Hardening Soft Information: Analyst Conservative Bias

Kerry Xiao

Amy Zang*

HKUST

November 2017

Abstract

In this paper, we examine whether sell-side financial analysts show a bias when translating their soft information into a hard format. Sell-side analysts produce both soft research output, in the form of a textual report, and hard research output, including earnings forecasts, target prices, and stock recommendations. In our study, we find evidence that analysts' hard outputs undershoot the neutral implication of their own soft output. Furthermore, our cross-sectional results show that our observed conservative bias increases when the underlying information signals are of poorer quality, which we measure by the forecast horizon, linguistic cues in the report, and characteristics of the firms' information environments. Consistent with the well-known analyst optimism, we find that hard outputs assimilate analysts' soft output more conservatively when their soft output conveys bad news. Our findings suggest that the fundamental distinctions between soft and hard information lead to a predicable bias when analysts harden their soft information.

*Corresponding author, amy.zang@ust.hk. We thank Allen Huang, Haifeng You, Tianyu Zhang, Rong Zheng, workshop participants at HKUST, and conference participants at 2017 CUHK Conference of Textual Analysis in Accounting and Finance for helpful comments. All errors are our own.

Hardening Soft Information: Analyst Conservative Bias

1. Introduction

Sell-side financial analysts are widely considered to be a crucial financial intermediary in the capital market. Indeed, it is well recognized that their research facilitates both information flow and price discovery in the market (Ramnath, Rock and Shane 2008; Bradshaw 2011; Kothari, So and Verdi 2016). One of their most important research outcomes is the analyst research report. A typical analyst report contains quantitative summary measures, including earnings forecasts, target prices, and stock recommendations, as well as a detailed, textual analysis of the company that runs, on average, seven pages (Huang, Zang and Zheng 2014, hereafter HZZ). Given the importance of the analyst report, it is useful to examine the relationship between the different components of the report. To identify the distinct components of the analyst report, we follow the terminology used by Liberti and Petersen (2017) and call the quantitative summary measures "hard outputs" and the textual analysis of the company "soft output."

As described in Liberti and Petersen (2017), there are fundamental distinctions between hard and soft information. Hard information is that which can be recorded in numbers, has no ambiguity in its interpretation, and is easy to store, transmit, compare and verify. By contrast, soft information is that which cannot be easily or accurately reduced to a numerical score. As such, its content can be unstructured, multi-dimensional, and ambiguous and its interpretation can be contextual depending on the information producers' private assessment of the information. Liberti and Petersen (2017) also point out that hard information requires a smaller "bandwidth" to transmit, and that soft information experiences a loss in richness when translated to a hard format.

In the analyst report, soft information is the textual portion that describes various aspects of firm performance. Analysts use this soft information to produce their hard information summaries (HZZ).[1] Given the combination of soft and hard information, analyst reports provide a valuable setting in which to investigate the fundamental distinctions between soft and hard information and the choices analysts make in their production and communication of information. That is, given a pool of information signals of varying qualities, a key question is how an analyst chooses to format the information. In particular, an analyst must make a choice between conveying information in a soft, textual form and presenting that information in a summary quantitative output consisting of earnings forecasts, target prices, and stock recommendations. In examining this choice, it is useful to ask if an analyst exhibits any predictable bias when translating soft information to a hard format, as well as whether any such bias may be economically motivated.[2] Since prior literature on financial analysts has traditionally focused on analyst hard outputs, the relationship between an analyst's soft and hard outputs is as yet unexplored.

More recently, a stream of literature in accounting and finance has started to examine the characteristics of the soft information provided by analysts, often drawing on computational linguistic analysis techniques. For example, HZZ use a naïve Bayesian machine learning approach and find that analysts' soft research output triggers an incremental market reaction beyond that triggered by their hard output. They further find that the informativeness of an analyst's soft output depends on the output's linguistic features, such as the assertiveness of its tone or its qualitative versus quantitative nature. In another study, Huang, Lehavy, Zang and Zheng (2017) find that analysts' soft output may cover an array of topics, including a firm's financial performance, recent corporate events, consumer surveys,

[1] HZZ find evidence inconsistent with the view that analysts provide a written report merely to justify their quantitative summary measures. In particular, they point out that analyst reports describe verbally various input variables involved in valuation. Although we cannot directly observe how analysts generate hard outputs, it is safe to assume that analysts generate their hard information based on an unobservable set of "raw" information collected through their private research. The raw information collected, such as conversations with division managers, survey results gathered from customers, impressions from visiting local stores, and cues picked up during attending conference calls, is unstructured and soft. In this paper, we use analysts' soft output in the reports to represent the unobservable set of raw information collected by analysts, and assume that analysts at the least take into account their own compilation of soft information when they generate their hard outputs.
[2] In our paper, the "bias" in the hard outputs is with respect to the analyst's own soft output.

potential litigation, management effectiveness, the competitive landscape, and the macroeconomic environment.

Several insights from the above-mentioned studies are particularly relevant to our investigation. First, HZZ's evidence suggests that the soft output provided by analysts contains relevant information beyond that provided by the hard output. Second, prior research suggests that soft output may draw on a diverse set of information signals of varying quality. As such, analysts may find it challenging to translate signals related to topics like management effectiveness or M&A rumors into hard information.[3] Based on these insights, we conjecture that there is a loss of information when analysts harden their soft information due to the following reasons: soft information is multi-dimensional whereas hard information is not, and analysts cannot convey the second moment of the underlying information signals in a hard output format.[4]

We investigate whether analysts' hard outputs reflect a bias in managing the loss of information, and whether any observed bias is due to analysts' economic incentives. Soft information provides a number of preferred characteristics from the analyst's perspective. If an information signal is uncertain or contingent, the analyst prefers to convey the information using a soft format because she can describe verbally how confident she is or the context in which readers should interpret the information. In this case, soft information helps to prevent any misunderstandings or misinterpretations. Another advantage of soft information is that it is more difficult for investors to verify a piece of soft information. Hence, an analyst takes less risk with her reputation when she delivers a highly uncertain signal in a soft format. To mitigate reputational risk when translating an uncertain signal to hard information, an analyst

[3] Take the following statement as an example. In this case, it is not clear how this piece of soft information is incorporated into the analyst's hard output, "*We believe the company has a reasonable good relationship with its employees, and as a result, we do not at this time assume there will be a strike, or any other work stoppage." (Dahlman Rose & Co.'s report for UPS issued on July 25, 2012)*

[4] Morgan Stanley analysts provide scenario-based forecasts (Joos, Piotroski, and Srinivasan 2016); such forecasts increase the dimensionality of their hard outputs in a parsimonious way. To simplify the analysis, we focus on a more extreme and common case in which analysts' hard outputs are single-dimensional.

may either choose to omit this signal or to incorporate it with a conservative bias. That is, we expect that an analyst's hard output will "undershoot" the soft signal in the sense that it will be too low (high) compared to the unbiased implication of the soft signal of good (bad) news.[5]

There is an intuitive and rational explanation for this expected conservative bias in translating one's soft information into a hard output. Let us assume that an analyst collects a piece of information that is good news but a noisy signal. In the soft information format, the analyst can describe the nature of the news as well as how confident she is about it. By contrast, an unbiased hard information output related to this news would carry an equal chance that the analyst's quantitative forecast errs on either side of the unbiased value implied by the soft information. If an analyst overshoots, she must subsequently revise her hard information downward in a later report, creating an impression of "bad news" that contradicts the good news conveyed by the soft information, as well as an impression of "self-correction" that casts doubt on the analyst's ability to provide accurate forecasts. To avoid this scenario, an analyst is likely to err on the conservative side when hardening a noisy soft signal — that is, she is better off undershooting her own soft information initially. This argument leads to our main hypothesis that analysts' hard outputs undershoot their own soft information – there is a conservative bias in hardening their soft information.

Note that the extent of the undershooting is driven by how noisy and ambiguous the underlying information signals are. If the underlying signals are of higher quality, the benefit to the analyst of being conservative in the hard output is smaller.[6] It is expected that a noisy signal will become less so over time. Thus, we expect that as the noise of a signal decreases, the analyst bias in subsequent reports should also decrease. We further expect that an analyst should move her hard output in one direction towards the neutral implication of the soft

[5] We use the terms "undershoot" and "underreact" interchangeably in this paper.

[6] In the extreme case when a signal becomes certain, that is, when the signal is a fact, an analyst will incorporate it into hard output with no bias.

signal bit by bit over time as a signal becomes less noisy – that is, the soft output leads the hard output.

To examine whether the hard information provided by analysts reflects a conservative bias relative to their soft information, we draw on a sample of analyst reports on S&P 500 firms from the period 1995 to 2012. To operationalize our empirical tests, we extract the textual favorableness of the soft output ($Tone$) of these reports. We then follow the data cleansing procedures and naïve Bayesian machine learning approach described in HZZ, classifying each sentence in the soft output based on whether its tone is positive, neutral, or negative. In particular, we determine the percentage of positive (negative) sentences as well as the overall tone ($Tone$, measured as $Pos\%$ minus $Neg\%$). We note that our reduction of soft information into a single dimension of $Tone$ allows us to extract only the overall level of favorableness contained within a report and thus presents a potential limitation of our study. However, this single dimension at the same time allows us to obtain an unbiased reduction of the soft output. To the extent that analysts may omit or downplay noisy signals in their hard output, our unbiased single-dimension $Tone$ provides us with a measure from which we can detect conservative bias when analysts translate their soft output into a hard format. Nonetheless, we acknowledge that our study contains the inevitable loss of power inherent in studies that measure the features of textual information quantitatively.

In our empirical tests, we first focus on the hard output of the analysts' annual earnings forecasts. The findings from our empirical tests suggest that individual analysts, on average, undershoot their soft information in their hard output. First, we find that the $Tone$ of the report text is significantly and negatively related to concurrent forecast error (measured as forecasted earnings minus actual earnings). By contrast, we should find no correlation if analysts assimilate soft information into earnings forecasts without any bias. Our finding of a negative correlation implies that an earnings forecast is too low (high) relative to actual earnings when the analyst's soft output is more (less) favorable, consistent with a conservative bias in the hardening process. Second, we find that our analysts' soft output leads the hard output of earnings forecasts. Specifically, we find that an analyst's $Tone$ in her prior two written reports can predict her forecast error in the current report in a direction consistent with the analyst undershooting her prior soft outputs. Third, we find that an analyst's soft output predicts her earnings forecast revision in the next report, consistent with an analyst assimilating her soft information into later earnings forecasts. Descriptive analyses show that soft information can precede hard information by up to two months. All of these findings support our prediction that analysts exhibit a conservative bias when hardening their own soft information.

In our study, we also examine the factors that may influence our observed conservative bias finding. To do so, we design three cross-sectional tests to examine whether hard outputs undershoot soft outputs more when the underlying signals are of poorer quality. In our first test, we focus on earnings forecasts and use the forecast horizon (i.e., the number of days between the forecast date and earnings announcement date) to measure signal quality. Consistent with our prediction, we find that the magnitude of the absolute value of the coefficient on $Tone$ in explaining analysts' own forecast errors increases with the forecast horizon. That is, we find that earnings forecasts assimilate soft outputs more conservatively earlier in the year when the quality of the underlying signal is relatively poor. In our second test, we examine whether we observe a difference in conservative bias related to the linguistic features of the textual report. Here, we find that earnings forecasts assimilate soft outputs more conservatively when the textual discussions avoid using numbers (i.e., sentences not containing "%," or "$"), and when the discussions are written less assertively. In our third test, we examine the relation between conservative bias and the quality of the information environment. Here, we find that earnings forecasts undershoot soft outputs more for firms with higher earnings volatility, higher return volatility, lower analyst following, and higher analyst forecast dispersion. In all the above regression analyses, we include controls for analyst, broker, and firm characteristics suggested in the prior literature as potential explanations for forecast revisions and errors. Moreover, we control for individual analysts' personal writing styles (i.e., whether an analyst is more positive than others, on average) by including analyst fixed effects, demeaning $Tone$ at the analyst-firm level, and replacing $Tone$ level with $Tone$ change (i.e., $\Delta Tone$ measured as the current report's $Tone$ minus that of the prior report), respectively.

In a final set of analyses regarding earnings forecasts, we find that earnings forecasts undershoot soft output more when $Tone$ is negative, and that negative sentences have a stronger predictive power for subsequent forecast revision than do positive statements. These results suggest that analysts are particularly conservative in hardening soft information when the news is bad, consistent with the well-known analyst optimism. In addition to examining conservative bias in earnings forecasts, we analyze bias in target prices and stock recommendations. Here, we find that $Tone$ predicts target price revisions and stock recommendations in an analyst's subsequent report.

Overall, our results show a general conservative bias when analysts translate their soft information into a hard format, and that this bias is more severe when the underlying signal is noisier or more ambiguous. These findings support our hypothesis that analysts find it challenging to translate noisy or ambiguous signals from soft information to hard in the production and communication of information.

To provide greater confidence in our results, we run several cross-sectional tests to rule out alternative explanations that our finding of a conservative bias reflects an analyst desire to tip in favor of clients or to obtain investment banking business for their employers. Our cross-sectional tests find no support for either explanation. Specifically, we find that undershooting is not more intense for analysts with stronger incentives to tip (i.e., firms with high-institutional ownership or All-Star analysts) or for analysts who are more likely to face investment-banking related conflicts of interest.[7]

Our study makes a number of contributions to the literature. To our knowledge, our study is the first to investigate how individual analysts harden their own soft information. In general, the literature provides limited attention to the question of how *individual* analysts, or *individual* investors in general, translate a mixture of soft signals of varying qualities into hard outputs, such as forecasts or trading decisions. The literature related to textual news and market reactions (e.g. Antweiler and Frank 2004; Tetlock 2007; Tetlock, Saar-Tsechansky and Macskassy 2008; Garcia 2013; Chen, De, Hu and Hwang 2014) aggregates investor reactions to soft information and as such does not reveal how individual decision makers harden their own soft information. Analyst reports provide us a unique opportunity to directly observe an individual analyst's set of soft information and compare it with her hard information. Our analyses show that the fundamental distinction between soft and hard information leads to a predictable bias in the process of hardening one's own soft information.

Second, our study contributes to the literature on analysts' forecast efficiency. The collective evidence in this literature suggests that analysts underreact to *external* information such as prior earnings (e.g., Mendenhall 1991; Abarbanell and Bernard 1992; Ali, Klein and Rosenfeld 1992; Abarbanell and Bushee 1997) and prior stock returns (i.e., Klein 1990; Lys and Sohn 1990; Abarbanell 1991; Elgers and Lo 1994). Recent studies have begun to investigate whether analyst forecasts incorporate the soft information found in the financial press and find evidence suggesting that analysts do read media news (Bradshaw, Wang and Zhou 2015; Huang and Mamo 2016). Our study differs from this stream of research in that

[7] Note that our results based on earnings forecasts cannot be explained by the conjecture that analysts tend to "walk down" their earnings expectations, as the conservative bias we document exists in both directions: hard outputs are too low (high) compared to soft outputs conveying more (less) favorable opinions. Moreover, our finding holds when the hard outputs are target prices and stock recommendations, respectively.

we examine the efficiency of individual analysts in assimilating their own set of soft information.[8] Our findings contribute to this literature by suggesting that the analyst under-reaction observed in previous studies could be driven or exacerbated by analysts' conservative bias in hardening their own soft information. However, we do not consider this issue directly in the current study. Moreover, our finding implies that investors with access to analysts' soft outputs, such as institutional investors, may enjoy an information advantage over investors who can access only analysts' hard outputs.

Third, our research contributes to the literature on market efficiency. One factor that has been documented as impacting market efficiency is the speed of information absorption. Our study suggests that it takes time for analysts to harden their soft information. To the extent that investors exhibit a similar process in hardening their information, or the extent to which investors over-rely on analysts' hard outputs in valuing firms, firms' share prices might not be efficient. Hence, our theory suggests an alternative explanation for the evidence in the literature that stock prices underreact to public news that does not rely on psychological judgment error (e.g., Barberis, Shleifer and Vishny 1998; Daniel, Hirshleifer and Subrahmanyam 1998; Zhang 2006b).[9]

2. Hypotheses Development

As mentioned, the question of how individuals convert their own soft information into hard information has received little attention in the literature. Part of this lack of attention may reflect the difficulty in obtaining samples that provide simultaneous observations of individual decision makers' qualitative and quantitative assessments. In our study, we are

[8] Our finding of a conservative bias when analysts harden their soft information holds after controlling for prior earnings and prior stock returns, suggesting that analysts are not efficient in incorporating information in their own soft outputs beyond that contained in the prior earnings or prior stock returns.

[9] Although investors do not bear any reputational risk in producing inaccurate forecasts, it is likely that the challenge of hardening ambiguous/noisy soft information leads to a similar bias. That is, an investor updates her valuation of a firm towards the unbiased implication of her soft information in one direction bit by bit over time to mitigate transaction costs.

able to obtain these observations through the analyst report. Analysts gather various information signals concerning different aspects of firms. They then incorporate this information into their reports through a combination of narrative and quantitative summary outputs, including earnings forecasts, target prices, and stock recommendations. The verbal discussion of various aspects of a firm's performance in the written report is what we refer to as analysts' soft output. By contrast, the earnings forecasts, target prices, and stock recommendations are considered hard research outputs because they are numerical or easily converted to numbers. We refer to the process of assimilating soft information into hard outputs as "hardening."

We develop our hypotheses using Figure 1 as illustration.[10] When an underlying information signal is noisy, communicating it in a soft format has the advantage of conveying its second moment. Phrases such as, "we are confident," "there is a remote chance," "it is likely," and words such as "approximately," "appear," and "seem" convey both the quality of a given signal and the analyst's confidence in it. Moreover, when a signal's likelihood is contingent on circumstances, a soft format allows the analyst to indicate the signal's dependence on the situation. Based on these advantages, we assume that a soft format allows an analyst to convey a signal's first and second moments at time t=1, as depicted by the yellow distribution in Figure 1-a. In Figure 1-a, let us assume the signal is a piece of good news and its neutral interpretation indicates an earnings increase of \$9.[11] However, when the analyst hardens this soft information into a single number, she has the incentive to err on the conservative side and forecast an earnings increase that is to the left of the mean of the distribution. Let us assume that the analyst provides a hard format forecast of \$5. This conservative estimate is the analyst's way of avoiding the cost associated with overshooting.

[10] For brevity, we illustrate only an example of good news in Figure 1. However, our theory works in both directions—that is, analysts' hard outputs undershoot both good and bad news when the underlying signal is noisy.

[11] Note that although the analyst knows and conveys the entire distribution of the underlying signal in the soft output, she might not know exactly the mean value of the distribution because soft information is unstructured.

To illustrate this point, let us assume that, with the passage of time, at t=2, the analyst has collected more evidence about this signal and as a result, can verbally describe the distribution of the signal with a narrower range, as depicted in Figure 1-b. When she hardens this soft information at t=2, she is conservative and provides a forecast of $6. By contrast, if the analyst converts the soft information in a neutral manner, she could forecast $6 or $12 (as shown in Figure 1-d), both of which are the same distance from the unbiased implication of $9. If she forecasts $12, she overshoots and faces the risk that she may need to revise her earnings forecast downward in the next period (t=3) when the noise of the signal reduces further, as shown in Figure 1-d. In this case, she must contradict her prior upward revision, resulting in a confusing report in which the soft and hard information are not consistent. This inconsistency would lead to a negative impact on her reputation. On the other hand, if she forecasts $6 at t=2, her forecast revision at t=3 would be in the same direction as the prior forecast revision, and consistent with the good news conveyed by her soft information. Therefore, she is better off undershooting versus overshooting when converting a noisy signal.

Based on the above logic, analysts prefer to convey an information signal in the soft format when it is noisy and ambiguous. When this signal must be converted to a hard format, they are likely to exhibit a conservative bias in which the hard output is too low (high) when the soft information conveys more (less) favorable news. This bias also makes it seem as if the analyst is waiting for the signal quality to improve before she hardens it in a neutral manner, bit by bit over time. This leads to our main hypothesis:

H1: Analysts' hard outputs undershoot their own soft output.

We also expect that the conservative bias that analysts exhibit in the hardening process will be more severe when the underlying signal is of poorer quality. To illustrate this, consider the extreme case in which the signal is certain. In this case, there is no risk of either over- or undershooting when hardening the soft information. As shown in Figure 1, when a signal is noisier, the distribution is flatter with fatter tails, and there is a greater chance that the hard output will be further away from what would be estimated under a neutral implication. We develop three hypotheses based on this intuition. First, because the hard output of an annual earnings forecast has horizon issues, that is, the annual earnings signal is noisier earlier in the year, we expect the following:

H2a: The conservative bias in earnings forecasts with respect to soft outputs increases with the forecast horizon.

Second, we expect that a soft output that avoids numbers or uses less assertive language is associated with greater conservative bias in the corresponding hard output. This leads to our next hypothesis:

H2b: Hard outputs undershoot soft output more when linguistic cues indicate a higher level of ambiguity/noise in the signal.

Third, when a firm's overall information environment is uncertain and volatile, the quality of the information signals collected by analysts tends to be poorer. Hence, in an uncertain and volatile information environment, we expect that analysts will prefer to use a soft format and that the process of hardening this soft information will be challenging and slow, resulting in a higher level of undershooting. This leads to our next hypothesis:

H2c: Hard outputs undershoot soft outputs more for firms that have a poorer information environment.

Our final hypothesis relates to the analyst optimism phenomenon, which indicates that analysts have incentives to issue optimistic forecasts to curry favor with management (Easterwood and Nutt 1999; Lim 2001), maintain or attract underwriting business (O'Brien, McNichols and Lin 2005; Chen and Jiang 2006), or generate commission revenue for their brokerage house (Barron, Byard and Liang 2013). Given analyst optimism incentives, we expect that analysts will be more conservative in hardening a piece of bad news:

H3: An analyst's hard outputs undershoot her soft information more when she is conveying bad as opposed to good news.

3. Sample and Variables

3.1. Sample selection

To obtain our sample, we start with 626,750 analyst reports from the Investext Database issued for S&P 500 firms during the period from 1995 to 2012. Table 1 illustrates our sample selection process. We then match the Investext analyst reports with the I/B/E/S database using analyst and broker names. This yields a set of 452,907 reports available in both databases. Because our study requires simultaneous observations of an analyst's soft and hard outputs, we focus on revision reports, in which analysts update one of their hard outputs. We follow HZZ and identify revision reports as those released within two days of the I/B/E/S forecast announcement dates. If there are multiple reports released within this window, we use the one that is issued earliest. Using this process, our final sample consists of 174,166 earnings forecast revision reports, 82,991 target price revision reports, and 29,070 recommendation revision reports.[12]

[Insert Table 1 here]

3.2. Variable measurement

As mentioned, we operationalize our empirical tests by measuring the overall textual favorableness of an analyst's soft information (i.e., $Tone$). Following the data cleansing procedures and the naïve Bayesian machine learning approach described in HZZ, we classify each sentence in a soft information observation as positive, neutral or negative. We then calculate the percentage of positive sentences ($Pos\%$) and negative sentences ($Neg\%$) as the

[12] Our percentage of revision reports is similar to that reported in HZZ. The majority of our report sample consist of reiteration reports, in which analysts do not revise their hard outputs. Only 6% of our analyst reports revise stock recommendations.

number of positive and negative sentences scaled by the total number of sentences after excluding brokerage disclosure, respectively. $Tone$ is calculated as $Pos\%$ minus $Neg\%$. For our cross-sectional tests, we include two measures of tone. First, we measure tone of the sentences that contain numbers (identified using "$" or "%"), $NumTone$, and tone of the sentences that do not include numbers, $NoNumTone$. Second, we calculate the tone of the sentences that are written with assertive language (i.e., sentences containing strong modal words, as in HZZ), $StrongTone$, and tone of the sentences that are not assertive, $NoStrongTone$.

Because we test whether soft output leads hard outputs, we identify the five reports prior to each revision report, based on the report dates. Note that these prior reports are not necessarily accompanied by a hard output.

In analyzing the hard output in a revision report, we focus on three pieces of information: earnings forecasts, target prices, and stock recommendations. We measure forecast error ($FError$) and forecast revision ($FRev$) by subtracting from the current forecast the actual earnings realized and the analyst's prior annual earnings forecast for the same firm-fiscal year, respectively. Our control variables include forecast horizon ($Horizon$) and boldness ($Bold$). We also measure target prices revisions ($TPRev$) and stock recommendation levels and revisions ($Recom$ and $RRev$, respectively). To eliminate any potential firm heterogeneity or an outlier effect, we scale our hard output continuous variables (i.e., $FError$, $FRev$, $TPRev$) by the firm's stock price 50 days prior to the revision and winsorize all our continuous hard and soft output measures at the top and bottom one percent.

Moreover, our regression analyses include control variables for those analyst, broker, and firm characteristics identified in the literature as possible explanations for analyst hard outputs (e.g. Cooper, Day and Lewis, 2001; Clement and Tse, 2005). These measures include the analyst's general and firm experience ($AnaGenExp$, $AnaFirmExp$), star status ($AnaAllStar$), firm coverage ($AnaNFirm$) and forecast frequency ($AnaNFore$); brokerage house size ($BHSize$) and investment bank status ($BHIB$); firm size ($FirmSize$), book-to-market ratio ($FirmBM$), and overall information environment measured by the amount of analyst coverage and total number of forecasts issued for a firm ($FirmNAna$; $FirmNFore$); and firm recent news, measured by the prior two-days cumulative abnormal returns ($PriorCAR$). Note that one of our cross-sectional predictions relies on a measure of the quality of a firm's information environment. We measure the information environment in this set of analyses by the amount of analyst coverage, analyst forecast dispersion, and earnings and stock return volatility. We draw all our hard output measures from the I/B/E/S, CRSP, Compustat and SDC databases. See Appendix for detailed definitions of our variables.

3.3. Descriptive statistics

The descriptive statistics for our main variables are reported in Table 2. From the statistics in Table 2, we see that the average $Tone$ for our sample report is 0.19 and that fewer than a quarter of the reports in our sample are comprised of a net negative opinion (i.e., $Tone$ < 0). By contrast, we find that the average percentage of positive sentences (Pos%) is approximately twice that of negative sentences (Neg%), indicating a general analyst optimism in the reports. We further see that $\Delta Tone$ distributes symmetrically around zero, suggesting that analysts are equally likely to revise their overall textual opinions in both directions. Finally, we see that tone is similar in number versus non-number sentences or assertive versus non-assertive sentences. Interestingly, we find that the tone in the assertive statements ($StrongTone$) has larger percentages in extreme values.[13]

[Insert Table 2 here]

[13] The extreme values of 1 and -1 mean that the content of assertive statements is full of positive or negative sentences, respectively. It is intuitive that assertive statements would be less ambiguous and would contain information signals of higher quality.

4. Research Design and Empirical Results

4.1. Test of H1: Analysts' hard outputs undershoot their own soft output.

In examining the results of our tests, we first focus on the hard output of the analyst's earnings forecast. Our hypothesis predicts that the earnings forecast will be too low (high) compared to an unbiased interpretation of the soft information when the soft information conveys good (bad) news. Using the subsequent realized earnings as our proxy for an unbiased interpretation of the soft information, H1 specifically suggests a negative relation between $Tone$ and forecast error ($FError$, calculated as forecast earnings minus actual earnings, scaled by stock price 50 days before). Note that an unbiased translation of an analyst's own soft information would mean that we cannot systematically predict the direction of the analyst's forecast error based on her soft information. H1 also predicts that an analyst's soft output should lead her earnings forecast as the analyst would be likely to revise her forecast in the direction implied by the soft output bit by bit when the underlying signal becomes less ambiguous over time. Hence, we expect the soft output to predict the direction of the subsequent forecast revision ($FRev$). We test H1 using the following regressions:

$$FError = \alpha + \beta_0 Tone + \beta_1 Tone\_P1 + \Sigma\theta_j{*}Controls_j + \varepsilon \qquad (1a)$$

$$FRev = \alpha + \beta_1 Tone\_P1 + \Sigma\theta_j{*}Controls_j + \varepsilon. \qquad (1b)$$

Here, H1 predicts a negative coefficient for $Tone$ (i.e., β_0 in Eq. 1a) in the forecast error regression, and a positive coefficient for the prior report's tone, $Tone\_P1$ (i.e., β_1 in Eq. 1b) in the forecast revision regression.

We first run our regressions without the inclusion of our control variables. The results in Table 3, Panel A are consistent with our predictions. Specifically, we find that the current report's $Tone$, reported in column 1, is negatively related to forecast error (significant at the 0.01 level). We further find that the prior report's tone ($Tone\_P1$), reported in column 7, is positively associated with forecast revision (significant at the 0.01 level). To test whether the current earnings forecast undershoots the soft output from an analyst's earlier reports, we include $Tone$ for one to five prior reports in our regression models. The results, reported in columns 2 to 6 (columns 8 to 11), show that the current forecast error (forecast revision) is significantly related to $Tone$ in the prior two (four) reports. In addition, after controlling for the current report's tone, we find a significant positive association between prior tone and forecast revision (column 12), suggesting that our observed undershooting is not driven by any persistence in tone.[14]

[Insert Table 3 here]

Next, we examine the above relation after controlling for the analyst, brokerage, and firm characteristics described in Section 3.2, as well as firm (or analyst) and year fixed effects. The results from this set of analyses are reported in Table 3, Panel B. Examining these results, we continue to find negative estimated coefficients for both $Tone$ and $Tone\_P1$ in the $FError$ regressions (significant at the 0.01 level); and positive estimated coefficients for *Tone_P1* in the *FRev* regressions (significant at the 0.01 level), supporting H1.[15] To examine if our observed forecast bias is partially driven by analyst characteristics, we also include analyst fixed effects (reported in column 2, 4, and 6 of Panel B) and find that doing so yields greater coefficients and t-stats for $Tone$ and $Tone\_P1$, as shown in columns 2, 4, and 6). In addition to analyst fixed effects, we examine whether an analyst's personal writing

[14] The average time gap between an analyst's current report and her prior report is 60 days. When we re-run our tests using subsamples for different time gaps, we find that our observed significant relation between tone and subsequent forecast error or forecast revision declines or disappears if the time gap is more than 60 days. That is, only analyst reports written recently (within two months) have significant predictive power for subsequent forecast revisions and errors.

[15] To address the concern that our results are driven by analysts' under-reactions to prior stock returns and prior earnings news, an observation that has been documented in the prior literature, we replace *PriorCAR* with prior stock returns using a longer window (measured as the value-weighted market-adjusted 60-days cumulative abnormal return prior to the forecast), and include prior annual earnings surprises into Eq. (1a) and (1b). We obtain similar results as those in Table 3, Panel B; that is, we find negative (positive) coefficients for *Tone* (*Tone_P1*) on *FError* (*FRev*) (significant at the 0.01 level), suggesting that analysts' conservative bias in hardening their own soft information remains after controlling for any under-reaction to public information prior to the report issuance.

style impacts our findings. To control for personal writing style, we first demean $Tone$ at the analyst-firm level ($ToneDM$), removing an individual analyst's average favorableness for the firm. The results, reported in Table 3, Panel C, are similar to those we find in Panel B. We also replace $Tone$ with a change in textual opinion ($\Delta Tone$) from an analyst's prior report to her current report, and find similar results, as seen in Table 3, Panel D. Therefore, we conclude that our H1 results are robust to different measures of tone.

4.2. Test of H2a: The conservative bias in earnings forecasts with respect to soft outputs increases with the forecast horizon.

One feature of the earnings forecast is that it has a "fixed underlying target", meaning that its information quality should improve over the year until it realizes on the earnings announcement date. This horizon feature of the earnings forecast allows us to conduct our first cross-sectional test of our undershooting hypothesis. Specifically, we expect that the extent of undershooting in the earnings forecast should be greater earlier in the year when the underlying signal is noisier. We further expect that the information lead of an analyst's soft output over the earnings forecast should dissipate as the earnings announcement date approaches.

To test H2a, we first partition our sample reports based on the forecast horizon (i.e. the number of days between the forecast and announcement dates) and report our results in Table 4, Panel A. From Panel A, we see that both the magnitude and significance of the estimated coefficients for $Tone$ in predicting $FError$ increase monotonically when the subsamples' respective forecast horizons increase from [0, 90) days to more than 270 days. In addition, we re-estimate Eq. (1a) after including the interaction term of $Tone \times Horizon$ and present the results in Table 4, Panel B. Here, we find that the interaction term is negatively associated with forecast error (significant at the 0.01 level), consistent with our prediction in

H2a that our observed undershooting in the earnings forecast with respect to soft information increases with the forecast horizon.

[Insert Table 4 here]

4.3. Test of H2b: Hard outputs undershoot soft output more when linguistic cues indicate a higher level of ambiguity/noise in the information signal.

Not surprisingly, it is more challenging to harden an ambiguous piece of soft information. We thus expect that hard outputs will undershoot an analyst's soft output more when there is greater ambiguity in the information signal. In testing H2b, we rely on linguistic cues to capture signal quality. Specifically, we define an underlying signal as being less ambiguous/noisy if the textual description of the signal contains a financial number or is more assertive in its tone. We measure the tone of the former as $NumTone$ and that of the latter as $StrongTone$ (see Section 3.2 for detailed measurement). The estimation regressions to test H2b are:

$$FError = \alpha + \beta_0 NumTone + \Upsilon_0 NoNumTone + \Sigma\theta_j{*}Controls_j + \varepsilon \qquad (2a)$$

$$FRev = \alpha + \beta_1 NumTone\_P1 + \Upsilon_1 NoNumTone\_P1 + \Sigma\theta_j{*}Controls_j + \varepsilon \qquad (2b)$$

$$FError = \alpha + \beta_0 StrongTone + \Upsilon_0 NoStrongTone + \Sigma\theta_j{*}Controls_j + \varepsilon \qquad (2c)$$

$$FRev = \alpha + \beta_1 StrongTone\_P1 + \Upsilon_1 NoStrongTone\_P1 + \Sigma\theta_j{*}Controls_j + \varepsilon. \qquad (2d)$$

[Insert Table 5 here]

From the results in Table 5, Panel A, we find that the magnitude of the estimated coefficient for $NumTone$ (β_0) is not significantly different from that of $NoNumTone$ (Υ_0) according to the F-test statistics (column 1). However, once we control for the horizon effect for the forecast error regression, our untabulated results show that statements without numbers ($NoNumTone$) have stronger explanatory power for $FError$ than do statements containing numbers ($NumTone$), consistent with H2b.[16] These results also suggest that number-including statements as a linguistic cue indicate the quality of an underlying information signal when the earnings horizon is large. The results in column (2) of Panel A show that the magnitude of the estimated coefficient for $NumTone\_P1$ (β_1) is smaller than that for $NoNumTone\_P1$ (Υ_1) (F-stat = 9.80, significant at the 0.01 level), again supporting H2b.

Finally, in Table 5, Panel B, we present our results regarding the magnitude of the absolute value of the coefficients for $NoStrongTone$ and $NoStrongTone\_P1$ (Υ_0 and Υ_1, respectively). We find that these are larger than those for $StrongTone$ and $StrongTone\_P1$ (β_0 and β_1, respectively; F-stat. = 21.91 and 39.85, respectively; both significant at the 0.01 level). It is worth noting that the coefficient for $StrongTone$ is insignificant in explaining forecast error, suggesting that analysts who use an assertive tone have fully incorporated their information signals into their soft output. Overall, the results based on linguistic cues support H2b that hard outputs undershoot the analyst's own soft information to a greater extent when the soft output reflects greater signal ambiguity.

4.4. H2c: Hard outputs undershoot soft outputs more for firms with a poorer information environment.

We next examine how the information environment of a firm impacts our observed conservative bias. To test H2c, we examine several characteristics of a firms' information environment to measure the overall quality of the underlying information signals. Specifically, we employ four indicator variables suggested by the prior literature as capturing

[16] It is important to control for horizon in the forecast error regression because the noise of the underlying signal decreases substantially when it approaches the earnings announcement date. This, in turn, impacts our regressions as signal noise is a necessary condition for observing the conservative bias in the hardening process. However, horizon is less of an issue in our forecast revision model as this model tests whether soft output predicts the direction of the subsequent revision in the hard output. For Eq. (2a), after we include the interaction terms $NumTone \times Horizon$ and $NoNumTone \times Horizon$, the absolute value of coefficient for the former is significantly smaller than that of the latter (F-stat. = 3.02, at the 0.1 level), suggesting that more ambiguous statements have a larger horizon-dependent predictive power for forecast errors than do less ambiguous statements.

the volatility and uncertainty of a firm's information environment (Barron, Kim, Lim and Stevens, 1998; Lim 2001; Zhang, 2006a, 2006b; Dichev and Tang, 2009): (1) high earnings volatility (*HighEPSVol*, equal to one if earnings volatility is above the sample median and zero otherwise); (2) high return volatility (*HighRetVol*, equal to one if return volatility is above the sample median and zero otherwise); (3) low analyst coverage (*FirmLowCov*, equal to one if analyst coverage is below the sample median and zero otherwise); and (4) high forecast dispersion (*FirmHighDisp*, equal to one if analyst forecast dispersion is above the sample median and zero otherwise). The estimation regressions are as follows:

$$FError = \alpha + \beta_0 Tone + \Upsilon_0 Tone \times Indicator\_Poor + \Sigma\theta_j * Controls_j + \varepsilon \qquad (3a)$$

$$FRev = \alpha + \beta_1 Tone\_P1 + \Upsilon_1 Tone\_P1 \times Indicator\_Poor + \Sigma\theta_j * Controls_j + \varepsilon. \qquad (3b)$$

[Insert Table 6 here]

The results for this set of analyses are reported in Table 6. In columns 1, 3, 5 and 7, we report our results for the forecast error regressions (Eq. 3a). These results show that the estimated coefficients for all the interaction terms ($Tone \times HighEPSVol$, $Tone \times HighRetVol$, $Tone \times FirmLowCov$ and $Tone \times FirmHighDisp$) are negative (significant at least at the 0.05 level).These results suggest that analysts undershoot their own soft outputs to a greater extent when a firm has a poorer information environment, supporting H2c.

In columns 2, 4, 6, and 8, we report our results for the forecast revision regressions (Eq. 3b). These results show that both *Tone_P1* and the four interaction terms (i.e., $Tone\_P1 \times HighEPSVol$, $Tone\_P1 \times HighRetVol$, $Tone\_P1 \times FirmLowCov$ and $Tone\_P1 \times FirmHighDisp$) have positive coefficients (all significant at the 0.01 level), indicating that hard outputs exhibit a higher level of conservative bias for firms with a poorer information environment, also supporting H2c.

4.5. H3: An analyst's hard outputs undershoot her soft outputs more when she conveys bad versus good news.

Prior literature has documented that analyst forecasts exhibit optimism for a number of economic reasons, including currying favor with management (Easterwood and Nutt 1999; Lim 2001), maintaining or attracting underwriting business (O'Brien, McNichols and Lin 2005; Chen and Jiang 2006), and generating commission revenue for the analyst's brokerage house (Barron, Byard and Liang 2013). Given these incentives, we predict that analysts should be more conservative when translating soft bad news to a hard format as there are greater negative consequences for the analyst if she overshoots bad news.

To test H3, we examine whether reports with an overall good news $Tone$ exhibit a lower level of conservative bias than reports with an overall bad news $Tone$. Because overall textual favorableness ($Tone$) is a continuous variable, we use zero as the natural cut-off point to separate our sample into reports with overall good versus bad news. That is, we designate reports containing fewer positive than negative sentences as having an overall bad news tone and designate these reports with the indicator variable, BN. From our descriptive statistics, we see that more than three quarters of our sample reports have a $Tone$ measure greater than zero (i.e., BN=0). Hence, reports with an overall bad news $Tone$ (i.e., BN=1) are relatively rare in our sample (i.e., less than a quarter).

After identifying good versus bad news reports, we conduct our forecast error regressions interacting bad news $Tone$ with $LongHorizon$, an indicator variable that equals one when $Horizon$ is above the sample median and zero otherwise. Note that when the underlying news is highly uncertain (i.e., when $LongHorizon$ = 1), H3 predicts that analysts will be more reluctant to harden bad news. That is, we expect the coefficient on $Tone \times BN \times LongHorizon$ (i.e., Υ_{0LH}) in the forecast error regression to be negative. In our forecast revision regression, our variable of interest is $Tone\_P1 \times BN\_P1$, where the coefficient captures the incremental effect of bad news $Tone$ (i.e., Υ_1). The estimated regressions here are as follows:

$$FError = \alpha + \beta_0 Tone + \Upsilon_0 Tone \times BN + \beta_{0LH} Tone \times LongHorizon$$
$$+ \Upsilon_{0LH} Tone \times BN \times LongHorizon + \Sigma\theta_j{*}Controls_j + \varepsilon \qquad (4a)$$
$$FRev = \alpha + \beta_1 Tone\_P1 + \Upsilon_1 Tone\_P1 \times BN\_P1 + \Sigma\theta_j{*}Controls_j + \varepsilon. \qquad (4b)$$

The results presented in Table 7, Panel A, column 1 show that the coefficient for Υ_{0LH} is negative in the regression of $FError$ (significant at the 0.05 level), indicating that when the forecast horizon is long, there is greater conservative bias under a bad news $Tone$. From the results in column 2, we see that the coefficient for Υ_1 is positive in the regression of $FRev$ (significant at the 0.05 level), also consistent with H3. Furthermore, we find that the magnitude of Υ_{0LH} is 154% of that of β_{0LH}, and the magnitude of Υ_1 is 56% of that of β_1, suggesting that the magnitude of the conservative bias is considerably higher when the soft output conveys bad rather than good news.

[Insert Table 7 here]

To further validate our findings, we conduct an alternative test by re-estimating Eq. (4a) and (4b) model and replacing overall $Tone$ with the percentages of positive and negative sentences, respectively:

$$FError = \alpha + \beta_0 Pos\% + \Upsilon_0 Neg\% + \beta_{0LH} Pos\% \times LongHorizon$$
$$+ \Upsilon_{0LH} Neg\% \times LongHorizon + \Sigma\theta_j{*}Controls_j + \varepsilon \qquad (5a)$$
$$FRev = \alpha + \beta_1 Pos\%\_P1 + \Upsilon_1 Neg\%\_P1 + \Sigma\theta_j{*}Controls_j + \varepsilon\ . \qquad (5b)$$

Note that we do not know whether the effect of an overall bad news $Tone$ (i.e., when BN = 1) is due to the presence of fewer positive sentences or more negative sentences. Thus, to investigate whether earnings forecasts undershoot negative and positive sentences to the same extent, we estimate and compare the absolute magnitudes of Υ_{0LH} and Υ_1 to those of β_{0LH} and β_1, respectively. From the results in Table 7, Panel B, we find that all of our estimated Υ_{0LH}, Υ_1, β_{0LH} and β_1 are significant at the 0.01 level, indicating that earnings forecasts undershoot both positive and negative statements. Further F tests show that the absolute magnitudes Υ_{0LH} and Υ_1 are significantly larger than those of β_{0LH} and β_1, respectively (F-stat. = 15.25 and 5.07, respectively, significant at least at the 0.05 level), supporting H3. In sum, we find evidence that hard outputs undershoot both analysts' positive and negative soft information, but that analysts are particularly conservative and wait longer to harden information when it conveys bad news.

5. Sensitivity Tests

5.1. Conservative bias in the hard outputs of target prices and stock recommendations

In addition to our main analyses, we conduct several sensitivity tests related to the type of hard output provided by analysts. Unlike earnings forecasts, where the underlying signal realizes on the earnings announcement date, target prices and stock recommendations can be considered to have moving targets. That is, every time an analyst revises her target price (stock recommendation), the underlying signals change to the stock price at the end of the next 12-month period (the overall firm prospects relative to current stock price). In this case, we do not have a clear realized underlying signal. Consequently, we test our conservative bias hypothesis by examining whether an analyst's soft output leads her target prices and stock recommendations. Specifically, we test if the level and change of $Tone$ in the analyst's prior report are positively related to her target price and stock recommendation in her current report. If so, such results would suggest that these two hard outputs incorporate the analyst's soft output in a delayed manner.

[Insert Table 8 & 9 here]

We present the results from these analyses in Table 8. From these results, we see that the estimated coefficients on prior report's level and change of tone ($Tone\_P1$ and $\Delta Tone\_P1$) are positively related with $TPRev$ (significant at the 0.01 level), consistent with our main hypothesis (H1). The results from our analyses using an analyst's stock recommendations, presented in Table 9, are also consistent with our prediction. Specifically, from the results in Panel A, we see that the estimated coefficients on the tone of the analyst's prior report ($Tone\_P1$) are positively related with the recommendation level ($Recom$) (significant at the 0.01 level). From the results in Panel B, we see that they are also positively related with the recommendation revision ($RRev$) in the current report when the prior recommendation is a "hold," "buy," or "strong buy" (significant at least at the 0.05 level). Our finding of no significant coefficient for $Tone\_P1$ in the recommendation revision regressions in Panel B when the prior recommendation is either "sell" or "strong sell" (columns 1 and 2, respectively) is unsurprising. A "sell" or "strong sell" recommendation is issued rarely and usually only when there is an unambiguous signal; thus, we do not observe any conservative bias for these recommendations. Overall, the results in Table 8 and 9 indicate that an analyst's soft information leads her hard outputs of target prices and stock recommendations. We therefore conclude that our finding of a conservative bias is applicable to all the hard outputs of an analyst, including earnings forecasts, target prices, and stock recommendations.[17]

5.2. Alternative explanations related to client and brokerage incentives

Analyst hard outputs are free and widely distributed to a broad base of investors. By contrast, their soft outputs tend to be expensive and more likely purchased by institutional clients. Given this distinction, it is possible that our observed conservative bias reflects an analyst incentive to delay the translation of soft output in order to provide institutional clients with an information advantage, a process referred to as tipping. To rule out this explanation, we conduct a similar cross-sectional analysis as that in Section 4.4 by interacting $Tone$ with

[17] We also examine that whether our H3 applies to target prices and stock recommendations by regressing $Pos\%\_P1$ and $Neg\%\_P1$ on $TPRev$ or $Recom$. Untabulated results show that the absolute value of the estimated magnitude of $Neg\%\_P1$ is larger than that of $Pos\%\_P1$ in tests of target price (F-stat. = 5.77, significant at the 0.05 level) and stock recommendations (F-stat. = 14.85, significant at the 0.01 level), supporting H3.

indicator variables that reflect whether an analyst covers firms with high-institutional ownership or is considered an All-Star analyst, respectively. Doing so, we find results (untabulated) similar to our main findings. Thus, we conclude that our results do not reflect a tipping bias.

In addition to client tipping, prior research has shown that analyst outputs are subject to investment-banking (IB) related conflict of interest. For example, analysts might strategically delay incorporating bad news into their hard outputs to curry favor with firms that are potential investment-banking clients. To address this possibility, we repeat our tests of H1 on subsamples of IB analysts and non-IB analysts to rule the possibility that our observed conservative bias is driven by IB analysts. Our untabulated results indicate that IB and non-IB analysts exhibit a similar level of conservative bias in hardening their own soft information. Furthermore, recall that we find that analysts undershoot even their good news when translating their soft information into a hard format. Therefore, we conclude that an IB-related incentive cannot explain our findings.

6. Conclusion

Sell-side financial analysts play an important intermediary role between firms and investors. Consequently, it is worthwhile to examine how they format and communicate their information. In particular, these analysts provide both soft and hard outputs in their reports. Their soft output is provided in a written report; their hard output is summarizing in nature and includes earnings forecasts, target prices, and stock recommendations. Prior research has largely focused on the hard outputs of analysts. By contrast, it has provided relatively little insight into soft information, specifically how an individual decision maker translates soft information that contains great dimensionality into a single-dimension hard signal. The analyst report provides us with a unique setting in which to study the process of how an analyst hardens her own soft information.

Using a sample of revision reports, we find that the hard outputs of analysts undershoot their own soft information. First, we find that hard outputs are conservatively biased, as compared to what the neutral integration of soft information would predict. Second, we find that analysts' soft output leads their hard outputs. Third, we find that our observed conservative bias is more severe when the quality of the underlying information signal is poorer. Specifically, we find that the conservative bias in the earnings forecast is greater when the forecast horizon increases, when linguistic cues indicate signal ambiguity, and when firms have a more volatile and uncertain information environment. Finally, we find that analysts' hard outputs underreact to their soft output more when the soft output conveys bad news, suggesting that analysts are particularly conservative when hardening bad news.

Overall, our paper is the first to document that analysts show a predictable bias when hardening their own soft information. This finding has important implications for investors, as it suggests that institutional investors with access to analyst soft outputs may enjoy an information advantage over investors who can access only an analyst's hard research outputs. It also raises the question of whether the well-documented finding that stock prices underreact to public news could be explained by a similar conservative hardening bias exhibited by investors, or by investor over-reliance on hard versus soft outputs of analysts. We leave this question to future research.

Reference

Abarbanell, J. S. (1991). Do analysts' earnings forecasts incorporate information in prior stock price changes?. *Journal of Accounting and Economics*, *14*(2), 147-165.

Abarbanell, J. S., & Bernard, V. L. (1992). Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *The Journal of Finance*, *47*(3), 1181-1207.

Abarbanell, J. S., & Bushee, B. J. (1997). Fundamental analysis, future earnings, and stock prices. *Journal of Accounting Research*, *35*(1), 1-24.

Ali, A., Klein, A., & Rosenfeld, J. (1992). Analysts' use of information about permanent and transitory earnings components in forecasting annual EPS. *The Accounting Review*, *67*(1) 183-198.

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, *59*(3), 1259-1294.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307-343.

Barron, O. E., Byard, D., & Liang, L. (2013). Analyst pessimism and forecast timing. *Journal of Business Finance & Accounting*, *40*(5-6), 719-739.

Barron, O. E., Kim, O., Lim, S. C., & Stevens, D. E. (1998). Using analysts' forecasts to measure properties of analysts' information environment. *The Accounting Review*, *73*(4) 421-433.

Bradshaw, M. T. (2011). Analysts' forecasts: what do we know after decades of work? Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1880339

Bradshaw, M., Wang, X., & Zhou, D. (2015). Analysts' assimilation of soft information in the financial press. Boston College: Working paper.

Chen, H., De, P., Hu, Y., & Hwang, B. H. (2014). Wisdom of crowds: The value of stock opinions transmitted through social media. *The Review of Financial Studies*, *27*(5), 1367-1403.

Chen, Q., & Jiang, W. (2006). Analysts' weighting of private and public information. *The Review of Financial Studies*, *19*(1), 319-355.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *The Journal of Finance*, *60*(1), 307-341.

Cooper, R. A., Day, T. E., & Lewis, C. M. (2001). Following the leader: a study of individual analysts' earnings forecasts. *Journal of Financial Economics*, *61*(3), 383-416.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *The Journal of Finance*, *53*(6), 1839-1885.

Dichev, I. D., & Tang, V. W. (2009). Earnings volatility and earnings predictability. *Journal of Accounting and Economics*, *47*(1), 160-181.

Easterwood, J. C., & Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism?. *The Journal of Finance*, *54*(5), 1777-1797.

Elgers, P. T., & Lo, M. H. (1994). Reductions in analysts' annual earnings forecast errors using information in prior earnings and security returns. *Journal of Accounting Research*, 290-303.

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, *68*(3), 1267-1300.

Hong, H., Lim, T., & Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies. *The Journal of Finance*, *55*(1), 265-295.

Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, *89*(6), 2151-2180.

Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2017). Analyst information discovery and interpretation roles: A topic modeling approach. *Management Science*. Forthcoming.

Huang, A. G., & Mamo, K. (2016). Do analysts read the news? Working paper. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2378418

Joos, P., Piotroski, J. D., & Srinivasan, S. (2016). Can analysts assess fundamental risk and valuation uncertainty? An empirical analysis of scenario-based value estimates. *Journal of Financial Economics*, *121*(3), 645-663.

Klein, A. (1990). A direct test of the cognitive bias theory of share price reversals. *Journal of Accounting and Economics*, *13*(2), 155-166.

Kothari, S. P., So, E., & Verdi, R. (2016). Analysts' forecasts and asset pricing: A survey. *Annual Review of Financial Economics*, *8*, 197-219.

Liberti, J. M., & Petersen, M. A. (2017). Information: Hard and soft. *The Review of Corporate Finance Studies*. Forthcoming.

Lim, T. (2001). Rationality and analysts' forecast bias. *The Journal of Finance*, *56*(1), 369-385.

Lys, T., & Sohn, S. (1990). The association between revisions of financial analysts' earnings forecasts and security-price changes. *Journal of Accounting and Economics*, *13*(4), 341-363.

Mendenhall, R. R. (1991). Evidence on the possible underweighting of earnings-related information. *Journal of Accounting Research*, *29*(1) 170-179.

O'Brien, P. C., McNichols, M. F., & Hsiou-Wei, L. (2005). Analyst impartiality and investment banking relationships. *Journal of Accounting Research*, *43*(4), 623-650.

Raedy, J. S., Shane, P., & Yang, Y. (2006). Horizon-dependent underreaction in financial analysts' earnings forecasts. *Contemporary Accounting Research*, *23*(1), 291-322.

Ramnath, S., Rock, S., & Shane, P. (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting*, *24*(1), 34-75.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, *62*(3), 1139-1168.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, *63*(3), 1437-1467.

Zhang, X. F. (2006a). Information uncertainty and analyst forecast behavior. *Contemporary Accounting Research*, *23*(2), 565-590.

Zhang, X. F. (2006b). Information uncertainty and stock returns. *The Journal of Finance*, *61*(1), 105-137.

Variable Definition

Earnings Forecast Characteristics

FError	Earnings forecast signed error, calculated as earnings forecast minus the actual earning in I/B/E/S, scaled by the stock price 50 days before the forecast date, winsorized at the top and bottom 1 percent.
FRev	Earnings forecast revisions, calculated as current earnings forecast minus the prior earnings forecast in I/B/E/S issued by the same analysts for the same firms same fiscal year, scaled by the stock price 50 days before the forecast date, winsorized at the top and bottom 1 percent.
Horizon	The number of days between forecast date and annual earnings announcement date, winsorized at the top and bottom 1 percent.
Bold	Dummy variable equal to one if the forecast is two standard deviations from the consensus of forecasts in prior 30 days, and zero otherwise.

Target Price and Recommendation Characteristics

TPRev	Target price revisions, calculated as current target price minus the analyst's prior target price in I/B/E/S issued by the same analysts for the same firms, scaled by the stock price 50 days before the announcement date, winsorized at the top and bottom 1 percent.
Recom	Current recommendation (high value indicates positive) in I/B/E/S.
RRev	Recommendation revisions, calculated as current recommendation (high value indicates positive) minus the analyst's prior recommendation in I/B/E/S issued by the same analysts for the same firm.

Analyst Report Characteristics

Pos%	Percentage of sentences in current report that are classified as positive by naïve Bayes approach (as in HZZ), winsorized at the top and bottom 1 percent.
Neg%	Percentage of sentences in current report that are classified as negative by naïve Bayes approach (as in HZZ), winsorized at the top and bottom 1 percent.
Tone	Textual opinion of current report, calculated as the percentage of positive sentences minus the percentage of negative sentences in the report as classified by the naïve Bayes approach (as in HZZ), winsorized at the top and bottom 1 percent.
ToneDM	Textual opinion demeaned by analyst average textual opinion, as *Tone* of current report minus the mean of *Tone* of all reports by the analyst, winsorized at the top and bottom 1 percent.
Δ*Tone*	Textual opinion revision, as *Tone* of current report minus *Tone* of the analyst's prior report, winsorized at the top and bottom 1 percent.
NumTone	Textual opinion of sentences containing "$" or "%" in current report, winsorized at the top and bottom 1 percent.
NoNumTone	Textual opinion of sentences not containing "$" or "%" in current report, winsorized at the top and bottom 1 percent.
StrongTone	Textual opinion of sentences containing any words in strong modal word list (Loughran and Mcdonald, 2011) in current report, winsorized at the top and bottom 1 percent.
NoStrongTone	Textual opinion of sentences not containing any words in strong modal word list (Loughran and Mcdonald, 2011) in current report, winsorized at the top and bottom 1 percent.
with suffix *_P1*	including *Pos%_P1, Neg%_P1, Tone_P1, ToneDM_P1,* Δ*Tone_P1, NumTone_P1, NoNumTone_P1, StrongTone_P1, NoStrongTone_P1:* Relevant variables of the analyst's prior report, winsorized at the top and bottom 1 percent.

Analyst and Broker Characteristics

AnaGenExp	The number of years that the analyst is in I/B/E/S.

AnaFirmExp	The number of years that the analyst covers the firm.
AnaAllStar	Dummy variable equal to one if the analyst is ranked as All-Star analyst by *Institutional Investor*, and zero otherwise.
AnaLFR	The Leader-Follower Ratio (Cooper, Day and Lewis, 2001) of the analyst in the year.
AnaNFore	The number of total forecasts the analyst makes.
AnaNFirm	The number of firms the analyst covers.
BHSize	The number of analysts in the brokerage house.
BHIB	Dummy variable equal to one if the brokerage house is investment bank, and zero otherwise.

Firm Characteristics

FirmSize	The log value of firm's market value of equity.
FirmBM	The firm's book-to-market ratio of equity.
FirmNAna	The number of analysts covering the firm.
FirmNFore	The number of forecasts for the firm in the year.
PriorCAR	Value-weighted market-adjusted two-day cumulative abnormal returns prior to the forecast.

Figure 1. Illustration of Analysts' Conservative Bias

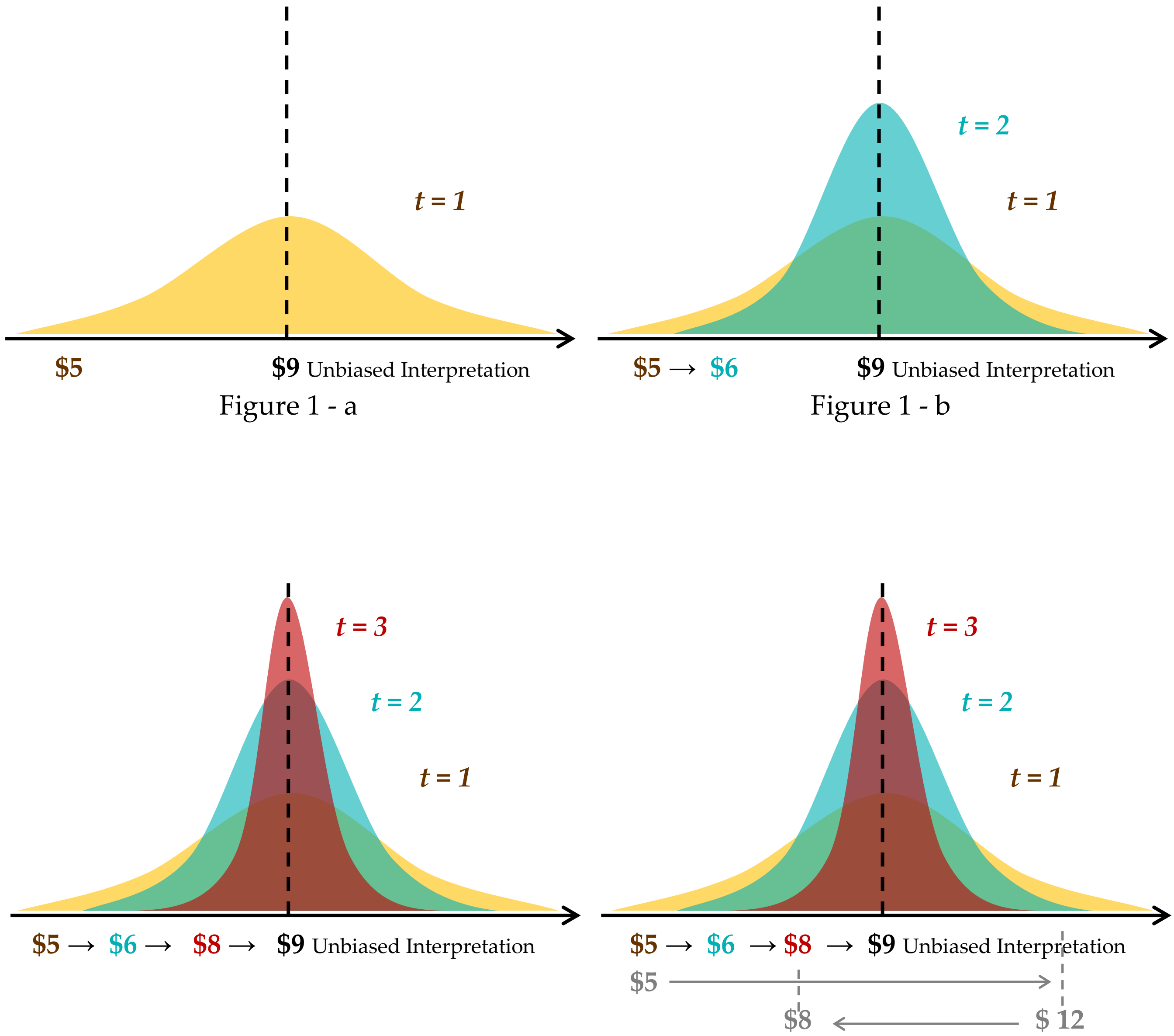

Figure 1 - a

Figure 1 - b

Figure 1 - c

Figure 1 - d

Table 1. Sample Selection

Analyst reports of S&P 500 firms during period from 1995 to 2012 (as in HZZ and Huang et al., 2017)		626,750
Less: reports without matched identifier	151,533	
Less: reports without identifiable textual opinion	22,310	
Report sample		452,907
Earnings Forecast sample*		174,166
Target price sample*		82,991
Recommendation sample*		29,070

** The matched Earnings Forecast/Target Price/Recommendation sample includes only those announced within a [-2, +2] window of each report date; for multiple announcements, we choose the earliest one.*

Table 2. Descriptive Statistics

This table presents the descriptive statistics for the main variables used in this study. Detailed variable definitions are listed in the appendix. N is the number of observations, and Mean and SD are sample means and standard deviations of variables, respectively. Min, 10th, 25th, 50th, 75th, 90th and Max show the sample distributions of variables.

Variable	N	Mean	SD	Min	10th	25th	50th	75th	90th	Max
Report Sample										
Tone	452,907	0.19	0.20	-0.33	-0.06	0.06	0.19	0.32	0.45	0.69
ΔTone	427,648	0.00	0.22	-0.60	-0.28	-0.13	0.00	0.13	0.27	0.58
Forecast Sample										
Pos%	174,166	0.35	0.16	0.00	0.15	0.24	0.34	0.46	0.56	0.75
Pos%_P1	160,028	0.34	0.16	0.00	0.13	0.22	0.33	0.44	0.55	0.73
Neg%	174,166	0.17	0.12	0.00	0.04	0.08	0.15	0.24	0.33	0.53
Neg%_P1	160,028	0.16	0.11	0.00	0.02	0.07	0.14	0.22	0.31	0.50
Tone	174,166	0.18	0.22	-0.40	-0.10	0.04	0.18	0.33	0.46	0.69
Tone_P1	160,028	0.18	0.21	-0.36	-0.08	0.05	0.18	0.32	0.44	0.67
Tone_P2	152,128	0.18	0.21	-0.36	-0.08	0.05	0.19	0.32	0.45	0.69
Tone_P3	144,501	0.19	0.21	-0.36	-0.07	0.05	0.19	0.32	0.45	0.69
Tone_P4	137,313	0.19	0.21	-0.36	-0.07	0.05	0.19	0.32	0.45	0.69
Tone_P5	130,645	0.19	0.21	-0.36	-0.07	0.06	0.19	0.32	0.45	0.68
NumTone	172,060	0.18	0.30	-0.67	-0.20	0.00	0.18	0.36	0.55	1.00
NumTone_P1	156,280	0.18	0.30	-0.67	-0.17	0.00	0.17	0.35	0.53	1.00
NoNumTone	173,616	0.18	0.25	-0.50	-0.13	0.00	0.18	0.33	0.50	0.81
NoNumTone_P1	159,735	0.18	0.23	-0.47	-0.11	0.03	0.18	0.33	0.48	0.75
StrongTone	141,073	0.24	0.52	-1.00	-0.50	0.00	0.25	0.57	1.00	1.00
StrongTone_P1	133,222	0.24	0.50	-1.00	-0.33	0.00	0.25	0.55	1.00	1.00
NoStrongTone	174,130	0.18	0.22	-0.40	-0.10	0.03	0.18	0.32	0.45	0.70
NoStrongTone_P1	160,000	0.18	0.21	-0.38	-0.09	0.04	0.17	0.31	0.44	0.68
FRev	134,895	-0.0009	0.0089	-0.0508	-0.0068	-0.0019	0.0003	0.0017	0.0049	0.0269
FError	171,922	0.0017	0.0203	-0.0593	-0.0098	-0.0032	-0.0004	0.0025	0.0125	0.1264
Bold	147,521	0.24	0.43	0.00	0.00	0.00	0.00	0.00	1.00	1.00
Horizon	172,102	201.73	101.87	7.00	86.00	104.00	192.00	283.00	361.00	371.00
AnaGenExp	174,157	9.04	5.55	1.00	3.00	5.00	8.00	12.00	17.00	31.00
AnaFirmExp	173,689	5.31	4.28	1.00	1.00	2.00	4.00	7.00	11.00	31.00
AnaAllStar	174,166	0.21	0.41	0.00	0.00	0.00	0.00	0.00	1.00	1.00
AnaLFR	139,342	1.81	2.50	0.00	0.27	0.61	1.19	2.13	3.75	100.33
AnaNFore	174,156	222.27	187.97	1.00	51.00	87.00	186.00	301.00	428.00	2242.00
AnaNFirm	174,157	16.49	7.10	1.00	9.00	12.00	16.00	20.00	25.00	124.00
BHSize	173,712	80.32	54.03	1.00	17.00	32.00	66.00	123.00	152.00	233.00
BHIB	174,166	0.75	0.43	0.00	0.00	0.00	1.00	1.00	1.00	1.00
FirmSize	174,052	9.53	1.14	4.83	8.18	8.77	9.43	10.21	11.09	13.35
FirmBM	168,821	0.45	0.42	-26.12	0.14	0.23	0.37	0.58	0.86	7.95
FirmNFore	173,703	331.86	231.03	3.00	97.00	162.00	274.00	443.00	643.00	1447.00
FirmNAna	173,703	27.08	10.30	1.00	15.00	20.00	26.00	33.00	41.00	70.00
PriorCAR	174,128	0.00	0.04	-0.73	-0.04	-0.02	0.00	0.02	0.04	1.20
Target Price Sample										
TPrc	82,876	1.19	0.28	0.54	0.91	1.04	1.17	1.30	1.47	2.53
TPRev	75,929	0.01	0.21	-0.87	-0.22	-0.08	0.04	0.11	0.22	0.61
Recommendation Sample										
Recom	29,070	3.54	0.89	1.00	3.00	3.00	3.00	4.00	5.00	5.00
RRev	21,291	-0.01	1.25	-4.00	-2.00	-1.00	0.00	1.00	2.00	4.00

Table 3. Tone on Forecast Error and Subsequent Forecast Revision

This table presents the results estimated by OLS regressions of tone in current and prior analyst reports on analyst earnings forecast properties, without (Panel A) or with (Panel B, C and D) control variables. Forecast properties include: forecast signed error as estimated value minus actual earnings (*FError*), and forecast revision from prior forecast by the same analyst (*FRev*). Panel A shows the tone(s) effect by gradually including current tone (*Tone*) and prior one-to-five tone(s) (*Tone_P1, Tone_P2, Tone_P3, Tone_P4, Tone_P5*). Panel B shows the effect of tone. Panel C shows the effect of tone demeaned by analyst-firm average tone. Panel D shows the effect of tone revisions. Detailed variable definitions are listed in the appendix. Firm and year (or analyst and year) fixed effects are included, and standard errors are two-way clustered by firm and year (or analyst and year). *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

Panel A. Tone, without Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	FError	FError	FError	FError	FError	FError	FRev	FRev	FRev	FRev	FRev	FRev
Tone	-0.0057***	-0.0050***	-0.0046***	-0.0044***	-0.0044***	-0.0044***						0.0092***
	(-4.92)	(-5.16)	(-5.40)	(-5.27)	(-5.14)	(-5.23)						(16.39)
Tone_P1		-0.0025***	-0.0022***	-0.0023***	-0.0023***	-0.0024***	0.0040***	0.0033***	0.0031***	0.0030***	0.0030***	0.0008***
		(-2.96)	(-3.43)	(-3.51)	(-3.69)	(-3.82)	(7.14)	(7.30)	(7.43)	(7.31)	(7.41)	(3.21)
Tone_P2			-0.0010	-0.0011	-0.0012*	-0.0011*		0.0017***	0.0015***	0.0014***	0.0015***	0.0001
			(-1.11)	(-1.49)	(-1.83)	(-1.82)		(4.99)	(5.68)	(5.87)	(6.77)	(0.71)
Tone_P3				-0.0004	-0.0005	-0.0006			0.0009***	0.0008***	0.0009***	-0.0000
				(-0.52)	(-0.75)	(-1.07)			(2.63)	(2.97)	(3.57)	(-0.23)
Tone_P4					0.0002	0.0000				0.0004	0.0005*	-0.0001
					(0.27)	(0.04)				(1.24)	(1.85)	(-0.52)
Tone_P5						0.0006					-0.0003	-0.0007***
						(1.00)					(-0.96)	(-2.59)
Constant	0.0099***	0.0050***	0.0061***	0.0067***	0.0058***	0.0058***	-0.0027***	-0.0026***	-0.0028***	-0.0027***	-0.0027***	-0.0027***
	(5.88)	(4.43)	(5.73)	(5.23)	(4.52)	(4.24)	(-8.67)	(-7.26)	(-7.43)	(-6.26)	(-5.40)	(-5.76)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.198	0.204	0.208	0.213	0.217	0.221	0.094	0.097	0.099	0.101	0.102	0.141
# of Obs.	171,922	158,000	150,196	142,655	135,536	128,955	130,598	124,500	118,446	112,672	107,276	107,276

Table 3.

Panel B. Tone, with Controls

	(1) FError	(2) FError	(3) FError	(4) FError	(5) FRev	(6) FRev
Tone	-0.0046***	-0.0066***	-0.0040***	-0.0056***		
	(-4.52)	(-4.97)	(-4.70)	(-5.09)		
Tone_P1			-0.0020***	-0.0038***	0.0030***	0.0038***
			(-3.20)	(-4.12)	(6.71)	(7.99)
Horizon	0.0000	0.0000*	0.0000	0.0000*	0.0000	0.0000
	(1.57)	(1.76)	(1.60)	(1.78)	(1.33)	(1.41)
Bold	-0.0009***	-0.0009***	-0.0010***	-0.0009***	-0.0013***	-0.0013***
	(-3.13)	(-3.13)	(-3.07)	(-2.94)	(-3.58)	(-3.47)
AnaGenExp	0.0001*	0.0002	0.0001*	0.0002	-0.0000***	-0.0000
	(1.67)	(0.74)	(1.65)	(0.75)	(-2.76)	(-0.54)
AnaFirmExp	0.0000	0.0001***	0.0000	0.0001***	-0.0000	-0.0000***
	(1.14)	(3.03)	(0.73)	(2.80)	(-0.13)	(-2.97)
AnaAllStar	-0.0004	-0.0005	-0.0005	-0.0005	0.0002	0.0000
	(-0.86)	(-0.63)	(-1.01)	(-0.65)	(1.53)	(0.01)
AnaLFR	-0.0001**	-0.0001**	-0.0001**	-0.0001***	0.0000**	0.0000***
	(-2.14)	(-2.45)	(-2.19)	(-2.59)	(2.18)	(3.71)
AnaNFore	0.0000	0.0000*	0.0000	0.0000*	0.0000	-0.0000*
	(1.08)	(1.94)	(0.97)	(1.92)	(0.76)	(-1.81)
AnaNFirm	-0.0000	-0.0001	-0.0000	-0.0001	-0.0000	0.0000
	(-0.84)	(-1.61)	(-0.65)	(-1.50)	(-0.86)	(0.13)
BHSize	-0.0000*	0.0000	-0.0000	0.0000	0.0000	-0.0000**
	(-1.70)	(0.69)	(-1.54)	(0.60)	(1.10)	(-2.04)
BHIB	0.0002	0.0003	0.0002	0.0003	0.0001	0.0001
	(0.81)	(0.42)	(0.56)	(0.42)	(0.71)	(0.34)
FirmSize	-0.0020	-0.0011***	-0.0021	-0.0012***	0.0019***	0.0008***
	(-1.15)	(-3.43)	(-1.19)	(-3.37)	(4.08)	(6.02)
FirmBM	0.0020	0.0021	0.0015	0.0018	-0.0013*	-0.0013***
	(0.78)	(1.48)	(0.61)	(1.26)	(-1.84)	(-2.91)
FirmNAna	-0.0001	-0.0002**	-0.0001	-0.0002**	0.0000	0.0000
	(-0.92)	(-2.11)	(-0.82)	(-1.97)	(0.60)	(1.45)
FirmNFore	0.0000**	0.0000***	0.0000**	0.0000**	-0.0000***	-0.0000***
	(2.37)	(2.64)	(2.32)	(2.55)	(-3.48)	(-4.03)
PriorCAR	-0.0209***	-0.0220***	-0.0215***	-0.0234***	0.0253***	0.0257***
	(-5.09)	(-5.21)	(-5.12)	(-5.55)	(10.06)	(9.72)
Constant	0.0256	0.0151***	0.0197	0.0057	-0.0188***	-0.0071***
	(1.40)	(2.71)	(1.11)	(1.03)	(-4.06)	(-4.69)
Firm FE	Yes	No	Yes	No	Yes	No
Analyst FE	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.227	0.178	0.233	0.183	0.125	0.099
# of Obs.	114,158	114,158	105,276	105,276	92,200	92,200

Table 3.

Panel C. Demeaned Tone

	(1) FError	(2) FError	(3) FError	(4) FError	(5) FRev	(6) FRev
ToneDM	-0.0034***	-0.0038***	-0.0034***	-0.0037***		
	(-4.25)	(-4.05)	(-4.62)	(-4.36)		
ToneDM_P1			-0.0013**	-0.0015**	0.0026***	0.0028***
			(-2.48)	(-2.28)	(5.97)	(6.14)
Constant	0.0259	0.0142**	0.0204	0.0037	-0.0191***	-0.0061***
	(1.41)	(2.56)	(1.14)	(0.66)	(-4.04)	(-3.89)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	No	Yes	No	Yes	No
Analyst FE	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.226	0.175	0.232	0.179	0.124	0.095
# of Obs.	114,158	114,158	105,276	105,276	92,200	92,200

Panel D. Tone Revisions

	(1) FError	(2) FError	(3) FError	(4) FError	(5) FRev	(6) FRev
ΔTone	-0.0011***	-0.0011***	-0.0015***	-0.0016***		
	(-3.56)	(-3.31)	(-3.42)	(-3.20)		
ΔTone_P1			-0.0011***	-0.0013***	0.0008***	0.0008***
			(-2.65)	(-2.86)	(5.67)	(5.00)
Constant	0.0218	0.0046	0.0238	0.0054	-0.0199***	-0.0070***
	(1.20)	(0.79)	(1.28)	(0.83)	(-4.10)	(-3.74)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	No	Yes	No	Yes	No
Analyst FE	No	Yes	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.233	0.179	0.237	0.184	0.124	0.092
# of Obs.	104,802	104,802	98,928	98,928	87,420	87,420

Table 4. Forecast Horizon

This table presents the variation of results estimated by OLS regressions in forecast horizon. Panel A partitions the forecast horizon by quarter. Panel B shows the interaction effect of tone and horizon. For brevity, results of control variables (same to Table 3, Panel B) are omitted. Detailed variable definitions are listed in the appendix. Firm and year fixed effects are included, and standard errors are two-way clustered by firm and year. *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

Panel A. Partition by Forecast Horizon, without Controls

Horizon	[0, 90)	[90, 180)	[180, 270)	[270, +)
	(1)	(2)	(3)	(4)
	FError	FError	FError	FError
Tone	-0.0021*	-0.0038***	-0.0074***	-0.0090***
	(-1.79)	(-3.66)	(-5.43)	(-4.69)
Constant	0.0015	0.0029***	0.0058***	0.0005
	(0.00)	(3.55)	(7.03)	(0.07)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Adj. R^2	0.173	0.213	0.251	0.268
# Obs	18,774	47,375	44,492	61,281

Panel B. Interaction with Forecast Horizon, with Controls

	(1)
	FError
Tone	0.0070***
	(2.66)
Tone × Horizon	-0.0001***
	(-4.00)
Horizon	0.0000***
	(2.60)
Constant	0.0225
	(1.27)
Controls	Yes
Firm FE	Yes
Year FE	Yes
Adj. R^2	0.230
# of Obs.	114,158

Table 5. Linguistic Cues

This table presents the results estimated by OLS regressions of tone in specific linguistic cues of current and prior analyst reports on analyst earnings forecast properties. Panel A shows the effect of tone in sentences containing number and sentences without numbers. Panel B shows the effect of tone in assertive and non-assertive sentences. For brevity, results of control variables (same to Table 3, Panel B) are omitted. Detailed variable definitions are listed in the appendix. Firm and year fixed effects are included, and standard errors are two-way clustered by firm and year. *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

Panel A. Tone of Number versus Non-Number Sentences

		(1) FError	(2) FRev
NumTone	β_0	-0.0024***	
		(-5.03)	
NoNumTone	Υ_0	-0.0022***	
		(-3.27)	
NumTone_P1	β_1		0.0010***
			(7.52)
NoNumTone _P1	Υ_1		0.0019***
			(5.50)
Constant		0.0238	-0.0183***
		(1.31)	(-4.10)
Controls		Yes	Yes
Firm FE		Yes	Yes
Year FE		Yes	Yes
Adj. R^2		0.228	0.125
# of Obs.		112,259	89,788
F-Test		$\beta_0 = \Upsilon_0$	$\beta_1 = \Upsilon_1$
		F-stat. = 0.07	F-stat. = 9.80
		p > F = 0.7855	p > F = 0.0017

Panel B. Tone of Assertive versus Non-Assertive Sentences

		(1) FError	(2) FRev
StrongTone	β_0	-0.0002	
		(-1.54)	
NoStrongTone	Υ_0	-0.0046***	
		(-4.80)	
StrongTone_P1	β_1		0.0002***
			(3.37)
NoStrongTone_P1	Υ_1		0.0029***
			(6.40)
Constant		0.0220	-0.0177***
		(1.23)	(-3.74)
Controls		Yes	Yes
Firm FE		Yes	Yes
Year FE		Yes	Yes
Adj. R^2		0.229	0.127
# of Obs.		92,815	76,712
F-Test		$\beta_0 = \Upsilon_0$	$\beta_1 = \Upsilon_1$
		F-stat. = 21.91	F-stat. = 39.85
		p > F = 0.0000	p > F = 0.0000

Table 6. Information Environment

This table presents the variation of results estimated by OLS regressions in firms' information environments. The dummy variables proxy for an uncertain and volatile information environment, and their interaction terms with current tone and prior tones, are added into the regression model. The dummy variables indicate a firms's five-year earnings volatility is above the median (*HighEPSVol*), one-year daily return volatility is above the median (*HighRetVol*), number of analysts following is lower than the median (*FirmLowCov*), and analyst forecast dispersion is larger than the median (*FirmHighDisp*). For brevity, results of control variables (same as specifications in Table 3, Panel B) are omitted. Detailed variable definitions are listed in the appendix. Firm and year fixed effects are included, and standard errors are two-way clustered by firm and year. *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

	(1) FError	(2) FRev		(3) FError	(4) FRev
Tone	-0.0018**		Tone	-0.0025***	
	(-2.15)			(-2.84)	
Tone_P1		0.0016***	Tone_P1		0.0011***
		(5.30)			(4.07)
Tone × HighEPSVol	-0.0054**		Tone × HighRetVol	-0.0040***	
	(-2.20)			(-2.59)	
Tone_P1 × HighEPSVol		0.0026***	Tone_P1 × HighRetVol		0.0036***
		(3.94)			(7.78)
HighEPSVol	0.0005	-0.0005**	HighRetVol	0.0020*	-0.0010***
	(0.65)	(-2.57)		(1.68)	(-3.11)
Constant	0.0241	-0.0185***	Constant	0.0222	-0.0173***
	(1.36)	(-4.12)		(1.27)	(-4.06)
Controls	Yes	Yes	Controls	Yes	Yes
Firm FE	Yes	Yes	Firm FE	Yes	Yes
Year FE	Yes	Yes	Year FE	Yes	Yes
Adj. R^2	0.229	0.126	Adj. R^2	0.229	0.126
# of Obs.	113,379	91,626	# of Obs.	108,898	87,884

	(5) FError	(6) FRev		(7) FError	(8) FRev
Tone	-0.0006		Tone	-0.0006	
	(-0.64)			(-0.77)	
Tone_P1		0.0016***	Tone_P1		0.0003
		(3.66)			(1.43)
Tone × FirmLowCov	-0.0076***		Tone × FirmHighDisp	-0.0078***	
	(-3.59)			(-3.18)	
Tone_P1 × FirmLowCov		0.0028***	Tone_P1 × FirmHighDisp		0.0049***
		(3.69)			(7.38)
FirmLessCov	0.0028**	-0.0005	FirmHighDisp	0.0014	-0.0018***
	(2.12)	(-1.46)		(1.61)	(-6.00)
Constant	0.0218	-0.0184***	Constant	0.0238	-0.0159***
	(1.27)	(-4.16)		(1.36)	(-3.74)
Controls	Yes	Yes	Controls	Yes	Yes
Firm FE	Yes	Yes	Firm FE	Yes	Yes
Year FE	Yes	Yes	Year FE	Yes	Yes
Adj. R^2	0.229	0.126	Adj. R^2	0.229	0.130
# of Obs.	114,158	92,200	# of Obs.	114,158	92,200

Table 7. Soft Output Conveying Bad News

This table reports the effect of bad news in soft information. In Panel A, the specifications include dummy variables (*BN* or *BN_P1* equal to one respectively if *Tone* or *Tone_P1* is nonpositive, and zero otherwise; *LongHorizon* equal to one if *Horizon* is longer than its median, and zero otherwise) interacting with *Tone* or *Tone_P1*. In Panel B, the specification includes the percentage of positive sentences (*Pos%* or *Pos%_P1*) and negative sentences (*Neg%* or *Neg%_P1*) in the current or prior report instead of aggregate textual opinions (*Tone* or *Tone_P1*), and a dummy variable (*LongHorizon* equal to one if *Horizon* is longer than its median, and zero otherwise) interacting with *Pos%* or *Neg%*. For brevity, results of control variables (same to Table 3, Panel B) are omitted. Detailed variable definitions are listed in the appendix. Firm and year fixed effects are included, and standard errors are two-way clustered by firm and year. *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

Panel A. Overall Opinion

		(1) FError	(2) FRev
Tone	β_0	-0.0007	
		(-0.91)	
Tone × BN	Υ_0	0.0018	
		(0.84)	
Tone × LongHorizon	β_{0LH}	-0.0068***	
		(-4.15)	
Tone × BN × LongHorizon	Υ_{0LH}	-0.0105**	
		(-2.22)	
Tone_P1	β_1		0.0027***
			(6.67)
Tone_P1 × BN_P1	Υ_1		0.0015**
			(2.37)
Constant		0.0249	-0.0187***
		(1.37)	(-4.04)
Controls		Yes	Yes
Firm FE		Yes	Yes
Year FE		Yes	Yes
Adj. R^2		0.230	0.125
# of Obs.		114,158	92,200

Table 7.

Panel B. Percentage of Negative and Positive Sentences

		(1) FError	(2) FRev
Pos%	β_0	-0.0013	
		(-1.47)	
Neg%	Υ_0	-0.0026	
		(-1.43)	
Pos% × LongHorizon	β_{0LH}	-0.0071***	
		(-3.77)	
Neg% × LongHorizon	Υ_{0LH}	0.0153***	
		(4.32)	
Pos%_P1	β_1		0.0027***
			(5.42)
Neg%_P1	Υ_1		-0.0035***
			(-7.51)
Constant		0.0252	-0.0186***
		(1.37)	(-4.00)
Controls		Yes	Yes
Firm FE		Yes	Yes
Year FE		Yes	Yes
Adj. R^2		0.230	0.125
# of Obs.		114,158	92,200
F-Test		$\beta_0 = - \Upsilon_0$	$\beta_1 = - \Upsilon_1$
		F-stat. = 5.33	F-stat. = 5.07
		p > F = 0.0210	p > F = 0.0244
		$\beta_{0LH} = - \Upsilon_{0LH}$	
		F-stat. = 15.25	
		p > F = 0.0001	

Table 8. Tone on Subsequent Target Price

This table presents the results estimated by OLS regressions of tone (*Tone_P1*) or tone revisions (*ΔTone_P1*) of prior analyst report on analyst target price revisions (*TPRev*). For brevity, results of control variables are omitted. Detailed variable definitions are listed in the appendix. Firm and year (or analyst and year) fixed effects are included, and standard errors are two-way clustered by firm and year (or analyst and year). *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

	(1) TPRev	(2) TPRev	(3) TPRev	(4) TPRev
Tone_P1	0.0580***	0.0659***		
	(5.40)	(5.64)		
ΔTone_P1			0.0318***	0.0303***
			(4.63)	(4.64)
Constant	-0.0523	-0.0227	-0.0690	-0.0093
	(-0.41)	(0.00)	(-0.52)	(-0.00)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	No	Yes	No
Analyst FE	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Adj. R^2	0.162	0.168	0.163	0.167
# of Obs.	52,632	52,632	50,324	50,324

Table 9. Tone on Subsequent Recommendation

This table presents the results estimated by OLS regressions of tone (*Tone_P1*) of prior analyst report on analyst recommendations (*Recom*, in panel A) or its revisions (*RRev*, in panel B and partitioned by prior recommendation level). For brevity, results of control variables are omitted. Detailed variable definitions are listed in the appendix. Firm and year (or analyst and year) fixed effects are included, and standard errors are two-way clustered by firm and year (or analyst and year). *t-statistics* are reported in parentheses. *, **, *** indicate the significance at 10%, 5%, and 1% level, respectively.

Panel A. Level of Recommendation

	(1) Recom	(2) Recom
Tone_P1	0.3314***	0.3026***
	(6.32)	(4.60)
Constant	2.3686***	2.3532***
	(11.96)	(7.39)
Controls	Yes	Yes
Firm FE	Yes	No
Analyst FE	No	Yes
Year FE	Yes	Yes
Adj. R^2	0.101	0.127
# of Obs.	15,730	15,730

Panel B. Revision of Recommendation

Prior Recommendation	Strong Sell (1) RRev	Sell (2) RRev	Hold (3) RRev	Buy (4) RRev	Strong Buy (5) RRev
Tone_P1	-0.5160	0.1163	0.5734***	0.3296***	0.2306**
	(-0.62)	(1.29)	(6.24)	(4.73)	(2.15)
Constant	4.2021	1.3866*	-0.5693*	-0.9156***	-3.0555***
	(1.60)	(1.81)	(-1.65)	(-2.86)	(-6.66)
Controls	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.023	0.018	0.115	0.169	0.165
# of Obs.	304	1,190	6,074	4,756	2,243