Understanding the Role of Managerial Textual Content in the Price Formation Process*

August 2013

Elizabeth Demers
Darden Graduate School of Business
University of Virginia
Charlottesville, VA 22903
demerse@darden.virginia.edu

Clara Vega
Board of Governors
of the Federal Reserve System
Washington, DC 20551
clara.vega@frb.gov

Abstract

Using automated textual analysis algorithms to quantify the content of more than 20,000 managerial press releases from 1998 to 2006, this study presents large sample evidence of the role of difficult-to-verify information in the equity price formation process. We extract two textual content measures, labeled textual net optimism and textual uncertainty, and we show that each measure is associated with numerous alternative proxies for the firm's valuation fundamentals, that the text of managerial press releases conveys "first-moment" and "second-moment" information that is excluded from the GAAP earnings summary figure, managers' earnings forecasts and pro-forma earnings news, and that the market relies upon this information in the price discovery process. We choose and validate an array of proxies for the informativeness of textual content and GAAP earnings, respectively, and use these conditioning variables to study how the price impact of the two signals varies with their expected informativeness about valuation fundamentals. We find that the price impact of textual content increases in settings where the text is more informative about fundamentals and decreases when GAAP earnings are noisy predictors, consistent with the tenets of a Bayesian learning model. However, our analysis also suggests that the relationship between these two sources of information is likely more complex than that of simple substitutes. We find that some firms seem to provide more informative text to complement what they understand to be informationally deficient GAAP earnings. In other words, in some cases, the informativeness of managerial text appears to be endogenously related to the informativeness of GAAP earnings.

JEL Classifications: G14; D82; M41

Keywords: Earnings announcements; textual analysis; textual content; post-earnings drift; cheap talk; earnings quality; information uncertainty; momentum; voluntary disclosure; language tone; linguistic algorithms; communication credibility

---

*Earlier versions of this paper were circulated under the title "Soft Information in Earnings Announcements: News or Noise?" We thank Cam Harvey, two anonymous referees, and an anonymous associate editor for their very valuable comments. We also thank workshop and conference participants at the Federal Reserve Board of Governors; Duke University; Lancaster University; the Swiss Finance Institute in Lugano; London Business School; the University of Barcelona; the University of Essex; the University of Fribourg; the National University of Singapore; the European Accounting Association and American Accounting Association annual meetings; the Conference on Forensic Accounting, Internal Control, and Corporate Governance at the University of Toronto; the Society for Economic Dynamics Istanbul 2009 meeting; as well as Alain Chaboud, Jennifer Francis, Dan Givoly, Denis Gromb, Stuart Mcleay, Per Olsson, Katherine Schipper, Steve Sharpe, Eva Sletten, and Jeremy Stein for valuable comments and discussions. We also appreciate the excellent research assistance of Armel Allard, Monique Bruet, Shannon Demers, Nicholas Klagge, Marcus Newman, and Hien Tran; we thank Craig Carroll for processing our text files using Diction software; and we thank PR Newswire for providing us with the earnings announcement data. The views presented in this paper are solely those of the authors and do not represent the views of the Federal Reserve Board or its staff.

Numerous studies have long shown that quantitative information about fundamentals explains only a small portion of asset price movements (e.g., Shiller (1981), Roll (1988), and Mitchell & Mulherin (1994)), and identifying additional sources of fundamental information that are incorporated into asset prices continues to be of basic importance to financial economics. In this paper, we study the impact of one potential key source of fundamental information, the text of the managerial press releases that accompany quarterly announcements of corporate earnings.

To analyze and quantify the textual content of managerial press releases, we apply several automated linguistic analysis algorithms to more than 20,000 corporate earnings press releases filed with the PR Newswire service from January 1998 to July 2006, extracting two variables that we label textual net optimism and textual uncertainty. Textual net optimism measures the net amount of positive words used in the announcement, which may relate to factual assertions (such as "we experienced record sales levels") or to managerial spin on the news reported (such as "we had a great quarter" or "we have a positive outlook"), whereas textual uncertainty captures the apparent degree of confidence in the assertions made in the press release (for example, the words "appear," "believe," "approximate," or "might" increase the uncertainty score of the text).

Using these variables, we investigate whether and how asset prices respond to the textual content of managerial press releases. We first find that, even after controlling for Generally Accepted Accounting Principles (GAAP) earnings news, managerial forecast news, pro-forma earnings news, and other variables, announcement period abnormal returns are positively associated with unexpected textual net optimism and negatively associated with textual uncertainty. We also find that GAAP earnings news that are accompanied by a managerial press release with higher textual uncertainty tend to have a smaller impact on the firm's equity price. These results suggest that the text of managerial press releases conveys "first-moment" and "second-moment" information that is excluded from the GAAP earnings summary figure, managers' earnings forecasts and pro-forma earnings news, and that the market relies upon this information in the price discovery process.

We then investigate how the asset price response to the textual net optimism varies with the perceived informativeness of both the managerial press release and the simultaneously released earnings figure. Our concept of perceived informativeness aggregates two potential sources. We expect information to be viewed as more informative if it is perceived to be more value-relevant and/or if it is perceived to be more credible. Although the issuance of text within an earnings

announcement is not voluntary per se, the content of the text is subject to managerial discretion and is non-binding and difficult-to-verify. We thus rely upon the voluntary disclosure literature to choose several proxies for settings where textual content is more likely to be viewed as incrementally informative for valuation. Specifically, we conjecture that textual net optimism will be more informative when it is supported by numerical data (that is more easily verified), when it is issued by firms with better past reporting reputations, or when it is subject to high levels of scrutiny by analysts. We test the validity of these empirical proxies for the expected informativeness of textual content in our sample, and we find that they do indeed have explanatory power for how well textual net optimism helps to predict future earnings.

As for the informativeness of the GAAP earnings figure, guided by a well-established literature, we conjecture that certain characteristics should be associated with a lower level of informativeness. These include the firm's inclusion in a high-tech industry, whether or not the firm has R&D spending, the firm's P/E ratio, the firm's EFKOS e-loading factor (a measure of earnings quality that is due to Ecker, Francis, Kim, Olsson, & Schipper (2006)), and, in addition, the textual uncertainty of the announcement. Again, we test the validity of these empirical proxies for the expected informativeness of GAAP earnings in our sample, and we confirm that each of these proxies helps us to distinguish between firms whose GAAP earnings are more informative for future earnings and those whose GAAP earnings are less informative.

Overall, when we then use these conditioning variables to study how the market impact varies with the perceived informativeness of the managerial press release and of the GAAP earnings, we find that the price impact of textual content generally increases in settings where it is more informative for future earnings and where historical GAAP earnings are a noisier indicator of expected future earnings. This is consistent, for example, with a basic Bayesian learning model, where the impact of a given piece of news should be positively related to its own informativeness but negatively related to the informativeness of other simultaneously released news.

We find, for instance, that textual net optimism has a larger impact on announcement period abnormal returns when the firm is subject to higher levels of analyst scrutiny or when the firm provides numerical support within the text. The findings are consistent with the view that higher levels of analyst coverage encourage more informative disclosures (e.g., Matsumoto, Pronk & Roelofsen (2011)) and that higher levels of analyst scrutiny and the provision of additional

verifiable numerical support limit managers' ability to release misleading text (Krishna & Morgan (2001) and Benabou & Laroque (1992)). For high-tech firms, firms with R&D spending, firms with high P/E ratios and high EFKOS e-loading factors, we confirm that the ability of current and past GAAP earnings to predict future earnings is lower. We then find that, for these firms, investors rely more heavily on the simultaneously released textual content to update prices, even if the textual content of these firms is, per se, not more informative for future earnings than that of other firms.

However, our analysis suggests that the relationship between these two sources of information used in price discovery, the GAAP earnings and the textual content of managerial press releases, is likely more complex than that of simple substitutes. For instance, for some firms, the informativeness of textual net optimism appears to be endogenously related to the informativeness of GAAP earnings. For firms that voluntarily release pro forma earnings or earnings forecasts, the textual net optimism variable has a higher predictive ability for future earnings while the GAAP earnings number itself has less predictive power. As expected, for these firms' releases, the price impact of GAAP earnings news is significantly lower, while the price impact of textual net optimism is higher. As Lev (2012) explains, managers are more likely to provide, and market participants are more likely to value, a voluntary disclosure of pro forma earnings and earnings forecasts when GAAP earnings are informationally deficient. Our findings are consistent with the view that these firms may also choose to provide more informative text to complement what they understand to be informationally deficient GAAP earnings.

The rest of the paper is organized as follows. Section I provides background by situating our paper in the related literature. Section II describes our sample, data sources, and the measurement of earnings and textual content variables. In Section III we provide robust evidence that managerial textual content is incrementally informative about the firm's valuation fundamentals. Section IV defines and validates our empirical proxies for the informativeness of textual net optimism and GAAP earnings for future earnings, and presents evidence that the market's response to textual net optimism generally varies across announcement characteristics in the manner predicted by a standard Bayesian learning model. Section V summarizes and concludes.

I. Background

In efficient markets, stock prices should respond to new information that is relevant to the assessment of the value of the firm, which is to say to any news signal that is incrementally informative about the firm's future cash flows and/or the uncertainties of those future cash flows. In the context of earnings announcements, extensive evidence has established that, while earnings realizations are informative about share price movements, they are not a sufficient statistic for valuation. For example, Francis, Schipper & Vincent (2002) and Brandt, Kishore, Santa-Clara, & Venkatachalam (2008) show that a significant portion of the market's reaction around the earnings announcement date is attributable to other information that managers release contemporaneously with bottom-line earnings, and that the proportion of the market's reaction that is attributable to this complementary information has increased over time.[1] This is not entirely surprising given that earnings reported under GAAP are historical and conservative in nature, whereas valuation is forward-looking and anticipates both gains and losses.[2] These well-understood limitations of GAAP earnings suggest that the textual content of the managerial press releases that accompany GAAP earnings may have a significant role to play in the price formation process.

We first briefly discuss the small but growing literature on the impact of textual content on asset prices and how our study fits within that literature. Without the benefit of automated linguistic analysis, some early work still made important contributions to our understanding of how the non-binding, difficult-to-verify component of the text of managerial releases impacts asset prices. For instance, Hutton, Miller & Skinner (2003) investigate whether and how the voluntary provision of supplementary statements with earnings forecasts affects asset prices, and Baginski, Hassell & Kimbrough (2004) study the impact of attributive statements accompanying managers' forecasts. Both sets of authors find that the market's response to this hard-to-verify information is affected by its likely credibility, where the measure of credibility is based on the assumption that bad news will be perceived as more credible than good news. These studies suffer from the inherent limitation that they use small samples of hand-collected and hand-

[1] Several studies further suggest that the explanatory power of earnings for stock prices has been declining over time (e.g., Collins, Maydew & Weiss (1997); Francis & Schipper (1999)).

[2] "Conservatism" in accounting is traditionally summarized by the old adage that one should "anticipate no profit but anticipate all losses" (Bliss (1924); Watts (2003a)). In practical terms, this means that increases in asset values that are not sufficiently verifiable are not recorded whereas decreases of similar verifiability are recorded (Watts (2003b)).

analyzed data in rather specific settings (i.e., the voluntary issuance of a forecast by managers) that necessarily entail a self-selected sample. Our study overcomes these limitations by using a large comprehensive sample of corporate reporting events, which computerized analysis allows, and by measuring the informativeness of hard-to-verify communications using an array of alternative, replicable proxies.

Some more recent studies use automated linguistic analysis to study how media-expressed pessimism about firms is related to abnormal returns and to future firm performance (e.g., Tetlock (2007); Tetlock, Saar-Tsechansky & Macskassy (2008); and Engelberg (2008)). There are obvious differences, of course, between media-issued communication and manager-issued communication. Relative to the media, managers have proprietary information about their firm's prospects that they may choose to reveal in the text of their announcements in order to complement other, more verifiable releases such as the mandatory GAAP earnings figure. The communication incentives facing managers and reporters are also different (Kothari, Li & Short (2009)). In particular, management may have incentives to send uninformative messages (Crawford & Sobel (1982)), to release good news and withhold bad news (Verrecchia (1983)), or even to report overly optimistic news (Watts & Zimmerman (1986); Fields, Lys & Vincent (2001); Kothari et al. (2009)). These richer information sets and different alignments of interest are one motivation for our investigations of the market's heterogeneous response to textual content. Furthermore, our broad sample of firms, which are not all subject to the high information environment of the S&P 500 companies underlying the previous media studies, provides us with the cross-sectional variation necessary to examine whether and how certain information characteristics affect the market's response to the textual content of managerial press releases.

Closest to our study, Davis, Piger & Sedor (2012) also apply automated linguistic algorithms to the text of management-issued communications. They find that asset prices respond to textual net optimism and that textual net optimism helps to predict the future performance of the firm, results that we confirm with our data using a longer sample period, a broader set of automated textual language algorithms and controls.[3] Our study differs from Davis et al. (2012) in several

[3] Davis et al. (2012) use the Diction software to measure net optimism. We use three different automated textual language algorithms, Diction, General Inquirer, and the Loughran & McDonald (2011) dictionaries, and in order to mitigate measurement error, we use principal components analysis to extract the common factor from these three measures.

important respects. We consider the role of textual uncertainty in the price discovery process, finding that it has a direct incremental impact upon returns and that it modifies the market's response to GAAP earnings news.[4] We study how certain characteristics affect the role of managerial textual content in the price discovery process. In particular, using numerous alternative proxies for the expected informativeness of the firm's GAAP earnings and textual content for the firm's valuation fundamentals, we document that the market responds differentially to unexpected net optimism depending upon the information properties of the two signals, a finding that is consistent with a Bayesian learning model. This latter result also helps to fill a void in a related literature. Some theoretical and experimental studies have investigated the question of what affects the response of economic agents to hard-to-verify information, but Cao, Ghysels and Hatheway (2000) emphasize that there is a paucity of large sample empirical evidence that speaks to this issue.

II. Sample and Data Description

A. Samples and Data

We use the text of quarterly earnings announcements for the period of January 1998 through July 2006 collected by PR Newswire. We are able to match, using the ticker symbol and the announcement date (allowing for a 3-trading-day window discrepancy), 27,705 of the PR Newswire observations with the CRSP/Compustat database (4,771 different firms), and 18,673 of these announcements are further matched to First Call (3,433 different firms). Hereafter we refer to these two samples as the "Compustat" and "First Call" samples, respectively. We include only those observations for which we can calculate earnings surprises and 3-trading-day abnormal returns surrounding the earnings announcement. We also drop observations with stock prices below $1 and above $10,000 and firms with negative or missing book values of equity. We drop earnings announcement days that are within two weeks of a dividend payment announcement or a merger and acquisition announcement, and we drop announcements that contain fewer than 100 words. After imposing all of the preceding restrictions, we are left with a final sample of 3,683 firms (2,729 firms) and 20,899 firm-quarter (14,649 firm-quarter) observations for the Compustat (First Call) sample.

[4] In related work, Li (2006) builds a measure of textual uncertainty using a frequency count of only two words, "risk" and "uncertainty," and finds that his measure is associated with lower annual firm performance.

We tabulate and discuss in this paper the results of all of our tests using only the broader Compustat sample; however, we present the results from all of the same tests using the First Call sample in the Internet Appendix. We choose to focus on the Compustat sample for several reasons. First, the First Call constraint imposes a bias in favor of larger firms that are subject to richer information environments, whereas we are interested in understanding the role of managerial textual content for the broader universe of firms, not all of which are subject to high exposure and associated analyst filtering mechanisms. Second, Graham, Harvey & Rajgopal (2006) report that 85 percent of CFO survey respondents considered earnings in the same quarter of the prior year to be the most important earnings benchmark, followed by the analysts' consensus estimate at 73.5 percent. The interviewed CFOs further noted that the first item in their press release is often a comparison of the current quarter's earnings with four-quarters-lagged earnings. Accordingly, we expect that the prior year's same quarter actual earnings provides the framing context for management's earnings announcement. We provide First Call results in the Internet Appendix in order to establish that our findings are robust to alternative measures of earnings surprises and to a large-firm-biased sample.

We obtain market values, stock returns, and trading volume from the Center for Research in Security Prices (CRSP) databases, while historical accounting data come from Compustat. First Call is our source for both management and analyst forecasts.[5]

Corporate earnings announcements were provided by PR Newswire, with each firm-quarter's announcement being furnished as an individual text file. Prior to subjecting these files to the algorithmic processing described below, we took a number of steps. First, we identified tabulated figures in the text by searching for strings of numbers, and we cut these elements from the files so that these tables of figures did not confound the textual analysis.[6] Next, using mechanical search algorithms that we designed based upon extensive manual review of the announcements,

[5]Our results are qualitatively similar if instead we use the Institutional Brokers' Estimate System (I/B/E/S) analyst forecasts rather than First Call analyst forecasts, and if instead we supplement the First Call analyst forecasts with I/B/E/S data where firms are covered by the latter but not by the former database. Because only First Call provides corresponding management forecasts, we chose to report the results that rely exclusively on this database.

[6] The textual analysis algorithms typically count each numerical expression as a word. Thus leaving numerical tables in the files will confound the measurement of net optimism and textual uncertainty by exaggerating both the total number of words as well as the numerical term scores.

we separately removed the company description and "safe harbor" paragraphs so that only the earnings announcements themselves remained in the text files to be analyzed.[7]

B. Variables Measurement

In our primary tests we use a seasonal random walk model to generate earnings expectations for the Compustat sample.[8] We standardize unexpected earnings by dividing the surprise by the firm-specific standard deviation of the forecast error, and we label the standardized unexpected earnings associated with firm j for quarter q released on day t as SUE_{jqt}. To calculate this measure, we require each firm to have non-missing earnings data for at least 10 quarters. To prevent a hindsight bias, we estimate the standard deviation of the forecast error using a maximum of 20 quarters of the firm's previous unexpected earnings data, following Bernard & Thomas (1989) and Tetlock et al. (2008). We also allow for a trend in the seasonal random walk for all firms with more than four years of earnings data.

In order to mitigate a potential correlated omitted variable bias, we control in our regressions for other information voluntarily released by managers on the earnings announcement date, such as surprises in management forecasts and pro forma earnings. To avoid unduly constraining the number of observations, we only include one-period-ahead management forecasts, although all of our results are qualitatively similar when we include every available forecast horizon.[9] We include both annual and quarterly management earnings forecasts. Following the literature, we define management forecast surprises as $UMF_{jt} = F_{jt} - E_{jt}$, where F_{jt} is the one-period-ahead management forecast of either annual or quarterly earnings per share of firm j on announcement day t obtained from the First Call Company Issued Guidelines and Summary Statistics files, and E_{jt} is the corresponding median analyst forecast of annual or quarterly earnings per share of firm j preceding the management forecast on day t taken from the same database. We standardize

[7] The company description sections typically describe the entity in extremely positive terms, whereas the safe harbor provisions include many uncertainty-related expressions. Thus, their inclusion would have the effect of increasing the net optimism and textual uncertainty scores in an artificial manner in the sense that neither of these sections is directly related to the managerial earnings announcement news per se that we seek to analyze.

[8] In the Internet Appendix, we report the results from all of the same tests using the First Call analysts' median as the market's expectation of earnings. To address concerns about stale forecasts being included in the First Call summary files, similar to the issue raised by Diether, Malloy & Scherbina (2002) in the context of the I/B/E/S summary files, we used the First Call Detail History files and we discarded stale forecasts following the methodology described in Diether et al. (2002). Similar to Diether et al. (2002), our empirical results are unaffected by this discarding of stale forecasts.

[9] For our Compustat sample, 15% of the earnings press releases include one-period-ahead annual earnings forecasts, 10% include one-period-ahead quarterly earnings forecasts, and 6% include both. We only consider point and range forecasts because we can unambiguously compare these forecasts to analysts' expectations and earnings realizations.

unexpected management forecasts by the firm-specific standard deviation of the forecast surprise, and we label the standardized unexpected management annual and quarterly earnings forecast surprise associated with firm j on day t as $SUMFA_{jt}$ and $SUMFQ_{jt}$, respectively. We identify earnings announcement dates where managers release pro forma earnings or earnings that exclude nonrecurring items, by doing a keyword search, and we create an indicator variable equal to 1 for announcements that include pro forma earnings.[10] We interact this indicator variable with the standardized First Call earnings surprise to estimate standardized unexpected pro forma earnings, $SUProForma_{jqt}$.[11, 12]

Various methods have been used in the prior literature to measure the optimism and uncertainty conveyed in public communications by government institutions, the media, and corporate entities. To ensure the robustness of our results to alternative measures of optimism, we applied three of the leading algorithms, General Inquirer (GI), the Diction text-analysis program (version 6.0), and the Loughran and McDonald (L&M) dictionaries. In order to minimize the potential for measurement error or bias from any one particular algorithmic source, and also for the sake of parsimony, we ultimately calculate and tabulate results using the first factor derived from these three measures of textual net optimism using principal components analysis. We similarly use two different algorithms, Diction and the L&M dictionaries, to estimate uncertainty (GI does not offer a textual analysis construct that is analogous to uncertainty), and we adopt the first factor from these as our textual uncertainty variable.

In general, each of the text analysis algorithms uses a series of dictionaries (i.e., word lists) to search text passages for different semantic features. For example, Diction defines *optimism* as

[10] Following Lougee & Marquardt (2004), our keyword search is "pro forma earnings/net income/loss" or "adjusted earnings/net income/loss."

[11] Similar to I/B/E/S, First Call reports earnings and analysts' forecasts that are adjusted to "exclude any unusual items that a majority of the contributing analysts deem non-operating and/or non-recurring" (First Call Historical Database User Guide). For a discussion of differences between pro forma or "street" earnings versus GAAP earnings and a discussion of the different forecast data providers, we refer the reader to Abarbanell & Lehavy (2002). As previously mentioned, our results are robust to using I/B/E/S data in place of First Call data, and to supplementing First Call data with I/B/E/S data.

[12] Studies that use small samples of hand-collected data, such as Lougee & Marquardt (2004); Johnson & Schwartz Jr (2005); and Bhattacharya, Black, Christensen, & Mergenthaler (2004), suggest that our indicator variable for pro forma earnings may be noisy. However, the general pro forma counts and industry-specific counts in these studies are consistent with ours. Furthermore, our results are robust to replacing our keyword-based indicator variable with an indicator variable set equal to 1 if the firm is covered by First Call. In other words, our results are robust to using a very conservative definition of pro forma reporting, namely that all First Call firms report pro forma earnings. This robustness is not surprising given that the First Call sample results tabulated in the Internet Appendix, which use analyst forecast errors as the measure of earnings surprise, are consistent with those reported in our main tables.

"language endorsing some person, group, concept or event or highlighting their positive entailments" (Digitext Inc. (2000)), and the Diction formula for *optimism* is [praise + satisfaction + inspiration] − [blame + hardship + denial].[13] We interpret the first and second components of the optimism formula as "*optimism*" and "*pessimism*," respectively, and we refer to the difference between the two as "*net optimism*." We obtain analogous measures of textual net optimism by using *positivity* minus *negativity* from GI, and *Fin-Pos* minus *Fin-Neg* from version 2 of the L&M dictionaries.[14] The measures of optimism and pessimism (or their analogues, positivity and negativity) are stated as a percentage of the total words in the text, which we then multiply by 100 in order to arrive at variables that are bounded by 0 and 100. Net optimism, being the difference between optimism and pessimism (or positivity and negativity), is thus bounded by −100 and 100.

Diction defines *certainty* as "language indicating resoluteness, inflexibility, and completeness and a tendency to speak *ex cathedra*" (Digitext Inc. (2000)). We redefine Diction's formula to include numerical terms as additive to certainty rather than subtracting them from the score. In other words, we define *certainty* as [tenacity + leveling + collectives + insistence + numerical terms] − [ambivalence + self-reference + variety]. In the context of earnings announcements, we view the provision of more ex-post verifiable quantitative information to be indicative of more direct and precise expression rather than the use of more obtuse language. In specification checks we find that our results are qualitatively similar when we redefine *certainty* to exclude numerical terms ([tenacity + leveling + collectives + insistence] − [ambivalence + self-reference + variety]), and they are also robust, albeit weaker, when we adopt Diction's original definition of *certainty*.

In order to match our measures we multiply the Diction *certainty* measure by −100, and to obtain measures for *uncertainty* that are of comparable magnitudes to optimism and pessimism, we normalize the calculated variable by adding the absolute value of the lowest (i.e., negative) valued raw certainty score. Hence our Diction-based *uncertainty* measure is bounded by 0 and 100. We also use the *Uncertainty v2* dictionary from L&M to generate an alternative measure of uncertainty expressed in the text of the earnings press release. Specifically, we multiply the

[13] The terms associated with each of the characteristics that generate the Diction variables are reproduced in Davis et al. (2012) and are available in extended detail in Digitext Inc. (2000).
[14] All of the L&M dictionaries are available at www.nd.edu/~mcdonald/Word_Lists.html.

percentage of L&M uncertainty-related words in the text passage by 100, such that this measure is also bounded by 0 and 100.

Our regression results are robust to using text-based measures derived from any of the three textual analysis algorithms; however, we generally find that the L&M-based unexpected textual net optimism has higher levels of statistical association with returns than that derived from GI, while the GI-based measure is in turn more statistically significant than that derived from Diction, which is the measure used by Davis et al. (2012). By contrast, Diction-based measures of uncertainty are more significantly associated with traditional measures of earnings uncertainty than the L&M-based uncertainty measure. We adopt the factor scores extracted from the set of algorithms as our text-based test variables in all of the primary regressions reported below. We label the textual net optimism factor $TNetOpt_{jqt}$, and the textual uncertainty factor $TUncertainty_{jqt}$ for firm j related to quarter q earnings released on date t.

Our regression specifications related to the asset price response to information signals are motivated by a basic Bayesian updating model. According to this model, asset prices should respond to the *unexpected* component of future cash flow signals, and to the *level* of uncertainty regarding cash flows. Consistent with this, in related prior literature the *unexpected* component of GAAP earnings information, SUE, is interacted with the *level* of analyst forecast dispersion (and not the *change* in analyst forecast dispersion). Similarly, the *unexpected* component of manager's earnings forecasts, SUMFA and SUMFQ, are interacted with the level of management forecast precision (and not the change in the precision) (see, for example, Baginski, Conrad & Hassell (1993), and Imhoff Jr & Lobo (1992), among others). Since previous literature has shown that textual net optimism is related to the *level* of firm's future cash flows, and since we conjecture that textual uncertainty is related to the *uncertainty* of future cash flows, our asset price response regression specifications uses the *unexpected* component of textual net optimism and the *level* of textual uncertainty. Accordingly, in our asset price regression specification we use the level of textual uncertainty, but we modify textual net optimism to capture the unexpected component of the information signal. Similar to the standard specification for earnings information surprises defined above, we use time-series econometric models to estimate its unexpected component. Results presented in the Internet Appendix show that the level of textual net optimism contained in management's most recent prior quarter's announcement is,

out of all the models we considered, the best expectation for this quarter's textual net optimism, and accordingly we used a non-seasonally-adjusted random walk model to calculate unexpected textual net optimism, $\Delta TNetOpt_{jqt} = TNetOpt_{jqt} - TNetOpt_{jq-1t-1}$. We choose not to standardize this variable because we do not have a long enough history with which to accurately estimate the standard deviation of the surprise without incurring a hindsight bias.

Panel A of Table I provides descriptive statistics for unexpected textual net optimism, as well as announcement period abnormal returns and the various measures of textual uncertainty to be discussed below, while Panel B presents their correlation matrix. As shown, the pairwise correlations between the three alternative raw measures of *ΔTNetOpt* are modest, ranging from about 32 percent between Diction and GI to 53 percent for the GI and L&M measures. By construction, the factor is highly correlated with each of the three raw measures (i.e., approximately 75 percent to 83 percent).

Insert Table I here

As shown in Panel B of Table I, the two original uncertainty scores that are available from L&M and Diction, respectively, have a pairwise correlation of only about 16%, while each has a correlation of 76% with the factor. One reason for the low correlation between the raw measures is that the L&M proxy is derived from a relatively small underlying dictionary of just 285 words, and as a consequence there are many zero-valued observations (and thus a lack of cross-sectional variation) for this measure in our sample.

III. The Informativeness of Textual Content About Valuation Fundamentals

In rational markets, textual content should only impact asset prices if it helps to predict the level of the firm's future cash flows and/or the uncertainty of the cash flows. In this section, we document the relation between managerial textual content and various proxies for these valuation fundamentals. The evidence presented here enhances our understanding of the economic content of the textual variables – particularly *TUncertainty*, which has not been widely studied in capital markets contexts.

A. The Informativeness of Textual Net Optimism for Future Cash Flows

To investigate the information content of managerial text we adopt the following pooled regression model, which is similar to that of Tetlock et al. (2008)[15]:

$$\begin{aligned} FutureCashFlows &= \beta_{10} + \beta_{11}Text_{jqt} + \beta_{12}ForecastQ_{jt} + \beta_{13}ForecastA_{jt} \\ &+ \beta_{14}EarnProforma_{jqt} + \beta_{15}Earn_{jt} + \beta_{16}\text{Size}_{jt} + \beta_{17}CAR_{jt} \\ &+ \beta_{18}\log(MB)_{jt} + \beta_{19}\log(Turnover)_{jt} + \beta_{110}FirstCall_{jqt} \\ &+ \beta_{111}Proforma_{jqt} + \beta_{112}AnalystDisp_{jq+1t} + \beta_{113}TUncertainty_{jqt} \\ &+ \beta_{114}AdditionalControl_{jqt} + \varepsilon_{1jt}. \end{aligned} \quad (1)$$

The measures we use for the dependent variable, *FutureCashFlows*, are, alternatively, standardized realized future earnings, $SEARN_{jq+1t+1}$, standardized analyst forecast revisions of future earnings, $SAFR_{jq+1t}$, and standardized unexpected earning, $SUE_{jq+1t+1}$. $Text_{jqt}$, $ForecastQ_{jt}$, $ForecastA_{jt}$ and $EarnProforma_{jqt}$ are the textual net optimism of the earnings press release, the standardized quarterly and annual management forecast variables, and the pro forma earnings corresponding to the dependent variable. In other words, the variables are measured in levels for the future earnings levels regression (i.e., $TNetOpt_{jqt}$, $SMFQ_{jt}$, $SMFA_{jt}$, $SProForma_{jqt}$), and in surprises for the earnings surprise and forecast change regressions (i.e., $\Delta TNetOpt_{jqt}$, $SUMFQ_{jt}$ $SUMFA_{jt}$, and $SUProForma_{jqt}$). Similarly, $Earn_{jt}$ is a vector of regression-specific control variables. For the future earnings levels regression, the controls include $SEARN_{jqt}$, $SEARN_{jq-1t-1}$, $SEARN_{jq-2t-2}$, and $SEARN_{jq-3t-3}$, the contemporaneous and lagged standardized earnings for firm *j*. For the forecast change and unexpected earnings model, the relevant control variable is SUE_{jqt}, the current quarter's earnings surprise. In the unexpected earnings model, the $AdditionalControl_{jqt}$ variable is $SAFR_{jq+1t}$. These and all other variables are defined in detail in the Appendix.

The results from estimations of equation (1) are presented in Table II. The first column shows the usefulness of the textual content variables in predicting the future earnings of the firm. Textual net optimism is positively associated with future earnings, even after controlling for

[15] Our results for the language variables of interest are robust to alternatively adopting the specifications proposed by Fama & French (2000) for the level of, and change in, annual profitability of the firm, where the latter is defined as annual earnings before interest and extraordinary items but after taxes, all divided by the book value of total assets.

current and past earnings of the firm, voluntary managerial forecasts of quarterly and annual earnings, pro-forma earnings, and other variables. This result is generally consistent with the findings in Davis et al. (2012), although their specification omits textual uncertainty and many of the earnings and forecast related variables for which our model controls. The finding suggests that managers use optimistic language to signal their expectations regarding the firm's positive future prospects, perhaps because these unrealized "gains" (or future profits) are precluded from recognition in the current period's GAAP earnings. In other words, managers are cognizant of the conservative bias in GAAP that precludes the recognition of even highly likely anticipated future positive outcomes, and they may be using language to credibly compensate for this informationally deficient aspect of earnings.

Insert Table II here

Textual uncertainty, *TUncertainty*, is negatively associated with the future earnings of the firm even after controlling for the dispersion across analysts' forecasts, *AnalystDisp*, a measure of uncertainty widely used in the literature. The negative coefficients on both variables are consistent with the notion that bad news is characterized by greater uncertainty. The -0.059 economic impact of *TUncertainty* is slightly higher than that of *AnalystDisp,* which is -0.033 (= -0.256×0.13) after standardizing coefficients. Importantly, *TUncertainty* is available for a larger sample of firms, whereas *AnalystDisp* is only available for the predominantly larger firms followed by First Call. *TUncertainty* therefore offers a potentially useful earnings prediction signal for smaller, more "neglected" firms. The results suggest that managers use more direct and forthright language when more positive future earnings are expected, or alternatively stated that they use more wavering diction when anticipating more negative future earnings. The relevance of this finding to investors and other market participants is that higher managerial uncertainty and less confident assertions in press releases foretell more negative future performance.

The results in the second column of Table II using analyst forecast revisions (*SAFR*) as the dependent variable enable us to examine whether the text of the earnings press release is informative about changes in *expectations* about future earnings. For this test we necessarily restrict our sample to firms covered by First Call and whose analysts provide at least a two-quarter-ahead forecast (so that we can calculate the forecast revision). The significant

coefficients on each of *ΔTNetOpt* and *TUncertainty* suggest that analysts incrementally incorporate the information conveyed in the text of the announcement into their revisions.

Finally, the results from the SUE specification in the third column show that text provides information regarding the subsequent quarter's earnings surprise, as once again both *ΔTNetOpt* and *TUncertainty* are significant. In other words, one can improve the naïve earnings forecast of last year's same quarter earnings by including textual content variables. The economic impact of *ΔTNetOpt* is 0.043, for example, which is about two-thirds the magnitude of the 0.06 impact of CAR ($=0.005 \times 12$ after standardizing coefficients), while the -0.034 economic impact of *TUncertainty* is of slightly larger magnitude than the -0.029 impact of Analyst Dispersion ($=-0.221 \times 0.13$ after standardizing coefficients). Our findings in relation to unexpected managerial net optimism are broadly consistent with those of Tetlock et al. (2008) in the context of media-sourced announcements. Those prior authors document that, after standardizing coefficients, the economic impact of *media* tone in firm-specific news stories over the 28 trading days ($t = -30$ to $t = -3$) prior to the earnings announcement date ($t = 0$) is 0.063. The findings from both studies clearly suggest that textual content contains incremental information that helps to predict the future earnings of the firm. Our findings complement those of Tetlock et al. (2008) by lending indirect support to their conjecture that the significance of the news stories that they study derives from media tone capturing elements of the firm's prior earnings announcement. However, it is particularly interesting to note the economic significance and predictive power of the textual content variables in our managerial announcement setting given the relative staleness of the news as compared to the media stories. Furthermore, we show in Table A1 in the Internet Appendix that this finding is robust to using the First Call sample, which suggests that analysts do not fully respond to available textual content regarding the firm's earnings expectations.

Overall, the findings across all three alternative measures and proxies for future cash flows strongly support the conjecture that textual net optimism and textual uncertainty are incrementally informative signals about this valuation fundamental.

B. The Information Content of Textual Uncertainty for Economic Uncertainty

The informational properties of textual uncertainty in financial contexts are not well understood, so we explore this variable further by examining its second moment information content. Specifically, we investigate the relation between *TUncertainty* and a frequently used measure of the perceived uncertainty of the firm's future cash flows, the dispersion in analyst forecasts (e.g., Bamber, Barron & Stober (1997), and Zhang (2006)), by regressing *changes* in analyst forecast dispersion on *changes* in *TUncertainty* and other variables that have been shown to affect changes in analyst forecast dispersion. We adopt this changes on changes specification because both variables in levels are likely to be persistent and hence any documented relations between them could be spurious. Regressing changes on changes eliminates the possibility that our economic inferences are based upon such spurious associations.

The positive association between $\Delta TUncertainty$ and the change in analyst forecast dispersion documented in Table III is consistent with *TUncertainty* capturing an element of the firm's future cash flow uncertainty, a notion that we investigate further below. In supporting analyses reported in the Internet Appendix, we also document that *TUncertainty* is positively associated with earnings volatility and the inverse of manager's forecast precision, and that it is negatively associated with the R^2 of a seasonal random walk model of earnings per share, all consistent with the notion that *TUncertainty* captures aspects of the second moment of earnings.[16] Given this documented informativeness of *TUncertainty* for the firm's economic uncertainties, in Section IV.C we study whether the market's response to each of the GAAP earnings news and textual net optimism news is modified by this dimension of textual content.

Insert Table III here

C. Baseline Market Response Regressions

In this section, we present additional results to support our conclusion that managerial textual content contains value relevant news. Specifically, Table IV shows that firm equity returns

[16] In the analyses reported in the Internet Appendix we also find that Diction's textual uncertainty measure is more highly associated than that of L&M with these various measures of earnings uncertainty. We interpret the weaker associations of the L&M variable to be due to its relatively small underlying dictionary, and the associated lack of cross-sectional variation (i.e., many zero-valued observations) and lower statistical power that this entails.

respond to Δ*TNetOpt* and *TUncertainty* (incremental to earnings, forecasts and other standard controls) during the 3-day quarterly earnings announcement window.[17]

Insert Table IV here

The results indicate that $\Delta TNetOpt$ is statistically significant, incremental to the earnings, management forecast, and pro forma earnings surprises. Consistent with the prior literature that documents an attenuated response to announcements for larger firms, which tend to have richer pre-announcement information environments, we also find that the responses to our news variables are attenuated for large firms as all of the *Size*-interacted terms have negative coefficients and all but the interaction with management annual earnings forecast surprises, *SUMFA*, are significant. The effect of $\Delta TNetOpt$ on abnormal returns for an average size firm is 0.749 (=4.591-0.194×19.8). The signs and significance of the coefficients on our textual content variables are generally consistent with the results of prior studies such as Tetlock et al. (2008) and Engelberg (2008) in the context of news stories, and Davis et al. (2012) in the context of management press releases. In terms of the magnitudes of the coefficient, the 0.749 effect of $\Delta TNetOpt$ on abnormal returns documented here is higher than the 0.427 (=0.0075×0.57×100) obtained using estimates reported by Davis et al. (2012), consistent with the view that our factor model reduces the measurement error in the linguistic content variable.[18]

While the findings for Δ*TNetOpt* in Table IV are broadly consistent with those in the extant literature, those for *TUncertainty* are new to our study and thus we elaborate further on possible interpretations of this textual content measure. The evidence in the previous section suggested that *TUncertainty* was capturing some aspect of the the firm's economic risk. In Table IV we see that the coefficient on the main effect of *TUncertainty* is negative and significant. The -0.109

[17] In related work using the same dataset as the current study, Demers, Francis & Vega (2013) document that the 3-day announcement period response shown here does not revert. Rather, there is a statistically and economically significant drift associated with *ΔTNetOpt* during the 60-day post-announcement period, which provides further evidence that this textual content signal contains value-relevant news rather than mere noise.

[18] Davis et al. (2012) do not report the standard deviation of their *NetOpt-LAGNETOPT* measure, so we use the reported interquartile range of 0.57 as a proxy for this in arriving at our estimate of the economic impact of textual content on returns for their study. Although our sample period is more recent than theirs, this does not explain the greater economic impact of *ΔTNetOpt* in our setting because we obtain a similar estimate to that of Davis et al. (2012) when we use their Diction-based variable with our data and regression specification. Specifically, the economic impact of *ΔTNetOpt* for our sample and specification using L&M, GI, and Diction, respectively, to capture textual content surprise is 0.73, 0.56, and 0.41, consistent with the latter proxies being subject to greater measurement error.

impact of *TUncertainty* is of slightly smaller magnitude than the -0.175 impact of Analyst Dispersion (= -1.351×0.13). The negative sign of the coefficients is consistent with our results in Table II as well as the findings of prior studies that measures of uncertainty are associated with lower future cash flows. The coefficient on the interaction term of $SUE \times TUncertainty$ is negative and significant, suggesting that *TUncertainty* captures elements of the noise or uncertainty in the GAAP earnings news variable. The coefficient on the interaction term of $\Delta TNetOpt \times TUncertainty$ is positive but statistically insignificant, however, it is positive and significant in the First Call sample (Table A2 in the Internet Appendix). Overall, the results are consistent with the view that *TUncertainty* may be related to the firm's future cash flow uncertainty. If this is indeed the case, then a Bayesian learning model would suggest that *TUncertainty* will modify the predictive content of current GAAP earnings and textual net optimism for future earnings. We explore this possibility below.

IV. Heterogeneous Informativeness of Textual Net Optimism

The results in the previous section establish that textual content variables help to predict the firm's future cash flows and the uncertainties of those cash flows, and that the market responds to these variables during the 3-day earnings announcement window. But is the textual content of different firms viewed similarly by the market? Certain conditions or firm characteristics may lead this textual content to be more informative about future earnings. But does the market weight managerial textual content accordingly? We investigate this question using the framework of a standard Bayesian learning model, wherein the market impact of one signal should be increasing in its own informativeness and in the noisiness of the second simultaneously released signal. In the current section, we appeal to prior empirical and theoretical research to select and validate proxies for the expected informativeness of textual net optimism and GAAP earnings signals, and then we investigate the heterogeneous market response to text across these sample partitions.

A. Informativeness of Textual Net Optimism

Textual content is both subject to managerial discretion and characterized by several features that typify "cheap talk" – it is nonbinding, costless to convey, and difficult-to-verify. Drawing

upon cheap talk theory and the voluntary disclosure literature, we choose empirical proxies to capture settings where textual net optimism is likely to be more informative about the firm's future cash flows. We consider a signal to be more informative if it is either more credible or more value relevant (or both), and we identify three firm or announcement characteristics that are expected be associated with heightened informativeness of managerial textual content: 1) the amount of numerical support provided with the text; 2) the extent of analyst coverage of the firm; and 3) the firm's past forecasting reputation. We motivate the rationale for these constructs and validate our multiple empirical proxies for each in turn.

Cheap talk theory (e.g., Benabou & Laroque (1992)) establishes the notion that verifiability is a key mechanism for inducing truthful revelation. Three variables in our setting capture managers' use of numerical data, information that increases the specificity of the disclosures in the announcement so that they are more readily subject to ex-post verification: the use of more numerical terms, the inclusion of pro forma earnings, and the inclusion of an earnings forecast. Furthermore, managers are more likely to provide (and market participants are more likely to appreciate) the voluntary disclosure of pro forma earnings and earnings forecasts when GAAP earnings are informationally deficient (Lev (2012)). For both of these reasons, we expect textual net optimism to be more informative for firms that provide more numerical support.

As for our second construct, we expect analyst coverage to influence the informativeness of text for at least three reasons. First, the presence of analysts following the firm is consistent with a setting in which "multiple experts" (i.e., together with the firm's management) are transmitting information, where the repeated announcements, conference calls, and other interactions between managers and analysts comprise an extended "dialogue." In such a repeated game, higher levels of analyst scrutiny will increase the likelihood that misleading textual content will be detected and subsequently punished, ultimately making the text of more heavily followed firms generally more informative. This reasoning is supported, for example, by Krishna & Morgan (2001) who model a perfectly informed expert (i.e., manager of the firm) who offers advice to a decision maker (i.e., an investor), under the assumption of misaligned preferences. They argue that the presence of multiple experts providing information, each with different biases, will help to

induce more informative disclosure.[19] Second, Matsumoto et al. (2011) show that the information content of the question and answer portion of earnings release conference calls increases with analyst coverage. Their findings suggest that analyst coverage produces incentives for managers to release more useful information. Thirdly, Barth, Kasznik & McNichols (2001) show that analyst coverage is increasing in the firm's intangible assets. These firms tend to be more complex and their earnings are less informative (Lev (2012)), so there is greater opportunity for analysts and/or managerial textual content to contribute information to the market about the firm's prospects. Thus, either because of the information revelation that they induce, or because their presence is known to be associated with less informative earnings, higher analyst coverage is expected to be associated with more informative net optimism.

Our third construct relates to prior reporting reputation. We conjecture that firms that have a history of providing more value relevant information are more likely to provide informative textual content in the future either because the firm cares more about its reputation or because the firm has other reasons (unobserved by econometricians) to provide informative textual content that are persistent over time. This reasoning is supported by Sobel (1985), who establishes that only in a repeated game will the sender truthfully reveal information, doing so both because he cares about his reputation and because the information is ex-post verifiable. Stocken (2000)'s model of managerial incentives to disclose private information in a multi-period setting similarly suggests that credible communication is less probable for managers with a poor reputation. We use two alternative measures of firm-level communication reputation. First, we use the earnings forecasting reputation measure proposed by Hutton & Stocken (2009), under the implicit assumption that past quantitative forecasting reputation is positively correlated with the informativeness of the current period's textual net optimism.[20] In other words, we expect informativeness in one sphere of communications (i.e., quantitative forecasting) to be associated with informativeness in the other (i.e., textual content). Second, we develop an alternative measure of informativeness that is designed to capture the incremental future earnings-related information conveyed by the firm's past textual content. Specifically, we regress one-quarter-

[19] The assumption that analysts are informed experts is not necessary, however, as Krishna & Morgan (2004) show that repeated two-way communication between uninformed agents and one informed agent (i.e., the firm or manager in our setting) will also improve the informational content of communication.

[20] Hutton and Stocken's (2009) managerial forecast accuracy measure is relative to the median analyst's estimate. Their performance measure therefore parsimoniously controls for many factors that affect forecast accuracy across firms, such as firm complexity, forecast horizon, and industry earnings volatility.

ahead standardized earnings on a set of non-text-based variables that have been shown to explain future earnings, and we extract the firm-specific residual from this regression. We then augment this regression with textual uncertainty and textual net optimism, and we extract the residual from the expanded regression.[21] The difference between the absolute values of the regression residuals from the two specifications is a measure of the improvement in explanatory power for future earnings provided by past textual content and serves an alternative proxy for reporting reputation.

We begin by presenting evidence in Table V of the differential informativeness of textual net optimism for future earnings across firms with high and low levels of each of the six previously described proxies. The specific regression that we run is equation (1), where the level of next quarter's earnings,$SEARN_{jq+1t+1}$, is used as the dependent variable and the independent variables are current textual net optimism, *TNetOpt*, and current and past realized earnings, *SEARN*, and other controls. In this specification we allow the effect of textual net optimism and current and past earnings to be different across the sample partitions described above: a pro forma versus non-pro forma indicator variable and a manager's forecast versus non-forecast indicator; and three dummy variables set equal to 1 if the numerical terms, analyst coverage, earnings forecast reputation or textual content reputation, respectively, belong to the bottom, middle, or top third of each variable's distribution. Operationally, we define noisy *TNetOpt* (GAAP earnings) firms to be those firms for which past *TNetOpt* (GAAP earnings) predict future earnings less well.[22]

We find in Panel A of Table V that the differences in the coefficients on *TNetOpt* between high and low values of the informativeness proxies is positive and significant in all cases. Firms that provide more numerical terms, pro forma earnings, or managerial forecasts to accompany the announcement of their realized GAAP earnings also have text that is more informative for future earnings. This is consistent with the notion that managers of these firms are using multiple

[21] The specific regression that we run is Equation (1), where $SEARN_{jq+1t+1}$ is used as the dependent variable. In order to avoid a hindsight bias, we only use information available at time *t* to estimate the firm's time-varying measure of text-based communication reputation.

[22] We note that this operational definition maps into a standard Bayesian learning model, wherein the quality of the information is determined by the informativeness of the signal itself (i.e., textual net optimism, TNetOpt, and standardized earnings, SEARN, in our context) rather than by the informativeness of the surprise (i.e., ΔTNetOpt or SUE). Accordingly, our dependent variable in equation (1) is future *SEARN* and our independent variables are current *TNetOpt* and current and past *SEARN*.

mechanisms to try to communicate in a more informative manner about their firm's prospects. Similarly, the net optimism of firms with better past textual content and forecasting reputations is also more informative, as is the text of firms that are more heavily followed by analysts. Our results also indicate that firms with better textual content reputation, as well as firms that provide pro forma earnings and earnings forecasts all have less informative GAAP earnings, which suggests the possibility that managers of these firms may be knowingly supplementing their informationally deficient GAAP earnings with more value-relevant textual net optimism.

Insert Table V here

Although the preceding evidence clearly supports the construct validity of our measures, we acknowledge that firm characteristics that are indicative of noisier textual net optimism may also be empirically correlated with noisier prior information (i.e., more uncertain business fundamentals). Zhang (2006) makes a similar point, and consistent with him we do not attempt to empirically distinguish between the two types of noisiness here. Most importantly, according to a standard Bayesian learning model, the extent to which our empirical proxies capture the uncertainty in investors' prior information, rather than the informativeness of net optimism that we target, will bias against finding statistically significant results for textual net optimism interacted with each of these proxies in the forthcoming market returns regressions.

B. Informativeness of GAAP Earnings

The noisiness of the earnings signal is also expected to affect the influence of textual net optimism on market prices. To examine this, we rely upon a well-established empirical literature to identify a set of proxies for "noisy" GAAP earnings, to which we also add our textual uncertainty variable. The proxies are: a high-tech versus non-tech indicator, and an R&D versus non-R&D indicator (Lev (2012)), and three dummy variables set equal to 1 if the P/E ratios (Lev & Zarowin (1999))[23], EFKOS e-loadings (Ecker et al. (2006))[24], or textual uncertainty of firm *j* on day *t* belong to the bottom, middle, or top third of each variable's distribution.

[23] Lev & Zarowin (1999), amongst others, suggest that firms that are heavily laden with intangible assets, such as those that are R&D-intensive or otherwise classified as "high-tech" or "growth" type firms have noisier historical earnings from a forward-looking valuation perspective.

[24] Ecker et al. (2006) provide a returns-based measure of earnings quality, termed an e-loading, which is estimated from firm-specific asset-pricing regressions augmented by an earnings quality mimicking factor. They present

Panel B of Table V shows that firms in each of these high earnings noisiness categories have realized earnings that are less informative about future earnings. The difference between the coefficients for high and low noisiness firms is highly significant in all five cases. Although this evidence clearly supports the construct validity of our measures, we acknowledge as we did for the noisy *TNetOpt* sample partitions above that firm characteristics that are indicative of noisier GAAP earnings may also be empirically correlated with noisier prior information. In this case, the distinction does not affect the results of our market pricing analyses that follow because noisiness in either (or both of) earnings and prior information lead to the same predictions: textual net optimism is expected to be more heavily weighted by the market.

We also note one further set of interesting results in Panel B of Table V: for the majority of our proxies, the informativeness of *TNetOpt* for future earnings does not significantly vary with the noisiness of GAAP earnings. Four out of five of the differences in the coefficients on *TNetOpt* between high and low earnings noisiness firms are insignificant, with the exception being for textual uncertainty. This result is relevant to our upcoming tests. If noisy GAAP earnings firms also had noisier textual net optimism, this would bias against our finding statistically significant differences in the pricing of net optimism across earnings noisiness partitions in the forthcoming returns regressions. This is not the case, however, allowing for sharper returns tests below.[25]

C. Hetereogeneous Market Response to Textual Net Optimism

The previous results establish that our proxies capture the informativeness of, respectively, earnings and textual net optimism for valuation fundamentals. But does the market discern this and respond heterogeneously to textual net optimism according to the informativeness of the two signals? We investigate this question by running announcement period returns regressions allowing for separate coefficients on unexpected textual net optimism for high and low values, in the case of dummy variables, or high, medium, and low values, in the case of continuous

---

empirical evidence to support the notion that firms with higher e-loadings are perceived by investors as having noisier earnings signals.

[25] Given the likelihood of endogeneity between the noisiness of earnings and the informativeness of textual content, the findings reported in Panel B of Table V may be somewhat surprising. One possible reason for this result is that the managers of firms having noisy earnings are making deliberate efforts to increase the informativeness of their text such that they are, on average, issuing text that is *as* informative as that of other firms, despite their working in more uncertain environments.

variables, of each of the established proxies, after controlling for other simultaneously released news items, size, and the main effects of the uncertainty parameters as in the following specification:[26]

$$
\begin{aligned}
CAR_{jt} = & \sum_{z=1}^{Z} \beta_{21\_z} SUE_{jqt} \times X_{zjt} + \sum_{z=1}^{Z} \beta_{22\_z} SUMFQ_{jt} \times X_{zjt} + \sum_{z=1}^{Z} \beta_{23\_z} SUMFA_{jt} \times X_{zjt} \\
& + \sum_{z=1}^{Z} \beta_{24\_z} SUProForma_{jqt} \times X_{zjt} + \sum_{z=1}^{Z} \beta_{25\_z} \Delta TNetOpt_{jqt} \times X_{zjt} \\
& + \beta_{26} AnalystDisp_{jqt} + \beta_{27} RangeMFQ_{jt} + \beta_{28} RangeMFA_{jt} + \beta_{29} TUncertainty_{jqt} \\
& + \beta_{210} SUE_{jqt} \times Size_{jt} + \beta_{211} SUMFQ_{jt} \times Size_{jt} + \beta_{212} SUMFA_{jt} \\
& \times Size_{jt} + \beta_{213} SUProForma_{jqt} \times Size_{jt} + \beta_{214} \Delta TNetOpt_{jqt} \times Size_{jt} + \beta_{215} Size_{jt} \\
& + \sum_{z=1}^{Z} \beta_{216\_z} X_{zjt} + \beta_{217} FirstCall_{jqt} + \beta_{218} ProForma_{jqt} + \varepsilon_{2jt},
\end{aligned}
\tag{2}
$$

where X_{zjt} are alternative proxies for the informativeness of textual net optimism and all other variables are as previously defined. Our tabulated results for the estimation of equation (2) present the total effect estimates of earnings surprises ($\beta_{21\_z} + \beta_{210}\overline{Size_{jt}}$) and unexpected textual net optimism ($\beta_{25\_z} + \beta_{214}\overline{Size_{jt}}$), each calculated using the within-group mean of size.

As shown in Panel A of Table VI, the market responds differentially to the unexpected net optimism of high versus low valued firms across all proxies for the ex-ante informativeness of textual content.[27] Specifically, the price impact of unexpected net optimism is larger for firms and announcements with higher textual content informativeness, consistent with the predictions of a standard Bayesian learning model. For example, the textual net optimism of firms that provide high levels of numerical terms affects asset prices by 27 basis points more than the textual net optimism of firms that provide low levels of numerical support.[28] Similarly, the

[26] Because of the high level of correlation among some of the explanatory variables, we consider the impact of each variable separately. Results from a single regression that includes all of the proxies that are not related to similar underlying constructs (and thus for which there is no strong theoretical correlation) yield consistent findings.

[27] For the sake of parsimony we present only the total effect in Table VI. In the Internet Appendix Tables A7 and A8 we present disaggregated results, allowing the reader to separately observe the effects of Size and the informativeness proxy on each of SUE and ΔTNetOpt. Importantly, the effect of the proxy is always statistically significant and in the same direction as the reported total effect.

[28] The numerical terms variable that we use here is also one component of the *TUncertainty* measure, which creates the potential for multicollinearity to affect our coefficient estimates. Notwithstanding this potential, all multicollinearity diagnostics are well below traditional thresholds of concern. We have also rerun the *Numerical Terms* regression reported in Table VI after excluding the *TUncertainty* main effect. Our results are robust, and indeed the difference between the high and low *Numerical Terms* announcements becomes more pronounced (economically and statistically). Because *TUncertainty* is more multifaceted than just *Numerical Terms*, and given that it should be controlled for in the other (i.e., non–*Numerical Terms*) equation (2) regressions reported in Table VI, we continue to control for the *TUncertainty* main effect in the *Numerical Terms* regression so as to present parallel results across our different proxies for textual informativeness.

inclusion of pro forma and management forecasts with the announcement leads the market to weight more heavily the text of those announcements. This could either be because managers' use numerical terms to enhance the informativeness of their textual net optimism and/or because managers are more likely to provide both additional numerical disclosures and more informative text when they realize that their GAAP earnings are informationally deficient. The latter explanation is also consistent with our finding in Table V that firms providing pro formas and/or earnings forecasts tend to have informationally deficient GAAP earnings. In each of the first three returns regressions shown in Table VI, the contemporaneously-released GAAP earnings surprise of firms that provide some form of numerical support in their text has a significantly smaller effect on abnormal returns. In other words, market participants appear to rely less on GAAP earnings news in the pricing of these firms both because the market has access to an informative alternative news source and because GAAP earnings are a noisy signal about the future performance of the firm.

Insert Table VI here

The market's response to textual net optimism is also increasing in the amount of analyst coverage, consistent with the presence of multiple experts inducing more informative disclosures,[29] and the market responds more to the text of firms that have better reputations for informative disclosures. The latter result holds regardless of whether reputation is captured by past quantitative forecasting accuracy or the information contained in past textual content.

We next investigate whether the price impact of textual content information is affected by the informativeness of the contemporaneously released GAAP earnings signal. Panel B of Table VI presents the results of re-estimating equation (3) alternatively using each of the previously described proxies for earnings noisiness for the X_{zjt} interactive terms. The results show that the market weights more heavily the text of firms with noisy GAAP earnings, consistent with the tenets of a standard Bayesian learning model.

Specifically, the *ΔTNetOpt* of high-tech firms affects asset prices by 24 basis points more than that of non-tech firms, and the text of R&D firms affects asset prices by 29 basis points more than the language of non-R&D firms. Similarly, the market's weighting of textual net

[29] In specification checks using media coverage and stock turnover as separate alternative proxies for the presence of multiple experts, we find that the differences in market response for low versus high media coverage and stock turnover firms are also reliably negative ($p = 0.01$ and $p = 0.05$, respectively).

optimism is monotonically increasing in firms' EFKOS e-loading factors and P/E multiples across the low, medium, and high groups, and the difference between the high and low groups is significant. The P/E finding suggests that when more of the firm's value resides in future growth prospects (i.e., when P/E ratios are higher), the text accompanying the announcement becomes a more important source of information regarding expected future cash flows. The results in the last column of Table VI confirm the finding in Table IV that *TUncertainty* modifies the pricing of SUE but not that of *ΔTNetOpt.* However, in the First Call sample *TUncertainty* modifies the pricing of both SUE and *ΔTNetOpt*.

Taken together, the results in Table VI provide robust evidence across numerous proxies that the market discerns not only the noisiness of GAAP earnings, but also the informativeness of difficult-to-verify textual content about the firm's valuation fundamentals. Our large-sample evidence indicates that the market weights firms' unexpected net optimism more heavily in the price formation process in settings where the text is more informative and where GAAP earnings are noisier indicators about fundamentals. These findings suggest that managers should expect the market to rely more heavily on textual cues when their firm's earnings are noisier, and furthermore that managers can increase the potential informativeness of this more difficult-to-verify aspect of communications by subjecting them to greater scrutiny, by corroborating them with numerical references, and by developing a reputation for credible communications.

V. Summary and Conclusion

There are significant limitations associated with GAAP earnings and related quantitative information (e.g., forecasts and pro forma earnings) that prevent managers from conveying their superior insights into their firm's prospects. Furthermore, prior literature has documented an increase over time in the importance of non-earnings information for explaining share prices. This suggests a potentially important informational role for managerial textual content in the price discovery process. The current study investigates whether and how such difficult-to-verify information is incorporated into asset prices.

Using a sample of more than 20,000 quarterly earnings announcements from the 1998 to 2006 period, we show that both textual net optimism and textual uncertainty are informative about the firm's valuation fundamentals, after controlling for other, standard, sources of

information, and that the market responds to this information during a 3-day earnings event window. Our findings support the view that the text of managerial press releases conveys "first-moment" and "second-moment" information that is excluded from the GAAP earnings summary figure, managers' earnings forecasts and pro-forma earnings news, and that the market relies upon this information in the price discovery process.

In addition, we choose and validate numerous alternative proxies for the expected informativeness of textual net optimism and GAAP earnings. Using these proxies, we document that the price response to unexpected textual net optimism varies in the manner that a basic Bayesian learning model would suggest. Specifically, we find that the market responds more to unexpected net optimism when it is more informative about valuation fundamentals, and when the other simultaneously released signal, historical GAAP earnings, is a noisier predictor of future earnings. We also find that for some firms, the informativeness of textual net optimism appears to be endogenously related to the informativeness of GAAP earnings. In particular, some firms seem to provide more informative text to complement what they understand to be informationally deficient GAAP earnings.

Appendix
Summary of Variable Definitions

Variable	Definition
Optimism	Percentage of words in the quarterly earnings announcements that are optimism increasing. Estimated using Diction 6.0 software, the General Inquirer word list, the Loughran and McDonald (2011) dictionaries, and a factor model of all three measures.
Pessimism	Percentage of words in the quarterly earnings announcements that are optimism decreasing. Estimated using Diction 6.0 software, the General Inquirer word list, the Loughran and McDonald (2011) dictionaries, and a factor model of all three measures.
$TNetOpt_{jqt}$	Optimism minus pessimism.
$\Delta TNetOpt_{jqt}$	$TNetOpt_{jqt} - TNetOpt_{jq-1t-1}$, where q indicates the fiscal quarter and t is the earnings announcement date for firm j.
$TUncertainty_{jqt}$	Certainty is a textual variable that indicates the degree of "resoluteness," "inflexibility," and "completeness" in the firm's quarterly earnings announcement. We redefine the Diction 6.0 definition of certainty to be [Tenacity + Leveling + Collectives + Insistence + Numerical Terms] − [Ambivalence + Self Reference + Variety] and use the negative of this measure to capture uncertainty. We also use the uncertainty measure estimated using the Loughran and McDonald (2011) dictionary, and a factor model of both measures.
SUE_{jqt}	Earnings surprise = $\frac{actual - forecast}{std(actual - forecast)}$ Our proxy for the market's expectation of earnings is last year's same-quarter earnings per share for the Compustat sample and the First Call median analyst forecast for the First Call sample. We use earnings per share (diluted) excluding extraordinary items (quarterly Compustat item data 9). We standardize the unexpected earnings by dividing the surprise by the firm-specific standard deviation of the forecast error. To calculate this measure, we require each firm to have nonmissing earnings data for 10 prior quarters. To prevent a hindsight bias, we estimate the standard deviation of the forecast error using a maximum of 20 quarters of the firm's previous unexpected earnings data following Bernard and Thomas (1989) and Tetlock et al. (2008). We also allow for a trend in the seasonal random walk used to calculate unexpected earnings for all firms with more than four years of earnings data.
$SMFQ_{jt}$, $SMFA_{jt}$	The management earnings forecast, MF_{jt} is the one-period-ahead management forecast of either annual or quarterly earnings per share of firm j on the earnings announcement day t from the First Call Company Issued Guidelines and Summary Statistics files. We standardize management earnings forecasts by the firm-specific standard deviation of the forecasts, and we label the standardized annual and quarterly earnings forecast associated with firm j at time t as $SMFA_{jt}$ and $SMFQ_{jt}$, respectively.
$SUMFQ_{jt}$, $SUMFA_{jt}$	We define management earnings forecast surprises as $UMF_{jt} = F_{jt} - E_{jt}$, where F_{jt} is the one-period-ahead management forecast of either annual or quarterly earnings per share of firm j on the earnings announcement day t from

	the First Call Company Issued Guidelines and Summary Statistics files, and E_{jt} is the corresponding median analyst forecast of annual or quarterly earnings per share of firm *j* preceding the management forecast on day *t*. We standardize unexpected management earnings forecasts by the firm-specific standard deviation of the forecast surprises, and we label the standardized annual and quarterly earnings forecast surprise associated with firm *j* at time *t* as $SUMFA_{jt}$ and $SUMFQ_{jt}$, respectively.
$ProForma_{jqt}$	Indicator variable created by identifying earnings announcement dates where managers release pro forma earnings, or earnings that exclude nonrecurring items, using a keyword search. Following Lougee and Marquardt (2004), our keywords are "pro forma earnings/net income/loss" or "adjusted earnings/net income/loss."
$SUProForma_{jqt}$	Interaction of the *ProForma* indicator variable with the standardized First Call earnings surprise.
$FirstCall_{jqt}$	Indicator variable set equal to 1 if the firm is covered by First Call.
$SEARN_{jq+1t+1}$	The earnings of firm *j* for quarter $q + 1$ released on date $t + 1$ divided by its standard deviation in the previous 20 quarters.
$SAFR_{jq+1t}$	The standardized analyst forecast revision of quarter $q + 1$ earnings on day *t,* defined as the First Call median analyst forecast for quarter $q + 1$ seven days after the earnings announcement of quarter *q* released on day *t* minus the median analyst forecast for quarter $q + 1$ earnings that prevailed one day prior to the earnings announcement of quarter *q* released on day *t*, all divided by its standard deviation (calculated using a maximum of 20 quarters of the firm's previous earnings forecast revisions).
$AnalystDisp_{jqt}$	Analyst forecast dispersion. The standard deviation across First Call analysts' forecasts divided by the absolute value of the median forecast. We require firms to have at least two forecast estimates, and we discard stale forecasts similar to the approach of Diether et al. (2002). We winsorize this variable by replacing values above the 99th percentile with the 99th percentile value and values below the 1st percentile with the 1st percentile value.
CAR_{jt}	Size- and book-to-market-adjusted cumulative abnormal returns defined over the earnings announcement window $[t - 1, t + 1]$ relative to the $t = 0$ earnings announcement day.
Size=Log(Market Cap.)	The natural logarithm of share price (item 199 in the annual Compustat database) as of the end of the previous fiscal year times the shares outstanding (item 25 in the annual Compustat database) at the end of the previous fiscal year. We winsorize this variable by replacing values above the 99th percentile with the 99th percentile value and values below the 1st percentile with the 1st percentile value.
Log(MB Ratio)	The natural logarithm of the market capitalization of the firm at the end of the previous fiscal year divided by the book equity of the firm at the end of the previous fiscal year. Following Pastor and Veronesi (2003), book equity is constructed as stockholders' equity plus balance sheet deferred taxes and investment tax credit (annual Compustat item 35) minus the book value of preferred stock. Depending on availability, stockholder's equity is computed as annual Compustat item 216, or 60 + 130, or 6 − 181, in that order, and

	preferred stock is computed as item 56, or 10, or 130, in that order. We winsorize this variable by replacing values above the 99th percentile with the 99th percentile value and values below the 1st percentile with the 1st percentile value.
$\log(Turnover)_{jt}$	The average of the natural log of de-trended turnover (i.e., the daily volume of shares traded divided by stock outstanding) cumulated over the pre-announcement period [$t-62$, $t-2$]. In order to present a pooled regression of NYSE/AMEX and Nasdaq firms, we follow the common heuristic of dividing the Nasdaq firms' volume by two (Atkins and Dyl (1997) and Dyl and Anderson (2005)). We de-trend turnover using the Campbell, Grossman and Wang (1993) method of calculating the turnover's trend as the rolling average of the prior 60 trading days, adding back the mean of turnover to our de-trended measure so that the units are economically meaningful. We winsorize this variable by replacing values above the 99th percentile with the 99th percentile value and values below the 1st percentile with the 1st percentile value.
RangeMFQ, RangeMFA	The ranges of the quarterly and annual management forecasts divided by the absolute value of the median forecast.
Numerical Terms	The simple count of the number of numerical terms in the announcement, estimated using Diction 6.0, divided by the total number of words excluding financial statements and other tables in the press release.
Managerial Forecast	Indicator variable equal to one if a managerial forecast accompanies the GAAP earnings release.
Analyst Coverage	The natural logarithm of 1 plus the number of First Call analysts posting an earnings estimate for the firm's current quarter.
Earnings Forecast Reputation	The managerial forecasting reputation measure proposed by Hutton and Stocken (2009), which is calculated as the number of "relatively accurate" forecasts divided by the total number of forecasts issued by management. A management forecast is deemed to be "relatively accurate" when management's forecast is strictly more accurate than the median analyst forecast prevailing on the day that the management forecast is released (|median analyst estimate – realized EPS| > |management forecast – realized EPS|). This relative performance measure parsimoniously controls for many factors that affect forecast accuracy across firms, such as firm complexity, forecast horizon, and industry earnings volatility. As in Hutton and Stocken (2009), the total number of forecasts that a firm has issued is determined by counting the firm's annual and quarterly earnings forecasts captured by First Call's CIG database since January 1994. To prevent a hindsight bias, we estimate this variable using only the data available prior to the earnings announcement date.
Textual Content Reputation	The difference between the absolute values of two regression residuals. The regression residual of equation (1) when $SEARN_{jq+1t+1}$ is the dependent variable and we do not include net optimism nor $TUncertainty_{jqt}$ in the regression. And the regression residual of equation (1) when $SEARN_{jq+1t+1}$ is the dependent variable and we do include net optimism and $TUncertainty_{jqt}$ in the regression. The difference between the absolute values of each of the

	former and the text-informed regression residuals provides a measure of the improvement in explanatory power for future earnings provided by the text.
High-Tech	Indicator set equal to 1 if dnum = 3570–3579, 3622, 3660–3692, 3694–3699, 3810–3839, 7370–7372, 7373–7379, 7391, 8730–8734.
R&D Expenses	Annual R&D expenses (data 4 in the quarterly Compustat tape) expressed as a percentage of total assets (data 44 in the quarterly Compustat tape).
EFKOS e-Loading	Variable obtained by regressing the daily excess return of firm *i* on the EFKOS factor as well as the Fama–French three factors (SML, HML, market return). We allow the loading to change over time, and we estimate the coefficient using all non-earnings announcement days in the previous 365 calendar days before the earnings announcement date (only for stocks with at least 100 data points during that period). We winsorize this variable by replacing values above the 99th percentile with the 99th percentile value and values below the 1st percentile with the 1st percentile value.
P/E Ratio	Share price (item 199 in the annual Compustat database) as of the end of the previous fiscal year divided by earnings as of the end of the current fiscal year. Earnings are calculated as income before extraordinary items available to common stockholders (annual Compustat item 237) plus deferred taxes from the income statement (annual Compustat item 50) plus the investment tax credit (annual Compustat item 51). We winsorize this variable by replacing values above the 99th percentile with the 99th percentile value and values below the 1st percentile with the 1st percentile value.

References

Abarbanell, Jeffery, & Lehavy, Reuven. (2002). Differences in Commercial Database Reported Earnings: Implications for Empirical Research. *unpublished working paper, University of Michigan*.

Baginski, S. P., Hassell, J. M., & Kimbrough, M. D. (2004). Why Do Managers Explain Their Earnings Forecasts? *Journal of Accounting Research, 42*(1), 1-29.

Baginski, Stephen, Conrad, Edward, & Hassell, John. (1993). The Effects of Management Forecast Precision on Equity Pricing and on the Assessment of Earnings Uncertainty. *Accounting Review, 68*(4), 913-927.

Bamber, Linda Smith, Barron, Orie E., & Stober, Thomas L. (1997). Trading volume and different aspects of disagreement coincident with earnings announcements. *The Accounting Review, 72*(4), 575-597.

Barth, Mary E., Kasznik, Ron, & McNichols, Maureen F. (2001). Analyst Coverage and Intangible Assets. *Journal of Accounting Research, 39*(1), 1-34.

Benabou, R., & Laroque, G. (1992). Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility. *Quarterly Journal of Economics, 107*(921-958).

Bernard, Victor, & Thomas, Jacob. (1989). Post-Earnings-Announcement Drift. *Journal of Accounting Research, 27*(3), 1-36.

Bhattacharya, Nilabhra, Black, Ervin L., Christensen, Theodore E., & Mergenthaler, Richard D. (2004). Empirical Evidence on Recent Trends in Pro Forma Reporting. *Accounting Horizons, 18*(1), 27-43.

Bliss, J.H. (1924). *Management through Accounts*. New York, NY: The Ronald Press Co.

Brandt, Michael W., Kishore, Runeet, Santa-Clara, Pedro, & Venkatachalam, Mohan. (2008). Earnings Announcements are Full of Surprises. *unpublished working paper, Duke University*.

Collins, Daniel W., Maydew, Edward L., & Weiss, Ira S. (1997). Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting & Economics, 24*(1), 39.

Crawford, Vincent, & Sobel, Joel. (1982). Strategic Information Transmission. *Econometrica, 50*(6), 1431-1451.

Davis, Angela, Piger, Jeremy, & Sedor, Lisa. (2012). Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. *Contemporary Accounting Research, 29*(3), 845-868.

Demers, Elizabeth, Francis, Jennifer, & Vega, Clara. (2013). Large sample evidence on the difficulty of processing textual content information. *unpublished manuscript*.

Diether, Karl B., Malloy, Christopher J., & Scherbina, Anna. (2002). Differences of Opinion and the Cross Section of Stock Returns. *Journal of Finance, 57*(5), 2113-2141.

Digitext Inc. (2000). Diction 5.0: The Text-Analysis Program. Austin, Texas.

Ecker, Frank, Francis, Jennifer, Kim, Irene, Olsson, Per, & Schipper, Katherine. (2006). A Returns-Based Representation of Earnings Quality. *The Accounting Review, 81*(4), 749-780.

Engelberg, Joseph. (2008). Costly Information Processing: Evidence from Earnings Announcements. *unpublished working paper, Northwestern University*.

Fama, Eugene F., & French, Kenneth R. (2000). Forecasting Profitability and Earnings. *Journal of Business, 73*(2), 161.

Fields, Thomas D., Lys, Thomas Z., & Vincent, Linda. (2001). Empirical research on accounting choice. *Journal of Accounting & Economics, 31*(1-3), 255-307.
Francis, Jennifer, & Schipper, Katherine. (1999). Have Financial Statements Lost Their Relevance? *Journal of Accounting Research, 37*(2), 319-352.
Francis, Jennifer, Schipper, Katherine, & Vincent, Linda. (2002). Expanded Disclosures and Increased Usefulness of Earnings Announcements. *The Accounting Review, 77*(3), 515-546.
Graham, John, Harvey, Campbell, & Rajgopal, Shiva. (2006). Value Destruction and Financial Reporting Decisions. *Financial Analysts Journal, 62*(6), 27-39.
Hutton, Amy P., Miller, Gregory, & Skinner, Douglas. (2003). The Role of Supplementary Statements with Management Earnings Forecasts. *Journal of Accounting Research, 41*(5), 867-890.
Hutton, Amy, & Stocken, Phillip. (2009). Prior Forecasting Accuracy and Investor Reaction to Management Earnings Forecasts *unpublished working paper*.
Imhoff Jr, Eugene A., & Lobo, Gerald J. (1992). The Effect of Ex Ante Earnings Uncertainty on Earnings Response Coefficients. *Accounting Review, 67*(2), 427-439.
Johnson, W. Bruce, & Schwartz Jr, William C. (2005). Are Investors Misled by "Pro Forma" Earnings? *Contemporary Accounting Research, 22*(4), 915-963.
Kothari, S. P., Li, Xu, & Short, James E. (2009). The Effect of Disclosures by Management, Analysts, and Business Press on Cost of Capital, Return Volatility, and Analyst Forecasts: A Study Using Content Analysis. *Accounting Review, 84*(5), 1639-1670.
Krishna, Vijay, & Morgan, John. (2001). A Model of Expertise. *Quarterly Journal of Economics, 116*, 747-775.
Krishna, Vijay, & Morgan, John. (2004). The Art of Conversation: Eliciting Information from Experts Through Multi-Stage Communication. *Journal of Economic Theory, 117*, 147-179.
Lev, Baruch. (2012). *Winning Investors Over*. Boston, MA: Harvard Business Review Press.
Lev, Baruch, & Zarowin. (1999). The Boundaries of Financial Reporting and How to Extend Them. *Journal of Accounting Research, 37*(2), 353-385.
Lougee, Barbara, & Marquardt, Carol. (2004). Earnings Informativeness and Strategic Disclosure: An Empirical Examination of "Pro Forma" Earnings. *Accounting Review, 79*(3), 769-795.
Loughran, T., & McDonald, B. (2011). When is a Liability Not a Liability? *Journal of Finance, 66*(February), 35-65.
Matsumoto, Dawn, Pronk, Maarten, & Roelofsen, Erik. (2011). What Makes Conference Calls Useful? The Information Content of Managers' Presentations and Analysts' Discussion Sessions. *The Accounting Review, 86*(4), 1383-1414.
Mitchell, Mark, & Mulherin, J. Harold. (1994). The Impact of Public Information on the Stock Market. *Journal of Finance, 49*(3), 923-950.
Roll, Richard. (1988). R^2. *Journal of Finance, 43*(3), 541-566.
Shiller, Robert J. (1981). Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends? *American Economic Review, 71*(3), 421.
Sobel, Joel. (1985). A Theory of Credibility. *Review of Economic Studies, 52*, 557-573.
Stocken, Phillip. (2000). Credibility of Voluntary Disclosure. *RAND Journal of Economics, 31*(2), 359-374.

Tetlock, Paul C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance, 62*(3), 1139-1168.

Tetlock, Paul C., Saar-Tsechansky, Maytal, & Macskassy, Sofus. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *Journal of Finance, 63*(3), 1437-1467.

Verrecchia, Robert. (1983). Discretionary Disclosure. *Journal of Accounting and Economics, 5*, 179-194.

Watts, Ross L. (2003a). Conservatism in Accounting Part I: Explanations and Implications. *Accounting Horizons, 17*(3), 207-221.

Watts, Ross L. (2003b). Conservatism in Accounting Part II: Evidence and Research Opportunities. *Accounting Horizons, 17*(4), 287-301.

Watts, Ross L., & Zimmerman, Jerold L. (1986). *Positive Accounting Theory*. Englewood Cliffs, New Jersey: Prentice-Hall.

Zhang, X. Frank. (2006). Information Uncertainty and Stock Returns. *Journal of Finance, 61*(1), 105-136.

Table I. Sample Statistics for Information Surprises, Uncertainty Measures, and Abnormal Returns

In this table we present in Panel A the mean, standard deviation, and skewness sample statistics and in Panel B the average quarterly bivariate correlations for the following variables: ΔTNetOpt, change in textual net optimism expressed in the text of the earnings statement from this quarter to the previous quarter calculated using Diction 6.0, General Inquirer (GI), and Loughran and McDonald's (2011) (L&M) method, respectively, and a factor model of all three measures; TUncertainty, a variable that indicates the degree of uncertainty expressed in the text of the earnings statement calculated using Diction 6.0 and the L&M method, respectively, and a factor model of both measures; SUE, standardized earnings surprise estimated using a seasonal random walk model and Compustat earnings per share data; SUMFQ and SUMFA, standardized management quarterly and annual earnings surprises, respectively, calculated using First Call data; SUProForma, standardized pro forma earnings surprises estimated using First Call data and an indicator variable equal to 1 if the earnings statement mentions pro forma earnings; CAR, cumulative abnormal returns for the three-day announcement period; and AnalystDisp, dispersion across analysts' forecasts estimated using First Call data. The summary statistics for all variables are calculated using all available earnings announcements from January 1998 to July 2006. For a detailed description of the variables, please refer to the Appendix.

Panel A: Summary Statistics

	ΔTNetOpt (Diction)	ΔTNetOpt (GI)	ΔTNetOpt (L&M)	ΔTNetOpt (Factor)	TUnc. (Dic.)	TUnc. (L&M)	TUnc. (Factor)	SUE	SUMFQ	SUMFA	SUProForma	CAR	Analyst Disp.
Mean	−0.017	−0.050	−0.022	−0.025	75.10	0.49	0.024	0.069	0.005	0.042	0.056	0.300	0.120
Std. Dev.	1.249	1.739	1.02	0.923	9.321	0.43	0.99	0.990	0.366	0.492	0.532	12.48	0.130

Panel B: Correlation Matrix

	ΔTNetOpt (Diction)	ΔTNetOpt (GI)	ΔTNetOpt (L&M)	ΔTNetOpt (Factor)	TUnc. (Dic.)	TUnc. (L&M)	TUnc. (Factor)	SUE	SUMFQ	SUMFA	SUProForma	CAR	Analyst Disp.
ΔTNetOpt (Diction)	1												
ΔTNetOpt (GI)	0.315	1											
ΔTNetOpt (L&M)	0.415	0.530	1										
ΔTNetOpt (Factor)	0.745	0.764	0.833	1									
TUnc. (Diction)	0.029	0.030	0.025	0.035	1								
TUnc. (L&M)	−0.019	−0.022	−0.027	−0.028	0.159	1							
TUnc. (Factor)	0.006	0.005	−0.001	0.004	0.758	0.764	1						
SUE	0.056	0.087	0.113	0.109	−0.031	−0.041	−0.047	1					
SUMFQ	0.005	0.017	0.026	0.021	−0.014	0.007	−0.004	0.043	1				
SUMFA	0.004	0.011	0.023	0.017	0.009	0.020	0.019	0.052	0.372	1			
SUProForma	0.020	0.042	0.038	0.042	−0.010	−0.004	−0.009	0.109	0.054	0.075	1		
CAR	0.056	0.082	0.103	0.102	−0.008	−0.028	−0.024	0.122	0.076	0.059	0.124	1	
Analyst Disp.	−0.027	−0.012	−0.031	−0.032	0.065	0.037	0.068	−0.072	−0.052	−0.047	−0.076	−0.036	1

Table II. Future Earnings and the Textual Content of Earnings Press Releases
In this table we present estimates of equation (1). The sample includes all available earnings announcements from January 1998 to July 2006. We use standard errors clustered by calendar quarter and firm to compute the one-sided p-values that are reported in parentheses below the coefficient estimates. The ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

	Future Earnings: $SEARN_{jq+1t+1}$	Analyst Forecast Revisions: $SAFR_{jq+1t}$	Future Unexpected Earnings: $SUE_{jq+1t+1}$
Intercept	0.061***	-0.107*	0.064***
	(0.004)	(0.06)	(0.001)
ΔTNetOpt		0.071***	0.043***
		(<0.001)	(<0.001)
TNetOpt	0.083***		
	(<0.001)		
TUncertainty	−0.059***	−0.023**	−0.034***
	(<0.001)	(0.03)	(<0.001)
SUE_{jqt}		0.261***	0.252***
		(<0.001)	(<0.001)
$\sum_{i=0}^{3} SEARN_{jq-it-i}$	0.935***		
	(<0.001)		
SUMFQ		0.260***	0.080***
		(<0.001)	(<0.001)
SMFQ	0.002***		
	(0.002)		
SUMFA		0.068***	0.008
		(0.004)	(0.27)
SMFA	0.001		
	(0.34)		
SUProForma			0.032**
			(0.023)
SProForma	0.0129*		
	(0.06)		
AnalystDisp	−0.256***	−0.336***	−0.221***
	(<0.001)	(<0.001)	(<0.001)
SAFR			0.103***
			(<0.001)
CAR	0.006***	0.017***	0.005***
	(<0.001)	(<0.001)	(<0.001)
Size	0.003	−0.027***	−0.008**
	(0.23)	(0.004)	(0.025)
Log(Market/Book)	−0.017**	0.019	0.018**
	(0.03)	(0.13)	(0.030)
Log(Turnover)	0.027***	0.059***	0.052***
	(<0.001)	(<0.001)	(<0.001)
Adjusted *R*-squared	72.45%	14.26%	9.58%
Observations	20,899	8,994	20,899

Table III. Change in Analyst Forecast Dispersion and the Textual Content of Earnings Press Releases

In this table we present estimates of regressing changes in dispersion across analysts' forecasts on earnings- and text-based variables, announcement-period abnormal returns, firm size, market-to-book ratio, and turnover. The sample includes all available earnings announcements from January 1998 to July 2006. We use standard errors clustered by calendar quarter and firm to compute the one-sided p-values that are reported in parentheses below the coefficient estimates. The ***, **, and * represent significance of a one-sided test at the 1, 5, and 10 percent levels, respectively.

Change in Analyst Forecast Dispersion	
Intercept	-0.124***
	(<0.001)
|ΔTNetOpt |	0.009**
	(0.05)
|ΔTNetOpt | × I(ΔTNetOpt < 0)	0.004
	(0.33)
ΔTUncertainty	0.001**
	(0.03)
|SUE|	0.067***
	(<0.001)
|SUE| × I(SUE < 0)	0.026*
	(0.09)
|SUMFQ|	0.016
	(0.24)
|SUMFQ| × I(SUMFQ < 0)	0.129**
	(0.01)
|SUMFA|	0.018*
	(0.08)
|SUMFA| × I(SUMFA < 0)	0.084**
	(0.02)
|CAR|	−0.0001
	(0.50)
|CAR| × I(CAR < 0)	0.008***
	(<0.001)
Size	0.013***
	(0.006)
Log(Market/Book)	0.013
	(0.11)
Log(Turnover)	0.0002
	(0.49)
Adjusted *R*-squared	1.81%
Observations	8,994

Table IV. Market Reaction to Earnings Surprises and Textual Content Variables
In this table we present estimates of regressing 3-trading-day abnormal returns on earnings surprises, textual content variables, firm size and measures of future cash flow uncertainty. The sample includes all available earnings announcements from January 1998 to July 2006. We use standard errors clustered by calendar quarter and firm to compute the one-sided p-values that are reported in parentheses below the coefficient estimates. The ***, **, and * represent significance of a one-sided test at the 1, 5, and 10 percent levels, respectively

Intercept	-0.167	-0.218	Intercept	-0.245
	(0.42)	(0.40)		(0.38)
ProForma	-0.341^{**}	-0.290^{**}	ProForma	-0.307^{**}
	(0.02)	(0.03)		(0.02)
FirstCall	-0.501^{***}	-0.380^{***}	FirstCall	-0.382^{**}
	(0.002)	(0.01)		(0.01)
SUE	6.824^{***}	6.795^{***}	SUE	6.891^{***}
	(<0.001)	(<0.001)		(<0.001)
SUMFQ	9.392^{***}	8.570^{***}	SUMFQ	9.303^{***}
	(<0.001)	(<0.001)		(<0.001)
SUMFA	2.802	2.743	SUMFA	2.725
	(0.18)	(0.19)		(0.18)
SUProForma	6.857^{**}	6.864^{**}	SUProForma	7.244^{**}
	(0.02)	(0.02)		(0.02)
ΔTNetOpt	4.591^{***}	4.603^{***}	ΔTNetOpt	4.585^{***}
	(<0.001)	(<0.001)		(<0.001)
Size	0.039	0.042	Size	0.043
	(0.19)	(0.17)		(0.16)
SUE × Size	-0.301^{***}	-0.299^{***}	SUE × Size	-0.305^{***}
	(<0.001)	(<0.001)		(<0.001)
SUMFQ × Size	-0.381^{***}	-0.346^{***}	SUMFQ × Size	-0.377^{***}
	(<0.001)	(0.002)		(<0.001)
SUMFA × Size	−0.107	−0.105	SUMFA × Size	−0.103
	(0.23)	(0.23)		(0.23)
SUProForma × Size	-0.237^{*}	-0.242^{*}	SUProForma × Size	-0.254^{*}
	(0.06)	(0.06)		(0.06)
ΔTNetOpt × Size	-0.194^{***}	-0.195^{***}	ΔTNetOpt × Size	-0.194^{***}
	(<0.001)	(<0.001)		(<0.001)

(Table continued on the next page)

Table IV (Continued). Market Reaction to Earnings Surprises and Textual Content Variables

AnalystDisp		−1.369***	AnalystDisp	−1.351***
		(0.002)		(0.002)
RangeMFQ		−0.161	RangeMFQ	−0.060
		(0.26)		(0.42)
RangeMFA		−0.239	RangeMFA	−0.160
		(0.40)		(0.49)
TUncertainty		−0.122**	TUncertainty	−0.109**
		(0.01)		(0.03)
SUE × AnalystDisp		−0.079	SUE × TUncertainty	−0.109**
		(0.44)		(0.04)
SUMFQ × RangeMFQ		−0.702	SUMFQ × TUncertainty	−0.074
		(0.26)		(0.36)
SUMFA × RangeMFA		−0.913	SUMFA × TUncertainty	−0.029
		(0.40)		(0.44)
SUProForma × AnalystDisp		0.640	SUProForma × TUncertainty	-0.142
		(0.27)		(0.12)
ΔTNetOpt × TUncertainty		0.026	ΔTNetOpt × TUncertainty	0.042
		(0.33)		(0.25)
Adjusted *R*-squared	5.03%	5.06%	Adjusted *R*-squared	5.07%
Observations	20,899	20,899	Observations	20,899

Table V. Informativeness of Textual Net Optimism and GAAP Earnings for Future Earnings Levels.

In this table we present estimates of equation (1), where future earnings, $SEARN_{jq+1t+1}$, is used as the dependent variable and the independent variables are current *TNetOpt*, current and past *SEARN*, and other controls. In this specification we allow the effect of textual net optimism and current and past earnings to be different across sample partitions which vary in the degree of informativeness of textual net optimism (Panel A) and the degree of noisiness of GAAP earnings (Panel B). The sample includes all available earnings announcements from January 1998 to July 2006. We use standard errors clustered by calendar quarter and firm to compute the one-sided p-values that are reported in parentheses below the coefficient estimates. The ***, **, and * represent significance at the 1, 5, and 10 percent levels, respectively.

	Panel A: Informativeness of Textual Net Optimism						Panel B: Noisiness of GAAP earnings				
	Numerical Terms	ProForma	Managerial Forecast	Analyst Coverage	Earnings Forecast Reputation	Textual Content Reputation	High-Tech	R&D Expenses	EFKOS e-Loading	PE Ratio	Textual Uncertainty
Earnings Impact											
High	0.954***	0.864***	0.919***	0.936***	0.863***	0.830***	0.833***	0.900***	0.911***	0.911***	0.902***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Medium	0.932***	NA	NA	0.938***	0.890***	0.918***	NA	NA	0.939***	0.968***	0.939***
	(<0.001)			(<0.001)	(<0.001)	(<0.001)			(<0.001)	(<0.001)	(<0.001)
Low	0.914***	0.951***	0.941***	0.935***	0.897***	0.973***	0.945***	0.947***	0.953***	0.988***	0.956***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
High-Low	0.040***	-0.087***	-0.021***	0.001	-0.034	-0.143***	-0.112***	-0.047***	-0.042***	-0.077***	-0.054***
	(<0.001)	(<0.001)	(0.005)	(0.41)	(0.12)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Textual Net Optimism Impact											
High	0.105***	0.092***	0.105***	0.109***	0.116***	0.473***	0.087***	0.092***	0.077***	0.092***	0.112***
	(<0.001)	(0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Medium	0.076***	NA	NA	0.080***	0.098***	0.117***	NA	NA	0.095***	0.069***	0.073***
	(<0.001)			(<0.001)	(<0.001)	(<0.001)			(<0.001)	(<0.001)	(<0.001)
Low	0.078***	0.056***	0.078***	0.078***	0.052**	-0.368***	0.085***	0.081***	0.089***	0.084***	0.067***
	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(0.01)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)	(<0.001)
High-Low	0.027**	0.036**	0.038**	0.031**	0.064**	0.841***	-0.002	0.011	-0.012	0.008	0.045**
	(0.04)	(0.03)	(0.05)	(0.05)	(0.05)	(<0.001)	(0.46)	(0.28)	(0.26)	(0.27)	(0.01)
Adj. R^2	72.46%	72.48%	72.53%	72.42%	59.45%	78.08%	72.49%	72.49%	72.36%	68.43%	72.44%
Observations	20,899	20,899	20,899	20,899	6,284	20,899	20,899	20,899	20,899	17,135	20,899

Table VI. Announcement Period Responses to Textual Net Optimism for Differentially Informative Earnings and Textual Content Firms
In this table we present estimates of equation (2) in the paper. The sample includes all available earnings announcements from January 1998 to July 2006. We use standard errors clustered by calendar quarter and firm to compute the one-sided p-values that are reported below the coefficient estimates. The ***, **, and * represent significance of a one-sided test at the 1, 5, and 10 percent levels, respectively.

	Panel A: Informativeness of Textual Net Optimism						Panel B: Noisiness of GAAP Earnings				
	Numerical Terms	Pro Forma	Managerial Forecast	Analyst Coverage	Earnings Forecast Reputation	Textual Content Reputation	High-Tech	R&D Expenses	EFKOS e-Loading	PE Ratio	Textual Uncertainty
Earnings Impact											
High	0.620*** (<0.001)	0.363*** (<0.001)	0.469*** (<0.001)	0.278*** (0.002)	0.609*** (<0.001)	0.878*** (<0.001)	0.753*** (<0.001)	1.07*** (<0.001)	1.357*** (<0.001)	0.545*** (<0.001)	0.729*** (<0.001)
Medium	0.936*** (<0.001)	NA	NA	0.409*** (<0.001)	0.551*** (<0.001)	0.744*** (<0.001)	NA	NA	0.661*** (<0.001)	0.533*** (<0.001)	0.960*** (<0.001)
Low	1.055*** (<0.001)	1.020*** (<0.001)	1.034*** (<0.001)	1.312*** (<0.001)	0.320** (0.01)	1.148*** (<0.001)	0.920*** (<0.001)	0.853*** (<0.001)	0.605*** (<0.001)	1.180*** (<0.001)	0.949*** (<0.001)
High-Low	-0.434*** (<0.001)	-0.657*** (<0.001)	-0.565*** (<0.001)	-1.034*** (0.01)	0.289* (0.09)	-0.271** (0.03)	-0.168 (0.15)	0.218* (0.07)	0.751** (0.03)	-0.635** (0.05)	-0.220** (0.04)
Textual Net Optimism Impact											
High	0.779*** (<0.001)	0.885*** (<0.001)	0.881*** (<0.001)	0.938*** (<0.001)	0.865*** (<0.001)	0.895*** (<0.001)	0.816*** (<0.001)	0.936*** (<0.001)	0.756*** (<0.001)	0.820*** (<0.001)	0.627*** (<0.001)
Medium	0.576*** (<0.001)	NA	NA	0.698*** (<0.001)	0.272 (0.14)	0.588*** (<0.001)	NA	NA	0.614*** (<0.001)	0.448*** (<0.001)	0.734*** (<0.001)
Low	0.558*** (<0.001)	0.573*** (<0.001)	0.557*** (<0.001)	0.476*** (<0.001)	0.434** (0.02)	0.303** (0.03)	0.574*** (<0.001)	0.647*** (<0.001)	0.458*** (<0.001)	0.390*** (<0.001)	0.529*** (<0.001)
High-Low	0.221* (0.09)	0.313** (0.02)	0.324** (0.03)	0.462** (0.01)	0.431* (0.06)	0.593*** (<0.001)	0.242** (0.04)	0.289** (0.02)	0.300** (0.04)	0.430*** (0.007)	0.098 (0.25)
Adjusted *R*-squared	4.76%	4.74%	4.73%	4.87%	6.68%	4.92%	5.19%	3.99%	4.89%	3.60%	4.77%
Observations	20,899	20,899	20,899	20,899	6,284	20,034	20,899	20,899	20,688	17,135	20,899