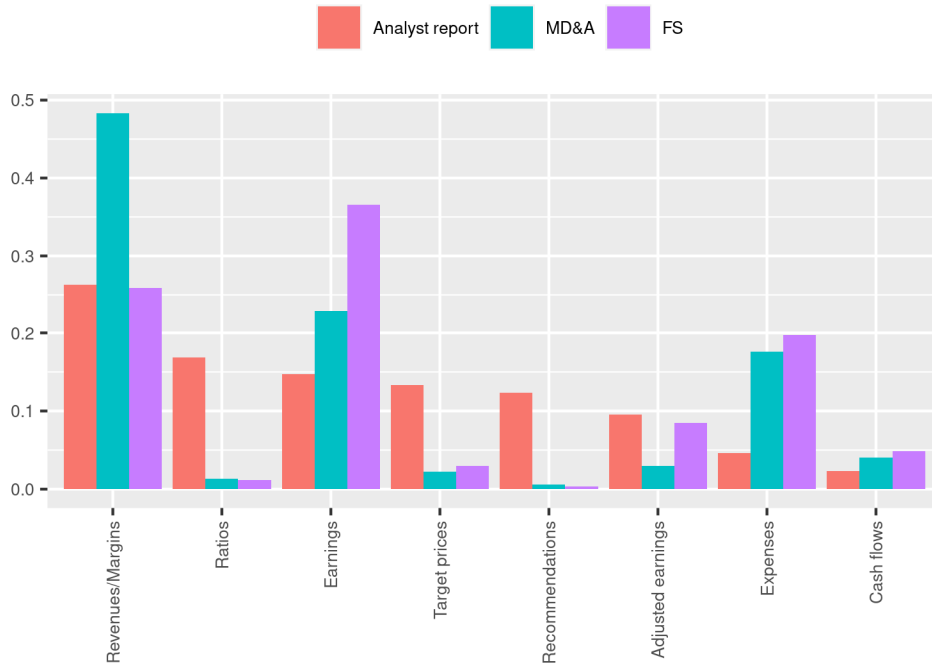
Figure 3
Analyst Report, MD&A, and Financial Statement Discussion by Topic Category



This figure presents the average discussion of each topic category in analyst reports (AR) and the Management’s Discussion and Analysis (MD&A) and Financial Statement (FS) sections of Form 10-K, as a proportion of the total discussion in each document. See Table A.1, Panels A through E, for the discussion topics in each category and the most frequent words in each topic. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Figure 4
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel A: Performance Category



Panel B: Analysis Category

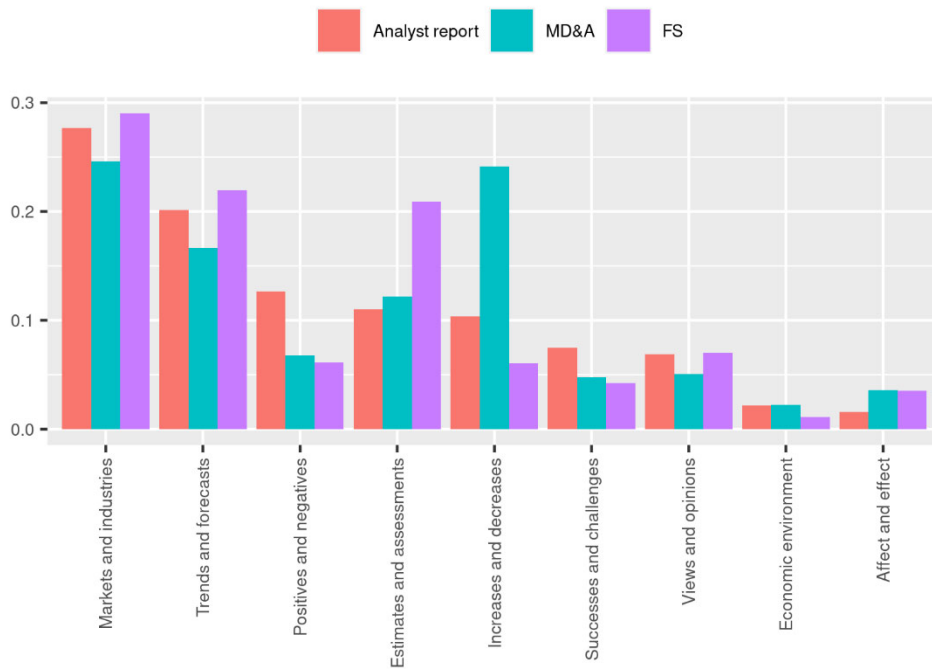


Figure 4
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel C: Business Category



Panel D: Financial Reporting Category

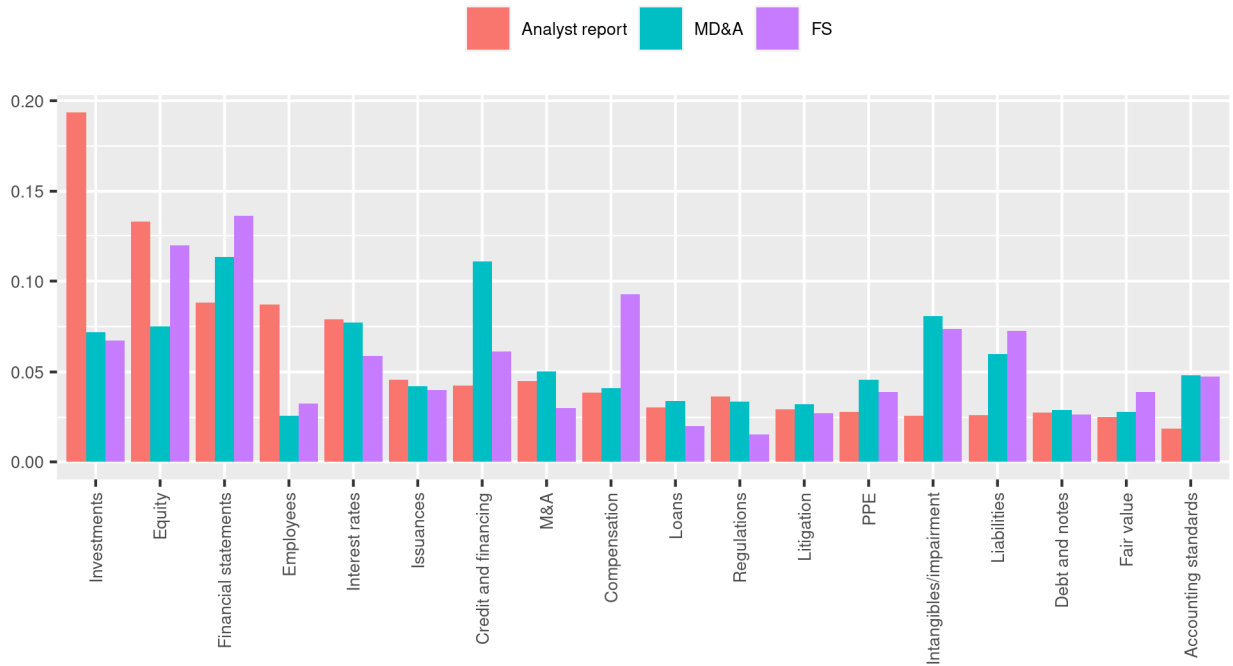
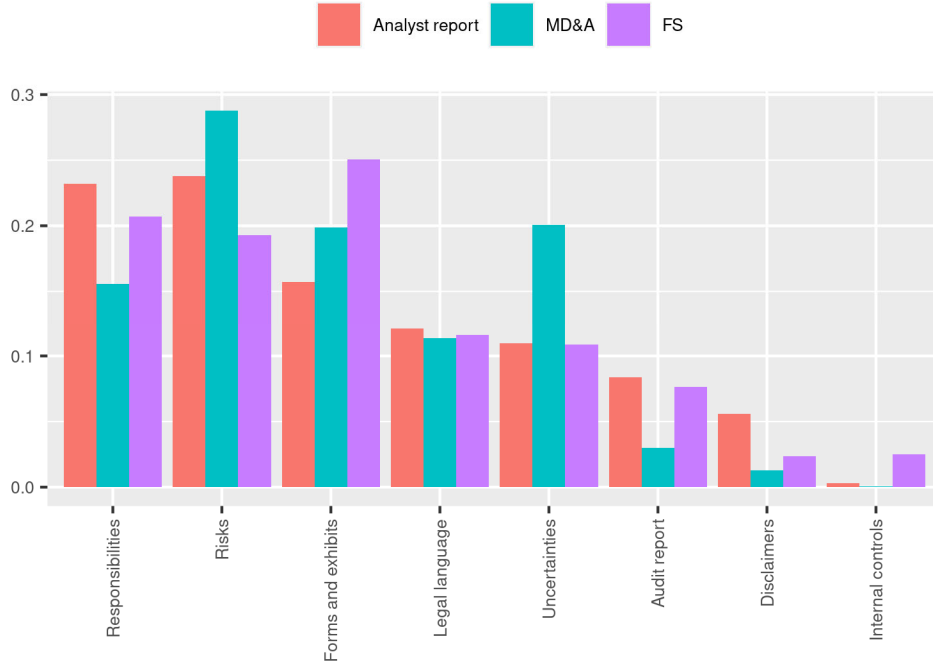


Figure 4
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel E: Regulatory Category



This figure presents the average discussion of each topic within each topic category in analyst reports (AR) and the Management’s Discussion and Analysis (MD&A) and Financial Statement (FS) sections of Forms 10-K, as a proportion of the total discussion in each document. Panels A through E present topics for the Performance, Analysis, Business, Financial Reporting, and Regulatory categories. See Table A.1, Panels A through E, for the most frequent words in each topic for each category. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 1
Analyst Report and Form 10-K Discussion by Topic Category

	AR	10-K	Diff.	t-stat.
Performance	14.11	10.08	4.03***	18.24
Analysis	12.90	10.03	2.87***	9.59
Business	11.71	12.84	-1.13***	-9.31
FinReporting	11.41	22.20	-10.79***	-101.18
Regulatory	6.75	10.26	-3.51***	-24.34
Other	43.11	34.59	8.53***	23.30
All Topics	100.00	100.00	23.20	

This table presents the average discussion of each topic category in analyst reports (AR) and Form 10-K as a proportion of the total discussion in each document. The All Topics row in the Diff. column represents the sum of the absolute differences in all 100 topics, divided by two. See Table A.1, Panels A through F, for the discussion topics in each category and the most frequent words in each topic. Tabulated amounts are in percentage points. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, based on standard errors double clustered by firm and by year. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 2
Analyst Report and Form 10-K Discussion by Topic

Panel A: Performance Category

	AR	10-K	Diff.	t-stat.
Revenues/Margins	26.33	35.22	-8.89***	-17.03
Ratios	16.85	1.69	15.16***	59.37
Earnings	14.71	29.69	-14.99***	-52.52
Target prices	13.33	4.10	9.23***	19.02
Recommendations	12.32	0.90	11.41***	21.09
Adjusted earnings	9.58	5.85	3.73***	25.03
Expenses	4.58	18.17	-13.59***	-128.54
Cash flows	2.30	4.37	-2.07***	-46.35
All Performance Topics	100.00	100.00	39.53	

Panel B: Analysis Category

	AR	10-K	Diff.	t-stat.
Markets and industries	27.15	32.09	-4.94***	-21.58
Trends and forecasts	19.71	17.30	2.41***	13.36
Positives and negatives	13.25	6.46	6.80***	61.93
Estimates and assessments	10.94	11.83	-0.88***	-2.79
Increases and decreases	10.92	11.14	-0.22***	-5.09
Successes and challenges	7.49	6.26	1.23***	18.23
Views and opinions	6.78	7.25	-0.47***	-6.53
Economic environment	2.24	3.33	-1.09***	-21.95
Affect and effect	1.51	4.34	-2.84***	-54.20
All Analysis Topics	100.00	100.00	10.44	

Table 2 (continued)
Analyst Report and Form 10-K Discussion by Topic

Panel C: Business Category

	AR	10-K	Diff.	t-stat.
Products and markets	16.08	16.27	-0.19	1.37
Servicies and technology	12.52	11.86	0.65***	6.51
Business and operations	12.40	14.57	-2.17***	-13.09
Banking	7.33	6.61	0.72**	2.51
Goods and retail	5.47	4.05	1.42***	19.52
Energy	5.43	3.34	2.10***	13.29
Prices and production	5.32	5.35	-0.02	-0.43
Insurance and healthcare	5.10	9.09	-3.99***	-27.92
States and regions	4.52	4.89	-0.38***	-4.80
Manufacturing and materials	4.39	4.62	-0.23***	-4.17
Cities and international locations	4.26	1.21	3.05***	15.62
Transportation and utilities	3.85	4.63	-0.78***	-13.17
Drugs and clinical trials	3.83	3.60	0.23	-0.35
Hotels and entertainment	3.55	3.01	0.54***	11.93
Projects and pipelines	3.16	3.19	-0.03**	-2.49
Partnerships and ventures	2.80	3.73	-0.93***	-14.49
All Business Topics	100.00	100.00	8.71	

Table 2 (continued)
Analyst Report and Form 10-K Discussion by Topic

Panel D: Financial Reporting Category

	AR	10-K	Diff.	t-stat.
Investments	19.34	6.38	12.96***	54.36
Equity	13.26	10.55	2.71***	11.96
Financial statements	8.82	10.29	-1.47***	-9.94
Employees	8.72	9.66	-0.94***	-4.25
Interest rates	7.88	5.51	2.37***	15.95
Issuances	4.57	3.70	0.87***	11.64
Credit and financing	4.24	7.65	-3.41***	-36.25
M&A	4.51	3.58	0.94***	8.69
Compensation	3.87	7.96	-4.09***	-34.17
Loans	3.05	2.64	0.42***	6.42
Regulations	3.64	4.78	-1.14***	-12.89
Litigation	2.94	4.61	-1.67***	-26.40
PPE	2.81	4.49	-1.68***	-19.37
Intangibles/impairment	2.58	5.09	-2.51***	-19.46
Liabilities	2.61	5.08	-2.47***	-31.56
Debt and notes	2.77	2.08	0.68***	12.84
Fair value	2.52	2.52	-0.00	-0.02
Accounting standards	1.88	3.45	-1.57***	-7.13
All Financial Reporting Topics	100.00	100.00	20.95	

Table 2 (continued)
Analyst Report and Form 10-K Discussion by Topic

Panel E: Regulatory Category

	AR	10-K	Diff.	t-stat.
Responsibilities	23.19	17.20	5.99***	20.01
Risks	23.80	13.84	9.96***	18.16
Forms and exhibits	15.70	28.03	-12.33***	-37.69
Legal language	12.11	18.58	-6.47***	-14.33
Uncertainties	10.97	12.92	-1.95***	-3.40
Audit report	8.35	5.06	3.30***	28.41
Disclaimers	5.59	2.90	2.69***	11.38
Internal controls	0.29	1.47	-1.18***	-11.25
All Regulatory Topics	100.00	100.00	21.93	

This table presents the average level of discussion of topic in each category in analyst reports (AR) and Form 10-K, as a proportion of the total discussion within the category. See Table A.1, Panels A through E, for the discussion topics in each category and the most frequent words in each topic. The All row in the Diff. column represents the sum of the absolute differences in each topic, divided by two. Tabulated amounts are in percentage points. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, based on standard errors double clustered by firm and by year. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 3
Analyst Report, MD&A, and Financial Statement Discussion by Topic Category

	AR	MD&A	FS	AR – MD&A		AR – FS		MD&A – FS	
				Diff.	tstat.	Diff.	tstat.	Diff.	tstat.
Performance	14.11	17.76	16.03	-3.65***	-12.68	-1.92***	-12.60	1.73***	8.75
Analysis	12.90	14.16	8.61	-1.26***	-6.95	4.29***	14.06	5.54***	38.36
Business	11.71	13.28	8.65	-1.57***	-14.41	3.06***	20.42	4.63***	33.54
FinReporting	11.41	20.04	30.59	-8.63***	-45.51	-19.18***	-101.03	-10.55***	-35.52
Regulatory	6.75	5.93	6.37	0.83***	6.82	0.38***	3.58	-0.44***	-7.82
Other	43.11	28.83	29.74	14.29***	42.48	13.38***	73.62	-0.91***	-5.26
All Topics	100.00	100.00	100.00	28.50		33.00		19.90	

This table presents the average discussion of each topic category in analyst reports (AR) and the Management’s Discussion and Analysis (MD&A) and Financial Statement (FS) sections of Forms 10-K as a proportion of the total discussion in each document. The All Topics row in the Diff. columns represent the sum of the absolute differences in all 100 topics, divided by two. See Table A.1, Panels A through F, for the discussion topics in each category and the most frequent words in each topic. Tabulated amounts are in percentage points. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, based on standard errors double clustered by firm and by year. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 4
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel A: Performance Category

	AR	MD&A	FS	AR – MD&A		AR – FS		MD&A – FS	
				Diff.	tstat.	Diff.	tstat.	Diff.	tstat.
Revenues/Margins	26.33	48.35	25.92	-22.02***	-72.73	0.42	0.58	22.43***	39.81
Ratios	16.85	1.30	1.16	15.55***	58.57	15.69***	60.90	0.14***	4.33
Earnings	14.71	22.90	36.54	-8.19***	-25.21	-21.83***	-88.49	-13.64***	-76.31
Target prices	13.33	2.22	2.99	11.11***	23.34	10.34***	20.80	-0.77***	-19.79
Recommendations	12.32	0.60	0.37	11.71***	22.15	11.95***	22.18	0.23***	8.02
Adjusted earnings	9.58	2.97	8.48	6.61***	37.50	1.10***	3.44	-5.51***	-12.79
Expenses	4.58	17.62	19.71	-13.05***	-92.31	-15.13***	-130.03	-2.09***	-17.05
Cash flows	2.30	4.04	4.83	-1.74***	-31.53	-2.53***	-40.57	-0.79***	-10.73
All Performance Topics	100.00	100.00	100.00	44.99		39.50		22.80	

Panel B: Analysis Category

	AR	MD&A	FS	AR – MD&A		AR – FS		MD&A – FS	
				Diff.	tstat.	Diff.	tstat.	Diff.	tstat.
Markets and industries	27.68	24.60	29.04	3.08***	20.03	-1.36***	-3.94	-4.44***	-17.20
Trends and forecasts	20.16	16.64	21.95	3.52***	17.42	-1.79***	-6.00	-5.32***	-29.66
Positives and negatives	12.65	6.76	6.15	5.89***	51.71	6.51***	51.40	0.62***	5.97
Estimates and assessments	11.02	12.21	20.89	-1.19***	-3.29	-9.88***	-27.37	-8.69***	-32.94
Increases and decreases	10.36	24.13	6.04	-13.77***	-57.93	4.32***	30.62	18.09***	66.15
Successes and challenges	7.46	4.80	4.24	2.67***	41.49	3.22***	45.42	0.55***	12.58
Views and opinions	6.90	5.04	7.01	1.86***	26.24	-0.11	-0.92	-1.97***	-19.81
Economic environment	2.16	2.25	1.12	-0.09	-0.89	1.04***	19.40	1.13***	11.90
Affect and effect	1.61	3.57	3.56	-1.97***	-28.30	-1.95***	-12.76	0.02	0.17
All Analysis Topics	100.00	100.00	100.00	17.02		15.09		20.40	

Table 4 (continued)
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel C: Business Category

	AR	MD&A	FS	AR – MD&A		AR – FS		MD&A – FS	
				Diff.	tstat.	Diff.	tstat.	Diff.	tstat.
Products and markets	16.96	17.45	16.90	-0.49**	-2.01	0.06	0.17	0.55***	2.71
Servicies and technology	12.28	10.99	10.28	1.29***	9.13	2.01***	18.90	0.71***	5.09
Business and operations	12.80	16.62	12.35	-3.83***	-27.75	0.44***	2.80	4.27***	37.89
Banking	7.96	5.15	8.00	2.81***	14.75	-0.04	-0.25	-2.85***	-41.00
Goods and retail	5.51	5.66	3.86	-0.15	-1.50	1.65***	19.07	1.80***	21.75
Energy	5.51	3.94	4.29	1.58***	8.73	1.22***	5.87	-0.36***	-5.52
Prices and production	4.86	6.94	4.98	-2.08***	-17.39	-0.12	-0.97	1.97***	17.73
Insurance and healthcare	5.32	8.47	15.25	-3.14***	-19.52	-9.93***	-37.45	-6.79***	-31.66
States and regions	4.51	2.80	4.36	1.71***	15.77	0.14	1.33	-1.57***	-19.22
Manufacturing and materials	4.15	4.07	3.51	0.08	1.06	0.64***	7.96	0.56***	9.97
Cities and international locations	4.42	0.76	0.91	3.66***	22.26	3.51***	21.30	-0.15***	-5.61
Transportation and utilities	3.71	4.49	4.07	-0.78***	-8.19	-0.36***	-3.36	0.42***	7.45
Drugs and clinical trials	3.02	2.67	1.80	0.35***	4.56	1.22***	12.46	0.88***	11.01
Hotels and entertainment	3.62	3.30	2.55	0.32***	4.22	1.07***	18.37	0.75***	11.72
Projects and pipelines	2.84	3.28	2.27	-0.43***	-7.47	0.57***	10.09	1.00***	20.90
Partnerships and ventures	2.51	3.41	4.61	-0.90***	-11.95	-2.09***	-22.03	-1.19***	-15.40
All Business Topics	100.00	100.00	100.00	11.80		12.54		12.90	

Table 4 (continued)
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel D: Financial Reporting Category

	AR	MD&A	FS	AR – MD&A		AR – FS		MD&A – FS	
				Diff.	tstat.	Diff.	tstat.	Diff.	tstat.
Investments	19.34	7.18	6.73	12.16***	42.16	12.61***	48.70	0.45**	2.31
Equity	13.26	7.49	11.98	5.76***	22.89	1.28***	4.08	-4.48***	-22.37
Financial statements	8.82	11.33	13.59	-2.51***	-17.54	-4.77***	-26.99	-2.26***	-10.63
Employees	8.72	2.60	3.26	6.11***	67.84	5.46***	61.35	-0.66***	-12.09
Interest rates	7.88	7.71	5.87	0.17	0.93	2.01***	10.40	1.84***	34.09
Issuances	4.57	4.23	3.99	0.34***	3.61	0.57***	8.16	0.24***	4.98
Credit and financing	4.24	11.07	6.14	-6.83***	-42.13	-1.90***	-14.44	4.93***	60.38
M&A	4.51	5.05	3.01	-0.53***	-5.59	1.51***	13.82	2.04***	16.07
Compensation	3.87	4.12	9.27	-0.24*	-1.86	-5.39***	-32.41	-5.15***	-32.99
Loans	3.05	3.42	2.01	-0.36***	-5.04	1.04***	12.37	1.41***	25.21
Regulations	3.64	3.36	1.58	0.27	1.44	2.06***	16.14	1.79***	17.74
Litigation	2.94	3.24	2.75	-0.30***	-2.96	0.19***	2.81	0.49***	5.29
PPE	2.81	4.56	3.90	-1.75***	-23.88	-1.09***	-14.55	0.66***	11.52
Intangibles/impairment	2.58	8.07	7.37	-5.50***	-25.75	-4.79***	-34.51	0.71***	5.99
Liabilities	2.61	6.00	7.25	-3.39***	-32.24	-4.64***	-48.48	-1.25***	-13.09
Debt and notes	2.77	2.91	2.68	-0.15*	-1.91	0.09*	1.77	0.24***	4.66
Fair value	2.52	2.82	3.88	-0.31***	-3.22	-1.37***	-23.53	-1.06***	-19.25
Accounting standards	1.88	4.83	4.75	-2.95***	-7.79	-2.87***	-8.98	0.08	0.42
All Financial Reporting Topics	100.00	100.00	100.00	24.82		26.82		14.86	

Table 4 (continued)
Analyst Report, MD&A, and Financial Statement Discussion by Topic

Panel E: Regulatory Category

	AR	MD&A	FS	AR – MD&A		AR – FS		MD&A – FS	
				Diff.	tstat.	Diff.	tstat.	Diff.	tstat.
Responsibilities	23.19	15.59	20.71	7.60***	22.36	2.48***	9.46	-5.12***	-33.86
Risks	23.80	28.80	19.27	-5.00***	-9.75	4.53***	8.06	9.53***	60.02
Forms and exhibits	15.70	19.88	25.06	-4.19***	-15.27	-9.36***	-25.57	-5.17***	-10.82
Legal language	12.11	11.36	11.62	0.75***	3.25	0.49	1.45	-0.26	-1.32
Uncertainties	10.97	20.07	10.85	-9.09***	-22.21	0.12	0.38	9.21***	15.24
Audit report	8.35	2.99	7.65	5.37***	48.50	0.71***	4.44	-4.66***	-28.99
Disclaimers	5.59	1.28	2.37	4.31***	16.44	3.22***	14.19	-1.08***	-21.78
Internal controls	0.29	0.03	2.48	0.26***	10.71	-2.19***	-10.61	-2.45***	-11.27
All Regulatory Topics	100.00	100.00	100.00	18.28		11.55		18.75	

This table presents the average discussion of each topic of each category in analyst reports (AR) and the Management’s Discussion and Analysis (MD&A) and Financial Statement (FS) sections of Forms 10-K as a proportion of the category discussion in each document. See Table A.1, Panels A through E, for the discussion topics in each category and the most frequent words in each topic. The All row in the Diff. columns represent the sum of the absolute differences in each topic, divided by two. Tabulated amounts are in percentage points. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, based on standard errors double clustered by firm and by year. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 5
Analyst Report Discussion by Topic Category: Paired Subsample Differences

	Profit – Loss		Large – Small		NonTech – Tech		NonFin – Fin	
	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.
Performance	0.50***	4.04	-1.43***	-14.11	0.07	0.56	-0.44***	-3.05
Analysis	0.70***	9.11	-0.21***	-2.81	0.11	1.11	0.42***	3.88
Business	-0.79***	-5.13	-0.22	-1.54	-0.81***	-4.67	2.09***	16.14
FinReporting	0.50***	4.73	0.23***	3.52	1.14***	13.63	-3.12***	-23.10
Regulatory	-0.26***	-4.88	0.51***	7.48	0.01	0.16	0.24***	3.80
Other	-0.66***	-6.22	1.12***	7.78	-0.54***	-4.01	0.81***	5.09
All Topics	3.37		4.60		5.29		7.06	

This table presents differences in analyst report (AR) discussions of each topic category between paired subsamples. The subsample pairs are profit versus loss firms, large versus small firms, non-technology versus technology firms, and non-financial versus financial firms. Loss firms report negative earnings. Large firms have market value of equity above the sample median. Technology firms have a primary three-digit SIC code of 283, 357, 360-368, 481, 737, or 873. Financial firms are in the Fama-French 48 industries of banking (44), insurance (45), and trading (47). See Table A.1, Panels A through F, in the Appendix for the discussion topics in each category and the most frequent words in each topic. Diff. represents average percentage point differences from the full sample mean of the discussion of each topic in AR as a proportion of the total discussion in each AR. The All row in the Diff. columns represent the sum of the absolute differences in all 100 topics, divided by two. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, based on standard errors double clustered by firm and by year. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 6
Analyst Report Discussion by Topic: Paired Subsample Differences

Panel A: Performance Category

	Profit – Loss		Large – Small		NonTech – Tech		NonFin – Fin	
	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.
Revenues/Margins	-0.21	-0.42	-2.93***	-9.67	-3.03***	-5.23	4.89***	10.51
Ratios	-0.88***	-3.01	-0.86***	-6.03	-0.34	-1.12	-1.72***	-5.78
Earnings	1.04***	6.59	-0.34**	-2.15	1.89***	12.00	-4.24***	-20.92
Target prices	0.63*	1.92	3.43***	13.79	1.11***	2.78	-0.13	-0.43
Recommendations	-0.26	-0.68	1.61***	4.44	0.15	0.32	0.52	1.48
Adjusted earnings	0.13	1.54	-0.59***	-5.34	0.15	1.12	0.11	0.80
Expenses	-0.12*	-1.73	-0.44***	-6.83	0.28***	3.43	-0.38***	-3.13
Cash flows	-0.33***	-6.81	0.12**	2.29	-0.22***	-3.29	0.96***	13.79
All Performance Topics	1.81		5.16		3.58		6.48	

Panel B: Analysis Category

	Profit – Loss		Large – Small		NonTech – Tech		NonFin – Fin	
	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.
Markets and industries	-1.70***	-7.74	-0.21	-1.14	-2.18***	-7.06	1.25***	5.19
Trends and forecasts	0.02	0.09	0.25	1.32	0.62**	2.03	-0.62***	-3.77
Positives and negatives	0.61***	5.21	-0.25***	-2.65	0.03	0.30	0.86***	4.91
Estimates and assessments	-0.07	-0.52	0.67***	3.66	0.50***	4.64	-0.39*	-1.86
Increases and decreases	1.78***	11.52	-0.56***	-4.22	1.48***	8.76	-1.32***	-6.76
Successes and challenges	-0.22***	-3.44	-0.16***	-2.78	-0.21***	-3.21	-0.13	-1.56
Views and opinions	-0.62***	-8.70	-0.03	-0.35	-0.74***	-7.60	0.46***	4.85
Economic environment	0.19***	3.68	0.21***	4.85	0.38***	7.10	-0.15***	-3.62
Affect and effect	-0.00	-0.01	0.08	1.49	0.12***	3.30	0.04	0.83
All Analysis Topics	2.61		1.21		3.13		2.61	

Table 6 (continued)
Analyst Report Discussion by Topic: Paired Subsample Differences

Panel C: Business Category

	Profit – Loss		Large – Small		NonTech – Tech		NonFin – Fin	
	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.
Products and markets	0.05	0.16	0.68***	3.18	-1.32***	-6.48	0.44*	1.76
Servicies and technology	-2.30***	-4.25	-1.37***	-4.45	-9.55***	-16.25	4.66***	10.31
Business and operations	0.04	0.22	-0.70***	-4.52	-0.97***	-4.14	-1.03***	-4.68
Banking	1.66***	8.14	0.86***	3.15	2.35***	13.59	-7.63***	-16.54
Goods and retail	1.07***	7.88	0.31*	1.67	0.95***	6.57	1.55***	9.86
Energy	0.80***	5.30	0.62***	3.09	2.29***	16.29	-1.75***	-8.50
Prices and production	0.92***	4.18	0.77***	3.98	3.00***	16.78	1.90***	8.77
Insurance and healthcare	0.67***	4.20	0.17	1.06	1.32***	6.92	-2.16***	-7.39
States and regions	1.03***	5.34	-0.10	-0.80	2.06***	11.97	-3.22***	-9.19
Manufacturing and materials	0.86***	6.00	0.02	0.12	1.23***	6.87	2.48***	15.11
Cities and international locations	0.05	0.45	0.82***	7.19	0.61***	4.55	-1.29***	-6.73
Transportation and utilities	1.04***	9.04	0.19	1.37	1.32***	9.42	1.50***	10.62
Drugs and clinical trials	-5.03***	-11.21	-2.16***	-9.68	-4.31***	-9.24	2.44***	15.04
Hotels and entertainment	0.81***	4.32	0.21	1.35	2.07***	11.83	0.17	0.79
Projects and pipelines	0.10	0.69	0.21	1.51	1.26***	9.10	0.80***	7.15
Partnerships and ventures	-1.78***	-9.44	-0.53***	-5.11	-2.31***	-10.39	1.14***	10.66
All Business Topics	9.10		4.86		18.47		17.08	

Table 6 (continued)
Analyst Report Discussion by Topic: Paired Subsample Differences

Panel D: Financial Reporting Category

	Profit – Loss		Large – Small		NonTech – Tech		NonFin – Fin	
	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.
Investments	-0.63***	-3.50	1.73***	7.98	-1.08***	-4.80	2.27***	7.94
Equity	0.23	1.61	-0.37***	-3.43	-0.20	-1.14	0.63***	3.78
Financial statements	-0.29***	-3.34	-0.44***	-5.41	-0.66***	-7.80	0.81***	6.08
Employees	-0.97***	-6.07	-0.14	-1.31	-1.11***	-7.15	2.06***	15.40
Interest rates	1.04***	9.78	0.43***	4.48	0.89***	7.87	-1.98***	-9.20
Issuances	0.07	1.14	0.38***	6.10	-0.12*	-1.84	0.81***	10.30
Credit and financing	-0.04	-0.32	-0.48***	-3.63	0.68***	6.92	-1.80***	-13.48
M&A	0.24***	3.52	-0.04	-0.41	0.11	1.15	0.54***	5.44
Compensation	-0.03	-0.27	-0.18	-1.63	-0.06	-0.48	0.86***	9.98
Loans	0.99***	5.48	-0.56***	-4.61	1.85***	17.28	-6.39***	-18.62
Regulations	-0.45***	-3.71	0.51***	6.89	-0.07	-0.80	0.91***	12.35
Litigation	-0.44***	-5.07	-0.08	-1.23	-0.46***	-5.14	0.93***	14.00
PPE	0.24***	3.21	-0.18*	-1.92	0.68***	7.55	0.06	0.33
Intangibles/impairment	0.17**	2.10	-0.30***	-4.02	0.18***	2.74	-1.24***	-9.75
Liabilities	-0.11*	-1.78	-0.53***	-7.30	-0.47***	-5.73	0.00	0.02
Debt and notes	-0.23***	-4.21	-0.44***	-6.75	-0.05	-0.80	0.99***	11.93
Fair value	0.04	1.16	0.23***	4.29	0.25***	5.44	-0.30***	-3.94
Accounting standards	0.17**	2.46	0.46***	3.60	-0.35***	-5.51	0.83***	10.63
All Financial Reporting Topics	3.19		3.73		4.64		11.71	

Table 6 (continued)
Analyst Report Discussion by Topic: Paired Subsample Differences

Panel E: Regulatory Category

	Profit – Loss		Large – Small		NonTech – Tech		NonFin – Fin	
	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.	Diff.	t-stat.
Responsibilities	-0.18	-0.82	0.82**	2.37	-0.49**	-1.97	0.54***	2.75
Risks	0.60**	2.20	-1.64***	-4.39	0.78***	2.66	-1.33***	-3.96
Forms and exhibits	-0.28**	-2.43	0.28	1.15	-0.58***	-4.16	0.68***	3.91
Legal language	-0.29*	-1.96	0.53***	3.46	-0.05	-0.26	0.06	0.32
Uncertainties	-0.42***	-3.60	-1.07***	-5.30	-0.21**	-2.03	0.30**	2.25
Audit report	0.19*	1.82	0.33***	2.67	0.22*	1.89	-0.28**	-2.29
Disclaimers	0.41***	3.55	0.87***	4.94	0.37**	2.31	0.01	0.06
Internal controls	-0.03	-1.40	-0.11***	-4.88	-0.04	-1.39	0.03	1.32
All Regulatory Topics	1.20		2.82		1.37		1.61	

This table presents differences in analyst report (AR) discussions of each topic between paired subsamples. The subsample pairs are Profit and Loss, Technology and Non-Technology, Large and Small, and Financial and Non-Financial. Loss firms report negative earnings before extraordinary items and Profit firms are all other firms. Large (Small) firms have market value of equity above (below) the sample median for that year. Technology (Non-Technology) firms are those whose primary three-digit SIC code is (is not) 283, 357, 360-368, 481, 737, or 873. Financial (Non-Financial) firms are those in (not in) the Fama-French 48 industries of banking (44), insurance (45), and trading (47). See Table A.1, Panels A through E, in the Appendix for the discussion topics in each category and the most frequent words in each topic. Diff. represents average percentage point differences from the full sample mean of the discussion of each topic in AR as a proportion of the total discussion in each AR. The t-stat column presents significance of differences in tabulated amounts between subsamples in the Diff. column. The All row in the Diff. columns represent the sum of the absolute differences in each topic, divided by two. *, **, and *** indicate significance at the 10, 5, and 1 percent levels, based on standard errors double clustered by firm and by year. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.

Table 7
Differences between Paired Subsamples in Topics Discussed in Analyst Reports and Forms 10-K

Panel A: Profit versus Loss Firms

	AR Profit	AR Loss	10-K Profit	10-K Loss
AR Profit	—	3.37	23.40	24.33
AR Loss		—	24.18	23.07
10-K Profit			—	5.93
10-K Loss				—

Panel B: Large versus Small Firms

	AR Large	AR Small	10-K Large	10-K Small
AR Large	—	4.60	24.03	23.92
AR Small		—	23.38	23.04
10-K Large			—	4.14
10-K Small				—

Panel C: Non-Technology versus Technology Firms

	AR NonTech	AR Tech	10-K NonTech	10-K Tech
AR NonTech	—	5.29	23.34	25.17
AR Tech		—	25.81	23.28
10-K NonTech			—	7.65
10-K Tech				—

Panel D: Non-Financial versus Financial Firms

	AR NonFin	AR Fin	10-K NonFin	10-K Fin
AR NonFin	—	7.06	23.49	26.86
AR Fin		—	24.88	22.72
10-K NonFin			—	10.27
10-K Fin				—

This table presents differences in topics discussed in analyst reports and Forms 10-K between paired subsamples of firms. The subsample pairs are Profit and Loss, Technology and Non-Technology, Large and Small, and Financial and Non-Financial. Loss firms report negative earnings before extraordinary items and Profit firms are all other firms. Large (Small) firms have market value of equity above (below) the sample median for that year. Technology (Non-Technology) firms are those whose primary three-digit SIC code is (is not) 283, 357, 360-368, 481, 737, or 873. Financial (Non-Financial) firms are those in (not in) the Fama-French 48 industries of banking (44), insurance (45), and trading (47). See Table A.1, Panels A through F, in the Appendix for the discussion topics in each category and the most frequent words in each topic. Tabulated amounts are the sum of one minus the sum of the absolute differences in AR discussions for between the paired firm groups for all 100 topics, presented in percentage points. The sample comprises 26,757 firm-year observations for 4,335 firms from 1996 to 2018.