Soft Information in Earnings Announcements: News or Noise?*

March 2008

Elizabeth Demers
Accounting & Control Area
INSEAD
Boulevard de Constance
77305 Fountainebleau Cedex
France
liz.demers@insead.edu

Clara Vega
Board of Governors
of the Federal Reserve System
International Finance Division
20th Street & Constitution Avenue, NW
Washington, DC 20551
clara.vega@frb.gov

JEL Classifications: G14; D82; M41
Keywords: Soft information; earnings announcements; post-earnings drift; cheap talk; earnings quality; information uncertainty; momentum; voluntary disclosure

*We thank workshop participants at Duke University, in addition to Per Olsson, Katherine Schipper, and especially Jennifer Francis for valuable comments and discussions. We also appreciate the excellent research assistance of Armel Allard, Shannon Demers, Marcus Newman, and Hien Tran, 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.

1. Introduction

Beginning with Ball and Brown (1968), many researchers have examined stock price responses to corporate earnings announcements. An extensive subsequent literature investigates the market responses to other elements of firms' voluntary disclosures issued simultaneously with the earnings news. With one exception that we are aware of (Davis, Piger and Sedor (2007)), all of these prior studies have examined market responses to quantitative, *hard*, largely verifiable information disclosed by management.[1] In this study, we consider the stock market's abnormal price, volume, and idiosyncratic volatility responses to managers' *soft* information disclosures during their quarterly earnings announcements, incremental to the impact of the hard earnings surprise. Specifically, we use a well-established linguistic algorithm to extract three dimensions of managerial sentiment - optimism, pessimism, and certainty - from over 20,000 corporate earnings announcements filed with the PR Newswire service during the period of January 1998 through July 2006. We find that the unexpected component of manager's "net optimism" (i.e., optimism less pessimism) is significantly associated with the short-window announcement period returns, and that it also significantly predicts post-earnings announcement drift. Our extended tests show that the pricing of this form of sentiment depends upon whether conditions are present to induce informative sentiment rather than "cheap talk" on the part of managers. We also introduce the formally measured linguistic construct of certainty into the finance and accounting literatures, and we document that managerial certainty is inversely associated with increased idiosyncratic volatility and the dispersion in investor beliefs during the short-window announcement interval, and it also predicts abnormal idiosyncratic volatility and volume during the intermediate-term post-announcement period.

A large prior literature documents the market's response to "hard" earnings announcements in the form of prices (Ball and Brown (1968); Skinner and Sloan (2002)), volume (e.g., Beaver (1968); Bailey, Li, Mao and Zhong (2003)), and volatility (e.g., Landsman and Maydew (2002); DeFond, Hung and Trezevant (2007)). Extensions of this literature examine the market response to other simultaneously

[1] Petersen (2004) defines hard information to be data that is quantitative, easy to store and transmit in impersonal ways, and whose content is independent of the collection process, while Stein (2002) characterizes soft information as that which is directly verifiable only by the person who collected and produced it.

disclosed "hard" information (e.g., Francis, Schipper and Vincent (2002)) and the relation between managers' voluntarily provided hard information to their corporate financial reporting incentives, such as the management of accounting accruals to meet earnings benchmarks (e.g., Burgstahler and Dichev (1997); Degeorge, Patel and Zeckhauser (1999)), the release of pro forma earnings contemporaneously with the mandated (audited) GAAP earnings announcements in order to guide or potentially mislead the market (e.g., Bradshaw and Sloan (2002); Hutton, Miller and Skinner (2003); and Bhattacharya, Black, Christensen and Larson (2003)), and the strategic disclosure of other information variables within earnings press releases for similar purposes (e.g., Schrand and Walther (2000)). Overall, the results from this research strongly suggest that the market responds in an economically and statistically significant manner to earnings and related credible, verifiable, hard information disclosures, while managers seem to behave in a strategic manner in order influence the market's response to their earnings and related news items. It is notable, however, that the explanatory power of the announcement period abnormal returns models is generally very low, suggesting that there is some additional news being released to which the market is responding but that has not been captured by the hard information explanatory variables that have been examined to date.[2] Our study extends this earnings announcement literature by investigating a potentially much subtler mechanism available to management to attempt to guide investors about their firm's prospects, the use of "soft" information in the form of sentimental tones conveyed in the language of their earnings press releases.[3]

"Cheap talk" models address the question of how much information can be credibly transmitted when communication is direct and costless (Krishna and Morgan (2008)). In the context of earnings announcements, we consider the sentiment expressed by management to be a form of direct and costless communication.[4] The

[2] An alternative explanation is that the explanatory power of the stock price response models is low due to researcher-induced specification errors. However, decades of research in this area with numerous improvements and alternative empirical approaches continue to produce similar results.

[3] Indeed, recent corporate accounting scandals have given rise to new legislation in relation to the issuance of pro forma figures ((SEC Release Nos. 33-8039, 34-45124, FR-59), fair disclosure (Regulation FD), and corporate accountability (the Sarbanes-Oxley Act), suggesting that managers are subject to much greater scrutiny in their accounting and reporting to the capital markets, and thus may need to resort to subtler forms of communication in order to convey subjective information about their firms to the market.

[4] Clearly there are securities laws which prevent managers from making blatantly false and misleading statements in their earnings announcements. However, the sentiment expressed in these announcements is subtle, and prosecution of managers who issue highly optimistic (in tone) but

classic theoretical work by Crawford and Sobel (1982) predicts that where the agent doing the communicating is biased (i.e., his incentives are not aligned with the party to whom he is communicating), as extensive prior research related to management incentives surrounding earnings announcements suggests to be the case in our setting, the informativeness of the communication will be decreasing in his bias. Theory also suggests that repeated games and the ability to punish or reward the communicator for informative rather than cheap talk help to induce less noisy communications. Outside of laboratory settings, there appears to be a paucity of empirical research testing cheap talk theories. Our study thus contributes to the cheap talk literature by providing empirical archival support for numerous propositions found in the extant theories.

There has been an increased interest in recent years in determining the sentiment, or soft information, conveyed in public communications by government institutions, the media, as well as corporate entities. Various methods have been employed to gauge the sentiment of communication and systematically analyze its impact on market forces and individual behavior. For example, Ehrmann and Fratzscher (2007) analyze the style of communication among central banks by manually classifying Reuters press releases in terms of economic outlook and policy inclinations. Das, Martinez-Jerez and Tufano (2005) examine the connection between on-line discussion, news activity, and movement in stock prices by developing their own sentiment index based upon five distinct language processing algorithms that classify discussion as bullish, bearish, or neutral, while Das and Chen (2007) use the same method to extract small investor sentiment from stock message boards. Li (2006) uses a simple count of the relative frequency of "risk" and "uncertainty" in corporate annual reports and relates this to future earnings and stock returns. Numerous studies use Diction software to extract sentiment from various texts (e.g., Bligh and Hess (2007); Ober, Zhao, Davis and Alexander (1999); Yuthas, Rogers and Dillard (2002); and Davis, *et al.* (2007)),[5] while Tetlock (2007), Tetlock, Saar-

otherwise truthful announcements would presumably be difficult to pursue. Accordingly, we assume that the tone, if not the factual content, of the announcements is costless to the manager.

[5] Bligh and Hess (2007) use the Diction software to measure "certainty, pessimism, optimism, activity, immediacy and jargon" in 45 FOMC statements, 44 congressional testimonies and 105 speeches given by the Chairman of the Federal Reserve, Alan Greenspan, between May 18th, 1999 and June 30th, 2004. They conclude that Greenspan's rhetoric predicts movements both in the Treasury forward rates and in the federal funds future rates. Ober, Zhao, Davis and Alexander (1999) use Diction to assess corporations' use of certainty in public communications by examining the "Management Discussion and Analysis" (MD&A) section of the10-K reports of the six Fortune 500 companies with the largest increases in profits and the six companies with the largest decreases in profits in 1996 in each of six

Tsechansky and Macskassy (2007), and Engelberg (2007) use *General Inquirer* ("GI"), an alternative linguistic algorithm, to measure the level of negativity in media content and relate this to securities returns.

Our study relates most closely to the soft information studies of Tetlock, *et al.* (2007), Engelberg (2007), Li (2006), and Davis, *et al.* (2007). Tetlock et al (2007) and Engelberg (2007) both examine whether a quantitative measure of negative language in firm-specific earnings news stories can be used to predict firms' future accounting earnings and stock returns. They conclude that linguistic media content captures otherwise hard to quantify aspects of firms' fundamentals which investors quickly incorporate into stock prices. Our study differs from, and is complementary to, Tetlock et al (2007) and Engelberg (2007) in several important respects. First, they examine the association between *media-expressed* sentiment and future measures of firm performance, while we examine the relation between *management-expressed* sentiment and both future stock returns and the dispersion in investor beliefs. Relative to the media, managers have different insights, motivations, biases, and fiduciary duties in their communications to parties who are external to the firm. This potential misalignment of interests, together with the subtlety of the language constructs derived from press releases, enable us to test cheap talk theories in the context of managerial earnings announcements.[6] Second, Tetlock et al (2007) and Engelberg (2007) examine only one dimension of language, negativity, while we consider the role of both unexpected net optimism as well as certainty in explaining various measures of market activity. Third, while Tetlock et al (2007) have a longer time series of observations for a sample that is restricted to very large, highly liquid S&P 500 firms, we have a shorter, more recent time series of observations that span a much broader sample of firms that are not all subject to high information

---

major industry groups. They find that corporations with large profit increases do not use rhetoric with significantly more certainty than companies that experience large decreases in their profits, and thus they conclude that managers "tell it like it is" and "avoid weasel words." Yuthas, Rogers and Dillard (2002) investigate the characteristics of corporate annual reports in order to ascertain whether corporate communication is ethical and conforms to Habermas' four principles of comprehensibility, truthfulness, sincerity, and legitimacy. They find that managers of firms with bad performance generally engage in "ethical discourse" by not distorting the truth about their companies' financial position. However, they also report that these firms strategically use "fewer self-referential terms," perhaps in order to disassociate bad performance with internal factors. Davis, Piger and Sedor (2007) examine the use of pessimistic and optimistic language in earnings press releases. A more extensive summary of published academic articles using the Diction software is provided at: http://www.dictionsoftware.com/files/dictionresearch.pdf.

[6] In addition, while Engelberg (2007) considers only the headline and first paragraph of the news article, we consider the entire text of the managers' announcements after excluding numerical tables, company descriptions, and the "Safe Harbor" provisions. Further details are provided in Section 2.2.

environments such as those in the S&P 500. This broader sample provides the cross-sectional variation necessary to examine how different firm characteristics affect the market's response to soft information.

Relative to Li (2006) who examines the relation between the risk sentiment expressed in corporate annual reports and future earnings and stock returns by adopting a count of a few researcher-specified "risk" and "uncertainty" words, our study uses a sophisticated, externally validated linguistic algorithm to extract management expressed certainty from corporate earnings announcements. Our results also differ from those of Li (2006) as he finds that risk sentiment can predict future returns in a cross-sectional setting while we find no relation between the certainty in manager's earnings releases and stock prices in either the short- or longer-term windows. Instead, we find that our measure of certainty is associated with dispersion in investor beliefs and abnormal idiosyncratic volatility during the announcement window and certainty is also a predictor of post-announcement abnormal volume and idiosyncratic volatility. Davis et al (2007) examine the association between the optimism and pessimism expressed in earnings announcements, and, respectively, future accounting performance and the market's response to the earnings announcements. Our study builds on theirs along several dimensions. First, using the same Diction linguistic software, we extract a third dimension of management sentiment, certainty, from the earnings announcements and we explore its relation to market activity. We find that more wavering (i.e., less certainty) in managements' communications, whether driven by an incentive to obfuscate or by the uncertainty inherent in the firm's economic circumstances, leads to greater dispersion in investors' beliefs as manifested in higher abnormal trading volume and volatility. Second, after replicating the short-window announcement abnormal returns results of Davis et al (2007), we extend their analysis by documenting that unexpected managerial net optimism also predicts post-earnings announcement returns. Third, we build on their general finding that the market responds to sentiment during the announcement period in two ways. First, we use empirical proxies for constructs suggested by the cheap talk theoretical literature to document that the market's response to sentiment is an increasing function of the firm's liquidity (a measure of market participants' ability to punish managers for misleading or uninformative cheap talk) as well as analyst coverage (also a proxy for market participants' ability to punish uninformative or misleading communication, as well as a mechanism to generate repeated two-way communications). Second, we show that the market's response to soft information is increasing in the relative lack of informativeness of the alternative, hard earnings information, such as for low earnings quality firms and for high-tech firms that have complex business models for which the standard accounting model less aptly captures the underlying economics of the firm.

The rest of this paper is organized as follows. Section 2 describes our samples, data sources, and the measurement of our sentiment variables. Section 3 explores the relation between unexpected component of managerial net optimism and short-window announcement period returns, while in Section 4 we examine the associations between managerial sentiment variables and the post-announcement drift phenomenon. The relation between managerial certainty and abnormal trading volume and idiosyncratic stock price volatilities, respectively, are documented in Section 5. Section 6 provides a summary and conclusion to our study.

2. Sample and Data Description

2.1 Samples

We obtain the text of quarterly earnings announcements for the period of January 1998 through July 2006 from PR Newswire. We are able to match using the ticker symbol and the announcement date (allowing for a 3-day window discrepancy) for 27,705 of the PR Newswire observations with the CRSP/Compustat database (4,771 different firms) and 17,484 of these are matched to IBES (3,372 different firms). Hereafter we refer to these two samples as the "Compustat" and "IBES" samples, respectively. We include only those observations for which we can calculate earnings surprises, 3-day abnormal returns surrounding the earnings announcement, and 60-day abnormal returns both prior to, and post, announcement. We also drop observations with stock prices below $1 and above $10,000 and firms with a negative book value. After imposing all of the preceding restrictions, we are left with a final sample of 4,521 firms (3,113 firms) and 25,481 firm-quarter (16,082 firm-quarter) observations for the Compustat (IBES) sample. In untabulated results we find that the firms in our Compustat sample report slightly higher earnings surprises, have slightly higher ROA, and are larger (on the basis of total sales, market capitalization and total assets) than the firms in the CRSP-Compustat universe. They also have higher P/E ratios (but not market-to-book ratios) and are more likely to report special items such as impairments and restructuring charges than firms in the corresponding population. Our IBES sample firms have slightly lower market-to-book ratios and are slightly larger than the CRSP-Compustat-IBES universe of firms. The IBES sample firms are not significantly different from the corresponding population on the basis of profitability (ROA), earnings surprises, trading volume, or P/E ratios.

Throughout this study, we tabulate and discuss the results of all of our tests using each of the Compustat and IBES samples, respectively, and we do so for several reasons. First, the IBES constraint imposes a bias in favor of the inclusion of firms that are larger and subject to richer information environments, while we are also interested in understanding the role of soft information for the broader universe of firms that are not subject to such high exposure and associated analyst filtering mechanisms. Second, Graham, Harvey and Rajgopal (2006) report that 85.1% of CFO survey respondents considered earnings in the same quarter of the prior year to be the most important earnings benchmark, followed secondly by the analyst consensus estimate at 73.5%. The CFOs interviewed in their study 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 current earnings announcement even if it is not the figure associated with the strongest market response for firms that are tracked by analysts.[7]

2.2 Data

We obtain market values, stock returns, and trading volume from the Center for Research in Security Prices (CRSP) databases. Historical accounting data are obtained from Compustat, while IBES provides the alternative source for historical earnings realizations that are matched to analyst estimates. We obtain media counts from the Factiva database.

Corporate quarterly earnings announcements are provided by PR Newswire, with each firm-quarter's announcement being furnished as an individual text file. Prior to subjecting these files to the linguistic algorithm processing described below, we undertake a number of analyses upon, and make a number of modifications to, the

[7] Although a recent study by Ljungqvist, Malloy and Marston (2007) suggests that the currently available IBES data may be subject to non-random ex post changes, the concerns that those authors raise relate to analyst *recommendations* rather than the analyst *estimates* that we use in our study.

announcements. First, we use keyword searches to develop indicator variables for the presence of an income statement, a balance sheet, and a statement of cash flows, respectively, in each announcement file. Next, we identify tabulated figures in the text (including the financial statements) by searching for strings of numbers, and where identified we cut these elements from the files so that tables of figures are not confounding the textual linguistic analysis. Third, using mechanical search algorithms that we designed based upon extensive manual review of the announcements, we separately remove the company description and "safe harbor" paragraphs from the announcements so that only the earnings announcements themselves remain in the text files to be analyzed.

2.3 Measuring Sentiment

In our primary reported tests, we use version 6.0 of the Diction text-analysis program to measure the level of optimism, pessimism, and certainty, respectively, in managers' earnings announcements. Diction is a well-established language processing algorithm that has been used extensively in prior research to analyze the speeches of Federal Reserve policymakers (Bligh and Hess (2007)), political speeches, corporate annual reports (e.g., Ober, *et al.* (1999); Yuthas, *et al.* (2002)), and earnings announcements (Davis, *et al.* (2007)).[8]

Diction uses a series of thirty-one dictionaries (word-lists) to search text passages for five semantic features: activity, optimism, certainty, realism, and commonality. For our study, we use Diction's "business" normative profile and analyze the earnings announcements using the *optimism* and *certainty* scores. *Optimism* is defined as, "language endorsing some person, group, concept or event or highlighting their positive entailments" while *certainty* is defined as, "language indicating resoluteness, inflexibility, and completeness and a tendency to speak *ex cathedra*" (Digitext Inc. (2000)). The terms associated with each of the characteristics that generate the optimism and certainty scores are reproduced in Appendix A. The Diction formula for *net optimism* is [praise + satisfaction + inspiration]-[blame + hardship + denial]. Similar to prior studies, 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.*" The Diction

[8] See http://www.dictionsoftware.com/files/dictionresearch.pdf for a more extensive summary of published academic studies using the Diction software.

program applies a scoring function which generates a measure designed to capture the average number of words per one hundred words in a firm's quarterly earnings announcement that are, in the case of *optimism,* optimism-increasing (i.e., contained in the praise, satisfaction, or inspiration dictionaries), and for *pessimism* that are optimism-decreasing (i.e., contained in the blame, hardship, or denial dictionary definitions). Thus, our measures of *optimism* and *pessimism* are each bounded by 0 and 100.[9]

The Diction formula for *certainty* is [tenacity + leveling + collectives + insistence] - [numerical terms + ambivalence + self reference + variety]. We redefine this formula to include numerical terms as additive to certainty rather than subtracting them from the score. In order to obtain measures for certainty 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 and dividing the sum through by the maximum value, and hence our *certainty* measure is bounded by zero and one.

In specification checks, we also use the General Inquirer (GI) program introduced into the finance literature by Tetlock (2007), Tetlock, *et al.* (2007), and Engelberg (2007). Specifically, we use GI to measure the *negative* sentiment in the earnings announcement text, which is simply defined as the percentage of negative words from the Harvard IV-4 psychological dictionary to the total words in the announcement text.[10] Simple dictionary definitions of the words "negative" and "pessimism" are different, and consistent with this the dictionary list of words that GI and Diction, respectively, associate with each of these linguistic sentimental constructs is also different. As one would expect, however, GI's *negative* sentiment and Diction's *pessimism* sentiment are correlated measures.

2.4 Descriptive Statistics

Table 1A provides descriptive statistics for the sentiment variables for each of the two samples. As shown, optimism has a mean value of about 1.3 for both samples, while pessimism has a mean value of about 0.6. Net optimism is slightly

[9] Technically speaking, the Diction program generates a score for each of the constructs that we label as "optimism," "pessimism," and "certainty" that represents the number of optimistic and pessimistic words, respectively, per 500 words of text. We divide the Diction scores for "optimism," "pessimism," and "certainty" by 5 in order to arrive at the metrics that we use.

[10] For a more information on the GI program, the reader is referred to Tetlock (2007) or the GI website at: http://www.wjh.harvard.edu/~inquirer/.

lower for the Compustat sample at 0.28 versus 0.30 for the IBES firms. The mean change in net optimism is slightly negative for both samples, consistent with moderate declines in the economy and in consumer confidence over our sample period. Each of the sentiment variables exhibits a considerable range of values in both samples. Table 1B presents the correlation matrix for the sentiment variables. As shown, certainty is not highly correlated with any of the other sentiment measures nor are optimism and pessimism highly correlated with one another or with the earnings surprise variable (SUE).[11] This combination of results seems to suggest that managers, on average, present a discussion in their earnings announcements that is directional (i.e., either optimistic or pessimistic) rather than balanced, and that the sentiment of the press release conveys different information from that conveyed in the hard earnings surprise.

Tables 2A and 2B present descriptive statistics for the firms in our Compustat and IBES samples, respectively. As shown, the quarterly earnings surprise is positive, on average, and larger for both samples when calculated using analyst earnings expectations (SUE_IBES) compared to the surprises generated using a Compustat-based seasonal random walk model (SUE). As expected, the IBES sample is constrained to relatively larger firms that, on average, are followed by more analysts, have higher levels of media coverage, are more liquid, and have higher earnings quality as captured by the lower value of the EFKOS e-loading factor.

3. The Relation Between Sentiment and Announcement Returns

We first investigate the announcement period response to the hard and soft information contained in the earnings announcement. Our dependent variable is defined as the 3-day, size- and book-to-market-adjusted cumulative abnormal returns (CARs) for the period [-1, +1] where 0 is the earnings announcement day, and our key test is whether sentiment is priced by the market incrementally to the hard information contained in the earnings surprise. Similar to the standard specification for earnings surprises, we adopt an expectations model for sentiment in order to attempt to capture the "surprise" element of the level of net optimism contained in management's press release. Only the unexpected component of sentiment should be reflected in the announcement period abnormal returns. Untabulated results show that the level of net

[11] SUE, the standardized earnings surprise, is defined in detail in Section 3.

optimism contained in management's most recent prior quarter's announcement is the best expectation for this quarter's net optimism, and accordingly we use a non-seasonally-adjusted random walk model to calculate the unexpected net optimism as $\Delta NetOpt_{jt}=NetOptimism_{jt}-NetOptimism_{jt-1}$.[12]

To make a meaningful comparison of the estimated surprises across firms and measures, we follow the literature and use standardized surprises. Specifically, we divide the surprise by the firm-specific standard deviation, defining standardized unexpected earnings (hard information) associated with firm *j* at time *t* as

$$SUE_{jt} = \frac{A_{jt} - E_{jt}}{\hat{\sigma}_j},$$

where A_{jt} is the announced earnings per share of firm *j* on day *t*, E_{jt} is either last year's same quarter earnings per share for the Compustat sample (E_{jt-4}) or the IBES median forecast for the IBES sample, and $\hat{\sigma}_j$ is the standard deviation of the forecast error, $A_{jt} - E_{jt}$, estimated using the entire Compustat and IBES sample of observations for each respective firm.[13] We require each firm to have non-missing earnings data in the Compustat and IBES databases, respectively, for 10 quarters.

We similarly define the standardized unexpected sentiment (soft information) associated with firm *j* at time *t* as follows,

$$Sentiment_{jt} = \frac{\Delta NetOpt_{jt} - \mu_{\Delta NetOpt}}{\sigma_{\Delta NetOpt}},$$

---

[12] The adjusted R-squared of the seasonally-adjusted random walk model is 19.55% compared to 24.12% for the non-seasonally-adjusted random walk model. The Akaike and Schwarz information criteria also favor the latter model. This result is in contrast to the earnings per share model selection for which our results are consistent with prior studies; the seasonally-adjusted random walk model outperforms the non-seasonally-adjusted model (the adjusted R-squared of the former is 23.97% compared to 13.90% for the latter). Although using the surprise component in sentiment is the theoretically correct specification and is consistent with the earnings surprise specification, in untabulated results we also find that raw sentiment affects asset prices. The latter result is consistent with the notion that in the announcements managers compare their earnings performance with last year's performance, and hence the raw net optimism expressed is already implicitly relative to an expectational benchmark. The market's response to raw sentiment is, however, weaker than the response to the "surprise" sentiment. We note that Engelberg (2007) and Tetlock (2007) and Tetlock et al. (2007) all analyze raw sentiment in news articles, presumably because the media is reporting "news," and hence it may not be necessary in their setting to estimate the "surprise" in the media's sentiment.

[13] Alternatively, to prevent a hindsight bias, we estimate the standard deviation of the forecast error using the firm's previous 20 quarters of unexpected earnings data following Bernard and Thomas (1989) and Tetlock et al. (2007). We also allow for a trend in unexpected earnings for all firms with more than four years of earnings data. Our results are qualitatively the same when we use these alternative measures.

where $\Delta NetOpt_{jt}$ is the difference between the net optimism of firm *j* in quarter *t*, estimated using either Diction or General Inquirer (GI), and the net optimism of the most recent prior quarter's announcement by firm *j*, $\mu_{\Delta NetOpt}$ and $\sigma_{\Delta NetOpt}$ are equal to the mean and standard deviation of *ΔNetOpt* across all firms and all quarters in our sample. We do not allow the mean and standard deviation of unexpected sentiment to be different across firms in the standardization process because doing so would reduce our sample size to an unacceptable level. Because $\mu_{\Delta NetOpt}$ and $\sigma_{\Delta NetOpt}$ are constant for any firm *j*, the standardization will not affect either the statistical significance of the response estimates or the fit of the regression, however the standardization will facilitate meaningful comparisons of asset price responses to different news items (i.e., soft versus hard information surprises).

3.1 Baseline Announcement Period Pricing Tests

We first examine the relative market responses to standardized earnings and managerial sentiment surprises using a pooled regression model. We expect that the hard earnings news will have a larger effect on asset prices relative to the soft information because the earnings surprises are more uniform, verifiable, have a greater likelihood of being understood, and on an overall basis are likely to be more credible than the non-standardized and somewhat nebulous soft information.[14] We use the following pooled regression to examine whether $\beta_1 > \beta_2$:

$$CAR_{jt} = \beta_0 + \beta_1 SUE_{jt} + \beta_2 \Delta NetOpt_{jt} + \varepsilon_{jt} \quad (1)$$

The results of these tests are reported in Panel A of Table 3. As shown, both the hard earnings surprise ("SUE") and the surprise component of net optimism ("Sentiment") are statistically very significant, both across the Compustat and IBES samples as well as across the alternative GI- and Diction-measures of sentiment. The adjusted R^2 and coefficients on the earnings surprise variables are generally similar to those reported in prior earnings response studies. Thus, consistent with the results of Tetlock, *et al.* (2007) and Engelberg (2007) who find that the stock prices respond to *media* sentiment, our results show that the market also considers sentiment expressed by *management* in their earnings announcements to be credible and informative. As

[14] Although the quarterly earnings figures are not audited, the annual financial results that are the composite of the quarterlies are audited at the end of the firm's fiscal year. Furthermore, the financial results are all prepared in accordance with generally accepted accounting policies ("GAAP"). This standardization combined with the ex post audit of the annual figures enhances understandability and credibility of the hard earnings measures reported by management.

expected, the coefficient on the standardized earnings surprise variable is considerably larger than that on the standardized sentiment surprise. The findings suggest that although the market clearly impounds managerial sentiment quickly into prices, investors nevertheless weight the more objective, hard earnings information much more heavily during the announcement window.

The results of interactively adding *size* to the equation (1) regression are shown in Panel B of Table 3, with size being defined as the natural log of the firm's market capitalization. As shown, the market response to both hard and soft information is lower for larger firms, consistent with the notion that these firms operate in richer information environments, with the result that both the soft and hard news embedded in the firms' earnings announcements are at least partially anticipated by market participants and thus generate a lower announcement period price response.

3.2 The Impact of Firm Characteristics on Announcement Period Response

Table 4 reports the results from a number of specifications that individually include in equation (1), in addition to the size-interacted terms, one more explanatory variable interacted with each of the hard earnings and soft information surprise variables. In what follows, we discuss only the coefficients on the newly added interactive terms unless the results for the earnings surprise, soft information surprise, and size-interacted surprise variables are inconsistent with those reported for the baseline regressions.

In our first test, we consider whether the liquidity of the company's stock impacts the intensity of the market's response to soft and hard information. *Liquidity* is measured as the average of the natural log of the daily volume of shares traded divided by stock outstanding over the 60-day pre- announcement period [-62, -2]. Theory leads us to expect an unambiguously positive sign on the interacted liquidity and sentiment variable in the asset pricing regression. For example, Stocken (2000) shows that the repeated interaction between a manager and investors may be sufficient to support full revelation of the manager's information. This is because the informed agent's possible current gains from opportunistic behavior can be wiped out by future losses in payoff from damaged reputation. However, if there are short-sale constraints on the firm's stock, then investors are constrained on the size of the punishment that they can inflict upon managers who issue confounded or misleading sentiment in their announcements, and accordingly managers of short-sale constrained firms may have an incentive to provide noisy soft information in their earnings releases. Our findings, shown in Table 4 Panel A, of significant positive coefficients on the *SentimentXLiquidity* variable for both samples are consistent with this notion that managerial press releases are repeated games with the punishment from issuing announcements with misleading or confounding sentiment being too high, on average, to elicit cheap talk from the managers of more liquid firms.[15] An alternative explanation for our findings is that liquidity enables more rapid and intensive price responses in general, independent of the notion of cheap talk that could be present in manager's relatively costless earnings discussions. However, the insignificant coefficient on liquidity interacted with SUE, a hard information variable that is not subject to cheap talk, is inconsistent with this alternative explanation. We therefore conclude that liquidity induces more truthful sentiment rather than cheap talk, and this sentiment is priced by the market.

We next examine whether *analyst coverage*, which we empirically operationalize as the log of the number of analysts covering the firm, has an impact the market's response to managerial sentiment. Table 4 Panel B shows that the interaction of analyst coverage with sentiment surprise is positive and significant in both the Compustat and IBES regressions, suggesting that firms that are more heavily followed have higher price responses to managerial sentiment incremental to the size effect. We interpret these results in a manner that is consistent with those for liquidity; we infer that greater scrutiny by analysts, combined with the potential for two-way communication between analysts and managers, induces more truthful sentiment rather than cheap talk. If the results were simply consistent with the notion that firms with higher levels of analyst following are informationally more efficient (i.e., impound information more quickly into prices), we would expect a symmetrical result on the SUE term interacted with analyst coverage as previously argued for liquidity. However, the SUE interacted term is once again insignificant for both samples.

Table 4 Panel C presents the results where *numerical terms*, measured as the simple count of the number of numerical terms in the announcement, are interactively

[15] An alternative interpretation of this result is that voluntary disclosure increases the liquidity of the stock, so that firms that provide useful soft information are also more liquid. However, we find that after controlling for other proxies for the degree of voluntary disclosure (i.e., the number of words in the announcement, the number of numerical terms, and the presence of financial statements), the soft information of more liquid firms still affects asset prices more than soft information of less liquid firms.

included in the announcement period returns regression.[16] For the Compustat sample, we find that sentiment interacted with numerical terms is positive and significant, consistent with the notion that providing more detailed, precise, and hard information (i.e., numbers) enhances the credibility of the sentiment concurrently expressed in the announcement, resulting in the sentiment being priced more.

Table 4 Panel D shows the differential market response to high-tech versus non-tech firms, where the high tech dummy variable is set equal to one for firms that fall into the Fama and French (1988) high tech industry portfolio definition.[17] The results suggest that there is a stronger market response to high tech firms' unexpected sentiment, although this is just barely significant for the Compustat sample. Our findings here are consistent with the notion that soft verbal data may be more important to conveying information to market participants in settings where the hard accounting data is less informative about the firm's economic realizations.[18] The finding is also consistent with the documented higher price responsiveness of "glamour" stocks to earnings news (Skinner and Sloan (2002)) and consistent with the results for our IBES sample's higher high tech firm response to earnings surprises.

Table 4E pursues the notion of earnings informativeness by examining the regression that interactively includes a measure of earnings quality, which we estimate using the data and methodology provided by Ecker, Francis, Kim, Olsson and Schipper (2006) (hereafter "EFKOS"). Higher values for the *EFKOS e-loading* factor represent lower levels of earnings quality. Lower earnings quality as captured by the EFKOS factor may result from factors that are innate to the firm's business model or from managerial attempts at obfuscation. Our results suggest that for lower earnings quality firms, the market response to sentiment is stronger, consistent with the notion that management provides soft information in order to compensate for hard accounting data that do not adequately capture the underlying economics of their

[16] Neither Diction nor GI consider numerical terms to be words, and hence the numerical terms are not included as part of the total word count. In untabulated results, we find that numerical terms and total words are negatively correlated.
[17] Specifically, high tech firms include those with dnums of 3570-3579, 3622, 3660-3692, 3694-3699, 3810-3839, 7370-7372, 7373-7379, 7391, or 8730-8734.
[18] Lev and Zarowin (1999) argue that accounting data are less useful for firms engaged in innovative activities or that operate in more dynamic environments. One reason for this is that US accounting regulations require that firms expense all expenditures on R&D, leading to depressed accounting earnings even though the R&D expenditures may represent valuable investments in intangible assets that will generate growth in revenues and earnings into the future.

firm's activities. Surprisingly, however, the market also weights earnings more heavily for firms that have higher EFKOS factors (i.e., lower quality earnings).

In untabulated results, we have also investigated the differential responsiveness of price to sentiment interacted with the market-to-book ratio as a proxy for "glamour" stocks[19], media coverage as an alternative proxy for the richness of the firm's information environment, and the level of competition in the industry as captured by a Herfindahl index. None of these sentiment-interacted variables are statistically significant.

4. The Predictive Role of Sentiment for Post-Announcement Drift

4.1 The Relation Between Sentiment and Post-Announcement Returns

In this section we examine whether the sentiment contained in the firms' earnings press release is related to the post-announcement drift phenomenon. For these tests, our dependent variable is defined as the 60-trading-day, size- and book-to-market-adjusted cumulative abnormal returns (CARs) for the period [+2, +62] relative to t=0 as the earnings announcement day. The results for the baseline post-earnings announcement drift ("PEAD") analyses are presented in Table 5. As shown, sentiment is positive in all of the panels although its significance is attenuated in the IBES sample regressions when the surprise in net optimism is measured using the Diction algorithm. Our results for management-expressed surprise net optimism complement those of Engelberg (2007) who finds that media expressed negative sentiment is associated with post-announcement returns. In untabulated analyses, we also consider the interactive role of liquidity, analyst coverage, numerical terms, high tech industry participation, and earnings quality ("EFKOS") on the predictive role of sentiment for post-announcement returns, however none of these interacted terms are significant.

4.2 Hedge Returns and the Sentiment-PEAD Relation

Based upon the previous results suggesting a predictive role for sentiment in the post-announcement period, we document the pseudo-hedge returns available from a

[19] We also considered using the price-to-earnings ratio as a proxy for glamour stocks, however earnings are often negative and thus this ratio is not well defined. Since our event window returns dependent variable is adjusted for the book-to-market effect, it's not entirely unexpected that the interaction of sentiment surprise and the market-to-book ratio turns up insignificant in the regression.

sentiment-based trading strategy. Panel A of Table 6 presents benchmark hedge returns from going long (short) in firms in the highest (lowest) SUE terciles. Panel B of Table 6 presents the hedge results from going long (short) in firms that fall into both the highest (lowest) hard earnings surprise and highest (lowest) soft sentiment surprise terciles. Both hedge strategies are implemented on a size (i.e., market capitalization) stratified basis, with large firms being defined as those in the 9th and 10th deciles, medium firms coming from deciles 6 through 8, and small firms consisting of the remaining firms from deciles 1 through 5. As shown in the furthest right-hand column of both panels, the hedge returns available from small- and medium-sized firm portfolios are statistically and economically significant for both the SUE and combined SUE-sentiment portfolio sorts (i.e., ranging from 1.7% to 4.1% for a 60-day holding period, or 10.6% to 27.3% annually). However, the returns available from the combined soft and hard earnings news strategy are considerably higher than those from the SUE only strategy for both the medium firm portfolio (13.3% versus 10.6% annualized) and especially for the small firm portfolio (19.4% versus 27.3% on an annualized basis).

5. The Association Between Sentiment and Investor Activity

We now consider whether unexpected net optimism as well as another linguistic characteristic, *certainty*, is related to two non-return measures of market activity, trading volume and volatility.

5.1 Sentiment and Trading Volume

Tetlock (2007) examines the relation between the pessimism score of a particular daily Wall Street Journal column and NYSE trading volume. At an aggregate level, he documents that higher levels of pessimism in this one media source are associated with increased trading volume. We investigate whether this result holds on a firm-specific basis in the context of management-initiated communications, after controlling for a number of additional determinants of firm-specific abnormal trading activity. In addition, we hypothesize that higher levels of wavering in management's communications will generate greater trading volume due to the greater opacity of the messages being sent. We test for each of these hypothesized relations using the following specification:

$$\sum_{i=-1}^{1} AbnormalTurnover_{jt+i} = \lambda_0 + \lambda_1 \sum_{i=2}^{4} AbnormalTurnover_{jt-i} + \lambda_2 \left|SUE_{jt}\right| + \lambda_3 \left|SUE_{jt}\right| \times I\left(SUE_{jt} < 0\right) \quad (2)$$
$$+ \lambda_4 \left|Sentiment_{jt}\right| + \lambda_5 \left|Sentiment_{jt}\right| \times I\left(Sentiment_{jt} < 0\right) + \lambda_6 certa\operatorname{int} y_{jt} + \overline{\lambda}_7 \overline{\mathrm{Y}}_{jt}$$

Following Bailey, Li, Mao and Zhong (2003), we define *Abnormal Turnover* to be the difference between the mean daily turnover during the event window (-1,+1) and the mean of daily turnover for that stock over the pre-announcement window (-200, -11), normalized by the mean turnover. 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.[20] Because abnormal turnover is highly serially correlated, we first control for past abnormal turnover. Consistent with prior literature, we also include the absolute value of the hard earnings surprise variable and analogously we include the absolute value of the competing soft information surprise measure.[21] Negative shocks may have a larger impact on the volatility of stock returns than positive shocks of the same absolute value, a phenomenon that is most often interpreted as the leverage effect unveiled by Black (1976)). Several GARCH volatility models allow for this effect, including the EGARCH model of Nelson (1991), and the GJR model of Glosten, Jagannathan and Runkle (1993), amongst others. Accordingly, we included a separate term capturing |SUE| interacted with an indicator variable for negative earnings surprises. We enter the competing soft information surprise variable into the regression in an analogous manner by including both the absolute value of *Sentiment* and allowing for a different coefficient on the |*Sentiment*| for negative surprises in sentiment. The *certainty* variable is as previously defined, and $\overline{\mathrm{Y}}_{jt}$ is a vector of additional control variables.

The results of the regression depicted by equation (2) are presented in Table 7. As expected from extensive prior literature, |SUE| is significantly positively associated with announcement period turnover. Of interest in this study is whether

[20] As a specification check, we estimate equation (3) separately for NYSE/AMEX and Nasdaq firms using unadjusted data and find that the separate sample results are similar in all material respects to the tabulated pooled results.

[21] Negative shocks may have a larger impact on the volatility of stock returns than positive shocks of the same absolute value, a phenomenon that is most often interpreted as the leverage effect unveiled by Black (1976)). Several GARCH volatility models allow for this effect, including the EGARCH model of Nelson (1991), and the GJR model of Glosten, Jagannathan and Runkle (1993), amongst others. In untabulated results we included a separate term capturing |SUE| interacted with an indicator variable for negative earnings surprises. The coefficient on this variable was far from significant, and hence for parsimony in the tabulated results we include only |SUE|. We enter the competing soft information surprise variable into the regression in an analogous manner by including both the absolute value of *Sentiment* and allowing for a different coefficient on the |*Sentiment*| for negative surprises in sentiment.

trading volume is also associated with unexpected sentiment. For the portfolio level media results of Tetlock (2007) to hold for management communications at the firm level, we would expect to observe a positive coefficient on the |*Sentiment*| variable. Consistent with this, we find that |*Sentiment*| is highly significant for both the Compustat and IBES samples, even in the presence of additional firm-level control variables. We also find that *certainty* is significantly negatively associated with abnormal turnover, suggesting that management communications that are more wavering in nature generate greater dispersion in investor beliefs as manifested in higher levels of trading volumes. Consistent with the notion of possibly obfuscated communications causing disagreement, we find that firms with lower earnings quality (i.e., a higher EFKOS e-loading factor) also exhibit higher abnormal turnover. Higher levels of media coverage, greater analyst following, and the provision of more financial statements also lead to more trading, while longer press releases (measured by the log of the total words in the announcement) are associated with lower levels of investor disagreement.

Table 8 presents the results from a parallel analysis that examines whether |*Sentiment*| and *certainty* are associated with post-announcement abnormal turnover. After controlling for past abnormal turnover and standard hard information control variables, we find that the soft information is not significant in the prediction model.

5.2 The Relation Between Sentiment and Announcement Idiosyncratic Volatility

Following prior research, we adopt the abnormal return variance around the earnings announcement date as an alternative measure of the information content of the earnings release, where higher variance is indicative of greater information content of the announcement (e.g., Beaver (1968); Landsman and Maydew (2002); DeFond, *et al.* (2007)). We define *volatility* as the log of the sum of the squared abnormal returns over the event window of [-1, +1], where abnormal returns are calculated as the firm's own daily return minus the contemporaneous return on a size- and book-to-market-matched portfolio. We test for the association between our measures of sentiment and volatility using the following specification:

$$\log(\sum_{i=-1}^{1} AR_{jt+i}^{2}) = \gamma_0 + \gamma_1 \log(\sum_{i=2}^{4} AR_{jt-i}^{2}) + \gamma_2 \left|SUE_{jt}\right| + \gamma_3 \left|SUE_{jt}\right| \times I\left(SUE_{jt} < 0\right) + \gamma_4 \left|Sentiment_{jt}\right| + \gamma_5 \left|Sentiment_{jt}\right| \times I\left(Sentiment_{jt} < 0\right) + \gamma_6 certainty_{jt} + \bar{\gamma}_7 \overline{\mathrm{Y}}_{jt} \quad (3)$$

The results from this regression are presented in Table 9. As expected, past volatility is the most important explanatory variable for announcement period volatility, and volatility is also an increasing function of the absolute value of SUE. Similar to the results for abnormal announcement period volume, we find that the soft information *certainty* variable is significantly negatively associated with variance. This finding once again suggests that higher levels uncertainty in the manager's press release, whether driven by the firm's business fundamentals or management's attempts at obfuscation using wavering language, lead to greater volatility in the firm's share price during the 3-day announcement period window. Also similar to the volume results, the surprise in net optimism is a significant determinant of announcement interval volatility, but only for the Compustat sample. Similar to the results for volume, firms with lower quality earnings (i.e., higher EFKOS e-loading factors), higher levels of media exposure, higher levels of analyst following, and that voluntarily provide more financial statements all have higher levels of abnormal volatility, while large firms, firms with longer press releases, and firms issuing announcements in the post-RegFD period (only for the Compustat sample) exhibit lower levels of volatility even after controlling for the general time trend in (declining) volatility.

5.3 Sentiment as a Predictor of Post-Announcement Idiosyncratic Volatility

In Table 10 we report the results for equation (3) estimated using the 60-day post-announcement period [+2, +62] stock volatility and using past volatility measured over [-62,-2] as the first control variable. Similar to the announcement period volatility results and the longer horizon turnover results, *certainty* takes a negative coefficient and is a significant predictor of volatility for both samples. In contrast to the post-announcement turnover results, however, we find here that unexpected net optimism is significantly associated with post-announcement idiosyncratic for both samples. Thus, not only are the soft information variables significantly related to contemporaneous announcement period idiosyncratic volatility, they are also leading indicators for post-announcement volatility. Most of the results for the control variables in the longer window are similar to those in the announcement period, except that the time trend variable is negative and significant (consistent with declining idiosyncratic volatility during the period covered by our data), and the *REG_FD* variable and the length of the announcement are no longer significant.

Our current tests do not attempt to distinguish between whether high *certainty* scores indicate deliberate managerial obfuscation that is manifested in higher levels of investor disagreement, or whether the managers' language is faithfully characterizing a more uncertain environment for the firm, which in turn drives dispersion in beliefs and hence abnormal volume and volatility. Future research will attempt to sort between these two competing explanations.

6. Summary and Conclusion

Prior research has established a general link between soft information, captured with the linguistic measures of negativity or pessimism, and stock returns. Most of this prior literature examines *media*-released information. Our study compliments and extends the existing studies by examining the conditions under which *management*-issued soft information is incorporated into prices, both in the short-window announcement period and in the intermediate term, post-announcement interval. We find that under conditions that induce truthful sentiment rather than cheap talk (i.e., stock liquidity and analyst scrutiny), the market responds more to the surprise sentiment in managerial announcements. We also find that in circumstances where the hard earnings information is less informative (i.e., for high tech firms with more complex business models and for firms with lower earnings quality as captured by the EFKOS e-loading factor), the market responds more to the complimentary soft information. We extend the soft information literature further by examining the role of another linguistic measure, certainty, in explaining announcement period and post-announcement abnormal trading volume (a measure of dispersion in investor beliefs) and abnormal idiosyncratic volatility. We find that the level of certainty expressed in management's earnings announcement is inversely related to trading volume and idiosyncratic volatility during the announcement period, and further that managerial uncertainty is a leading indicator for post-announcement 60-day abnormal trading and volatility. Taken together, our findings suggest that the soft information conveyed by management plays an important role in the price discovery process by complementing the simultaneously released hard earnings news that has been the subject of decades of prior research.

Appendix A

Diction 5.0 *Optimism* and *Certainty* Terms and Dictionary Definitions

Reproduced from <u>Diction 5.0: The Text-Analysis Program</u>, © 2000 Digitext, Inc.

<u>The *Optimism* Score</u>

Definition: Language endorsing some person, group, concept or event or highlighting their positive entailments.

Formula: [Praise + Satisfaction + Inspiration] – [Blame + Hardship + Denial]

PRAISE: Affirmations of some person, group, or abstract entity. Included are terms isolating important social qualities (*dear, delightful, witty*), physical qualities (*mighty, handsome, beautiful),* intellectual qualities (*shrewd, bright, vigilant, reasonable),* entrepreneurial qualities (*successful, conscientious, renowned*), and moral qualities (*faithful, good, noble*). All terms in this dictionary are adjectives.

SATISFACTION: Terms associated with positive affective states (*cheerful, passionate, happiness*), with moments of undiminished joy (*thanks, smile, welcome*) and pleasurable diversion (*excited, fun, lucky),* or with moments of triumph (*celebrating, pride, auspicious).* Also included are words of nurturance: *healing, encourage, secure, relieved.*

INSPIRATION: Abstract virtues deserving of universal respect. Most of the terms in this dictionary are nouns isolating desirable moral qualities (*faith, honesty, self-sacrifice, virtue*) as well as attractive personal qualities (*courage, dedication, wisdom, mercy).* Social and political ideals are also included: *patriotism, success, education, justice*.

BLAME: Terms designating social inappropriateness (*mean, naive, sloppy, stupid*) as well as downright evil (*fascist, blood-thirsty, repugnant, malicious*) compose this dictionary. In addition, adjectives describing unfortunate circumstances (*bankrupt, rash, morbid, embarrassing*) or unplanned vicissitudes (*weary, nervous, painful, detrimental)* are included. The dictionary also contains outright denigrations: *cruel, illegitimate, offensive, miserly.*

HARDSHIP: This dictionary contains natural disasters (*earthquake, starvation, tornado, pollution),* hostile actions (*killers, bankruptcy, enemies, vices*) and censurable human behavior (*infidelity, despots, betrayal*). It also includes unsavory political outcomes (*injustice*, *slavery, exploitation, rebellion*) as well as normal human fears (*grief, unemployment, died, apprehension*) and incapacities (*error*, *cop-outs, weakness*).

DENIAL: A dictionary consisting of standard negative contractions (*aren't, shouldn't, don't*), negative functions words (*nor, not, nay*), and terms designating null sets (*nothing, nobody, none*).

The *Certainty* Score

Definition: Language indicating resoluteness, inflexibility, and completeness and a tendency to speak *ex cathedra*.

Formula: [Tenacity + Leveling + Collectives + Insistence]
- [Numerical Terms + Ambivalence + Self Reference + Variety][22]

TENACITY: All uses of the verb to be (is, am, will, shall), three definitive verb forms (has, must, do) and their variants, as well as all associated contractions (he'll, they've, ain't). These verbs connote confidence and totality.

LEVELING: Words used to ignore individual differences and to build a sense of completeness and assurance. Included are totalizing terms (everybody, anyone, each, fully), adverbs of permanence (always, completely, inevitably, consistently), and resolute adjectives (unconditional, consummate, absolute, open-and-shut).

COLLECTIVES: Singular nouns connoting plurality that function to decrease specificity. These words reflect a dependence on categorical modes of thought. Included are social groupings (crowd, choir, team, humanity), task groups (army, congress, legislature, staff) and geographical entities (county, world, kingdom, republic).

INSISTENCE: This is a measure of code-restriction and semantic contentedness. The assumption is that repetition of key terms indicates a preference for a limited, ordered world. In calculating the measure, all words occurring three or more times that function as nouns or noun-derived adjectives are identified (either cybernetically or with the user's assistance) and the following calculation performed: [Number of Eligible Words x Sum of their Occurrences] ÷ 10. (For small input files, high frequency terms used two or more times are used in the calculation).

NUMERICAL TERMS: Any sum, date, or product specifying the facts in a given case. This dictionary treats each isolated integer as a single word and each separate group of integers as a single word. In addition, the dictionary contains common numbers in lexical format (one, tenfold, hundred, zero) as well as terms indicating numerical operations (subtract, divide, multiply, percentage) and quantitative topics (digitize, tally, mathematics). The presumption is that Numerical Terms hyper-specify a claim, thus detracting from its universality.

AMBIVALENCE: Words expressing hesitation or uncertainty, implying a speaker's inability or unwillingness to commit to the verbalization being made. Included are hedges (allegedly, perhaps, might), statements of inexactness (almost, approximate, vague, somewhere) and confusion (baffled, puzzling, hesitate). Also included are

[22] The formula we use for our "certainty" variable is [Tenacity + Leveling + Collectives + Insistence + Numerical Terms] - [Ambivalence + Self Reference + Variety]

words of restrained possibility (could, would, he’d) and mystery (dilemma, guess, suppose, seems).

SELF-REFERENCE: All first-person references, including I, I’d, I’ll, I’m, I’ve, me, mine, my, myself. Self-references are treated as acts of indexing whereby the locus of action appears to reside in the speaker and not in the world at large (thereby implicitly acknowledging the speaker s limited vision).

VARIETY: This measure conforms to Wendell Johnson’s (1946) Type-Token Ratio which divides the number of different words in a passage by the passage’s total words. A high score indicates a speaker’s avoidance of overstatement and a preference for precise, molecular statements.

Appendix B
Summary of Variable Definitions

Variable	Definition
Sentiment Variables	
Optimism	Percentage number of words in a firm's quarterly earnings announcement that are optimism-increasing (i.e., contained in the praise, satisfaction, or inspiration dictionaries)
Pessimism	Percentage number of words in a firm's quarterly earnings announcement that are optimism-decreasing (i.e., contained in the blame, hardship, or denial dictionary definitions)
Net Optimism	Optimism minus pessimism
Certainty	Certainty, is a normalized variable that indicates the degree of "resoluteness", "inflexibility", and "completeness" in the firm's quarterly earnings announcement. We redefine the Diction 2.0 definition of certainty to be [Tenacity + Leveling + Collectives + Insistence + Numerical Terms] - [Ambivalence + Self Reference + Variety]
Other Variables	
SUE	Earnings surprise = $\frac{actual - forecast}{std(actual - forecast)}$
CARs	Size- and book-to-market-adjusted cumulative abnormal returns defined alternatively over the earnings announcement window [t-1,t+1] or the post-announcement drift period [t+2,t+62] relative to the *t=0* earnings announcement day.
Abnormal turnover	Following Bailey, Li, Mao and Zhong (2003), we define *Abnormal Turnover* to be the difference between daily turnover and the mean of daily turnover for that stock over the pre-announcement window (-200, -11), normalized by the mean turnover. We then sum abnormal turnover over the event windows [t-1,t+1] and [t+2, t+62].
Volatility	We measure the volatility of abnormal returns during the event window as the logarithm of the sum of squared abnormal returns during the [t-1, t+1] and [t+2, t+62] event windows.
Time Trend	= 1 for 1st calendar quarter of 1998, increased by 1 for each calendar quarter thereafter
RegFD	Indicator=1 for firm quarters that occur after October 23, 2000
Financial Statements	We create a count variable, *Financial Statements*, which is incremented by one for each voluntary disclosure of the following items within the earnings announcements: a cash flow statement, an income statement, a balance sheet, and a tabulated summary of financial highlights
Total Words	Natural log of the total number of words contained in the earnings announcement
Log(Analyst+1)	Analyst is computed using IBES data and it is equal to the number of analysts that post an earnings estimate for the current quarter
Forecast Dispersion	We use IBES to estimate this variable, and define it as the standard deviation of forecasts across analysts divided by the absolute value of the median forecast. We require firms to at least have two forecast estimates.
Liquidity	The average of the natural log of (the daily volume of shares traded divided by stock outstanding) 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.
Recent Media Coverage	The number of times a firm is mentioned in the headline or lead paragraph of an article from newswire services in the previous ten days before the earnings announcement date [t-12,t-2]. We only take into account publications that have over 500,000 current subscribers using Factiva. The list of data sources is: The Wall Street Journal (all editions), Associated Press Newswire, the Chicago Tribune, the Globe and Mail, Gannett News Service, the Los Angeles Times, the New York Times, the Washington Post, USA Today and all Dow Jones newswires.
High Tech	Indicator = 1 if dnum 3570-3579, 3622, 3660-3692, 3694-3699, 3810-3839, 7370-7372, 7373-7379, 7391, 8730-8734
EFKOS e-Loading	Is obtained by regressing the daily excess return of firm *i* on EFKOS factor as well as the Fama-French three factors (SML, HML, Market Return). We allow the loading to change over time and estimate the coefficient using all non-earnings announcement days in the previous 365 calendar days (only for stocks with at least 100 data points in that period) before the earnings announcement.

References

Bailey, Warren, Haitao Li, Connie Mao, and Rui Zhong, 2003, Regulation Fair Disclosure and Earnings Information: Market, Analyst, and Corporate Responses, *Journal of Finance* 58, 2487-2514.

Ball, Ray, and Phillip Brown, 1968, An Empirical Evaluation of Accounting Income Numbers, *Journal of Accounting Research* 6, 159-78.

Beaver, William H., 1968, The Information Content of Annual Earnings Announcements, *Journal of Accounting Research* Supplement, 67-92.

Bhattacharya, Nilabhra, Ervin Black, Theodore Christensen, and Chad Larson, 2003, Assessing the Relative Informativeness and Permanence of Pro Forma Earnings and GAAP Operating Earnings, *Journal of Accounting and Economics* 36, 285-319.

Bligh, Michelle, and Gregory Hess, 2007, The Power of Leading Subtly: Alan Greenspan, Rhetorical Leadership, and Monetary Policy, *The Leadership Quarterly* 18, 87-104.

Bradshaw, Mark, and Richard Sloan, 2002, GAAP versus The Street: An Empirical Assessment of Two Alternative Definitions of Earnings, *Journal of Accounting Research* 40, 41-66.

Burgstahler, David C., and Ilia D. Dichev, 1997, Earnings Management to Avoid Earnings Decreases and Losses, *Journal of Accounting and Economics* 24, 99-126.

Crawford, Vincent, and Joel Sobel, 1982, Strategic Information Transmission, *Econometrica* 50, 1431-1451.

Das, Sanjiv, and Mike Chen, 2007, Yahoo! for Amazon: Sentiment Extraction From Small Talk on the Web, *Management Science* 53, 1375-1388.

Das, Sanjiv, Asis Martinez-Jerez, and Peter Tufano, 2005, e-Information: A Clinical Study of Investor Discussion and Sentiment, *Financial Management* 34, 103-137.

Davis, Angela, Jeremy Piger, and Lisa Sedor, 2007, Beyond the Numbers: Managers' Use of Optimistic and Pessimistic Tone in Earnings Press Releases, *unpublished working paper*.

DeFond, Mark, Mingyi Hung, and Robert Trezevant, 2007, Investor Protection and the Information Content of Annual Earnings Announcements: International Evidence, *Journal of Accounting & Economics* 43, 37-67.

Degeorge, Francois, Jayendu Patel, and Richard Zeckhauser, 1999, Earnings Management to Exceed Thresholds, *Journal of Business* 72, 1-33.

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

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

Ehrmann, Michael, and Marcel Fratzscher, 2007, Communication by Central Bank Committee Members: Different Strategies, Same Effectiveness, *Journal of Money, Credit & Banking* 39, 509-541.

Engelberg, Joseph, 2007, Costly Information Processing: Evidence from Earnings Announcements, *unpublished working paper, Northwestern University*.

Fama, Eugene, and Kenneth French, 1988, Permanent and Temporary Components of Stock Prices, *Journal of Political Economy* 96, 246-273.

Francis, Jennifer, Katherine Schipper, and Linda Vincent, 2002, Expanded Disclosures and Increased Usefulness of Earnings Announcements, *The Accounting Review* 77, 515-546.

Graham, John, Campbell Harvey, and Shiva Rajgopal, 2006, Value Destruction and Financial Reporting Decisions, *Financial Analysts Journal* 62, 27-39.

Hutton, Amy P., Gregory Miller, and Douglas Skinner, 2003, The Role of Supplementary Statements with Management Earnings Forecasts, *Journal of Accounting Research* 41, 867-890.

Krishna, Vijay, and John Morgan, 2008, Cheap Talk, in Steven N. Durlauf, and Lawrence E. Blume, eds.: *The New Palgrave Dictionary of Economics* (Palgrave Macmillan).

Landsman, Wayne, and Edward Maydew, 2002, Has the Information Content of Quarterly Earnings Announcements Declined in the Past Three Decades?, *Journal of Accounting Research* 40, 797-808.

Lev, Baruch, and Zarowin, 1999, The Boundaries of Financial Reporting and How to Extend Them, *Journal of Accounting Research* 37, 353-385.

Li, Feng, 2006, Do Stock Market Investors Understand the Risk Sentiment of Corporate Annual Reports?, *unpublished working paper, University of Michigan*.

Ljungqvist, Alexander, Christopher Malloy, and Felicia Marston, 2007, Rewriting History, *working paper, NYU*.

Ober, S., J. Zhao, R. Davis, and M. Alexander, 1999, Telling It Like It Is: The Use of Certainty In Public Business Discourse, *The Journal of Business Communication* 36, 280-300.

Petersen, Mitchell, 2004, Information: Hard and Soft, *unpublished working paper, Northwestern University*.

Schrand, Catherine M, and Beverly Walther, 2000, Strategic Benchmarks in Earnings Announcements: The Selective Disclosure of Prior-Period Earnings Components, *The Accounting Review* 75, 151-177.

Skinner, Douglas, and Richard Sloan, 2002, Earnings Surprises, Growth Expectations, and Stock Returns or Don't Let an Earnings Torpedo Sink Your Portfolio, *Review of Accounting Studies* 7, 289.

Stein, Jeremy C., 2002, Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms, *Journal of Finance*.

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

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy, 2007, More Than Words: Quantifying Language to Measure Firms' Fundamentals, *Journal of Finance* forthcoming.

Yuthas, K., R. Rogers, and J. Dillard, 2002, Communicative Action and Corporate Annual Reports, *Journal of Business Ethics* 41, 141-157.

Table 1A: Sentiment Sample Statistics

In this table we present summary statistics for the following variables estimated using earnings press releases and Diction 5.0 textual-analysis program: Optimism, the average number of words per one hundred words in a firm's quarterly earnings announcement that are optimism-increasing; pessimism, the average number of words per one hundred words in a firm's quarterly earnings announcement that are optimism-decreasing; Netopt, optimism minus pessimism; ΔNetOpt, change in Netopt from this quarter to the previous quarter; and certainty, is a normalized variable that indicates the degree of "resoluteness", "inflexibility", and "completeness" in the firm's quarterly earnings announcement. The summary statistics are calculated using 3,764 (2,610) firms sampled during earnings announcement dates from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. For a detailed description of the variables please refer to the Appendix.

	Mean	Std. Dev.	25 Percentile	75 Percentile
		Compustat Sample		
Optimism	1.3032	0.9806	0.6250	1.7654
Pessimism	0.6435	0.6321	0.2000	0.9000
NetOpt	0.2825	0.4663	0.0350	0.5202
ΔNetOpt	-0.0071	0.4850	-0.2388	0.2216
Certainty	0.3110	0.1013	0.2383	0.3728
		IBES Sample		
Optimism	1.3151	0.9686	0.6444	1.7708
Pessimism	0.6065	0.5867	0.2000	0.8494
NetOpt	0.3006	0.4503	0.0514	0.5304
ΔNetOpt	-0.0088	0.4676	-0.2282	0.2152
Certainty	0.3120	0.0985	0.2413	0.3721

Table 1B: Sentiment Correlation Matrix

We estimate the average quarterly bi-variate correlations of the variables used in the empirical tests. The bivariate correlations are calculated using 3,764 (2,610) firms sampled during earnings announcement dates from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. Optimism is the average number of words per one hundred words in a firm's quarterly earnings announcement that are optimism-increasing; pessimism is the average number of words per one hundred words in a firm's quarterly earnings announcement that are optimism-decreasing; Netopt is optimism minus pessimism; ΔNetOpt is the change in Netopt from this quarter to the previous quarter; and certainty, is a normalized variable that indicates the degree of "resoluteness", "inflexibility", and "completeness" in the firm's quarterly earnings announcement. For a detailed description of the variables please refer to the Appendix.

	Optimism	Pessimism	NetOpt	ΔNetOpt	Certainty	SUE
			Compustat Sample			
Optimism	1					
Pessimism	-0.1075	1				
NetOpt	0.8262	-0.6272	1			
ΔNetOpt	0.4376	-0.297	0.5125	1		
Certainty	-0.1639	-0.0596	-0.1045	-0.0643	1	
SUE	0.0406	-0.0822	0.0728	0.0558	0.0403	1
			IBES Sample			
Optimism	1					
Pessimism	-0.1097	1				
NetOpt	0.8394	-0.6103	1			
ΔNetOpt	0.4261	-0.2887	0.4983	1		
Certainty	-0.1513	-0.0658	-0.0946	-0.0518	1	
SUE	0.0639	-0.1075	0.1025	0.0580	0.0319	1

Table 2A: Descriptive Statistics – Compustat Sample

In this table we present summary statistics for the variables used in the empirical tests. The variables are defined in Appendix B and the summary statistics are calculated using 3,764 firms sampled during earnings announcement dates from January, 1998 to July, 2006, and in total there are 21,580 firm-quarter observations.

	Mean	Std. Dev.	25 Percentile	75 Percentile
3-Day CAR	0.0031	0.0875	-0.0312	0.0352
Post announcement 60-Day CAR	0.0008	0.1865	-0.0788	0.0765
SUE	0.0559	1.0005	-0.3147	0.4577
SUE_IBES	0.2455	1.1522	-0.1618	0.7201
NEGV	0.2310	0.4215	0.0000	0.0000
Log(Market Capitalization)	19.8413	1.9443	18.4475	21.1119
Analyst following	3.7087	5.0108	0.0000	6.0000
Log(1+analyst)	1.0612	0.9808	0.0000	1.9459
Liquidity	0.6292	0.6715	0.2013	0.7987
Media Coverage	1.1914	1.7018	0.0000	2.0000
EFKOS e-Loading factor	0.0970	0.5024	-0.1951	0.2877
Hightech indicator variable	0.1627	0.3691	0.0000	0.0000
Numerical Terms	81.3385	30.2652	58.9110	100.4960
Financial Statements	1.7582	0.9497	1.0000	2.0000
Total Words	824.7232	556.2340	440.0000	1064.0000
Log(Total Words)	6.5124	0.6503	6.0868	6.9698

Table 2B: Descriptive Statistics – IBES Sample

In this table we present summary statistics for the variables used in the empirical tests. The variables are defined in Appendix B and the summary statistics are calculated using 2,610 firms sampled during earnings announcement dates from January, 1998 to July, 2006, and in total there are 13,907 firm-quarter observations.

	Mean	Std. Dev.	25 Percentile	75 Percentile
3-Day CAR	0.0020	0.0813	-0.0316	0.0360
Post announcement 60-Day CAR	0.0027	0.1615	-0.0695	0.0750
SUE	0.0666	0.9993	-0.3084	0.4742
SUE_IBES	0.2459	1.1507	-0.1625	0.7212
NEGV	0.2063	0.4047	0.0000	0.0000
Log(Market Capitalization)	20.4897	1.6846	19.2942	21.5673
Analyst following	5.8174	5.1841	2.0000	8.0000
Log(1+analyst)	1.6694	0.7009	1.0986	2.1972
Liquidity	0.7380	0.6798	0.2941	0.9410
Media Coverage	1.3809	1.7857	0.0000	2.0000
EFKOS e-Loading factor	0.0444	0.4554	-0.2345	0.2260
Hightech indicator variable	0.1689	0.3747	0.0000	0.0000
Numerical Terms	83.3340	30.3325	60.8610	102.5450
Financial Statements	1.9013	0.9105	1.0000	3.0000
Total Words	888.1782	580.4106	483.0000	1151.0000
Log(Total Words)	6.5932	0.6448	6.1800	7.0484

Table 3. Announcement Period CARs defined over [*t-1, t+1*]

In this table we present estimates of the following two equations:

$$CAR_{jt} = \beta_0 + \beta_{SUE} SUE_{jt} + \beta_{Sent} Sentiment_{jt} + \varepsilon_{jt},$$

$$CAR_{jt} = \beta_0 + \beta_{SUE} SUE_{jt} + \beta_{SUE\_Size} SUE_{jt} \times Size_{jt} + \beta_{Sent} Sentiment_{jt} + \beta_{Sent\_Size} Sentiment_{jt} \times Size_{jt} + \varepsilon_{jt},$$

where SUE_{jt} is the standardized unexpected earnings, $Sentiment_{jt}$, is the standardized unexpected sentiment in the earnings statement. The sample period includes all available earnings announcements from January, 1998 to July, 2006, for a total of 21,580 (13,907) firm-quarter observations for the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
	Panel A: Baseline Results					
Diction						
SUE	0.01051	13.19	0	0.02024	16.41	0
Sentiment	0.00421	5.38	0	0.00201	2.77	0.0056
Intercept	0.00248	2.97	0.003	-0.00302	-2.65	0.008
Adjusted R-squared	1.72%			8.32%		
General Inquirer						
SUE	0.01020	12.72	0	0.02003	16.1	0
Sentiment	0.00627	8.00	0	0.00349	3.79	0.0001
Intercept	0.00247	2.87	0.0042	-0.00299	-2.60	0.0093
Adjusted R-squared	2.00%			8.43%		
	Panel B: Baseline Results Controlling for Firm Size					
Diction						
SUE	0.06686	7.26	0	0.05407	3.93	0.0001
SUE×Size	-0.00285	-6.55	0	-0.00163	-2.65	0.0081
Sentiment	0.02972	2.85	0.0043	0.02005	2.58	0.01
Sentiment ×Size	-0.0013	-2.58	0.0098	-0.00088	-2.39	0.0167
Intercept	0.00281	3.37	0.0008	-0.00248	-2.33	0.0199
Adjusted R-squared	2.19%			8.52%		
General Inquirer						
SUE	0.06491	7.31	0	0.05174	3.71	0.0002
SUE×Size	-0.00277	-6.61	0	-0.00153	-2.45	0.0141
Sentiment	0.03818	5.00	0	0.03281	3.6	0.0003
Sentiment ×Size	-0.00163	-4.53	0	-0.00144	-3.43	0.0006
Intercept	0.00283	3.30	0.001	-0.00242	-2.26	0.0241
Adjusted R-squared	2.52%			8.68%		

Table 4. Announcement Period CARs with Firm Characteristics

In this table we present estimates of the following equation:

$$CAR_{jt} = \beta_0 + \beta_{SUE} SUE_{jt} + \beta_{SizeSUE} SUE_{jt} \times Size_{jt} + \beta_{XSUE} SUE_{jt} \times X_{jt}$$
$$+ \beta_{Sent} Sentiment_{jt} + \beta_{SizeSent} Sentiment_{jt} \times Size_{jt} + \beta_{XSent} Sentiment_{jt} \times X_{jt} + \varepsilon_{jt},$$

where SUE_{jt} is the standardized unexpected earnings, $Sentiment_{jt}$, is the standardized unexpected sentiment in the earnings statement. The sample period includes all available earnings announcement dates from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
	Panel A: Liquidity					
SUE	0.06679	7.17	0	0.05207	4.11	0
SUE×Size	-0.00284	-6.36	0	-0.00167	-2.72	0.0065
SUE×Liquidity	-0.00007	-0.13	0.8986	0.00327	1.48	0.1397
Sentiment	0.03008	2.91	0.0036	0.02038	2.61	0.009
Sentiment×Size	-0.00136	-2.72	0.0064	-0.00099	-2.59	0.0097
Sentiment×Liquidity	0.00115	2.66	0.0078	0.00242	3.57	0.0004
Intercept	0.00281	3.35	0.0008	-0.00252	-2.31	0.0209
Adjusted R-squared	2.22%			8.88%		
	Panel B: Analyst Coverage					
SUE	0.06873	6.65	0	0.04831	3.05	0.0023
SUE×Size	-0.00296	-5.87	0	-0.00131	-1.76	0.0783
SUE×Analyst Coverage	0.00007	0.66	0.5094	-0.00015	-1.12	0.2648
Sentiment	0.03647	3.01	0.0026	0.0308	2.74	0.0061
Sentiment×Size	-0.00169	-2.8	0.0052	-0.0015	-2.6	0.0093
Sentiment× Analyst Cov.	0.00025	1.88	0.0603	0.00032	1.76	0.0788
Intercept	0.00283	3.37	0.0008	-0.00247	-2.31	0.0207
Adjusted R-squared	2.20%			8.53%		
	Panel C: Numerical Terms					
SUE	0.06939	7.41	0	0.05409	4.28	0
SUE×Size	-0.00272	-6.28	0	-0.00164	-2.51	0.0119
SUE×Numerical Terms	-0.00006	-3.24	0.0012	0.000001	0.02	0.9871
Sentiment	0.0278	2.57	0.0101	0.01879	2.45	0.0143
Sentiment×Size	-0.00132	-2.66	0.0078	-0.00087	-2.38	0.0171
Sentiment×Numerical Terms	0.00003	1.82	0.0687	0.00001	0.69	0.4905
Intercept	0.00302	3.67	0.0002	-0.00245	-2.24	0.0248
Adjusted R-squared	2.24%			8.51%		

Table 4. Announcement Period CARs with Firm Characteristics (Continued)

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
	Panel D: High Tech					
SUE	0.06661	7.63	0	0.0469	3.57	0.0004
SUE×Size	-0.00284	-6.78	0	-0.00138	-2.35	0.0189
SUE×High Tech	0.00043	0.15	0.8799	0.01013	3.48	0.0005
Sentiment	0.02853	2.75	0.0059	0.01898	2.37	0.0176
Sentiment×Size	-0.00128	-2.53	0.0114	-0.00088	-2.27	0.0231
Sentiment×High Tech	0.00358	1.58	0.113	0.00371	1.73	0.0829
Intercept	0.00282	3.38	0.0007	-0.00263	-2.38	0.0173
Adjusted R-squared	2.21%			8.87%		
	Panel E: EFKOS e-loading factor					
SUE	0.05971	6.85	0	0.04719	3.50	0.0005
SUE×Size	-0.0025	-6.10	0	-0.00131	-2.16	0.0305
SUE×EFKOS e-loading	0.00448	1.80	0.0725	0.00709	3.02	0.0025
Sentiment	0.01885	1.75	0.0802	0.01223	1.45	0.1473
Sentiment×Size	-0.00079	-1.52	0.1279	-0.00053	-1.32	0.1857
Sentiment× EFKOS e-loading	0.00742	4.01	0.0001	0.00589	3.03	0.0024
Intercept	0.00288	3.63	0.0003	-0.00231	-2.20	0.0278
Adjusted R-squared	2.43%			8.88%		

Table 5. Long Horizon CARs defined over [t+2, t+62]

In this table we present estimates of the following equation:

$$CAR_{jt} = \beta_0 + \beta_{SUE} SUE_{jt} + \beta_{Sent} Sentiment_{jt} + \varepsilon_{jt},$$

where SUE_{jt} is the standardized unexpected earnings, $Sentiment_{jt}$, is the standardized unexpected sentiment in the earnings statement. The sample period includes all available earnings announcements from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
Diction						
SUE	0.0072	4.15	0	0.0041	2.68	0.0074
Sentiment (Diction)	0.0027	1.85	0.065	0.0027	1.44	0.1493
Intercept	0.0004	0.21	0.8357	0.0018	0.89	0.3744
Adjusted R-squared	0.17%			0.10%		
General Inquirer						
SUE	0.0069	3.93	0.0001	0.0038	2.48	0.013
Sentiment (GI)	0.0054	3.7	0.0002	0.0053	2.94	0.0032
Intercept	0.0004	0.2	0.8381	0.0018	0.92	0.3592
Adjusted R-squared	0.23%			0.17%		

Table 6A. Post-Announcement Drift

Each calendar quarter stocks are classified in one of three groups according to their earnings announcement surprise terciles. The surprise tercile for firm *i* in quarter *t* is a ranking from 1 to 3 of the earnings surprise, SUE_{it} , based on the previous quarter's surprise tercile cutoffs. The 3-day abnormal return is the cumulative size and B/M adjusted return over trading days [-1,+1], where day 0 is the earnings announcement date. The 60-day abnormal return after the announcement is the cumulative size and B/M adjusted return over trading days [+2, +62]. The cumulative returns are multiplied by 100. Firms in market capitalization deciles 9 and 10 are assigned to the large-firm group, firms in deciles 6 through 8 are assigned to the medium-firm group and those in deciles 1 to 5 are assigned to the small-firm group. Three, two and one asterisk denote, respectively, that the estimates are statistically significant at the one, five and ten percent level.

Tercile	SUE	3-Day CAR	60-Day CAR	3-1
		Small		
1	-0.961***	-1.847***	-1.998***	
2	0.065***	-0.112	-0.274	
3	1.075***	2.857***	1.017***	3.015***
		Medium		
1	-0.986***	-0.475***	0.161	
2	0.078***	0.269*	0.903***	
3	1.066***	1.45***	1.845***	1.683***
		Large		
1	-0.938***	-0.159	-0.393	
2	0.078***	0.346**	-0.056	
3	0.974***	1.155***	0.327	0.719

Table 6B. Post-Announcement Drift Revisited

Each calendar quarter stocks are classified in one of three groups according to their earnings announcement surprise terciles and sentiment surprise terciles. The surprise tercile for firm *i* in quarter *t* is a ranking from 1 to 3 of the earnings surprise, SUE_{it} , and sentiment surprise, $Sent_{it}$, based on the previous quarter's surprise tercile cutoffs. We label tercile 1 if both surprises fall in the first tercile, tercile 2 if both surprises fall in the second tercile and tercile 3 if both surprises fall in the third tercile. The 3-day abnormal return is the cumulative size and B/M adjusted return over trading days [-1,+1], where day 0 is the earnings announcement date. The 60-day abnormal return after the announcement is the cumulative size and B/M adjusted return over trading days [+2, +62]. The cumulative returns are multiplied by 100. Firms in market capitalization deciles 9 and 10 are assigned to the large-firm group, firms in deciles 6 through 8 are assigned to the medium-firm group and those in deciles 1 to 5 are assigned to the small-firm group. Three, two and one asterisks denote, respectively, that the estimates are statistically significant at the one, five and ten percent level.

Tercile	SUE	ΔNetOpt	3-Day CAR	60-Day CAR	3-1
			Small		
1	-0.999***	-1.104***	-2.436***	-2.217***	
2	0.068***	0.005	-0.079	0.185	
3	1.085***	1.035***	3.956***	1.885***	4.102***
			Medium		
1	-1.049***	-1.041***	-1.053***	0.745	
2	0.077***	0.016**	0.238	0.938**	
3	1.079***	1.014***	1.701***	2.84***	2.095**
			Large		
1	-0.981***	-0.995***	-0.21	-0.872	
2	0.081***	0.004	0.313	-0.074	
3	0.94***	0.986***	1.259***	-0.042	0.830

Table 7: Announcement period abnormal turnover defined over [t-1, t+1]
In this table we present estimates of the following equation:

$$\sum_{i=-1}^{1} AbnormalTurnover_{jt+i} = \gamma_0 + \gamma_1 \sum_{i=2}^{4} AbnormalTurnover_{jt-i} + \gamma_2 \left|SUE_{jt}\right| + \gamma_3 \left|SUE_{jt}\right| \times I(SUE_{jt} < 0)$$
$$+\gamma_4 \left|Sentiment_{jt}\right| + \gamma_5 \left|Sentiment_{jt}\right| \times I(Sentiment_{jt} < 0) + \gamma_6 certa\operatorname{int} y_{jt} + \overline{\gamma}_5 \overline{\mathrm{Y}}_{jt}$$

Following Bailey, Li, Mao and Zhong (2003), we define *Abnormal Turnover* to be the difference between turnover and the mean of daily turnover for that stock over the pre-announcement window (-200, -11), normalized by the mean turnover. The rest of the variables are defined in Appendix B. The sample period includes all available earnings announcements from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
Past Abnormal Turnover	0.87801	21.07	0	0.94545	18.56	0
|SUE|	0.43492	6.36	0	0.52106	11.16	0
|SUE |×I(SUE<0)	-0.61541	-8.58	0	-0.00968	-0.12	0.9082
|Sentiment|	0.14196	2.79	0.0053	0.07399	1.92	0.0543
|Sentiment|×I(Sentiment<0)	-0.01071	-0.2	0.8423	0.02453	0.35	0.7229
Certainty	-0.17199	-5.02	0	-0.14677	-3.67	0.0002
Adjusted R-squared	8.84%			10.33%		
Past Abnormal Turnover	0.87825	20.82	0	0.94935	18.77	0
|SUE|	0.45159	6.03	0	0.46758	9.82	0
|SUE |×I(SUE<0)	-0.58022	-8.2	0	0.03446	0.38	0.7042
|Sentiment|	0.13579	2.72	0.0066	0.07165	1.88	0.0603
|Sentiment|×I(Sentiment<0)	-0.01422	-0.26	0.795	0.02636	0.38	0.7056
Certainty	-0.06661	-2.22	0.0262	-0.06529	-2.28	0.0228
Log(Market Cap.)	-0.11904	-3.4	0.0007	-0.0703	-1.76	0.0776
Log(1+analyst)	0.20193	4.86	0	0.357	5.42	0
EFKOS e-Loading	0.51865	4.01	0.0001	0.62987	4.77	0
REG_FD	0.0411	0.18	0.8589	0.09666	0.43	0.6678
Time Trend	0.03669	3.46	0.0005	0.02337	2.45	0.0144
Financial Statements	0.165	4.35	0	0.22697	5.28	0
Log(Total Words)	-0.15515	-3.08	0.002	-0.09683	-1.94	0.0524
Recent Media Coverage	0.03207	1.76	0.0779	0.02943	1.57	0.1163
Forecast Dispersion				0.2221	1.93	0.0534
Adjusted R-squared	10.42%			12.18%		

Table 10: Long horizon abnormal turnover defined over [t+2, t+62]
In this table we present estimates of the following equation:

$$\log(\sum_{i=2}^{62} AbnormalTurnover_{jt+i}) = \gamma_0 + \gamma_1 \log(\sum_{i=2}^{62} AbnormalTurnover_{jt-i}) + \gamma_2 \left|SUE_{jt}\right|$$
$$+\gamma_3 \left|SUE_{jt}\right| \times I(SUE_{jt} < 0) + \gamma_4 \left|Sentiment_{jt}\right| + \gamma_5 \left|Sentiment_{jt}\right| \times I(Sentiment_{jt} < 0) + \gamma_6 certa\operatorname{int} y_{jt} + \overline{\gamma}_5 \overline{\mathrm{Y}}_{jt}$$

Following Bailey, Li, Mao and Zhong (2003), we define *Abnormal Turnover* to be the difference between turnover and the mean of daily turnover for that stock over the pre-announcement window (-200, -11), normalized by the mean turnover. The rest of the variables are defined in Appendix B. The sample period includes all available earnings announcements from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
Past Abnormal Turnover	0.08168	7.83	0	0.11288	9.57	0
|SUE|	1.95559	5.63	0	1.10365	4.96	0
|SUE|×I(SUE<0)	-2.85381	-7.6	0	-0.21476	-0.84	0.3983
|Sentiment|	0.23313	0.75	0.4549	-0.00596	-0.02	0.987
|Sentiment|×I(Sentiment<0)	0.35253	1.1	0.2698	0.38278	1.09	0.2742
Certainty	0.07891	0.44	0.6578	0.11579	0.55	0.5856
Adjusted R-squared	0.97%			1.44%		
Past Abnormal Turnover	0.07837	7.68	0	0.10952	9.13	0
|SUE|	2.03478	6.19	0	1.31463	6.25	0
|SUE|×I(SUE<0)	-2.98669	-8.87	0	-0.43678	-1.57	0.1169
|Sentiment|	0.19049	0.64	0.5194	-0.00374	-0.01	0.9915
|Sentiment|×I(Sentiment<0)	0.36729	1.06	0.2913	0.41288	1.07	0.2868
Certainty	0.19906	1.08	0.281	0.17469	0.68	0.4953
Log(Market Cap.)	-0.6983	-2.63	0.0084	0.13262	0.53	0.5947
Log(1+analyst)	-1.24422	-6.24	0	-2.49547	-7.12	0
EFKOS e-Loading	-1.32017	-1.71	0.0874	-0.82762	-1.17	0.2432
REG_FD	1.70583	0.7	0.4871	1.46558	0.64	0.5238
Time Trend	0.0047	0.05	0.9632	-0.03228	-0.3	0.761
Financial Statements	-0.59388	-3.43	0.0006	-0.1502	-0.61	0.5416
Log(Total Words)	0.25824	1.04	0.2974	0.32095	1.47	0.1422
Recent Media Coverage	0.04007	0.31	0.7588	0.07563	0.67	0.5006
Forecast Dispersion				1.38777	2.97	0.003
Adjusted R-squared	1.85%			2.19%		

Table 9: Announcement period volatility defined over [t-1, t+1]

In this table we present estimates of the following equation:

$$\log(\sum_{i=-1}^{1} AR^2_{jt+i}) = \gamma_0 + \gamma_1 \log(\sum_{i=2}^{4} AR^2_{jt-i}) + \gamma_2 \left|SUE_{jt}\right| + \gamma_3 \left|SUE_{jt}\right| \times I(SUE_{jt} < 0)$$
$$+\gamma_4 \left|Sentiment_{jt}\right| + \gamma_5 \left|Sentiment_{jt}\right| \times I(Sentiment_{jt} < 0) + \gamma_6 certa\mathrm{int}\, y_{jt} + \overline{\gamma}_5 \overline{\Upsilon}_{jt}$$

We measure the volatility of abnormal returns during the event window as the logarithm of the sum of the absolute value of abnormal returns during the [t-1, t+1] event window. The rest of the variables are defined in Appendix B. The sample period includes all available earnings announcement dates from January, 1998 to July, 2006, and in total there are 21,580 (13,907) firm-quarter observations in the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
Past Volatility	0.34305	36.88	0	0.33429	28.73	0
|SUE|	0.0554	3.27	0.0011	0.19459	12.53	0
|SUE|×I(SUE<0)	-0.01426	-0.64	0.5243	-0.01954	-0.72	0.4698
|Sentiment|	0.03813	1.88	0.0606	0.00513	0.19	0.8484
|Sentiment|×I(Sentiment<0)	0.04276	1.55	0.1203	0.03922	1.31	0.1897
Certainty	-0.09791	-6.94	0	-0.07999	-5.29	0
Adjusted R-squared	10.95%			10.57%		
Past Volatility	0.28868	28.89	0	0.26365	20.56	0
|SUE|	0.08397	4.75	0	0.22655	13.92	0
|SUE|×I(SUE<0)	-0.02817	-1.22	0.2242	-0.05687	-2.19	0.0288
|Sentiment|	0.02753	1.37	0.1721	-0.0014	-0.06	0.9543
|Sentiment|×I(Sentiment<0)	0.0339	1.21	0.2282	0.02449	0.79	0.4276
Certainty	-0.07088	-5	0	-0.04325	-2.74	0.0061
Log(Market Cap.)	-0.12318	-10.37	0	-0.19069	-12.43	0
Log(1+analyst)	0.20598	11.1	0	0.35469	11.86	0
EFKOS e-Loading	0.41754	10.45	0	0.46112	10.27	0
REG_FD	-0.19248	-1.77	0.0762	-0.18792	-1.49	0.1364
Time Trend	0.00233	0.55	0.5794	0.00159	0.31	0.7532
Financial Statements	0.10184	5.34	0	0.10129	4.5	0
Log(Total Words)	-0.08337	-3.66	0.0003	-0.09758	-3.38	0.0007
Recent Media Coverage	0.03524	3.79	0.0002	0.01659	1.65	0.0989
Forecast Dispersion				0.0406	0.92	0.356
Adjusted R-squared	14.28%			15.30%		

Table 10: Long horizon volatility defined over [t+2, t+62]

In this table we present estimates of the following equation:

$$\log(\sum_{i=2}^{62} AR_{jt+i}^2) = \gamma_0 + \gamma_1 \log(\sum_{i=2}^{62} AR_{jt-i}^2) + \gamma_2 \left|SUE_{jt}\right| + \gamma_3 \left|SUE_{jt}\right| \times I(SUE_{jt} < 0)$$
$$+\gamma_4 \left|Sentiment_{jt}\right| + \gamma_5 \left|Sentiment_{jt}\right| \times I(Sentiment_{jt} < 0) + \gamma_6 certainty_{jt} + \overline{\gamma}_5 \overline{\Upsilon}_{jt}$$

We measure the volatility of abnormal returns during the event window as the logarithm of the sum of the absolute value of abnormal returns during the [t-1, t+1] event window. The rest of the variables are defined in Appendix B. The sample period includes all available earnings announcement dates from January, 1998 to July, 2006, and in total there are 14,937 (11,168) firm-quarter observations in the Compustat (IBES) sample. We use standard errors clustered by calendar quarter and firm to compute the t-statistics and the p-values that are reported next to the coefficient estimates.

	Compustat Sample			IBES Sample		
	coefficient	t-stat	p-value	coefficient	t-stat	p-value
Past Volatility	0.74256	65.81	0	0.76192	61.5	0
|SUE|	0.00572	0.5	0.6178	-0.02618	-1.81	0.0705
|SUE|×I(SUE<0)	0.00856	0.77	0.4395	0.04293	2.54	0.0112
|Sentiment|	0.03555	3.15	0.0016	0.03073	2.35	0.019
|Sentiment|×I(Sentiment<0)	-0.02419	-1.99	0.0463	-0.03399	-2.44	0.0145
Certainty	-0.03109	-4.3	0	-0.0366	-4.94	0
Adjusted R-squared	55.05%			58.28%		
Past Volatility	0.56897	35.56	0	0.59619	34.24	0
|SUE|	0.02959	2.84	0.0045	0.0198	1.9	0.058
|SUE|×I(SUE<0)	-0.0161	-1.66	0.0969	-0.00202	-0.16	0.8725
|Sentiment|	0.03248	2.86	0.0042	0.02735	2.22	0.0265
|Sentiment||×I(Sentiment<0)	-0.02637	-2.39	0.0168	-0.03775	-2.94	0.0033
Certainty	-0.02954	-4.45	0	-0.0337	-4.97	0
Log(Market Cap.)	-0.12626	-13	0	-0.12381	-10.62	0
Log(1+analyst)	0.04384	4.72	0	0.07715	5.48	0
EFKOS e-Loading	0.17702	8.01	0	0.15241	7.58	0
REG_FD	-0.12273	-1.3	0.1947	-0.16701	-1.51	0.1321
Time Trend	-0.01349	-3.23	0.0012	-0.01338	-2.68	0.0073
Financial Statements	0.02076	3.06	0.0022	0.02261	3.24	0.0012
Log(Total Words)	-0.01143	-1.05	0.2949	0.00002	0	0.9988
Recent Media Coverage	0.02418	5.93	0	0.01702	4.4	0
Forecast Dispersion				0.03593	1.34	0.1814
Adjusted R-squared	59.63%			61.90%		