Predictability of Analyst Stock Recommendation Revisions

Mark Bradshaw* *Boston College*

Jared Flake *Boston College*

Mark Piorkowski *Indiana University*

This Draft: August 2022

PRELIMINARY DRAFT
PLEASE DO NOT QUOTE OR POST

Abstract

On average, analysts' stock recommendation revisions have immediate effects on stock prices. However, recent research indicates that only a small subset of recommendations are influential in the sense that they are associated with significant returns. Given the relative infrequency of recommendation revisions, we predict that analysts signal future recommendation revisions through changes in the tone of sequential research reports, which preempts market reactions when revisions are announced. We find that the signed change in tone of analysts' reports is positively associated with future recommendation revisions. Using the change in research report tone and other determinants, we construct a measure of the likelihood of a recommendation revision. This predictability measure is associated with attenuated market reactions to recommendation upgrades and a lower likelihood that the upgrades are classified as influential. We also find that our recommendation revision prediction model is better at predicting upgrades than downgrades and has similar performance to fraud prediction models in the literature.

---

*Corresponding author contact information: 140 Commonwealth Avenue, Chestnut Hill, MA 02467; email: mark.bradshaw@bc.edu. We thank Farzana Afrin, Ted Christensen, Eric Gelsomin, Amy Hutton, Stephannie Larocque (discussant), Josh Lee, Miao Liu, Sugata Roychowdhury, Jon Underwood, and workshop participants at the 2021 BYU Accounting Research Symposium, 2022 FARS Midyear Meeting, and Deakin University for helpful comments. We also thank Dean Andy Boynton for access to the sample of analyst reports, and Steve Lacey for expert research assistance. We gratefully acknowledge financial support from the Boston College Carroll School of Management and the Indiana University Kelley School of Business.

1. Introduction

Accounting and finance research includes a well-established literature on sell-side financial analysts, spanning decades and hundreds of studies (Brown 2000). A majority of these studies focus on analysts' earnings forecasts, despite these being of relatively low-order importance among analysts' primary activities. Nevertheless, earnings forecasts are an important input into analysts' valuations of companies they follow (Bradshaw 2004; Loh and Mian 2006), which provide the basis for buy, hold, or sell recommendations on the stocks of those companies. Accordingly, these stock recommendations reflect what Schipper (1991, p. 106) concludes are "the ultimate analyst judgment." In other words, an analyst's stock recommendation is the culmination of their research on a firm and is a natural focal point of investors who rely on analysts for their expertise in capital markets (Lawrence, Ryans, and Sun 2017).

Stickel (1995) and Womack (1996) provide the seminal studies on recommendations and Womack (1996) concludes that changes in these recommendations are equivalent to an analyst saying, "I have analyzed the public information, and the current stock price is not 'right'" (p. 164). Accordingly, recommendation revisions are valuable signals for investors, and prior research finds significant market reactions to recommendation revisions. Womack (1996) finds mean three-day abnormal returns to upgrades (downgrades) of 3.0% (-4.7%), a remarkably large short-window spread of 7.7%. Barber et al. (2001) examine an extensive set of consensus recommendations to determine whether investors could implement a portfolio investment strategy based on the cross-sectional distribution of recommendation levels. They find that an active strategy of rebalancing portfolio holdings daily produces extensive abnormal returns before considering hypothetical trading costs, and conclude that, "Ceteris paribus, an investor would be better off purchasing shares in firms with more favorable consensus recommendations and selling shares in those with less

favorable consensus ratings." Jegadeesh and Kim (2006) examine recommendation revisions in G7 countries and find that market reactions are not significant outside of the U.S., leading them to conclude that U.S. analysts are more skilled.

In contrast to the interpretation that recommendation revisions are valuable signals, Altinkilic and Hansen (2009, p. 17) argue that recommendation revisions are "usually information-free for investors," arguing that analysts simply piggyback off price-relevant news. Altinkilic, Balashov, and Hansen (2013) examine intraday pricing and draw the same conclusion. Further, Altinkilic, Hansen, and Ye (2016) attribute the lack of market reactions to analysts' recommendation revisions to the prevalence of algorithmic trading. Alternatively, in a similar but less extreme inference, Loh and Stulz (2011) argue that the average association between recommendations and returns is attributable to only a small subset (12%) of what they classify as "influential" revisions, with the remaining deemed by them as "non-influential."

In this study, our primary objective is to reconcile the disparate conclusions regarding the value-relevance of analysts' stock recommendations. In contrast to these sardonic views of analyst research in the studies noted above, we conjecture that the information content of recommendation revisions is likely pre-empted by analysts in reports preceding the one that accompanies a recommendation revision. Note that the majority of analysts' research reports are either explicit or implicit reiterations of their outstanding recommendation. Our premise is that the recommendation revision itself necessarily incorporates pre-revision reports that accompany the outstanding recommendation. This view is analogous to the information content of earnings as first documented by Ball and Brown (1968), where information about annual earnings is released throughout the period leading up to earnings announcement. Similarly, given mixed evidence on whether recommendation revisions are on average informative in short announcement windows,

we conjecture that much of the news in a recommendation revision may be preempted through the qualitative (e.g., report tone) and quantitative (e.g., target prices) content of analyst reports released prior to and leading up to the revision. Our primary prediction is that the observed short-window information content of analysts' recommendation revisions can be thus muted by analysts preempting information in recommendation changes, especially for upgrades. If so, then the appropriate basis for assessing the value of analysts' recommendations extends to the analyses that are often the prelude to the revisions.

Recommendations are constrained to a discrete range of either five categories (i.e., strong sell, sell, hold, buy, or strong buy) or three (i.e., sell, hold, buy). Empirically, recommendations are revised relatively infrequently (Womack 1996; Huang, Zang, and Zheng 2014; Chen, Jung, and Ronen 2017), and analysts are aware of the influence of their research on stock prices. Thus, they exhibit a reluctance to release a volatile series of 'see-saw' recommendations, consistent with evidence of an underreaction in analysts' earnings forecasts in Raedy, Shane, and Yang (2006). Accordingly, if analysts prefer to minimize the volatility of their research output, they may signal impending recommendation revisions through their narratives and other quantitative forecast revisions.

We investigate whether differential market reactions to analysts' recommendation revisions are associated with the predictability of the revision, where the predictability of the revision is based on changes in tone across analysts' reports and other determinants that precede the revision. Our prediction is predicated on the incentives analysts face, largely stemming from their institutional clients. Until recently, research has been largely funded by allocations of trading by institutional investors, as well as investment banking fees paid by corporate clients. Neither

institutional investors nor corporate clients respond favorably to surprise announcements by analysts. Indeed, Goldman Sachs Asset Management (2016) explicitly states,

> "Analysts are reticent to change their buy, sell or hold recommendations frequently, so will often subtly shift the tone of their report to indicate a potential change in view while keeping their headline rating the same. By utilizing natural language processing to parse through an analyst's writing, investors can help predict changes in their headline forecasts before those shifts take place."

We examine three hypotheses that link prior research suggesting strong average reactions to recommendation changes (Womack 1996) to more recent research showing that the average reactions are attributable to a small subset of influential recommendations (Loh and Stulz 2011). First, we corroborate our baseline assumption that the change in tone across a series of analysts' reports is associated with recommendation revisions. Second, we investigate whether market reactions to recommendation revisions are attenuated for more predictable recommendation revisions, based on changes in the tone of sequential reports and other determinants preceding the revision. Finally, we examine whether recommendations deemed influential by Loh and Stulz (2011) are concentrated among the least predictable recommendation revisions.

In addition to documenting large short-window market reactions, Womack (1996) also finds a significant post-recommendation revision drift for upgrades, consistent with the initial three-day abnormal returns being incomplete.[1] In the one-month following an upgrade, Womack (1996) finds a mean abnormal return of 2.4%, whereas the mean abnormal return for downgrades is insignificantly different from zero. The asymmetry is consistent with evidence in different settings that market reactions to bad news are swifter than reactions to equivalent measures of good news (e.g., Patell and Wolfson 1982; Skinner and Sloan 2002; Kothari, Shu, and Wysocki

[1] Other studies finding long-term abnormal returns associated with recommendations include Barber, Lehavy, McNichols, and Trueman (2001) and Jegadeesh, Kim, Krische, and Lee (2004).

2009; Hutton, Marcus, and Tehranian 2009). Thus, we supplement our primary hypotheses by examining upgrades and downgrades separately.

Our main results indicate a strong association between changes in the tone of analysts' reports and the likelihood of a recommendation revision. Further, we find an attenuated association between the predictability of recommendation revisions and market reactions, but only for upgrades, consistent with downgrades being more difficult to anticipate. Finally, recommendations deemed influential under Loh and Stulz's (2011) procedures are less predictable based on changes in the tone of analysts' prior reports and other determinants, but again, only for upgrades, consistent with analysts being more likely to signal good news than bad news. We supplement our analyses by examining the performance of our recommendation revision predictability model against fraud prediction models, because fraud is a similarly infrequent event which accounting researchers have studied extensively to develop strong predictive models (e.g., Dechow, Ge, Larson, and Sloan 2011; Bao et al. 2020). With the ability to predict fraud as a benchmark, we show that our model has similar predictive performance, especially for upgrades.

Overall, our results are consistent with changes in analysts' report tone and other determinants being leading indicators of recommendation revisions, and that these leading indicators preempt the information content at the recommendation revision announcement date. This result is an important addition to the literature on analysts' research and how it affects capital markets. Our results suggest that studies that examine recommendations on a stand-alone basis (e.g., Loh and Stulz 2011) provide a floor to any estimate of the usefulness of recommendations to investors, and that recommendations are often predictable based on the information in analysts' previous research reports. More (less) predictable recommendation upgrades are associated with attenuated (stronger) market reactions and are less (more) likely to be construed as influential

without considering previously released information by analysts that effectively preempts information in a recommendation revision. While prior research appeals to the small fraction of influential recommendation revisions as evidence of the limited usefulness of analysts' recommendations, we recharacterize this finding. The lack of influence does not suggest that the event is information free per se, but rather that analysts have already released a portion of the information to the market through the content in their prior reports.

2. Background

2.1 Analyst Recommendations

Research on stock recommendations dates back to Cowles (1932), who examined the profitability of recommendations by a columnist for *The Wall Street Journal*. The sample size was small, and he concluded that any profitable recommendation track record is as likely chance as it is skill. Additional studies in the 1970s also concluded that stock recommendations by institutional investors and research departments exhibited no systematic ability to forecast returns (e.g., Logue and Tuttle 1973; Bidwell 1977; Shepard 1977).

Subsequently, using more extensive data and precise measurement dates, numerous studies find that stock recommendations are associated with abnormal returns (e.g., Groth et al. 1979; Copeland and Mayers 1982; Bjerring, Lakonishok, and Vermaelen 1983; Bauman, Datta, and Iskandar-Datta 1995). The seminal study by Womack (1996) provides strong evidence that analysts' stock recommendations are associated with both short- and longer-window returns. Importantly, he finds that less than ten percent of recommendation revisions occur close to earnings announcements, so his results cannot be attributed to other information. Womack (1996) finds mean size-adjusted announcement returns of 3.0% for new buy ratings and -4.7% for new sell ratings. In addition, he finds evidence of a mean six-month drift of 2.4% for new buys and -

9.1% for new sells. Womack (1996) concludes that the short-term and long-run market responses to recommendation changes are consistent with analysts having the ability to identify undervalued stocks.

More recent research reinforces the conclusions in Womack (1996). For example, Kim, Lin, and Slovin (1997) examine buy recommendations released after trading hours. They find that most returns occur within the first five minutes of trading for NYSE stocks and the first fifteen minutes for Nasdaq stocks. Barber et al. (2001) find that investors can use analyst recommendation revisions to form portfolios and generate significant abnormal returns. Green (2006) examines the profitability to investors who might receive early access to analyst stock recommendation revisions. He concludes that, even considering transactions costs, investors with early access could generate a 1.02% (1.50%) abnormal two-day return following upgrades (downgrades), and that the investors who take positions within the first few hours after a recommendation revision would also generate significant abnormal returns.

The market reaction to recommendation revisions is incremental to earnings forecast revisions (Francis and Soffer 1997), and recommendation revisions accompanied by earnings forecast revisions generate greater market reactions than recommendation revisions without earnings forecast revisions (Kecskés, Michaely, and Womack 2017). The market response may also be attributable in part to attention recommendations receive. For example, stock recommendations are the most frequently discussed analyst research output in media articles (Bonner, Hugon, and Walther 2007).[2] Institutional investors' trading is also consistent with recommendations having investment value. Using consensus analyst recommendation data,

[2] In a sample of media articles, Bonner, Hugon and Walther (2007) find that 11% of media observations mention stock recommendations, compared to 4% for earnings forecasts and 2% for target prices.

Brown, Wei, and Wermers (2014) find that mutual funds herd into (out of) stocks with consensus upgrades (consensus downgrades).[3]

Recent research, however, has called into question the extent to which recommendation revisions are significant market events. Most notably, as discussed previously, Loh and Stulz (2011) examine variation in market reactions to recommendation revisions and find that only a small subset (12%) of revisions can be deemed "influential."[4] They consider an individual recommendation revision to be influential if it has a short-window abnormal return that is statistically significant and in the same direction as the recommendation revision. In addition, Loh and Stulz (2011) suggest that many recommendations are potentially confounded by firm-specific news, and therefore exclude recommendation revisions taking place on days with firm-specific news events.[5]

2.2 Analyst Reports

We use a large sample of analyst reports to investigate whether analysts preempt recommendation revisions through the research conveyed to investors in their reports leading up to the recommendation revision. We use the text of these reports to generate a change-in-tone measure, which we conjecture will be a predictor of subsequent recommendation revisions. This conjecture follows Asquith, Mikhail, and Au (2005), who reference the following quote from *Business Week Online* on the value of analyst reports: "In the end, stock ratings and target prices

[3] Prior research also finds that analysts herd in stock recommendations (Welch 2000; Jegadeesh and Kim 2010).

[4] Altınkılıç and Hansen (2009) find that almost 80% of revisions are in response to corporate events, and that the average return to revisions is small on average when using intraday returns. However, Bradley et al. (2014) find that time stamps in I/B/E/S for recommendation revisions released during trading hours are systematically delayed. Using intraday returns with newswire-reported time stamps, Bradley et al. (2014) find large average announcement returns following recommendation changes.

[5] Loh and Stulz (2011) exclude recommendation changes occurring in the three-day window around earnings announcements and management earnings guidance days, as well as days with multiple analysts releasing recommendations for the same firm.

are just the skin and bones of analysts' research. The meat of such reports is in the analysis, detail, and tone." Indeed, prior research documents that these reports include detailed discussions of nonfinancial topics such as industry competition and management quality (Previts et al. 1994; Asquith, Mikhail, and Au 2005; Huang et al. 2014).[6]

Prior research reveals that report sentiment is associated with contemporaneous returns and is incremental to contemporaneously issued recommendations, earnings forecasts, and target prices (Asquith, Mikhail, and Au 2005; De Franco and Hope 2011; Twedt and Rees 2012; Huang et al. 2014).[7] Huang et al. (2014) find that there is a stronger market reaction for text with a negative tone, text emphasizing nonfinancial topics, and text that is more assertive and concise. In addition, Huang et al. (2014) find that the market response to report text is supported by fundamentals, as tone is associated with future earnings growth for up to the subsequent five years.

2.3 Frictions Impeding Timely and Complete Recommendation Revisions

The significant average market reaction to recommendation revisions may be partly attributable to revisions being relatively infrequent events. The vast majority of analyst reports contain a reiteration of an existing recommendation (e.g., Womack 1996; Chen, Jung, and Ronen 2017; Huang et al. 2014). For example, Chen et al. (2017) find that over 90% of analyst reports in their sample contain a reiteration (as opposed to an initiation, termination, upgrade, or downgrade). Similarly, over 94% of sample reports in Huang et al. (2014) are reiterations.[8]

[6] Insight into analysts' views on the quality of firm management is important for investors to understand. For example, Barker (1999) finds that analysts rank their assessment of management quality as their most important information source.

[7] De Franco, et al. (2015) find that analyst report readability is positively associated with short-term trading volume surrounding the report release date.

[8] Refer to Table 2, Panel B of Huang et al. (2014). Approximately 5.5% of sample reports contain a recommendation change (2.7% upgrades, 2.8% downgrades).

Prior research provides several explanations for why recommendation revisions are relatively infrequent. First, unlike the continuous nature of earnings forecasts and target prices, recommendations are discrete measures. In addition, while brokerage houses historically used a five-tier rating system, following the reforms to equity research in the early 2000s, the majority of large research departments shifted to a three-tier system (Barber et al. 2006; Kadan et al. 2008).[9] Kadan et al. (2008) find that the overall informativeness of recommendation revisions decreases in the post-reform era especially for brokers using a three-tier system, consistent with the coarser set of recommendation levels reducing the information available to investors.

In addition, several studies find that analysts are sluggish to impound negative earnings forecasts into recommendations. In contrast, Skinner (1994) examines managers and finds managers tend to disclose bad news before earnings announcements. He suggests that expected litigation and reputation concerns incentivize managers to preempt bad news. Although managers and analysts are both concerned about their reputations, litigation risk has limited influence on analysts. Instead, analysts are largely concerned about repercussions from institutional investor clientele (Agrawal and Chen 2008). Institutional investors who hold stocks, potentially on the recommendation of analysts, frown on downgrades that will negatively influence returns, and similarly prefer to have some indication prior to analysts upgrading stocks, when portfolio managers might strategically invest.

[9] NASD Rule 2711, which was approved by the SEC in May 2002 with an effective date of September 2002, required the public disclosure of research departments' rating distributions. The rule required the departments to report the percentage of their recommendations classified using three categories (buys, holds/neutrals, and sells), regardless of how many categories the departments used in their rating systems. Barber et al. (2006) and Kadan et al. (2008) note that many brokers shifted from a five-tier rating system to a three-tier one in the days and weeks leading up to the effective date of NASD Rule 2711. The group of brokers switching to a three-tier system included all ten of the original Global Settlement sanctioned investment banks, and three-tier brokers represent 75% of recommendations in I/B/E/S in the post-reform period, compared to 17% of I/B/E/S recommendations in the pre-reform period (Kadan et al. 2008).

Consistent with pressures from institutional investors preventing analysts from making unannounced bold research calls, Raedy, Shane and Yang (2006) find an underreaction in analyst earnings forecasts, with the magnitude of the underreaction positively associated with the forecast horizon. Using an analytical model, they argue that a horizon-dependent underreaction is a rational response to analysts' asymmetric loss function.[10] As recommendations generally have a longer time horizon than earnings forecasts and limited discrete levels, analysts' underreaction to bad news should be even greater for recommendations.

Similarly, Xiao and Zang (2017) provide evidence of another form of analyst underreaction – that analysts exhibit a conservative bias in "hardening" their soft information into quantitative summary measures. Using a large sample of analyst reports for S&P 500 firms, they find evidence consistent with earnings forecast revisions undershooting the revisions implied by the analyst report tone.[11] The extent of the conservative bias is stronger when the information signals are of lower quality (e.g., longer forecast horizon) and for analyst reports with negative tone.

Bernhardt, Wan and Xiao (2016) find evidence consistent with analysts being reluctant to issue frequent recommendation revisions in response to small changes in analysts' valuation assessments. They discuss several sources of stickiness in analyst recommendations, such as public information arriving to the analyst in a lumpy manner (e.g., quarterly earnings announcements) and strategic recommendation revision decisions made by analysts (e.g., analyst requiring a significant amount of information to upgrade or downgrade to a neutral rating). Analysts may also

[10] The asymmetric loss function arises under the assumption that analysts experience larger reputational costs when new information results in a change in investor expectations in the opposite direction of the prior forecast revision, relative to the reputational cost arising from a change in investor expectations in the same direction as the prior revision.

[11] Xiao and Zang (2017) focus on earnings forecasts in their main analysis. In supplemental analysis, they also find that conservative bias exists in target prices and stock recommendations.

be reluctant to issue revisions to avoid internal review. For example, the Global Settlement mandated that sanctioned banks establish oversight committees to review all recommendation changes and material target price revisions (e.g., Piorkowski 2021).

Overall, these prior studies suggest that because analysts are reluctant to revise their recommendations frequently, they may use other channels to communicate information pertinent to an ultimate revision of their existing recommendation.

2.4 Prediction Models for Infrequent Events

Our research setting aims to use both quantitative and qualitative information to predict infrequent recommendation revisions. Our research methodology is related to other studies that also have an objective of predicting rare events. Most notably, the accounting literature includes studies that attempt to predict bankruptcy, earnings manipulation, and accounting misstatements, such as Altman (1968), Beneish (1999), and Dechow et al. (2011), respectively. Like these studies, our objective is to identify the determinants of an infrequent event – a recommendation revision. Accordingly, we adopt similar approaches to those in these studies, but include qualitative rather than quantitative information as our primary explanatory variable. This approach is similar to Perols et al. (2017) and Bao et al. (2020), who both expand explanatory variables in fraud prediction to those based on alternative data.

3. Hypotheses

3.1 Recommendation Revision Predictability

Significant average market reactions associated with recommendation revisions (e.g., Womack 1996) provide investors with an incentive to predict recommendation revisions before they occur. Moreover, the relatively infrequent nature of revisions due to the three-tier rating

systems (e.g., Kadan et al. 2008) and analysts' horizon-dependent underreaction (Raedy, Shane, and Yang 2006) suggests that there is an opportunity to predict revisions. Alternatively, if Loh and Stulz (2011) are correct that most recommendation revisions are not associated with significant returns, it is possible analysts preempt such revisions earlier, leading to muted market reactions. Thus, whether revisions are predictable is an empirical question.

We argue that changes in analyst report sentiment provide a valuable signal that is associated with future recommendation revisions. The informativeness of analyst report sentiment may be attributable in part to the implications of changes in the nature of the tone with which analysts discuss firm fundamentals. In particular, because there are, on average, over nine analyst reports per recommendation revision (Huang et al. 2014; Chen, Jung, and Ronen 2017), the evolution of report tone over time may provide insight into the likelihood that the analyst's evolving views materialize into a recommendation revision, which leads to our first hypothesis.

H1: The change in analyst report tone predicts future recommendation revisions.

3.2 Recommendation Revision Predictability and Market Reactions

The market reactions to analyst reports documented in prior research that controls for quantitative elements suggest investors view the textual discussion as value-relevant. If investors perceive changes in sentiment of analysts' qualitative discussions as a signal of future recommendation revisions, reactions to subsequent revisions would reflect attenuated market reactions because the news has been preempted. The market reaction at the time of a recommendation revision is then influenced by the extent to which the revision was anticipated in advance. Therefore, we hypothesize that there is an attenuated relationship between the

predictability of a recommendation revision (based on ex-ante changes in tone and other determinants) and the market reaction to the recommendation revision.[12]

> **H2**: The predictability of a recommendation revision is associated with an attenuated short-window market reaction centered on the recommendation revision.

3.3 Predictable Recommendation Revisions and Influential Status

Finally, as an extension to the previous hypothesis, we also expect that predictability of recommendation revisions influences the likelihood they will be deemed "influential." Loh and Stulz (2011) classify a recommendation revision as influential if it is associated with a short-window abnormal return that is statistically significant and in the same direction as the revision. We measure the predictability of a revision, and triangulate our results with Loh and Stulz (2011) by testing whether a more predictable revision is associated with lower market reaction, and hence, lower probability of being labeled as influential under Loh and Stulz's (2011) methodology. Evidence consistent with our second hypothesis might tautologically result in a recommendation revision being less likely classified as influential, we formally test this hypothesis to triangulate our results, Womack (1996), and Loh and Stulz (2011).

> **H3**: The predictability of the recommendation revision is negatively associated with the likelihood that the recommendation revision is classified as influential.

An implication of this hypothesis is that attenuated market reactions to the non-influential recommendation revisions are attributable in part to the market anticipating the revisions. Accordingly, most recommendation revisions appear non-influential not because they are uninformative per se, but because the revisions are anticipated by the market based on the evolution

[12] As recommendation upgrades (downgrades) are associated with positive (negative) market reactions on average (e.g., Womack 1996), we expect that the predictability of upgrades (downgrades) is negatively (positively) associated with the short-window market reactions to the revisions.

of sentiment conveyed qualitatively by the analyst, along with other determinants (e.g., difference between the analyst's outstanding target price and current stock price).

4. Data and Research Design

4.1 Data and Sample

Our sample consists of the intersection the following data: (i) analyst reports downloaded from Thomson One; (ii) recommendation, target price, and earnings forecast data from I/B/E/S; (iii) financial statement data from Compustat; and (iv) stock return characteristics from CRSP. The main sample consists of 191,017 observations including 470 firms, 1,436 analysts, 6,771 analyst-firm pairs, and 14 years from 2004-2017. See Table 1 for a summary of our sample selection procedure. Details are discussed below.

From Thomson One, we collect 693,333 analyst reports issued for S&P 500 firms as of 2017. As these reports contain a significant amount of boilerplate disclosure, we follow Huang et al. (2014) and hand check several reports for each broker to identify the pattern that distinguishes their boilerplate disclosures. We then use that pattern to identify the boilerplate sections in the remaining reports and we remove them from the analysis. Due to the limitation in resources, we keep only reports written by top research brokers by output. This filter leaves us with 422,000 reports. We drop reports that contain fewer than 5 sentences and firms for which we do not have a valid I/B/E/S ticker match. As we create a panel of reports for each firm-analyst pair, the order of the report matters. We thus aggregate the number of sentences and tone in reports written on the same day to a single observation per analyst-firm day. We merge in recommendation data for which we do not have reports, because in some cases the I/B/E/S detail recommendation file has recommendation changes for which we do not have reports. Because we use the changing

sentiment in the reports *prior* to the revision to predict recommendation revisions, we can include these recommendation changes in our sample.

To merge our sample of reports with I/B/E/S, we match the Thomson ticker with the official ticker on I/B/E/S. For the remaining unmatched firms, we use the I/B/E/S ticker. We use Thomson and I/B/E/S company names to corroborate our merge on ticker is valid. We find a valid merge for 470 firms. We obtain a list of unique author names from I/B/E/S and Thomson and perform a fuzzy match on the authors' last names and corroborate that they are from the same broker in the I/B/E/S and Thomson dataset. We use a conservative approach and keep only those matches with a similarity score greater than 90 and from the same broker.

After we merge the datasets and require non-missing control variables, we have a sample of 191,017 observations, 7,343 of which are recommendation revisions.[13] We also filter recommendation revisions following Loh and Stulz (2011) to classify them as influential. To classify recommendation revisions as influential, we require that the recommendation does not occur on the same day as other information events: firms' earnings announcement, issuance of guidance, or recommendation revisions by other analysts for the same firm. This restriction and the additional control variable requirements lead to 5,549 recommendation revisions remaining in our sample. Table 5 and Table 6 exclude observations prior to 2005 to incorporate the use of rolling windows to create the *Upgrade (Downgrade) Predictability*. This step reduces our sample to 4,939 recommendation revisions.

[13] We define recommendation revisions using five tiers. While Kadan et al. (2008) suggest using three tiers, some brokers in our sample use five tiers. Forcing a five-tier system to a three-tier system would incorrectly classify some recommendation changes as reiterations.

4.2 Parsing Analyst Reports to Capture Analyst Sentiment

Our main measure of tone uses a refined version of the language model BERT, specifically trained to classify sentences from financial texts as positive, negative, or neutral. BERT was originally developed by Google (Devlin et al. 2018) to incorporate the context surrounding words and sentences from the left and the right to improve classification tasks. Researchers have trained BERT on a set of 10,000 sentences extracted from analyst reports and hand-classified as positive, negative, or neutral (Huang et al. 2014; Huang, Wang, and Yang 2021; Yang, Uy, and Huang 2020). Huang et al. (2020) Table 1 reports the overall accuracy for several classification techniques used in prior research. They find that FinBERT has an accuracy of 88.2%, BERT has an accuracy of 85.2%, Loughran-McDonald Dictionary has an accuracy of 62.1% (with substantially worse performance for positive and negative sentences), and the naïve Bayes model, which was trained on the same set of sentences from analyst reports, has an accuracy of 73.6%.[14]

We first keep only standard English characters by removing non-ascii characters from the text. Each broker has one or more standard boiler plate disclosure sections at the end of the report. We manually examine the reports to find the patterns that distinguish these disclosure sections from the relevant content in the reports. These delimiters are used to create regular expressions to find and remove the boiler plate disclosures from the text. We then split the report into individual sentences and use the FinBERT algorithm to classify each sentence as positive, negative, or neutral. Tone is then aggregated at the daily level and measured as the number of positive sentences minus the number of negative sentences, scaled by the total number of sentences.

$$Tone = \frac{\#\ Positive\ Sentences - \#\ Negative\ Sentences}{\#\ Sentences}$$

[14] Our results remain qualitatively similar if we use the Loughran-McDonald dictionary to classify analyst report sentiment.

We capture analysts' evolving sentiment toward a firm as the change in tone since their last recommendation change (*ΔTone*) to represent the total shift in an analyst's sentiment since the previous recommendation revision. We measure *ΔTone* by subtracting the prior report tone ($Tone_{ijt-1}$) from the analyst's baseline tone ($Tone_{ijt-T}$).

$$\Delta Tone_{ijt-1} = Tone_{ijt-1} - Tone_{ijt-T}$$

where $t - T$ is the date of the prior recommendation change, i is the index for firms, j is the index for individual analysts, and t is the index for the report date. The analyst's baseline tone is the tone of the first available report written on a firm since the analyst's recommendation change for that firm. If we have the report that includes the prior recommendation change, the recommendation change report is the baseline tone. If we do not have the recommendation change report, we use the tone from the first available report following the prior recommendation change as the baseline tone. The baseline tone may also be the initiating coverage report tone.

4.3 Recommendation Revision Prediction Model

We develop a model to predict recommendation revisions. The primary variable of interest in our prediction model is *ΔTone*. We also include an array of analyst-firm and firm characteristics that have been found to be determinants of valuations and recommendations. We measure all values as of the *prior* report, given our model is meant to be predictive. We collect data for the momentum and volatility of stock returns for the six months leading up to and ending 5 days prior to the report date to ensure *ΔTone* is incremental to other information (e.g., firm news).

The additional analyst-firm variables include the analysts' prior recommendation level, *Rec*; the change in target price or one-year ahead earnings forecast, if any, during the month leading up to the prior report date as *ΔTP* and *ΔEF*, respectively; the ratio of outstanding target price minus

stock price five days prior to the report date, scaled by stock price prior to report date as *(TP-P)/P*; and the natural log of the number of days since the analysts last changed their recommendation (*Days Since Rec Change*). The additional firm characteristics include the firms' return on assets, *ROA*; the natural log of the firms market capitalization, *Ln(MVE)*; the ratio of the firms' book value to equity value, *BTM*; *Leverage*; year-over-year sales growth, *Sales Growth*; and the fraction of the shares held by institutional investors, *Institutional Ownership*. To control for additional time invariant analyst and firm characteristics related to the propensity for recommendation revisions, we include analyst and firm fixed effects. To control for any time trends related to *ΔTone* and *ΔRec* (e.g., financial crisis), we include year fixed effects. To identify the variables related to *ΔRec* and measure *Upgrade (Downgrade) Predictability*, we estimate the OLS regression in Equation 1 and alternate the dependent variable with *Upgrade* and *Downgrade*.[15] In the specification with *Upgrade (Downgrade)* as the dependent variable, we restrict the sample to reports with a prior recommendation that is not already a buy (sell), because such recommendations cannot be further upgraded (downgraded).[16]

$$\begin{aligned} Rec\ Change_{ijt} &= \beta_1 \Delta Tone_{ijt-1} + \beta_2 Rec_{ijt-1} + \beta_3 (TP_{ijt-1} - P_{it})/P_{it} + \beta_4 \Delta TP_{ijt-1} \\ &+ \beta_5 \Delta EF_{ijt-1} + \beta_6 SixMonth\ BHAR_{it} + \beta_7 SixMonth\ Volatility_{it} \\ &+ \beta_8 Ln(Days\ Since\ Rec\ Change)_{ijt-1} + \beta_9 ROA_{it-1} + \beta_{10} Ln(MVE)_{it-1} \\ &+ \beta_{11} BTM_{it-1} + \beta_{12} Leverage_{it-1} + \beta_{13} Sales\ Growth\ (y/y)_{it-1} \\ &+ \beta_{14} Institutional\ Ownership_{it-1} + \gamma_{YEAR} + \omega_{ANALYST} + \eta_{FIRM} + \epsilon_{ijt} \end{aligned} \quad (1)$$

4.4 The Predictability of Recommendation Revisions and their Influential Status

[15] We use a linear probability model (LPM) in place of a logit specification following prior research (e.g., Hanlon and Hoopes 2014; Yost 2021) and corroborated by Angrist and Pischke (2008). In untabulated analyses, we corroborate that our results are robust to the use of logit models. Table 7 provides additional evidence consistent with the use of OLS models as they perform comparably well to the Logit models.

[16] While a buy (sell) recommendation can be upgraded (downgraded) to a strong buy (strong sell), it happens infrequently in our sample (see Panel C of Table 2).

The final part of our main analysis focuses on the market reaction to the recommendation revision and examines whether the revision is influential using the predictability of recommendation revisions. We modify Equation 1 by replacing *ΔRec* with *Upgrade* (*Downgrade*) on the left-hand side and removing the fixed effects. We then estimate a regression for each year that uses all observations prior to that year. This method ensures we only use ex ante data to create predictability measures. We remove the fixed effects because, particularly in the early years, we have few observations in each group and are unable to obtain good fixed effect estimates.[17] We then use the resulting predicted values as *Upgrade (Downgrade) Predictability* to estimate the OLS regression in Equation 2.

$$\begin{aligned} CAR_{ijt} \text{ or } Influential_{ijt} \\ &= \beta_1 Predictability_{ijt-1} + \beta_2 \Delta Tone_{ijt-1} + \beta_3 (TP_{ijt-1} - P_{it})/P_{it} + \beta_4 \Delta TP_{ijt-1} \\ &+ \beta_5 \Delta EF_{ijt-1} + \beta_6 SixMonth\ BHAR_{it} + \beta_7 SixMonth\ Volatility_{it} \\ &+ \beta_8 Ln(Days\ Since\ Rec\ Change)_{ijt-1} + \beta_9 ROA_{it-1} + \beta_{10} Ln(MVE)_{it-1} \\ &+ \beta_{11} BTM_{it-1} + \beta_{12} Leverage_{it-1} + \beta_{13} Sales\ Growth_{it-1} \\ &+ \beta_{14} Institutional\ Ownership_{it-1} + \beta_{15} |\Delta Rec|_{ijt} + \beta_{16} AwayFromCons_{ijt-1} \\ &+ \beta_{17} AvgLn(Turnover)_{it-1} + \gamma_{YEAR} + \omega_{ANALYST} + \eta_{FIRM} + \epsilon_{ijt} \end{aligned} \tag{2}$$

CAR is defined as the two-day cumulative abnormal return around the recommendation revision, adjusted for the value-weighted return in CRSP. A recommendation is deemed *Influential* if the two-day CAR is in the same direction and significantly greater than 1.96 standard deviations of the expectation given the prior three-month volatility of daily returns (Loh and Stulz 2011). To control for any remaining correlation between the independent variables from Equation 1 with the error term in Equation 2, we include the explanatory variables from Equation 1 in Equation 2.[18]

4.5 Validation of Tone Measurement

[17] The results in Table 5 and Table 6 are robust to the use of a variety of fixed effect structures to create the predictability measures.
[18] We exclude *Rec* from Equation 2 due to high collinearity with *Upgrade (Downgrade) Predictability*.

Panel A of Table 2 reveals report tone increases monotonically from downgrade to upgrade. Also, reports containing upgrades or downgrades are longer than reiteration reports, and upgrade reports are longer than downgrade reports. Panel B of Table 2 shows a similar pattern for recommendation levels. These results are consistent with Huang et al. (2014; hereafter HZZ). Finally, Panel C of Table 2 presents the prior and new recommendation levels for recommendation changes.

Figure 1 tracks the mean level of tone and the number of sentences in each year. The figure indicates that analyst report sentiment is related to market performance. For example, Figure 1 shows a decrease in tone during 2007, a sharp drop in 2008 and a rebound in 2010. Tone stays relatively constant throughout the remainder of the sample. We also find that reports in the early period of the sample are longer than reports in the later period. The movement of tone and sentence length over time is consistent with Figure 1 in HZZ.

Table 3 indicates that the average report has a tone of 0.107, consistent with optimistic tilt as suggested by HZZ. Although our average tone is smaller than the average tone in HZZ of 0.187, we find similar numbers of positive and negative sentences. Thus, our lower level of tone is likely due to us having a greater number of sentences classified as neutral.[19] In our sample, the average report contains 79 sentences while HZZ Table 2 Panel A reports an average of 57 sentences.

4.6 Descriptive Statistics

Table 3 Panel A presents descriptive statistics for the inputs for our regression analysis. *ΔRec* has a mean of -0.002 with slightly more downgrades than upgrades: 1.8 (2.0)% of reports

[19] The relatively high number of sentences classified as neutral is likely due to an imperfect parsing of the analyst reports so that tables and other non-relevant portions of the report are not fully extracted. These sentences are generally classified as neutral. An additional explanation of these differences is our restriction to the largest brokers, which are on average less optimistic relative to smaller brokers (e.g., Jacob, Rock, and Weber 2008).

contain upgrades (downgrades). The remaining reports do not contain recommendation revisions ("reiteration" reports). While the average report tone is 0.107, the mean *ΔTone* is -0.004. The mean *Rec* is 3.571, which is consistent with a greater number of buy recommendations (1=Strong Sell, 2=Sell, 3=Hold, 4=Buy, 5=Strong Buy). *ΔTP* is slightly positive (mean of 0.009), while *ΔEF* is slightly negative (mean of -0.001). The mean (median) number of days since the prior recommendation revision is 698 (476), which is similar to the time between recommendation revisions for large brokers in Boulland, Ornthanalai, and Womack (2017). *ROA* is on average 0.015 while the mean *MVE* is 32,093 million which is consistent with a sample of S&P 500 firms. Each analyst has on average 133 reports in our sample with an average of 28 reports per analyst-firm pair. There are on average 406 reports per firm.

Panel B of Table 3 presents descriptive statistics for the 5,549 recommendation changes we use in our analysis of the predictability and influential status of recommendation revisions. *CAR[0,1]* has a mean of 0.000 and we classify 15.8% of recommendation changes as influential. This result is in line with Loh and Stulz (2011), who find that 12% of recommendation changes are influential. Given our sample of recommendation changes from the S&P 500 firms and the largest brokers, we expect our sample of recommendation revisions to contain a higher percentage of influential revisions than those in Loh and Stulz (2011). *Upgrade* (*Downgrade*) *Predictability* has a mean value of 0.028 (0.022) and a standard deviation of 0.028 (0.019). There are on average 12.9 reports and 552.2 days between recommendation revisions.

5. Predicting Recommendation Changes

5.1 Association Between ΔTone and Future Rec Change

We find that the association between *ΔRec* and *ΔTone* is significantly positive.[20] Table 4 presents results from our regression analysis. Column 1 is a regression of *ΔRec* on *ΔTone* and a set of control variables including year, analyst, and firm fixed effects. Under this strict fixed effect structure, we are interested in the association between the within analyst *ΔRec* and the within analyst *ΔTone*. We find that a one standard deviation change in *ΔTone* is associated with a 0.004 increase in *ΔRec* relative to the unconditional mean of -0.002.[21] A one standard deviation change in *ΔTone* is associated with a 0.005 (0.002) increase in *Upgrade* (*Downgrade*) likelihood which is 12.7% (10.8%) of the mean of 0.039 (0.021).[22] While the coefficient and R^2 in Table 4 are small, our model has significant predictive power when compared to models that predict similarly rare events, such as the bankruptcy, earnings manipulation, and accounting misstatement prediction literatures (e.g., Dechow et al. 2011; Bao et al. 2020).

The recommendation level (*Rec*) is negatively associated with *ΔRec*, which is mechanical because (in a three-tier ranking) a buy (sell) recommendation can only by downgraded (upgraded). Prior six-month buy-and-hold returns are not associated with the recommendation change. *Ln(Days Since Revision)* is positively associated with *ΔRec*. In Column 1, the associations offset when predicting a change, but in Columns 2 and 3 where we predict upgrades and downgrades

[20] We find that using the Loughran-McDonald sentiment dictionary to measure *ΔTone* produces similar results; however, *ΔTone* using FinBERT produces the strongest results both in economic magnitude and statistical significance. Similarly, results are robust to measuring *ΔTone* by subtracting the prior report tone ($Tone_{ijt-1}$) from the previous report tone ($Tone_{ijt-2}$). Not surprisingly, using this shorter measurement window for *ΔTone* decreases the predictive power of our model. First, shortening the forecast horizon decreases the signal-to-noise. Second, if tone change is non-linear, using fewer reports increases the volatility of the tone measure. When we include both the cumulative tone change and one-period tone change measures in the model, only the cumulative tone change measure loads significantly.

[21] Result are robust to a variety of fixed effect structures (e.g., analyst-firm) and to the use of a logit model.

[22] The upgrade (downgrade) frequency is calculated from the sample of reports where the analyst does not already hold a buy (sell) or strong buy (strong sell) recommendation.

separately, the coefficient is significantly positive. Columns 2 and 3 provide similar results when we replace the dependent variable with *Upgrade* and *Downgrade*, respectively.

5.2 Recommendation Revision Predictability and Influential Status

Table 5 displays the regression results for the test of the relation between the market reaction to the *ΔRec* and the predictability of the *Upgrade (Downgrade)*.[23] We restrict the sample to upgrades and downgrades and calculate *Upgrade (Downgrade) Predictability* as the predicted value from rolling window regressions of a modified version of Equation 1 that incorporates only ex-ante data. The results reveal a muted market response to predictable recommendation changes with a coefficient of -0.139 for our baseline model (column 1) and t-statistic of -2.534 in the upgrade sample. We also find a negative association between upgrade predictability and market reaction when we add an array of control variables in column 2 (coef. = -0.123; t-stat = -1.748). We find that a one standard deviation change in *Upgrade Predictability* is associated with a 0.22% decrease in *CAR [0,1]* which is 13.4% of the sample mean for upgrades of 1.66%. In the downgrade sample, we obtain a coefficient of 0.027 with a t-statistic of 0.345 (column 3). The decreased significance in the downgrade sample likely results from the fact that our model has a more difficult time predicting downgrades than it does upgrades (see Table 7), which is consistent with downgrades being less predictable to investors. These results suggest that analysts' sentiment along with the other factors included in the first stage are leading indicators of future recommendation revisions, and investors have largely priced in this preempted information prior to the recommendation revision.

[23] Table 3 Panel B presents descriptive statistics for the sample used in Tables 5 and 6.

We classify recommendation revisions as influential following Loh and Stulz (2011). Specifically, a revision is classified as influential if the two-day CAR around the recommendation change is significantly greater than 1.96 standard deviations given the prior three months of daily returns and in the same direction as the recommendation change. We present results from regressions of *Influential* on *Upgrade (Downgrade) Predictability* and a set of control variables in Table 6. As in Table 5, Columns 1 and 2 include the results in the upgrade sample and Columns 3 and 4 include the results in the downgrade sample. Columns 1 and 3 contain the baseline results with only analyst, firm, and year fixed effects. In Columns 2 and 4 we add in various controls for the relation between *Upgrade (Downgrade) Predictability* and our *Influential* indicator. The results are consistent with those in Table 5 suggesting that upgrades that are more predictable are less likely to be influential. We find that a one standard deviation increase in *Upgrade Predictability* is associated with a 2.98% decrease in *Influential* likelihood which is 18.2% if the sample mean for upgrades of 16.4%.

This evidence indicates that recommendation revisions may be classified as non-influential, not because the analyst does not provide valuable research, but because the market has already priced in the recommendation revision due to the analysts' research and other factors. Again, our results are stronger in the upgrade sample with *Upgrade Predictability* because (based on our Table 7 results) our models are better able to predict upgrades relative to downgrades.

5.3 Performance Evaluation of Predictive Model

To better understand how well our recommendation revision prediction model performs, we compare our model's performance to the performance of state-of-the-art models used to predict fraud (e.g., Dechow et al. 2011; Bao et al. 2020). Bao et al. (2020) provide a summary of their metrics and discussion of the difficulty of evaluating predictive models with significant class

imbalance. We use the fraud prediction setting as a comparison, because it is a well-studied prediction task in the accounting literature. See Appendix B for detail on how we create the metrics and how we interpret them. Upgrades (Downgrades) comprise only 1.8 (2.0)% of the total sample and 1.6 (1.8)% in the test period from 2015-2017. In the test period, upgrades (downgrades) represent 3.3 (1.9)% of the subsample of the reports with an outstanding non-buy (non-sell) recommendation. As in Bao et al. (2020) we split our sample into training and test subsamples. We use 2004-2014 to train our model and 2015-2017 to test it. We use each report with the necessary data available to predict whether the next report will contain an upgrade (downgrade). The reports with predicted values in the top 3% (2%) are classified as predicted upgrades (downgrades) to match the frequency in our sample. The Logit and OLS models exhibit similar performance, and both outperform the fixed effects models.

Our models are better at predicting upgrades than downgrades, except with the AUC metric, as previously discussed (see Table 7).[24] The significant difference in the performance of the models predicting upgrades and the models predicting downgrades in part explains the differences in the results when we split our sample by upgrades and downgrades in Tables 5 and 6. Our simple OLS predictive model has a similar performance as the sophisticated RUSBoost model in Bao et al. (2020). In their fraud prediction model, Bao et al. (2020) report sensitivity measures of 3.99% for their logit model that uses the 14 financial ratios from Dechow et al. (2011) and 4.88% for their model that applies sophisticated machine learning techniques. Sensitivity measures of 7.0% for upgrades and 6.1% for downgrades may seem small, but we use more than

[24] One difference in our evaluation is that Bao et al. (2020) classifies the top 1% of observations in each year as fraud, whereas we pool our predictions and classify the top 3% (2%) of observations over the entire period as an upgrade (downgrade).

190,000 observations to predict approximately 7,000 events, which strikes us as surprisingly efficient.

6. Conclusion

Recommendation revisions are rare events, which have strong effects on stock prices. We examine whether analysts preempt recommendation revisions through their qualitative and quantitative output leading up to the observed revisions. We use a large sample of analyst research reports to measure changes in the tone across sequential reports. We find a positive association between change in tone and the probability of a subsequent recommendation revision. More importantly, recommendation upgrades that are more predictable using our model are associated with attenuated market reactions. Loh and Stulz (2011) find that only a small subset of recommendations are deemed influential, in the sense that they result in returns consistent with the significant overall results in Womack (1996). Our results extend our understanding of the role of analysts, and the usefulness of recommendations specifically, by providing evidence that the informativeness of recommendations revisions are preempted through changes in tone across research reports. In addition to extending the literature, we provide evidence that investors can track qualitative changes in analysts' research reports to anticipate infrequent changes in recommendations.

References

Agrawal, A., and Chen, M. A. (2008). Do analyst conflicts matter? Evidence from stock recommendations. *The Journal of Law and Economics*, 51(3), 503-537.

Altınkılıç, O., Balashov, V. S., and Hansen, R. S. (2013). Are analysts' forecasts informative to the general public? *Management Science*, 59(11), 2550-2565.

Altınkılıç, O., and Hansen, R. S. (2009). On the information role of stock recommendation revisions. *Journal of Accounting and Economics*, *48*(1), 17-36.

Altınkılıç, O., Hansen, R. S., and Ye, L. (2016). Can analysts pick stocks for the long-run? *Journal of Financial Economics*, 119(2), 371-398.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), pp 598-608.

Angrist, J. D., and Pischke, J. S. (2008). *Mostly harmless econometrics*. Princeton University Press.

Asquith, P., Mikhail, M. B., and Au, A. S. (2005). Information content of equity analyst reports. *Journal of Financial Economics*, *75*(2), 245-282.

Ball, R., and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 159-178.

Bao, Y., Ke, B., Li, B., Yu, Y. J., and Zhang, J. (2020). Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research*, *58*(1), 199-235.

Barber, B., Lehavy, R., McNichols, M., and Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance*, *56*(2), 531-563.

Barber, B. M., Lehavy, R., McNichols, M., and Trueman, B. (2006). Buys, holds, and sells: The distribution of investment banks' stock ratings and the implications for the profitability of analysts' recommendations. *Journal of Accounting and Economics*, *41*(1-2), 87-117.

Barker, R. G. (1999). The role of dividends in valuation models used by analysts and fund managers. *European Accounting Review*, *8*(2), 195-218.

Bauman, W. S., Datta, S., and Iskandar-Datta, M. E. (1995). Investment analyst recommendations: a test of 'the announcement effect' and 'the valuable information effect.' *Journal of Business Finance & Accounting*, 22(5), 659-670.

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24-36.

Bernhardt, D., Wan, C., and Xiao, Z. (2016). The reluctant analyst. *Journal of Accounting Research*, *54*(4), 987-1040.

Beyer, A., Cohen, D. A., Lys, T. Z., and Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2-3), 296-343.

Bidwell, C.M. III (1977). How good is institutional brokerage research? *The Journal of Portfolio Management*, Winter, 26-31.

Bjerring, J. H., Lakonishok, J., and Vermaelen, T. (1983). Stock prices and financial analysts' recommendations. *The Journal of Finance*, 38(1), 187-204.

Bonner, S. E., Hugon, A., and Walther, B. R. (2007). Investor reaction to celebrity analysts: The case of earnings forecast revisions. *Journal of Accounting Research*, *45*(3), 481-513.

Boulland, R., Ornthanalai, C., & Womack, K. L. (2017). Speed and expertise in stock picking: Older, slower, and wiser? Rotman School of Management Working Paper.

Bradley, D., Clarke, J., Lee, S., and Ornthanalai, C. (2014). Are analysts' recommendations informative? Intraday evidence on the impact of time stamp delays. *The Journal of Finance*, *69*(2), 645-673.

Bradshaw, M. T. (2004). How do analysts use their earnings forecasts in generating stock recommendations? *The Accounting Review*, 79(1), 25-50.

Brown, L., ed. (2000). *I/B/E/S Research Bibliography: Sixth Edition*. I/B/E/S International Inc., New York, NY.

Brown, N. C., Wei, K. D., and Wermers, R. (2014). Analyst recommendations, mutual fund herding, and overreaction in stock prices. *Management Science*, *60*(1), 1-20.

Chen, J., Jung, M. J., and Ronen, J. (2017). The confirmation effect of analyst recommendation reiterations. *Journal of Accounting, Auditing & Finance*, *32*(4), 576-592.

Copeland, T.E., and D. Mayers (1982). The Value Line Enigma (1965-1978): A case study of performance evaluation issues. *Journal of Financial Economics* 10, 289-321.

Cowles, A. III (1932). Can stock market forecasters forecast? *Econometrica*, 309-324.

Dechow, P. M., Ge, W., Larson, C. R., and Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, *28*(1), 17-82.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

De Franco, G., Hope, O. K., Vyas, D., and Zhou, Y. (2015). Analyst report readability. *Contemporary Accounting Research*, *32*(1), 76-104.

De Franco, G., and Hope, O. K. (2011). Do analysts' notes provide new information? *Journal of Accounting, Auditing & Finance*, *26*(2), 229-254.

Francis, J., and Soffer, L. (1997). The relative informativeness of analysts' stock recommendations and earnings forecast revisions. *Journal of Accounting Research*, *35*(2), 193-211.

Goldman Sachs Asset Management (2016). Text, tone and topic. https://www.gsam.com/content/gsam/global/en/market-insights/gsam-insights/gsam-perspectives/2016/big-data/natural-language-processing.html.

Green, T. C. (2006). The value of client access to analyst recommendations. *Journal of Financial and Quantitative Analysis*, *41*(1), 1-24.

Groth, J. C., Lewellen, W. G., Schlarbaum, G. G., and Lease, R. C. (1979). An analysis of brokerage house securities recommendations. *Financial Analysts Journal*, 35(1), 32-40.

Hanlon, M., and Hoopes, J. L. (2014). What do firms do when dividend tax rates change? An examination of alternative payout responses. *Journal of Financial Economics*, *114*(1), 105-124.

Huang, A. H., Wang, H., and Yang, Y. (2021). FinBERT—A Deep Learning Approach to Extracting Textual Information. *Available at SSRN: https://ssrn.com/abstract=3910214*.

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

Hutton, A., Marcus, A. and Tehranian, H. (2009). Opaque financial reports and the distribution of stock returns. *Journal of Financial Economics*, 94(1), pp.67-86.

Jacob, J., Rock, S., & Weber, D. P. (2008). Do non-investment bank analysts make better earnings forecasts? *Journal of Accounting, Auditing & Finance*, 23(1), 23-61.

Järvelin, K., and Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)*, *20*(4), 422-446.

Jegadeesh, N., and Kim, W. (2006). Value of analyst recommendations: International evidence. *Journal of Financial Markets*, 9(3), 274-309.

Jegadeesh, N., and Kim, W. (2010). Do analysts herd? An analysis of recommendations and market reactions. *The Review of Financial Studies*, *23*(2), 901-937.

Jegadeesh, N., Kim, J., Krische, S. D., and Lee, C. M. (2004). Analyzing the analysts: When do recommendations add value? *The Journal of Finance*, *59*(3), 1083-1124.

Kadan, O., Madureira, L., Wang, R., and Zach, T. (2008). Conflicts of interest and stock recommendations: The effects of the global settlement and related regulations. *The Review of Financial Studies*, *22*(10), 4189-4217.

Kecskés, A., Michaely, R., and Womack, K. L. (2017). Do earnings estimates add value to sell-side analysts' investment recommendations? *Management Science*, *63*(6), 1855-1871.

Kim, S. T., Lin, J. C., and Slovin, M. B. (1997). Market structure, informed trading, and analysts' recommendations. *Journal of Financial and Quantitative Analysis*, *32*(4), 507-524.

Kothari, S.P., Shu, S. and Wysocki, P.D., 2009. Do managers withhold bad news? *Journal of Accounting Research*, 47(1), 241-276.

Lawrence, A., Ryans, J. P., & Sun, E. Y. (2017). Investor demand for sell-side research. *The Accounting Review*, 92(2), 123-149.

Logue, D. E., and Tuttle, D. L. (1973). Brokerage house investment advice. *Financial Review*, 8(1), 38-54.

Loh, R. K., and Mian, G. M. (2006). Do accurate earnings forecasts facilitate superior investment recommendations? *Journal of Financial Economics*, 80(2), 455-483.

Loh, R. K., and Stulz, R. M. (2011). When are analyst recommendation changes influential? *The Review of Financial Studies*, *24*(2), 593-627.

Patell, J.M. and Wolfson, M.A. (1982). Good news, bad news, and the intraday timing of corporate disclosures. *The Accounting Review*, 509-527.

Perols, J. L., Bowen, R. M., Zimmermann, C., & Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review*, 92(2), 221-245.

Piorkowski, M. (2021). Banks' internal review policies and analysts' information production: Evidence from oversight committees. Working Paper.

Previts, G. J., Bricker, R. J., Robinson, T. R., and Young, S. J. (1994). A content analysis of sell-side financial analyst company reports. *Accounting Horizons*, *8*(2), 55.

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

Schipper, K. (1991). Analysts' forecasts. *Accounting Horizons*, *5*(4), 105.

Shepard, L. (1977). How good is investment advice for individuals? *The Journal of Portfolio Management*, Winter, 32-36.

Skinner, D. J. (1994). Why firms voluntarily disclose bad news. *Journal of Accounting Research*, 32(1), 38-60.

Skinner, D. J., and Sloan, R. G. (2002). Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies*, *7*(2), 289-312.

Stickel, S. E. (1995). The anatomy of the performance of buy and sell recommendations. *Financial Analysts Journal*, *51*(5), 25-39.

Twedt, B., and Rees, L. (2012). Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. *Journal of Accounting and Public Policy*, *31*(1), 1-21.

Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics*, *58*(3), 369-396.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance*, *51*(1), 137-167.

Xiao, K., and Zang, A. (2017). Hardening soft Information: Analyst conservative bias. *Available at SSRN 3081777*.

Yang, Y., Uy, M. C. S., and Huang, A. (2020). Finbert: A pretrained language model for financial communications. *arXiv preprint arXiv:2006.08097*.

Yost, B. (2021). Does increased visibility to the IRS cause public firms to go private? Evidence from FIN 48. Working paper.

Appendix A: Variable Definitions

Variable	Definition and Data Source
ΔRec	The analyst's recommendation at t minus the recommendation at t-1 as reported in I/B/E/S recommendation detail file prior to the significant changes made in 2018.
Upgrade	An indicator variable equal to 1 for observations with a positive *ΔRec*.
Downgrade	An indicator variable equal to 1 for observations with a negative *ΔRec*.
Reiteration	An indicator variable equal to 1 for observations with a *ΔRec* equal to 0.
Tone	Analyst tone is equal to the number of positive sentences minus the number of negative sentences divided by the total number of sentences written on a given day. Sentences are classified using finBERT. See Huang et al. (2020) and their GitHub repository (https://github.com/yya518/FinBERT) for their model and description. $Tone = \frac{\#\ Pos\ Sentences - \#\ Neg\ Sentences}{\#\ Sentences}$.
ΔTone	The difference in the tone of the prior report ($Tone_{ijt-1}$) and the analysts' baseline tone ($Tone_{ijt-T}$ where $t-T$ is the date of the prior recommendation change). The analysts' baseline tone is the tone of the first available report since her recommendation change. If we have the report that includes the prior recommendation change, the recommendation change report is the baseline tone. If we do not have the recommendation change report, we use the tone from the first available report following the recommendation change as the baseline tone. The baseline tone may be the initiating coverage report tone. $\Delta Tone_{ijt-1} = Tone_{ijt-1} - Tone_{ijt-T}$.
Rec	The level of the analysts' stock recommendation as of their prior report (1=Strong Sell, 2=Sell, 3=Hold, 4=Buy, 5=Strong Buy). We flip the order of the recommendation from I/B/E/S.
(TP – P) / P	Defined as the percentage difference between the most recent target price forecast prior to the revision and the stock price 5 days prior to the recommendation change.
ΔTP	Measured in the month leading up to the prior report as the latest change in target price, if any, scaled by price 50 days. If there are no target price changes ΔTP is equal to 0.
ΔEF	Measured in the month leading up to the prior report as the latest change in the 1-year ahead earnings forecast, if any, scaled by price 50 days. If there are no earnings forecast changes ΔEF is equal to 0.

Six-Month BHAR	The buy and hold return calculated using the daily stock return from the CRSP daily file over the 180 days ending 5 days prior to the current report date, adjusted for the value-weighted market return. $e^{\sum(\log(1+ret))} - e^{\sum(\log(1+vwretd))}$.
Six-Month Volatility	The standard deviation of the daily stock return from the CRSP daily file over the 180 days ending 5 days prior to the current report date.
Days Since Rec Change	The number of days since the prior analyst recommendation revision as of the prior report date.
Ln(Days Since Rec Change)	The natural log of the number of days since the prior analyst recommendation revision as of the prior report date.
ROA	Equal to earnings divided by assets as reported in the most recent earnings announcement preceding the prior report. Compustat variables IBQ/ATQ.
MVE	The firm's market capitalization as of the quarter end as reported in the most recent earnings announcement preceding the prior report. Compustat variables $CSHOQ * PRCCQ$.
Ln(MVE)	The natural log of MVE.
N Sentences	The number of sentences in the analysts' report.
BTM	The firm's book value scaled by its market value as reported in the most recent earnings announcement preceding the prior report. Compustat variables $\frac{CEQQ}{CSHOQ*PRCCQ}$.
Leverage	The firm's total liabilities scaled by its total assets as reported in the most recent earnings announcement preceding the prior report. Compustat variables $\frac{LTQ}{ATQ}$.
Sales Growth	The firm's year over year sales growth as reported in the most recent earnings announcement preceding the prior report. Compustat variables $\frac{SALEQ_t}{SALEQ_{t-4}} - 1$.
Institutional Ownership	The fraction of shares owned by institutional shareholders as reported in the quarter preceding the prior report.
CAR [0,1]	The two-day cumulative abnormal return around the recommendation change adjusted by the value-weighted return in CRSP. Calculated as $\sum_0^1(ret_t - vwretd_t)$.
Influential	$Influential = 1$ if $\lvert CAR[0,1]\rvert > 1.96 * \sigma(residual) * \sqrt{2}$ and *CAR* and *ΔRec* are in the same direction, following Loh and Stulz (2011). The residual comes from the following regression: $ret_t = \alpha + vwretd_t + \epsilon_t$.
Upgrade Predictability	Predicted value from the regression results in Table 4, Column 2.
Downgrade Predictability	Predicted value from the regression results in Table 4, Column 3.

Influential Before (Any Firm)	An indicator variable equal to 1 for recommendation changes by analysts who have previously issued an influential recommendation change for any firm and 0 otherwise.
Influential Before (Same Firm)	An indicator variable equal to 1 for recommendation changes by analysts who have previously issued an influential recommendation change for this same firm and 0 otherwise.
$\lvert \Delta Rec \rvert$	The absolute value of the change in recommendation.
Away From Consensus	An indicator variable equal to 1 if the analyst's recommendation change is further from the mean consensus recommendation than their prior recommendation and 0 otherwise.
All-Star Analyst	An indicator variable equal to 1 for analysts who receive all-star status as indicated in Institutional Investor.
General Experience	The number of years the analyst has issued EPS forecasts from the I/B/E/S EPS US detail file for any firm.
Firm Experience	The number of years the analyst has issued EPS forecasts for the firm from the I/B/E/S EPS US detail file.
Avg. Ln(Turnover)	The mean of the natural log of the daily turnover over the prior three months.

Appendix B: Prediction Performance Evaluation Metrics

See Bao et al. (2020) Section 4 for more details. (1) AUC is the area under the receiver operating characteristics (ROC) curve. The ROC curve is a visual representation of the true positive rate for given false positive rates. AUC is the probability that a randomly chosen upgrade (downgrade) will be ranked higher than a randomly chosen reiteration or downgrade (upgrade). In evaluating the performance using the AUC, a random guess would produce 0.50. Thus, any sensible model would have an AUC greater than 0.50. (2) NDCG@k or the Normalized Discounted Cumulative Gain at k is a metric developed to evaluate ranking and recommendation tasks as in Järvelin and Kekäläinen (2002). Many search engines use NDCG@k to evaluate the performance of the returned results as it is important for the most relevant results to be listed at the top. k is the top 1% of observations based on the predicted value of the upgrade or downgrade. $DCG@k = \sum_1^k (2^{rel_i} - 1) / log_2(i+1)$. $rel_i = 1$ if the ith observation is an upgrade (downgrade) and 0 otherwise. The actual upgrades (downgrades) that have higher predictability scores are discounted less than those with lower predictability scores. We then scale $DCG@k$ by the $ideal\ DCG@k$, which is equal to $\sum_1^k (2^1 - 1) / log_2(i+1)$. This calculation represents the value if all the actual upgrades (downgrades) are ranked in the top 1%. $NDCG@k = DCG@k / ideal\ DCG@k$. This metric is higher as more actual upgrades (downgrades) are ranked above non-upgrades (downgrades). If all actual upgrades are at the top of the distribution, the metric will equal 1. (3) $Sensitivity = TP/(TP + FN)$. TP is the true positive rate or the number of upgrades (downgrades) that we correctly classify as an upgrade (downgrade). FN is the false negative rate, or the number of actual upgrades (downgrades) not predicted to be an upgrade (downgrade). $TP + FN$ is the total number of actual upgrades (downgrades). (4) $Precision = TP/(TP + FP)$. TP is as defined previously in equation (3). FP is the number of observations that were incorrectly predicted to be upgrades (downgrades). $TP + FP$ is equal to the total number of observations predicted to be an upgrade (downgrade). Frequency is the frequency of actual upgrades (downgrades) in the sample period from 2015-2017. Upgrades (Downgrades) comprise 1.6 (1.8)% in the test period from 2015-2017, while 3.3% (1.9%) of the subsample of the reports with an outstanding non-buy (non-sell) recommendations contain an upgrade (downgrade).

Table 1: Analyst Report Sample Composition

Description	Dropped Observations	Sample Size
Number of total analyst reports		693,333
Keep reports from Top 19 brokers	(271,333)	422,000
Drop reports with fewer than 5 sentences	(2,100)	419,900
Drop reports with missing I/B/E/S ticker match	(34,447)	385,453
Keep one report per analyst-firm day (aggregate tone change at daily level)	(40,770)	344,683
Drop reports that do not merge with broker-analyst-firm per I/B/E/S	(76,542)	268,141
Merge in recommendation initiations and changes for which we do not have accompanying reports	11,809	279,950
Drop observations with missing Compustat/CRSP identifying information	(51)	279,899
Drop initiations of coverage	(9,288)	270,611
Keep observations from 2004-2017	(13,125)	257,486
Drop observations with missing lagged $\Delta Tone$	(28,168)	229,318
Drop observations with missing control variables	(38,301)	191,017
Final sample for regression analyses		**191,017**
Recommendation revisions only	(183,674)	**7,343**
Recommendation revisions identified as influential	(1,794)	**5,549**
Number of firms		470
Number of analysts		1,436
Number of analyst-firms		6,771

Table 2: Descriptive Statistics

Panel A displays the average number of sentences in analyst reports and the average *Tone* and Δ*Tone* split by downgrades, reiterations, and upgrades. Panel B displays descriptive statistics for *Tone* by the level of the analysts' recommendation.

Panel A: Tone and Sentence Length by Recommendation Change

	Downgrade Mean	*Reiteration* Mean	*Upgrade* Mean
ΔTone	-0.046	-0.005	0.051
Tone	-0.030	0.108	0.174
N Sentences	99.081	79.411	121.254
Observations	3,857	183,674	3,486

Panel B: *Tone* by Recommendation Level

	N	Mean	Std. Dev.	Min	P25	P50	P75	Max
1	662	-0.06	0.16	-0.60	-0.16	-0.06	0.04	0.54
2	9,528	-0.00	0.15	-0.63	-0.09	0.00	0.08	0.55
3	76,943	0.06	0.16	-0.84	-0.03	0.06	0.16	0.86
4	87,856	0.15	0.16	-0.67	0.05	0.14	0.25	0.80
5	16,028	0.17	0.16	-0.57	0.07	0.16	0.27	0.81
Total	191,017	0.11	0.17	-0.84	0.00	0.10	0.21	0.86

Panel C: *ΔRec*

		New *Rec*					
		1	2	3	4	5	Total
Prior *Rec*	1	0	0	70	0	3	73
	2	3	0	654	57	0	714
	3	68	698	0	2,202	394	3,362
	4	0	41	2,474	0	106	2,621
	5	2	1	500	70	0	573
	Total	73	740	3,698	2,329	503	7,343

Figure 1. Tone and Sentence Length by Year

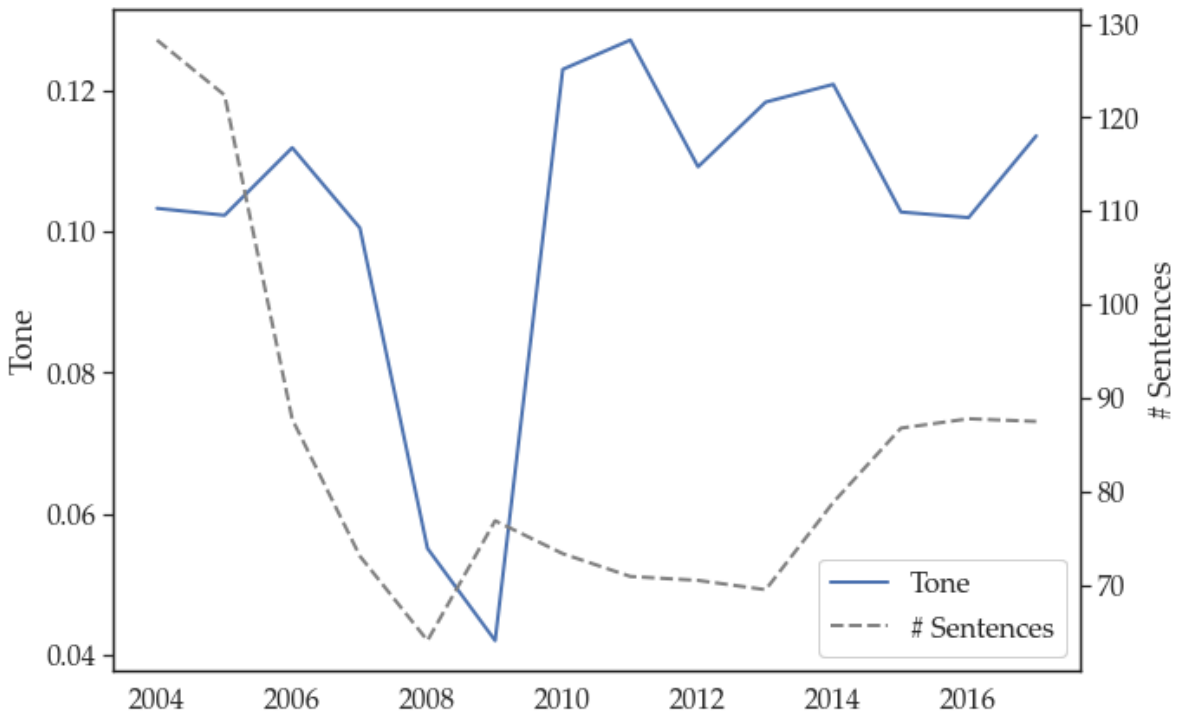

Table 3: Descriptive Statistics

Note: Panel A displays basic descriptive statistics for the set of variables used in our main analysis. Panel B displays basic descriptive statistics for the regressions of *CAR* and *Influential* on *Upgrade* (*Downgrade*) Predictability. We follow Loh and Stulz (2011) and classify recommendation changes as influential if the magnitude of the two-day *CAR* is significant ($|CAR[0,1]| > \sqrt{2} * 1.96 * \sigma$) and in the same direction as the recommendation change. Where σ is the standard deviation of the daily returns over the prior three months. See the appendix for our complete set of variable definitions.

Panel A: Main Sample

	N	Mean	Std. Dev	Min	25th Pctl	50th Pctl	75th Pctl	Max
ΔRec	191,017	-0.002	0.238	-4.000	0.000	0.000	0.000	4.000
Upgrade	191,017	0.018	0.134	0.000	0.000	0.000	0.000	1.000
Downgrade	191,017	0.020	0.141	0.000	0.000	0.000	0.000	1.000
Tone	191,017	0.107	0.166	-0.840	0.000	0.100	0.213	0.857
N Sentences	191,017	79.387	77.163	5	33	55	101	3,813
ΔTone	191,017	-0.004	0.191	-0.495	-0.123	-0.006	0.114	0.485
Rec	191,017	3.571	0.730	1.000	3.000	4.000	4.000	5.000
(TP – P) / P	191,017	0.207	0.889	-0.729	-0.008	0.099	0.211	7.650
ΔTP	191,017	0.009	0.096	-0.433	0.000	0.000	0.025	0.380
ΔEF	191,017	-0.001	0.008	-0.052	-0.000	0.000	0.000	0.024
Six-Month BHAR	191,017	0.011	0.182	-0.495	-0.095	0.006	0.107	0.670
Six-Month Volatility	191,017	0.019	0.011	0.007	0.012	0.016	0.022	0.072
Reports Since Rec Change	191,017	18.956	19.452	2	6	12	25	188
Days Since Rec Change	191,017	698.907	673.831	1	206	476	968	3,210
ROA	191,017	0.015	0.019	-0.070	0.005	0.013	0.024	0.073
Ln(MVE)	191,017	9.639	1.198	6.727	8.844	9.575	10.367	12.464
MVE	191,017	32,093	48,352	834	6,929	14,405	31,779	258,782
BTM	191,017	0.447	0.360	-0.209	0.213	0.361	0.593	2.070
Leverage	191,017	0.628	0.205	0.139	0.482	0.621	0.777	1.175
Sales Growth	191,017	0.051	0.195	-0.511	-0.032	0.040	0.115	0.934
Institutional Ownership	191,017	0.730	0.261	0.000	0.678	0.799	0.887	1.107

Panel B: Descriptive Statistics for the Recommendation Changes used in the CAR and Influential Regressions

	N	Mean	Std. Dev	Min	25th Pctl	50th Pctl	75th Pctl	Max
CAR [0,1]	5,549	0.000	0.033	-0.106	-0.017	0.000	0.018	0.105
Influential	5,549	0.158	0.364	0.000	0.000	0.000	0.000	1.000
Upgrade Predictability	5,549	0.028	0.028	-0.033	0.006	0.033	0.046	0.087
Downgrade Predictability	5,549	0.022	0.019	-0.014	0.009	0.020	0.034	0.068
ΔTone	5,549	0.007	0.194	-0.495	-0.113	0.004	0.127	0.485
(TP – P) / P	5,549	0.190	0.899	-0.729	-0.038	0.070	0.191	7.650
ΔTP	5,549	0.003	0.105	-0.433	0.000	0.000	0.027	0.380
ΔEF	5,549	-0.001	0.009	-0.052	-0.001	0.000	0.001	0.024
Six-Month BHAR	5,549	0.010	0.206	-0.496	-0.113	-0.002	0.119	0.670
Six-Month Volatility	5,549	0.021	0.013	0.007	0.013	0.017	0.025	0.072
Reports Since Rec Change	5,549	12.902	13.312	2	5	9	16	170
Days Since Rec Change	5,549	552.194	541.265	1	187	373	731	3,210
ROA	5,549	0.013	0.020	-0.070	0.003	0.011	0.022	0.073
Ln(MVE)	5,549	9.368	1.110	6.727	8.637	9.372	10.018	12.464
BTM	5,549	0.506	0.387	-0.209	0.246	0.417	0.669	2.070
Leverage	5,549	0.629	0.202	0.139	0.485	0.622	0.773	1.175
Sales Growth	5,549	0.042	0.215	-0.511	-0.046	0.035	0.114	0.934
Institutional Ownership	5,549	0.732	0.260	0.000	0.677	0.800	0.892	1.107
|ΔRec|	5,549	1.153	0.363	1.000	1.000	1.000	1.000	4.000
Away From Consensus	5,549	0.545	0.498	0.000	0.000	1.000	1.000	1.000
Avg. Ln(Turnover)	5,549	-4.831	0.656	-6.242	-5.276	-4.890	-4.424	-2.988

Table 4: Predicting Recommendation Changes

Note: This table displays the results from regressions of *ΔRec*, *Upgrade*, or *Downgrade* on *ΔTone* and several control variables with analyst, firm, and year fixed effects. Standard errors are calculated using firm and analyst clusters. Significance levels are represented by * p<0.10, ** p<0.05, *** p<0.01. t-statistics are reported in parentheses.

	ΔRec	*Upgrade*	*Downgrade*
ΔTone	0.023***	0.025***	-0.012***
	(6.464)	(5.571)	(-5.516)
Rec	-0.080***	-0.060***	0.027***
	(-28.468)	(-15.035)	(21.240)
(TP – P) / P	0.008***	0.008***	-0.003**
	(3.651)	(3.873)	(-2.317)
ΔTP	0.074***	0.033***	-0.047***
	(8.621)	(3.988)	(-9.261)
ΔEF	-0.035	0.017	0.067
	(-0.377)	(0.180)	(1.170)
Six-Month BHAR	0.001	0.005	0.002
	(0.247)	(0.909)	(0.623)
Six-Month Volatility	-0.352***	-0.019	0.341***
	(-2.829)	(-0.144)	(3.897)
Ln(Days Since Rec Change)	-0.000	0.008***	0.003***
	(-0.710)	(11.002)	(8.716)
ROA	-0.080	-0.063	0.029
	(-1.531)	(-1.088)	(0.890)
Ln(MVE)	0.007**	0.003	-0.005**
	(2.369)	(0.709)	(-2.566)
BTM	-0.021***	-0.013**	0.012***
	(-3.823)	(-2.294)	(3.455)
Leverage	0.008	0.013	-0.004
	(0.637)	(0.809)	(-0.547)
Sales Growth	0.004	0.003	-0.003
	(0.802)	(0.517)	(-1.222)
Institutional Ownership	-0.003	-0.001	0.001
	(-0.914)	(-0.247)	(0.418)
Year FE	Yes	Yes	Yes
Analyst FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes
Observations	190,959	87,068	180,770
Adjusted R^2	0.041	0.034	0.027

Table 5: Upgrade and Downgrade Predictability and Attenuation of Market Reaction

Note: Table 5 presents the results of regressing CAR [0,1] around the recommendation change on the predictability of the recommendation change and a set of control variables. Columns 1 and 2 include the sample of positive recommendation changes. Columns 3 and 4 present the results using the sample of negative recommendation changes. Columns 1 and 3 present the baseline regression with analyst, firm, and year fixed effects with no other control variables while Columns 2 and 4 present the results with an array of controls in addition to the analyst, firm, and year fixed effects. Standard errors are calculated using analyst clusters. Significance levels are represented by * p<0.10, ** p<0.05, *** p<0.01. t-statistics are reported in parentheses.

	Upgrades		Downgrades	
	CAR [0,1]	*CAR [0,1]*	*CAR [0,1]*	*CAR [0,1]*
Upgrade Predictability	-0.139**	-0.123*	-	-
	(-2.534)	(-1.748)		
Downgrade Predictability	-	-	0.027	0.049
			(0.345)	(0.485)
ΔTone		0.002		0.000
		(0.582)		(0.072)
(TP – P) / P		-0.001		0.002
		(-0.622)		(0.629)
ΔTP		-0.007		0.018*
		(-0.829)		(1.671)
ΔEF		0.122		-0.040
		(1.259)		(-0.295)
Six-Month BHAR		-0.007		-0.009
		(-1.391)		(-1.600)
Six-Month Volatility		0.296**		0.047
		(2.041)		(0.295)
Ln(Days Since Rec Change)		0.000		0.000
		(0.449)		(0.191)
ROA		-0.081		0.020
		(-1.179)		(0.298)
Ln(MVE)		-0.004		0.004
		(-1.365)		(1.131)
BTM		0.003		0.000
		(0.696)		(0.006)
Leverage		-0.011		0.009
		(-1.024)		(0.800)
Sales Growth		-0.001		0.000
		(-0.184)		(0.035)
Institutional Ownership		-0.002		-0.005
		(-0.611)		(-0.891)
|ΔRec|		0.008**		0.002
		(2.573)		(0.441)
Away From Consensus		0.003*		-0.003
		(1.660)		(-1.546)

Avg. Ln(Turnover)		-0.007**		-0.001
		(-2.368)		(-0.245)
Year FE	Yes	Yes	Yes	Yes
Analyst FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	2,434	2,434	2,505	2,505
Adjusted R^2	0.131	0.145	0.062	0.065

Table 6: Influential Changes in Recommendations and Recommendation Change Predictability

Note: Table 6 presents the results of regressing *Influential* on the predictability of the recommendation change and a set of control variables. Columns 1 and 2 include the sample of positive recommendation changes. Columns 3 and 4 present the results using the sample of negative recommendation changes. Columns 1 and 3 present the baseline regression with analyst, firm, and year fixed effects and no other control variables while Columns 2 and 4 present the results with an array of controls in addition to the analyst, firm, and year fixed effects. Standard errors are calculated using analyst clusters. Significance levels are represented by * p<0.10, ** p<0.05, *** p<0.01. t-statistics are reported in parentheses.

	Upgrades		Downgrades	
	Influential	*Influential*	*Influential*	*Influential*
Upgrade Predictability	-0.989**	-1.657**	-	-
	(-1.975)	(-2.202)		
Downgrade Predictability	-	-	-0.862	-0.825
			(-1.166)	(-0.775)
ΔTone		0.034		-0.016
		(0.616)		(-0.299)
(TP – P) / P		-0.034*		-0.008
		(-1.654)		(-0.396)
ΔTP		-0.067		-0.041
		(-0.702)		(-0.386)
ΔEF		1.634*		1.041
		(1.691)		(1.044)
Six-Month BHAR		-0.056		0.024
		(-1.205)		(0.416)
Six-Month Volatility		2.825*		-1.968
		(1.827)		(-1.356)
Ln(Days Since Rec Change)		-0.013		-0.007
		(-1.410)		(-0.703)
ROA		-1.316*		-0.520
		(-1.886)		(-0.916)
Ln(MVE)		-0.076**		-0.047
		(-2.277)		(-1.360)
BTM		-0.070		-0.004
		(-1.292)		(-0.084)
Leverage		-0.291*		-0.160
		(-1.958)		(-1.043)
Sales Growth		0.022		0.031
		(0.488)		(0.674)
Institutional Ownership		-0.002		0.027
		(-0.042)		(0.544)
|ΔRec|		0.006		-0.043
		(0.150)		(-1.045)
Away From Consensus		-0.008		0.003

		(-0.302)		(0.157)
Avg. Ln(Turnover)		-0.207***		-0.055
		(-5.314)		(-1.504)
Year FE	Yes	Yes	Yes	Yes
Analyst FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Observations	2,434	2,434	2,505	2,505
Adjusted R^2	0.053	0.069	0.099	0.099

Table 7: Prediction Performance Evaluation

Note: This table displays four evaluation metrics for the various models used to predict recommendation upgrades and downgrades. These four metrics follow the fraud prediction literature for benchmarking purposes. We follow a similar approach to that in Bao et al. (2020) in training and testing our models. In the model evaluation, we use their same metrics and below have a similar explanation of the four metrics. Bao et al. (2020) provides a more in-depth discussion of these metrics. The models were trained using the sample period from 2004-2014 and tested in the sample period 2015-2017.

(1) AUC is the area under the receiver operating characteristics (ROC) curve. The ROC curve is a visual representation of the true positive rate for a given false positive rate. A random guess would have an AUC of 0.50.

(2) NDCG@k or the Normalized Discounted Cumulative Gain at k. k is the top 3% or 1.5% of observations based on the predicted value of the upgrade or downgrade, respectively. $DCG@k = \sum_1^k (2^{rel_i} - 1)/\log_2(i+1)$. We then scale $DCG@k$ by the $idealDCG@k = \sum_1^k (2^1 - 1)/\log_2(i+1)$, this is the value if all the actual upgrades (downgrades) are ranked in the top 1%. $NDCG@k = DCG@k/idealDCG@k$. This metric is higher as more actual upgrades (downgrades) are ranked above non-upgrades (downgrades). If all actual upgrades are at the top, the metric will equal 1.

(3) $Sensitivity = TP/(TP + FN)$. TP is the true positive rate or the number of upgrades (downgrades) that we correctly classify as an upgrade (downgrade). FN is the false negative rate, or the number actual upgrades (downgrades) not predicted to be an upgrade (downgrade). $TP + FN$ is the total number of actual upgrades (downgrades).

(4) $Precision = TP/(TP + FP)$. TP is as defined previously in equation (3). FP is the number of observations predicted to be upgrades (downgrades) but were not. $TP + FP$ is equal to the total number of observations predicted to be an upgrade (downgrade).

Frequency is the frequency of actual upgrades (downgrades) in the sample period from 2015-2017. For (3) sensitivity and (4) precision, we predict observations in the top 3% (2%) for *Upgrade (Downgrade) Predictability* to be upgrades (downgrades).

Model	AUC	NDCG@k	Sensitivity	Precision	Actual Frequency
Upgrade (Logit)	0.591	0.084	0.069	0.077	0.033
Upgrade (OLS)	0.590	0.085	0.070	0.078	0.033
Upgrade (Analyst and Firm FEs)	0.495	0.030	0.028	0.031	0.033
Upgrade (Analyst and Firm and Year FEs)	0.495	0.031	0.029	0.032	0.033
Downgrade (Logit)	0.685	0.066	0.061	0.059	0.019
Downgrade (OLS)	0.685	0.065	0.061	0.059	0.019
Downgrade (Analyst and Firm FEs)	0.496	0.019	0.019	0.018	0.019
Downgrade (Analyst and Firm and Year FEs)	0.496	0.019	0.019	0.018	0.019