Do Sell-side Analyst Reports Have Investment Value?

Linying Lv*

First draft: January 2025. This draft: March 2025.

Abstract

This paper documents new investment value in analyst reports. Analyst narratives embedded with large language models strongly forecast future stock returns, generating significant alpha beyond established analyst-based and fundamental-based factors. The return predictability arises primarily from reports that convey negative sentiment but forecast favorable long-term prospects, suggesting systematic market overreaction to near-term negative news. The effect is more pronounced for large, mature firms and for reports authored by skilled, experienced analysts. A Shapley value decomposition reveals that analysts' strategic outlook contributes the most to portfolio performance, especially forward-looking discussions on fundamentals. Beyond demonstrating untapped value in qualitative information, this paper illustrates the broader potential of artificial intelligence to augment—rather than replace—expert human judgment in financial markets.

JEL Classification: G12, G14, G24

Keywords: Analyst Report, Investment Value, Large Language Models, Explainable Artificial Intelligence

*Olin Business School, Washington University in St. Louis, llyu@wustl.edu. I thank John Barrios, Justin Birru, Leland Bybee, Allaudeen Hameed, Songrun He, and David Hirshleifer for their helpful comments. Feedback and discussions are welcome!

1 Introduction

Can narratives in sell-side analyst reports predict stock returns? While analysts provide both quantitative forecasts (earnings estimates, target prices, and recommendations) and qualitative analyses in written reports, decades of research have focused primarily on the former. Studies consistently show that quantitative forecasts generate profitable trading signals (e.g., Womack, 1996; Brav and Lehavy, 2003; Jegadeesh et al., 2004; Birru et al., 2022; Farago et al., 2023). Yet the narrative components of analyst reports remain an underutilized information source that may contain significant investment value.

The main challenge in analyzing report content has been the unstructured and high-dimensional nature of text data. Traditional approaches often condense text data to a single sentiment measure, but this dimensional reduction raises two critical concerns. First, when sentiment measures diverge from numerical forecasts, it becomes difficult to distinguish whether these gaps represent measurement error or genuine analyst insight. Second, dictionary-based sentiment methods face fundamental challenges in context-dependent interpretation, as highlighted by Loughran and McDonald (2011). The recent emergence of large language models (LLMs) offers a promising solution to these challenges by enabling more nuanced and comprehensive analysis of textual information.

In this paper, I analyze 1.2 million analyst reports from Thomson Reuters Investext between 2000 and 2023 to address two fundamental questions: whether analyst reports contain incremental investment value and, if so, what drives this value. Using this comprehensive dataset, I propose an integrated framework that examines investment value through three lenses: information representation, information relevance, and information interpretation.

First, I map analyst reports into a high-dimensional vector space using LLM transformer embeddings. These embeddings capture the LLM's deep linguistic understanding and semantic interpretation of the text, representing a sophisticated "thinking process" that converts textual input into structured output.[1] Second, I use machine learning models to extract value-relevant information to predict future 12-month stock returns and construct trading strategies based on these predictions. Finally, I conduct attribution analysis using Shapley value decomposition to identify which report components contribute most significantly to portfolio performance.

I begin by examining the return predictability of analyst reports with Ridge regressions. Projecting future 12-month stock returns on embeddings generated by the LLaMA3-8B model reveals significant out-of-sample predictive power. This predictability remains

[1]This approach of text representation has gained traction in recent studies (Jha et al., 2020; Chen et al., 2022; Chen et al., 2023b).

robust when controlling for quantitative analyst forecasts (recommendation revisions, earnings forecast revisions, and target price revisions) and when using DGTW-adjusted returns. Cross-sectionally, the predictive power is strongest for large, mature firms and reports authored by skilled, experienced analysts. This pattern suggests that text embeddings capture unique, longer-horizon insights that even well-covered stocks fail to fully price, and that superior information production requires both analytical capability and deep institutional knowledge.

To assess economic significance, I construct a monthly rebalanced portfolio sorted on analyst report return predictions. A self-financing trading strategy that buys stocks with the most positive predictions and shorts those with the most negative predictions generates an average monthly return of 1.04% with 68 basis points alpha (t = 2.64) relative to the Fama and French (2015) five factors plus momentum. To further evaluate the incremental value, I compare the strategy's performance to 94 fundamental-based factors from Gu et al. (2020) and 18 analyst-based factors from Chen and Zimmermann (2022). The report-based strategy consistently generates significant information ratios ranging from 0.73 to 1.41, suggesting that analyst narratives contain substantial investment value beyond established predictive signals.

A natural question is whether analysts are simply piggybacking on known drift effects—such as the post-earnings-announcement-drift (PEAD) documented by Livnat and Mendenhall (2006)—and thus capturing documented price momentum rather than genuine new information. To investigate this, I examine how negative versus positive earnings shocks (SUE) interact with analysts' predicted returns. Strikingly, most return predictability arises from reports predicting positive future returns but issued after negative news, suggesting a pattern distinct from simple price drift hypotheses.

This result raises a key question about where the market inefficiency originates: Are analysts attempting to correct market overreaction to negative news, or are they themselves contributing to the overreaction? To differentiate, I examine the sentiment and recommendations of reports with the highest and lowest return predictions. The results reveal that reports predicting positive returns tend to accompany downgrades, negative text sentiment, and negative contemporaneous returns. This pattern complicates the source of market inefficiency. If analysts were simply correcting the market's overreaction, one would expect their recommendations to go against the prevailing negative sentiment. Instead, the reports show mixed signals: official downgrades coexist with bullish commentary on long-run prospects. The contradiction between these signals likely reduces the market's ability to fully extract the positive information. Consequently, analysts both fuel short-term negativity and provide valuable insight about future rebounds—a dual role that highlights the importance of

dissecting analysts' qualitative content rather than relying solely on quantitative recommendations.

Another interesting question is which specific topics within analyst reports primarily drive investment value. To address this, I rely on a Shapley value decomposition framework that attributes a report's predictive power to the constituted content. The basic idea is straightforward: after splitting the report into different topics, I test how trading strategies based on various combinations of these topics perform. By exhaustively examining all possible topic combinations, I obtain a fair estimate of how much each topic contributes to the final outcome.

The first step involves ensuring that each topic classification is both exclusive and meaningful, while avoiding arbitrary assumptions. I leverage the language capabilities of the ChatGPT-4o model to identify a hierarchy of topics, ultimately consolidating them into five meta-categories: Company and Industry Overview, Financial Analysis, Strategic Outlook, Risk and Governance, and Additional Content. I then fine-tune a BERT model to classify all 52 million sentences into these categories. Once the sentences are labeled, I construct topic-specific embeddings and generate return forecasts for each topic. By examining how portfolio performance changes when each topic is included or excluded—i.e., across all combinations—I derive Shapley values that quantify how much each category contributes to the Sharpe ratio and return.

I find that the Strategic Outlook discussions, despite comprising only 15% of report content, generate the largest contribution to investment value, accounting for 41% of the portfolio Sharpe ratio. By contrast, Financial Analysis—the largest category by distribution—contributes only about 16% of the Sharpe ratio, suggesting that standard financial data are quickly priced in, leaving less scope for future return prediction. Meanwhile, Company and Industry Overview and Risk and Governance show moderate contributions. Additional Content adds little. Univariate portfolio performance validates the Shapley value attribution analysis. A trading strategy based solely on strategic content generates a monthly long-short return of 1.41% with a Sharpe ratio of 0.93, significantly outperforming other content categories, which yield more modest returns between 0.52% and 0.58% monthly with Sharpe ratios of 0.37 to 0.41.

Strategic Outlook includes forward-looking elements such as growth potential, valuation, and investment opportunities. Consider this excerpt from a report with a positive future return prediction despite a downgrade: "Although the shares are an attractive long-term investment, we must caution that the severity of the current industry downturn increases the chance of further near-term earnings surprises." Linking the attribution result with the overreaction story, the investment opportunity appears to arise when analysts identify strong

long-term prospects that are temporarily obscured by near-term challenges. Supportively, when dividing the strategic discussion further into three dimensions of timeframe, sentiment, and focus, the Shapley value decomposition emphasizes the contribution of sentences related to positive long-term fundamentals.

A potential concern about return predictability is look-ahead bias with LLMs (Sarkar and Vafa, 2024). Since pre-trained language models could have learned future stock performance information from their training data, the embedding-generating process could introduce information leakage. However, I argue that the results are robust to this concern for three reasons. First, LLMs are fundamentally designed to optimize language understanding rather than financial prediction capabilities. In other words, the pre-training process focuses on capturing semantic meaning through next-token prediction, not predicting stock returns. Second, if LLMs were indeed learning and exploiting future market information, we would expect to see the strongest predictability with the Company and Industry Overview content, as it contains the most comprehensive stock-specific information. Instead, this category generates insignificant alpha. Third, I directly test for look-ahead bias by examining portfolio performance after each model's knowledge cutoff date using four different LLMs: BERT, RoBERTa, LLaMA2, and LLaMA3, which were trained on data ending between December 2018 and March 2023. The earliest models (BERT and RoBERTa) provide over five years of true out-of-sample testing. If the predictability stemmed from information leakage, performance should deteriorate in these post-knowledge cutoff periods. Instead, the strategies generate even superior performances, yielding monthly long-short returns between 1.76% and 5.28% with Sharpe ratios ranging from 0.74 to 2.37.

The paper contributes to the literature on the value of information from brokerage analysts. Previous papers generally focus on the cross-sectional return predictability of analyst quantitative outputs, such as stock recommendations (Green, 2006; Christophe et al., 2010; Chen et al., 2023a), earnings forecasts (Lys and Sohn, 1990; Bordalo et al., 2019), target price forecasts (Brav and Lehavy, 2003; Farago et al., 2023) and short-term trade ideas (Birru et al., 2022). I add to this strand of literature by documenting significant investment value in analyst report narratives that is distinct from and incremental to their quantitative forecasts. Furthermore, a contemporaneous complementary work by Lv (2024) employs similar embedding and Shapley value decomposition analytical framework to study analyst report information content and value for strategic investors. In contrast, this paper focuses on the investment value of report information in a longer horizon that is not immediately incorporated in stock prices. These papers complement each other in providing a comprehensive understanding of the dynamics of how analyst research content is processed by markets and creates value for investors.

This paper also contributes to the extensive overreaction literature by uncovering a novel pattern in analyst texts: negative sentiment paired with an overall positive long-run outlook.[2] A related paper by Bordalo et al. (2019) examines long-term growth expectations (LTG) and finds that analysts overreact to short-term news, with return predictability arising as prices gradually correct these expectations. A major distinction between our findings lies in the asymmetry of overreaction. While the LTG factor is mainly driven by analysts being over-optimistic and overvaluing stocks in high-LTG portfolios, my findings reveal that analysts often exhibit cautionary hedge through downgrades or negative tones, while simultaneously embedding optimistic forecasts about longer-term fundamentals in their written discussions. The mismatch between near-term pessimism and constructive long-term insights provides a textual signature of market overreaction, where analysts behave cautiously with conservative narratives and forecasts despite being aware of positive fundamentals. By detecting these subtle, dual-layer signals in analyst reports, this work extends prior studies of De Bondt and Thaler (1990) on analyst overreaction and clarifies how overreaction dynamics are expedited with analyst written reports.

Third, this study contributes to the expanding literature on text factors, which has predominantly extracted sentiment or topics from corporate disclosures, earnings-call transcripts, and news media using dictionary-based or topic-modeling approaches (e.g., Tetlock, 2007; Chen et al., 2022; Meursault et al., 2023; Bybee et al., 2023; Hirshleifer et al., 2023). By contrast, I deploy state-of-the-art LLM embeddings to the nuanced, forward-looking narratives in sell-side analyst reports. These embeddings not only capture richer semantic and contextual relations but also have an additive property. The latter allows me to isolate specific content categories, such as financial analysis, risk factors, and strategic outlook, and then quantify each category's standalone contribution to investment value through Shapley value decomposition. This approach offers a general framework for interpreting granular insights and attributing text predictive power to interested components.

The rest of the paper is organized as follows. Section 2 describes the data and methodology, including the prediction models, topic classification approach, and Shapley value decomposition framework. Section 3 presents the empirical results. Section 4 concludes.

[2]For example: De Long et al. (1990); Porta et al. (1997); Hong and Stein (1999); Gennaioli et al. (2016); Weber (2018).

2 Data and Methodology

2.1 Data

I combine analyst data from two primary sources. The first source is Thomson One's Investext database, where I download 1,194,330 analyst reports covering S&P 1500 constituent firms from 2000 to 2023. Table 1 summarizes the annual statistics of analyst report characteristics, brokerage coverage, and unique analysts. The second source is the I/B/E/S database, which contains analyst qualitative forecasts including announcement and revision data. To merge these datasets, I follow a two-step process. First, I match lead analysts in Investext reports to I/B/E/S Analyst IDs (AMASKCD) following Li et al. (2024). Second, I link individual reports to their corresponding EPS forecasts, price targets, and recommendations in I/B/E/S using a five-day window around the announcement date ([-2, +2] days) following Huang et al. (2014).

Following Gu et al. (2020), I incorporate an array of 94 firm-level predictive characteristics. This dataset captures the market activity of all firms listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ from January 1957 through December 2023. The monthly individual stock returns are sourced from the Center for Research in Security Prices (CRSP) database. Table A1 describes the details of anomaly variables.

To identify the incremental signal relative to quantitative analyst information, I compile 18 factors from Chen and Zimmermann (2022) with significant investment values in analyst-produced information documented in the literature. These factors use analysts' forecasts, recommendations, or coverage changes to derive signals that predict stock performance. Table A2 provides a detailed description of the factors.

2.2 Methodology

2.2.1 Prediction Models

In transforming the report information into structured representations, I input each report text into the LLaMA3-8B model and take the average of 32 transformer layer embeddings as text representation.

To identify value-relevant information, I employ Ridge regression models to predict the return of stocks in the next 12 months following the report release using report embeddings. I deliberately employ linear Ridge regression models to emphasize the inherent information value of analyst reports rather than giving credit to sophisticated machine learning models. The Ridge regression models with an expanding training window are given by:

$$Ret_{it,12m} = \beta_0 + \beta' y_{ijt}^{\text{AI}} + \epsilon_{ijt}, \tag{1}$$

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ \left\| Ret_{it,12m} - \beta_0 - y_{ijt}^{\text{AI}} \beta \right\|_2^2 + \theta \|\beta\|_2^2 \right\}, \tag{2}$$

where $Ret_{it,12m}$ is the next 12 months return of stick i following reports' release day t, and y_{ijt}^{AI} is the structured representation (4,096-dimension embeddings generated with the LLaMA3-8B model) of content from analyst report j on stock i released at day t. I use the first five-year data as the initial training sample. The out-of-sample test starts in 2005 and ends in 2023.

2.2.2 Topic Modeling

To identify information sources, I first conduct a topic modeling exercise to group the sentences in analyst reports into clusters based on their content. Ideally, I would like to group sentences into exclusive topic categories that capture the full spectrum of information in analyst reports while avoiding subjective classification biases. I leverage the natural language understanding capabilities of state-of-the-art LLMs to extract exclusive and meaningful topics.

The hierarchical clustering algorithm begins with the following prompt applied to a random sample of 100 analyst reports:

> **Prompt 1:** Please read the provided text file of sell-side analyst reports carefully. What are high-level mutually exclusive topics covered in these reports? Make sure that each sentence from the text file can be assigned to one of the topics. Here is the report content: {text}.

This prompting approach identifies 16 distinct topics: Executive Summary, Company Overview, Industry Analysis, Competitive Landscape, Income Statement Analysis, Balance Sheet Analysis, Cash Flow Analysis, Financial Ratios, Business Segments, Growth Strategies, Risk Factors, Management and Governance, ESG Factors, Valuation, Investment Thesis, and Appendices and Disclosures. To ensure complete coverage, I add a residual category "None of Above," bringing the total to 17 topics.

For analytical tractability, I consolidate these topics into five broad categories that represent the main dimensions of analyst research: 1) company and industry overview; 2) financial analysis; 3) strategic outlook; 4) risk and governance; 5) additional content. Table A3 provides detailed descriptions of these categories and their constituent topics. Figure A1 presents word clouds of the categories.

I classify all 52,350,385 sentences into specific topics using a fine-tuned BERT model.

I choose BERT for this classification task because its encoder architecture is specifically suitable for sequence classification compared to generative models like GPT. To train the classifier, I fine-tune BERT using a manually labeled dataset of 17,028 sentences extracted from 100 randomly selected analyst reports. The model assigns each sentence to one of the 17 identified topics, which are then mapped to the five meta-categories.

2.2.3 Shapley Value Decomposition

A follow-up question is where the investment value of analyst reports comes from. I attribute the investment value of the full report to five categories defined in Section 2.2.2 using the Shapley Additive exPlanations (SHAP) methodology proposed by Lv (2024).

SHAP, originally developed by Shapley (1953) in cooperative game theory, provides a framework for fairly allocating contributions among multiple players. The method satisfies key properties of efficiency and additivity, ensuring that the sum of all contributions equals the total value. By treating each word as a player, I can attribute the investment value contribution of individual words and their corresponding topics.

I implement this framework by mapping tokens to players, topics to teams, and monthly investment performance to individual games. The Shapley value decomposition process aggregates tokens within their respective topics, evaluates each topic's performance across all possible coalitions in each monthly period, and computes the final contribution as the average marginal contribution of each topic across all periods. Below, I present the mathematical formulations.

I first aggregate token embeddings into topic-level representations. LLMs generate token-level embeddings, which I combine into a report-level representation through averaging:

$$y^{emb} = \frac{1}{N} \sum_{i=1}^{N} e_i, \tag{3}$$

where e_i is the average embedding of token i across all transformer layers, N is the total number of words in the report. Since each token is assigned to a specific topic, I can decompose the report embedding into topic-level representations with the average embeddings of all tokens that fall into the topic:

$$y^{emb} = \sum_{p=1}^{P} y_p^{emb} = \frac{1}{N} \sum_{p=1}^{P} \sum_{i_p=1}^{N_p} e_{i_p}, \tag{4}$$

where P is the number of topics, y_p^{emb} is the embedding for topic p, N_p is the number of tokens belonging to topic p, and i_p is the index of tokens within topic p.

With topic representations, I can generate Ridge predictions for all possible topic combinations and evaluate their investment performance. I calculate investment value outcomes in terms of Sharpe ratios or portfolio returns for each combination.

With outcomes from all possible topic combinations, I then compute the Shapley value using the following formula:

$$\varphi_p\left(SR\right) = \sum_{S \subseteq P \setminus \{p\}} \frac{|S|!\left(P - |S| - 1\right)!}{P!} \left[SR\left(y_s^{\text{emb}} + y_p^{\text{emb}}\right) - SR\left(y_s^{\text{emb}}\right)\right], \tag{5}$$

where the sum extends over all subsets S of P not containing topic p, including the empty set. Note that $\binom{n}{k}$ is the binomial coefficient. y_p^{emb} is the token-weighted sentence embedding of topic p, and y_s^{emb} is the token-weighted sentence embedding of all topics in subset S. For each combination of topic embeddings, I calculate the Sharpe ratio accordingly.[3] This approach fully distributes the portfolio Sharpe ratio among topics, revealing their relative contribution to report investment value.

Referring to Lv (2024), a potential concern in the topic representation stage is that token embeddings inherently contain information from other tokens, a feature designed into the LLM architecture to enable context understanding. One solution is to process embeddings at the sentence level rather than the report level to ban inter-sentence communications. The token-weighted mean of sentence embeddings can serve as an alternative report representation:

$$y^{emb} = \sum_{i=1}^{n} \frac{Token_i}{\sum_{i=1}^{n} Token_i} y_i^{emb}, \tag{6}$$

where $Token_i$ represents the number of tokens in i-th sentence and y_i^{emb} represents the embedding of the i-th sentence.

While this approach theoretically sacrifices contextual information across the entire report, I demonstrate in empirical tests that the performance deterioration is minimal.

3 Empirical Results

3.1 Investment Value

In this section, I examine the investment value of analyst reports through multiple lenses. Section 3.1.1 analyzes return predictability and portfolio performance. Section 3.1.2 compares the embedding-based model predictions with traditional report sentiment measures.

[3]Shapley value requires an input for X when it is missing. I assign a value of zero, treating the embedding as having all dimensions set to zero for model estimation.

Section 3.1.3 studies the cross-sectional variation in investment value across firm characteristics and analyst attributes.

3.1.1 Return Predictability

I begin with a univariate event-day approach to examine the stock price impact of analyst reports. This analysis is predicated on the notion that stock prices adjust when analysts release new information reflecting changes in their prior beliefs. The presence of abnormal returns following report releases would suggest that the market does not immediately fully incorporate analyst information. The price dynamic would indicate the investment value of the information.

To investigate these price dynamics, I use the predicted 12-month ahead stock returns from regression (1) as a measure of analysts' outlook. I sort reports into decile groups based on their predictions for each trading day.[4] To calculate buy-and-hold abnormal returns, I follow the characteristics-based benchmark approach proposed by Daniel et al. (1997):

$$CAR_{0,T} = \prod_{t=0}^{T} (1 + R_{it}) - \prod_{t=0}^{T} \left(1 + R_{it}^{DGTW}\right), \tag{7}$$

where R_{it} represents the raw return of stock i on trading day t, R_{it}^{DGTW} denotes the value-weighted return of stock i's benchmark portfolio on day t, and T indicates the end date of the accumulation period. 0 represents the first trading day when the market is open at or after a report is released.

Figure 1 compares the abnormal returns between reports with the most positive and most negative outlooks over a 252-trading-day horizon (one calendar year). Reports with the highest predicted returns exhibit a cumulative DGTW-adjusted abnormal return of 3.75% over the subsequent 252 trading days, while those with the lowest predictions show an abnormal return of -0.56% over the same horizon. The return spread between High and Low groups is statistically significant at the 1% level. The price dynamics following report releases reveal an asymmetric pattern between positive and negative outlooks. This asymmetry suggests two potential explanations: either analysts are more skilled at identifying stocks with positive potential, or investors are not fully incorporating analysts' positive narratives into prices.

As a second test of the investment value of analyst outlook, I use the following report-level regression model to explore return predictability:

$$RET_{it,12m} = \alpha_t + \beta' x_{i,t} + \varepsilon_{i,t+12}, \tag{8}$$

[4] In the sample, a trading day contains an average of 167 reports.

where $RET_{it,12m}$ represents the realized 12-month ahead return of stock i. $x_{i,t}$ denotes analyst information from each report, consisting of four measures: $\widehat{RET_{12m}}$ is the predicted 12-month ahead return from Ridge regression (1); REC_{rev} represents recommendation revision, calculated as the difference between the current report's recommendation and the analyst's last recommendation for the same stock in I/B/E/S; EF_{rev} denotes earnings forecast revision; and TP_{rev} represents target price revision. The model includes year-month fixed effects α_t to control for systematic market conditions. Standard errors are clustered two ways at both stock and year-month levels.

I report the regression results in Table 2. The estimated returns derived from analyst report information demonstrate statistically significant predictive power for future stock returns. After controlling for analyst-year-month and industry-year-month fixed effects, the coefficient in column (2) (0.06, t-statistics=4.37) suggests that a one standard deviation increase in estimated returns corresponds to a 10.41% increase in realized returns relative to the sample mean over the following year. This finding aligns with the evidence in Figure 1, confirming the consistency between analyst report outlook and subsequent stock market movements.

I further compare the predictive power of numerical analyst forecasts at the one-year horizon. Columns (3)-(6) show that the predictive power of text-based estimates remains robust when controlling for analyst recommendation revisions, earnings forecast revisions, and target price revisions. All coefficients maintain their statistical and economic significance. While these numerical forecasts are documented as valuable predictors in the literature, their coefficient estimates become statistically insignificant when competing with text predictions for one-year returns. This indicates that the report text contains unique quantitative information in forecasting stock market returns.

I proceed with a portfolio analysis to evaluate the economic value of analyst report predictions. The trading strategy design incorporates two key elements. First, portfolios are rebalanced monthly to balance the tradeoff between capturing new information and minimizing transaction costs. Second, motivated by the persistence in return predictability documented in Figure 1, I aggregate analyst report predictions over extended historical windows to capture sustained components in analyst outlooks.

Specifically, at the beginning of each month, I construct portfolios using the following procedure: First, for each stock, I calculate the average predicted return from all analyst reports issued during the past LB months, where $LB \in \{9, 12, 18, 24\}$. I then sort stocks into decile portfolios based on these averaged predictions. Within each decile portfolio, stocks are weighted by their previous month-end market capitalization to mitigate the influence of microcaps. I then construct a zero-cost long-short portfolio that takes a long position in the

highest predicted return decile (decile 10) and a short position in the lowest predicted return decile (decile 1).

Table 3 presents performance statistics of value-weighted decile portfolios formed based on Ridge models' return forecasts using different lookback periods (LB) of 9, 12, 18, and 24 months. While the decile performance shows some variation, the highest decile consistently outperforms the lowest decile. The long-short portfolio (H-L) that buys the highest decile and shorts the lowest decile generates positive returns ranging from 0.87% to 1.16% per month, with corresponding Sharpe ratios between 0.62 and 0.69. The portfolio alphas relative to the Fama-French five-factor and momentum factor are positive and statistically significant, ranging from 0.52 to 0.79 (t-statistics between 1.99 and 2.64). The alpha across decile portfolios reveals that the long-short portfolio performance is primarily driven by the long leg rather than the short leg, consistent with the asymmetric pattern documented in Figure 1. For subsequent analyses, I adopt the 12-month lookback period as the benchmark specification, as it yields the highest t-statistic (2.64) for the long-short portfolio alpha and aligns with Ridge models' 12-month return prediction target.

Figure 2 plots the cumulative returns over the period of January 2005 to June 2024 for the 12-month lookback period trading strategy. A self-financing value-weighted portfolio that buys stocks with the most positive outlook and shorts those with the most pessimistic outlook generates a cumulative return of 846% over this period. For comparison, the value-weighted market return in excess of the risk-free rate during this period is 399%. Notably, the long lag earns 2274% over the 20 years, while the short leg delivers a cumulative excess return of 151%. Thus, the investment value stems from picking stocks with strong future prospects, which is consistent with the portfolio statistics in Table 3. The strategy experiences significant drawdowns during the global financial crisis and again post-2022, conditions in which fundamental signals can be temporarily overshadowed by systemic risk aversion and forced liquidations.

3.1.2 Short-term Overreaction

Analysts often release reports in response to major news or corporate announcements. This raises the question of whether the text predictability stems from documented price drift phenomena, such as the post-earnings-announcement-drift (PEAD), rather than reflecting genuine new information about firms' long-term fundamentals.

To distinguish between the two drivers, I conduct a double-sorting analysis based on earnings surprises and return predictions, presented in Figure 3. First, reports are classified into 'SUE(Small)' or 'SUE(Large)' groups based on whether their stocks' most recent earnings surprise falls below or above the monthly sample median. Following Livnat and Mendenhall

(2006), earnings surprise (SUE) is calculated as the actual EPS minus the last consensus EPS forecast before the earnings announcement, with higher values indicating positive news. Second, within each SUE group, I sort reports into deciles based on their predicted returns. Recall that the return predictability is driven by long lag (positive predictions), if the observed predictability were attributable to a price-drift like PEAD, then reports with high SUE (i.e., good earnings news) should exhibit a greater spread between the 'High' and 'Low' groups.

Surprisingly, the data reveal the opposite pattern. As shown in Figure 3 panel (a), the return spread is 7.52% for reports issued after low earnings surprises, compared to 4.30% for those issued following high surprises. Thus, it appears that weak earnings announcements in fact contribute more to return predictability, contradicting a drift-based explanation and posing its own puzzle.

To gather more evidence, I examine several measures across the return prediction deciles: (1) $CAR_{[0,+1]}$, the two-day DGTW characteristic-adjusted buy-and-hold abnormal returns following report release dates, which captures immediate market reactions; (2) REC_{rev}, analyst recommendation revisions that reflect changes in analyst opinions; and (3) two text sentiment measures generated using a fine-tuned BERT model that classifies each sentence as positive, negative, or neutral. $Tone_{Head}$ evaluates report headlines, while $Tone_{Body}$ assesses the report body content. Specifically, $Tone_{Body}$ is calculated as:

$$Tone_{Body,i} = \frac{N_i^+ - N_i^-}{N_i}, \tag{9}$$

where N_i^+ (N_i^-) represents the number of positive (negative) sentences in report i as classified by the BERT model, and N_i denotes the total number of sentences in the report.

Table 4 presents two noteworthy findings. First, the realized returns (R_{12m}) exhibit a significant spread of 5%, which closely aligns with the 'H-L' return spread documented in Figure 1. Second, the contemporaneous market reaction ($CAR_{[0,+1]}$) following report releases shows a negative relationship with predicted returns, with a significant high-minus-low spread of -0.25%. Together with the evidence in Figure 3, these results present an intriguing puzzle: The most optimistic forward-looking reports tend to follow bad news, and vice versa.

I propose three potential explanations for this pattern. First, both analysts and the market may overreact to short-term negative news (e.g., a weak quarter or a disappointing press release). Yet the analyst's deeper analysis (e.g., the model-based 12-month forecast) might conclude that the fundamentals remain solid, eventually driving future price appreciation. Under this scenario, we should observe consistent negative relationships between analyst recommendations, narrative tones, and predicted returns.

An alternative story is that analysts intentionally address short-term negativity but emphasize the firm's longer-term strengths in order to reassure overly pessimistic investors. Here, the analyst acknowledges current problems but stresses the company's strategy or fundamentals as a counterpoint to the short-run overreaction. If this is the driving mechanism, there should be a positive correlation between the analyst's optimistic return forecasts and the tone of the report text.

A third explanation posits that analysts systematically include disclaimers or cautious language—even when they maintain a constructive long-term view—to hedge official recommendations and manage liability. In this case, the overall textual tone might appear muted or slightly negative despite a fundamentally bullish numerical recommendation.

The empirical evidence points to systematic overreaction to near-term negative signals, rather than a brief misalignment or benign hedging of recommendations. Table 4 shows that the most optimistic long-horizon forecasts consistently coincide with negative short-run sentiment (in both the text and recommendation revisions). Meanwhile, Figure 3 indicates that this predictability is strongest specifically among reports issued after weak earnings news and downgrades, suggesting analysts are not offsetting short-term setbacks with unequivocally bullish outlooks.

If overreaction is the primary driver of return predictability, does this imply that analysts' forward-looking outlooks are irrelevant? To explore this, I conduct an analysis by sorting portfolios based on sentiment measures to assess whether such strategies can generate alpha. Table A4 reveals that simple sentiment-based sorting fails to produce significant alpha, suggesting that sentiment alone does not represent the informational value embedded in analyst reports. Rather, analysts' assessments of fundamental value remain crucial for predicting subsequent return reversals.

Taken together, the analyses reveal that analysts and the market place disproportionate weight on immediate setbacks, even while the analyst's expertise anticipates subsequent stock outperformance. The investment value arises from analysts' ability to recognize longer-term fundamentals despite their own cautious or negative language in the short run. This systematic mismatch, where near-term negativity coincides with strong future returns, ultimately generates the portfolio alpha.

3.1.3 Cross-sectional Analyses

I next turn to the question of reports from what kinds of analysts on which stocks generate investment value. Figure 4 presents DGTW-adjusted buy-and-hold abnormal returns sorted on analyst and firm characteristics. The results reveal substantial heterogeneity in the investment value of research reports across different segments of the market.

I first examine the return predictability of analyst reports across large vs small stocks and mature vs young stocks. Panel (a) shows that reports on large-capitalization stocks generate significantly higher abnormal return spreads compared to small-cap stocks (4.14% vs 1.74%), with the difference particularly pronounced for stocks predicted to have high returns. This pattern suggests that despite the greater analyst coverage and presumed market efficiency in the large-cap segment, reports from skilled analysts still contain meaningful investment opportunities among these stocks. The life cycle results in Panel (b) indicate that mature firms offer greater potential for profitable analyst recommendations compared to young firms. Reports on mature firms predicted to have high returns generate cumulative abnormal returns of approximately 5.15% over the 252-day horizon, while similar reports on young firms yield a substantially lower return spread of 1.51%.

In addition, I also test the predictability in the cross-section of analyst skills and experience. Analyst skill, measured by historical EPS forecast accuracy and shown in Panel (c), appears to be a crucial determinant of report value. High-skill analysts generate significantly larger abnormal return spreads (5.74%), particularly for their most optimistic recommendations, while reports from low-skill analysts show significantly smaller investment value in terms of return spreads (4.44%). Finally, Panel (d) demonstrates that analyst experience plays a vital role in report effectiveness. Experienced analysts' reports, especially those predicting high returns, generate substantial abnormal returns approaching 6.14% over the holding period. In contrast, reports from less experienced analysts show less ability to identify profitable investment opportunities.

These findings suggest that the investment value of research reports is concentrated among experienced, skilled analysts covering large, mature firms. This pattern is consistent with the idea that superior information production requires both analytical capability and deep institutional knowledge. Moreover, large, mature businesses with more predictable cash flows and stable competitive moats often snap back from negative news—leading to a positive return drift that skilled analysts may hint.

3.2 What content matters?

I now investigate the sources of investment value in analyst reports by decomposing portfolio returns and Sharpe ratios across five topic categories, as described in Section 2.2.2. Using yearly fitted Ridge models, I calculate each topic's contribution to portfolio performance by measuring the average decrease in performance across all possible topic combinations. I group together topic Shapley values with specific categories and report the sum of their constitute importance.

Table 5 presents both the distribution of content and its contribution to portfolio performance. Financial analysis and company/industry overview dominate the content, comprising 36.56% and 28.53% of total coverage respectively. Strategic outlook and risk and governance form a second tier of coverage at 15.13% and 14.14% of sentences, while additional content accounts for the remaining 5.63%.

The Shapley value decomposition reveals a different pattern. Strategic Outlook emerges as the most valuable category, accounting for 41.34% and 31.43% of the portfolio's Sharpe ratio and returns respectively, despite not being the most extensively discussed component. Company and Industry Overview maintains proportional importance, contributing 27.61% to the Sharpe ratio and 26.92% to returns. Risk and Governance shows outsized impact on returns relative to its discussion frequency, accounting for 11.21% of Sharpe ratio value and 21.36% of returns. Additional Content has minimal impact on performance.

Perhaps the most surprising result is the low contribution of Financial Analysis given findings in Lv (2024) that emphasize high information content in income statement analysis. Despite constituting the most extensively covered category (36.56% of content), Financial analysis discussions contribute only to 16.39% of the Sharpe ratio and 19.53% of returns. One reason may be that financial data are more rapidly priced in, leaving less room for drift or mispricing to persist. In contrast, strategic and forward-looking insights may only gradually be reflected in valuations, thus creating greater opportunity for abnormal returns.

The findings in Table 5 suggest that the investment value of analyst reports stems primarily from their strategic insights and industry analysis rather than traditional financial analysis. The substantial contribution of strategic content to portfolio performance, despite its relatively lower coverage, indicates that analysts' forward-looking assessments and strategic perspectives may be particularly valuable for investment decisions.

Table 6 presents a detailed analysis of strategic outlook content along three key dimensions - timeframe, sentiment, and analytical focus - using Shapley value decomposition. For each dimension, I classify sentences using a structured prompt that categorizes timeframe (long-term, short-term, or both), sentiment (negative, neutral, or positive), and focus (risk versus fundamental analysis) with the following prompt:

Prompt 2:
Please read the following sentence from a sell-side analyst report (investment thesis section) carefully and classify it into three numeric labels:
Timeframe: 1 = Long-term (multi-quarter or multi-year outlook); 2 = Short-term (next quarter or near-term); 3 = Both (if it clearly addresses both short- and long-term).
Sentiment: 1 = Positive potential; 0 = Neutral; 1 = Negative/Bearish.
Focus: 1 = Fundamentals (e.g., belief in growth strategy, industry tailwinds, earnings); 0 = Cautious risk (e.g., warnings, near-term headwinds, legal/regulatory risk).
Important note: If the sentence mentions both short-term risk and a fundamental (longer-term) driver, prioritize the fundamental aspect and mark the third label as **1** (Fundamentals).
Output format: Only output the three numbers in the format: `[X, Y, Z]`.
Here is the sentence: {sentence}

The results reveal several notable patterns. First, long-term oriented discussions dominate the investment value, accounting for approximately half of both Sharpe ratio (49.51%) and return (50.20%) contributions. Short-term discussions contribute about a quarter of the predictive power, while discussions spanning both timeframes account for the remaining portion.

Second, the sentiment analysis shows that positive outlooks generate the largest share of investment value, contributing 49.94% to Sharpe ratio and 50.86% to returns. Neutral and negative sentiments have roughly equal contributions, each accounting for about a quarter of the portfolio performance metrics.

Perhaps most strikingly, the focus decomposition shows that fundamental analysis overwhelmingly drives investment value, accounting for 87.00% of Sharpe ratio and 87.75% of return contributions. Risk-focused discussions, while important for comprehensive analysis, contribute relatively little to portfolio performance metrics.

These findings suggest that the predictive power of strategic outlook sections stems primarily from long-term, positive discussions focused on fundamental analysis, rather than from risk warnings or short-term forecasts. This pattern aligns with the notion that durable competitive advantages and long-term growth opportunities require more analytical skill to identify and value correctly than short-term fluctuations or risk factors.

A follow-up question would be, since Shapley value attribution pins down the content contributing the most to investment value, can we separate this most valuable information and form a portfolio free of text noise? In Table 7, I report the value-weighted decile portfolio performance sorted by category-specific content. I focus on four categories, excluding

additional content that primarily consists of latent sentences.

Strategic Outlook content generates the strongest portfolio performance, with the long-short strategy delivering a mean return of 1.41% per month and a Sharpe ratio of 0.93. The outperformance appears to be driven by both the long and short legs, with the highest decile earning 1.87% monthly and showing significant alpha (0.72, t = 2.85). This superior performance aligns with the earlier Shapley value decomposition, which identifies Strategic Outlook as the most valuable category.

The remaining categories, Company and Industry Overview, Financial Analysis, and Risk and Governance, generate more modest performance metrics. Long-short returns range from 0.52% to 0.58% monthly and Sharpe ratios are between 0.37 and 0.41. While the highest deciles of these categories show some predictive power, particularly Risk and Governance with an alpha of 0.32 (t = 1.97), their overall performance is substantially weaker than Strategic Outlook. This pattern is again consistent with the Shapley value decomposition. It suggests that traditional financial analysis, industry insights, and risk assessments are less informative about future returns than forward-looking strategic insights.

In general, the results again highlight the long-term investment opportunities embedded in analysts' forward-looking strategic insights. Additionally, the univariate analysis confirms the effectiveness of the attribution analysis with Shapley values, showing that focusing on the most valuable component of Strategic Outlook content indeed helps form more efficient portfolios.

3.3 Incremental Investment Value

Literature has documented "zoo of factors" with investment value. A key question is whether the investment value of analyst reports is incremental to existing predictive signals. I therefore extend my analysis to examine whether analyst reports contain unique information beyond two comprehensive sets of established predictors: 18 analyst-based factors including EPS forecast revisions and recommendation changes from the Chen and Zimmermann (2022) database, and 94 market and fundamental-based factors documented to have cross-sectional predictability in Gu et al. (2020).[5] Detailed descriptions of these analyst-based and fundamental-based factors can be found in Tables A2 and A1, respectively.

In Table 8, I evaluate the predictive power of analyst reports relative to and in combination with established return predictors. The analysis examines four main signals: analyst report full content predictions (RP), strategic outlook content predictions (SO), the average return series of 18 analyst-based factors (ANA), and predictions from a 4-layer artificial

[5]The replication code is publicly available on Dacheng Xiu's website.

neural network (ANOM) - the best-performing machine learning model in Gu et al. (2020). For each strategy and combination, I report mean returns, Sharpe ratio (SR), and information ratio (IR), with corresponding t-statistics in parentheses. To assess the robustness of the predictive signals, I compute risk-adjusted returns (α) relative to four prominent factor models: the Fama and French (2015) five factors (α_{F5}), the five factors plus momentum (α_{F6}), the Hou et al. (2015) q-factor model (α_{HXZ}), and the Daniel et al. (2020) behavioral factor model (α_{DHS}). The Information Ratio of the portfolio i relative to a benchmark b is calculated as:

$$IR_{i,b} = \frac{\bar{r}_{i,b}}{\sigma_{i,b}}, \tag{10}$$

where $\bar{r}_{i,b}$ is the average active return (difference between portfolio i and benchmark b returns) and $\sigma_{i,b}$ is the standard deviation of these active returns, also known as tracking error. In Table 8, I compute the IR of combined strategies relative to individual components to measure the incremental value of information sources. For example, the IR of "RP + ANA" measures the incremental gain from combining report predictions with analyst-based factors, relative to using "ANA" signals only.

Panel A first establishes the standalone performance of each information source. Both full report content (RP) and strategic outlook sections (SO) generate significant risk-adjusted returns across various factor models, with alphas ranging from 0.73 to 1.21 for RP and 0.95 to 1.46 for SO. The numerical analyst-based factors (ANA) show modest but significant predictive power (alphas between 0.16 and 0.28). The machine learning-based anomaly portfolio (ANOM) exhibits strong performance with alphas between 1.15 and 1.40, suggesting effective capture of fundamental signals.

Panels B and C address the central question of incremental value. When combining report predictions with analyst-based factors (RP + ANA), the portfolio maintains significant alphas (0.50-0.68) and shows meaningful incremental gain with an IR of 0.73. The combination with fundamental factors (RP + ANOM) yields even stronger results, with alphas between 1.03 and 1.18 and an IR of 1.16. Most notably, incorporating both analyst and fundamental factors (RP + ANA + ANOM) produces the highest Sharpe ratio (1.60) and consistent alphas across factor models (0.77-0.92), with an IR of 1.23 (t=4.05). The patterns are similar but even stronger when using strategic outlook content instead of full reports (IR=1.41, t=4.35), consistent with my earlier findings about the particular value of analyst strategic outlook discussions.

These results suggest that analyst reports contain statistically and economically significant incremental information not captured by traditional analyst outputs like forecast revi-

sions or by machine learning models trained on fundamental data. The strong incremental performance and information ratios indicate that textual analysis of research reports, particularly their strategic content, offers a distinct and complementary source of investment value.

3.4 Look ahead Bias

A key concern in demonstrating out-of-sample profitability from analyst reports is look-ahead bias. Because LLMs are pre-trained on massive text corpora, there is a risk that they could have indirectly learned from future stock performance data. While Table 7 shows that providing corporate and industry information to LLMs does not generate significant profits—mitigating concerns about entity- or industry-level information leakage—one might argue that strategic outlooks contain forward-looking information that could align with the LLMs' pre-training corpus. The fact that these outlooks can later be confirmed by actual outcomes may introduce another source of information leakage.

To further address this concern, I conduct a stringent analysis restricted to post-knowledge cutoff samples. This analysis encompasses a broad range of language models with diverse architectures and scales. BERT and RoBERTa are relatively smaller, pioneering transformer-based NLP models (with 110 million to 355 million parameters). In contrast, LLaMA2-13B is a significantly larger model at 13 billion parameters, while LLaMA3-8B represents a more recent architecture with 8 billion parameters. Specifically, I evaluate portfolio performance only after each model's training data cutoff date: BERT (post-June 2024), RoBERTa (post-June 2024), LLaMA2 (post-June 2024), and LLaMA3 (post-April 2024). Table 9 presents these results across two dimensions: analysis of full reports (Panel A) and strategic outlooks (Panel B).

Strikingly, the performance in this true out-of-sample period is even stronger than in the full sample. Examining the high-minus-low (H-L) portfolio returns in Panel A, all models generate substantial positive returns ranging from 1.59% (LLaMA2) to 2.70% (LLaMA3) with corresponding Sharpe ratios between 0.74 and 1.28. The performance amplifies further when focusing solely on strategic outlooks (Panel B), with H-L returns more than doubling for some models, ranging from 1.81% (RoBERTa) to an impressive 5.28% (LLaMA3). The enhanced performance in the post-cutoff period provides compelling evidence that the LLMs' ability to extract profitable information from analyst reports is not driven by potential information leakage from their training text.

4 Conclusion

Investment value from sell-side analyst forecasts has been well-documented in the literature. Beyond offering simple numerical projections, analysts also produce research reports that detail their fundamental analyses and valuation methods—reports that institutional investors often deem more valuable than forecasts alone. Building on this notion, I conduct a comprehensive examination of a large corpus of analyst reports spanning 2000 to 2023, extracting embedded value-relevant information using state-of-the-art LLMs and machine learning techniques.

My findings reveal a new source of investment value. The textual information, when represented as structured vectors, predicts future stock returns. This return predictability offers significant incremental investment value beyond existing analyst- and fundamental-based factors, and it remains robust in a period free from look-ahead bias.

I also observe that return predictability is predominantly driven by reports conveying negative sentiment yet predicting positive future returns. This pattern suggests a systematic overreaction to negative news, such as weak quarterly earnings, even when a stock's fundamentals remain strong. Further, the predictability is more pronounced for large, mature stocks and is particularly evident in reports authored by skilled and experienced analysts.

Next, I investigate which specific content within analyst reports contributes most to their investment value. Using a Shapley value decomposition, I show that Strategic Outlook content, especially statements revealing positive future fundamentals in the investment thesis and valuation models, dominates portfolio performance. Remarkably, a trading strategy constructed univariately on strategic outlook content outperforms trading strategies based on full content.

This study provides the first direct evidence of the investment value embedded in analyst report text. These reports reveal analysts' beliefs and expertise regarding particular stocks and convey valuable information to the market. More broadly, the findings illustrate the potential of artificial intelligence to enhance human judgment in financial contexts. Rather than replacing analyst insights, language models and machine learning add the most value when scaling and systematically extracting value-relevant information from natural language.

In addition, the Shapley value decomposition approach proves to be a powerful tool for interpreting textual information, helping researchers and investors understand how different content types affect outcomes of interest. Future research could extend these methods to other forms of financial narratives and examine the mechanisms through which qualitative information is gradually incorporated into asset prices.

References

Birru, Justin, Sinan Gokkaya, Xi Liu, and René M Stulz, 2022, Are analyst short-term trade ideas valuable?, *The Journal of Finance* 77, 1829–1875.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer, 2019, Diagnostic expectations and stock returns, *The Journal of Finance* 74, 2839–2874.

Brav, Alon, and Reuven Lehavy, 2003, An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics, *The Journal of Finance* 58, 1933–1967.

Bybee, Leland, Bryan Kelly, and Yinan Su, 2023, Narrative asset pricing: Interpretable systematic risk factors from news text, *The Review of Financial Studies* 36, 4759–4787.

Chen, Andrew Y., and Tom Zimmermann, 2022, Open source cross-sectional asset pricing, *Critical Finance Review* 27, 207–264.

Chen, Hailiang, Byoung-Hyoun Hwang, and Zhuozhen Peng, 2023a, Revisiting the cross-section of expected stock returns: Evidence from a textual analysis of buy recommendations, *Nanyang Business School Research Paper* .

Chen, Luyang, Markus Pelger, and Jason Zhu, 2023b, Deep learning in asset pricing, *Management Science* .

Chen, Yifei, Bryan T Kelly, and Dacheng Xiu, 2022, Expected returns and large language models, *Available at SSRN 4416687* .

Christophe, Stephen E, Michael G Ferri, and Jim Hsieh, 2010, Informed trading before analyst downgrades: Evidence from short sellers, *Journal of Financial Economics* 95, 85–106.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *The Journal of Finance* 52, 1035–1058.

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short-and long-horizon behavioral factors, *The Review of Financial Studies* 33, 1673–1736.

De Bondt, Werner FM, and Richard H Thaler, 1990, Do security analysts overreact?, *The American Economic Review* 52–57.

De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Fama, Eugene F, and Kenneth R French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Farago, Adam, Erik Hjalmarsson, and Ming Zeng, 2023, Analysts are good at ranking stocks, *Available at SSRN 4495163* .

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer, 2016, Expectations and investment, *NBER Macroeconomics Annual* 30, 379–431.

Green, T Clifton, 2006, The value of client access to analyst recommendations, *Journal of Financial and Quantitative Analysis* 41, 1–24.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *The Review of Financial Studies* 33, 2223–2273.

Hirshleifer, David, Dat Mai, and Kuntara Pukthuanthong, 2023, War discourse and the cross section of expected stock returns, Technical report, National Bureau of Economic Research.

Hong, Harrison, and Jeremy C Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of Finance* 54, 2143–2184.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *The Review of Financial Studies* 28, 650–705.

Huang, Allen H, Amy Y Zang, and Rong Zheng, 2014, Evidence on the information content of text in analyst reports, *The Accounting Review* 89, 2151–2180.

Jegadeesh, Narasimhan, Joonghyuk Kim, Susan D Krische, and Charles MC Lee, 2004, Analyzing the analysts: When do recommendations add value?, *The Journal of Finance* 59, 1083–1124.

Jha, Manish, Hongyi Liu, and Asaf Manela, 2020, Does finance benefit society? a language embedding approach, *A Language Embedding Approach (July 18, 2020)* .

Li, Kai, Feng Mai, Rui Shen, Chelsea Yang, and Tengfei Zhang, 2024, Dissecting corporate culture using generative ai–insights from analyst reports, *Available at SSRN 4558295* .

Livnat, Joshua, and Richard R Mendenhall, 2006, Comparing the post–earnings announcement drift for surprises calculated from analyst and time series forecasts, *Journal of Accounting Research* 44, 177–205.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *The Journal of Finance* 66, 35–65.

Lv, Linying, 2024, The value of information from sell-side analysts, *arXiv preprint arXiv:2411.13813* .

Lys, Thomas, and Sungkyu Sohn, 1990, The association between revisions of financial analysts' earnings forecasts and security-price changes, *Journal of Accounting and Economics*

13, 341–363.

Meursault, Vitaly, Pierre Jinghong Liang, Bryan R Routledge, and Madeline Marco Scanlon, 2023, Pead. txt: Post-earnings-announcement drift using text, *Journal of Financial and Quantitative Analysis* 58, 2299–2326.

Porta, Rafael La, Josef Lakonishok, Andrei Shleifer, and Robert Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *the Journal of Finance* 52, 859–874.

Sarkar, Suproteem K, and Keyon Vafa, 2024, Lookahead bias in pretrained language models, *Available at SSRN* .

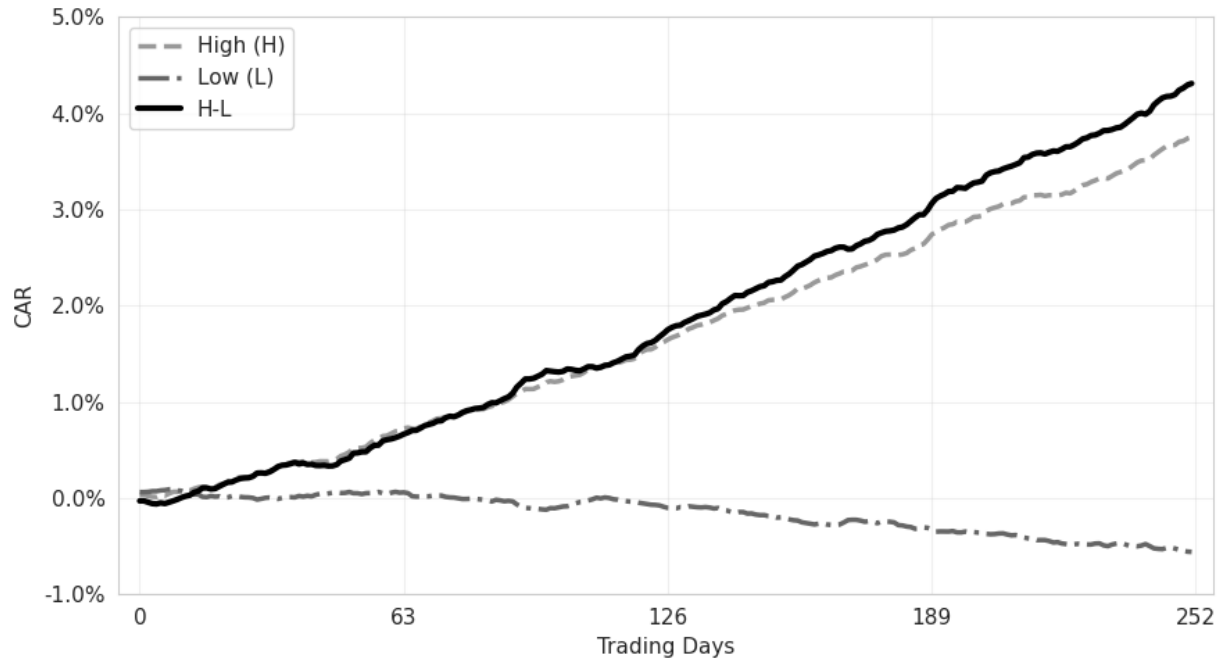
Shapley, Lloyd S, 1953, A value for n-person games, *Contributions to the Theory of Games* 2, 307–321.

Tetlock, Paul C, 2007, Giving content to investor sentiment: The role of media in the stock market, *The Journal of Finance* 62, 1139–1168.

Weber, Michael, 2018, Cash flow duration and the term structure of equity returns, *Journal of Financial Economics* 128, 486–503.

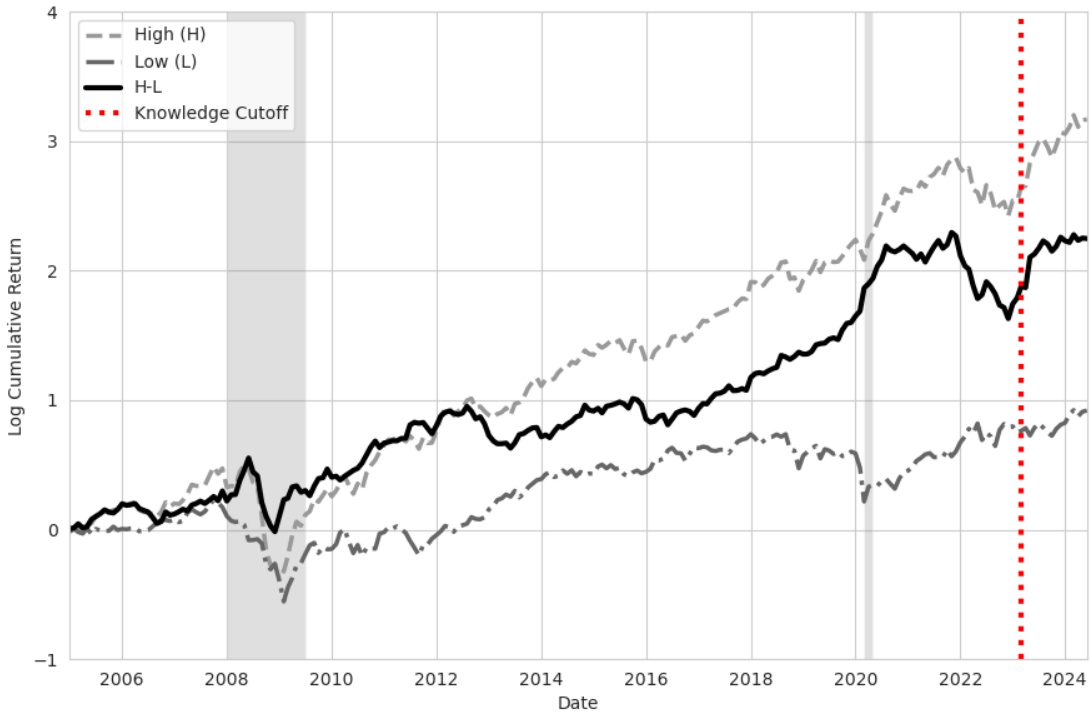
Womack, Kent L, 1996, Do brokerage analysts' recommendations have investment value?, *The Journal of Finance* 51, 137–167.

FIGURE 1: Report Outlooks and Abnormal Stock Returns



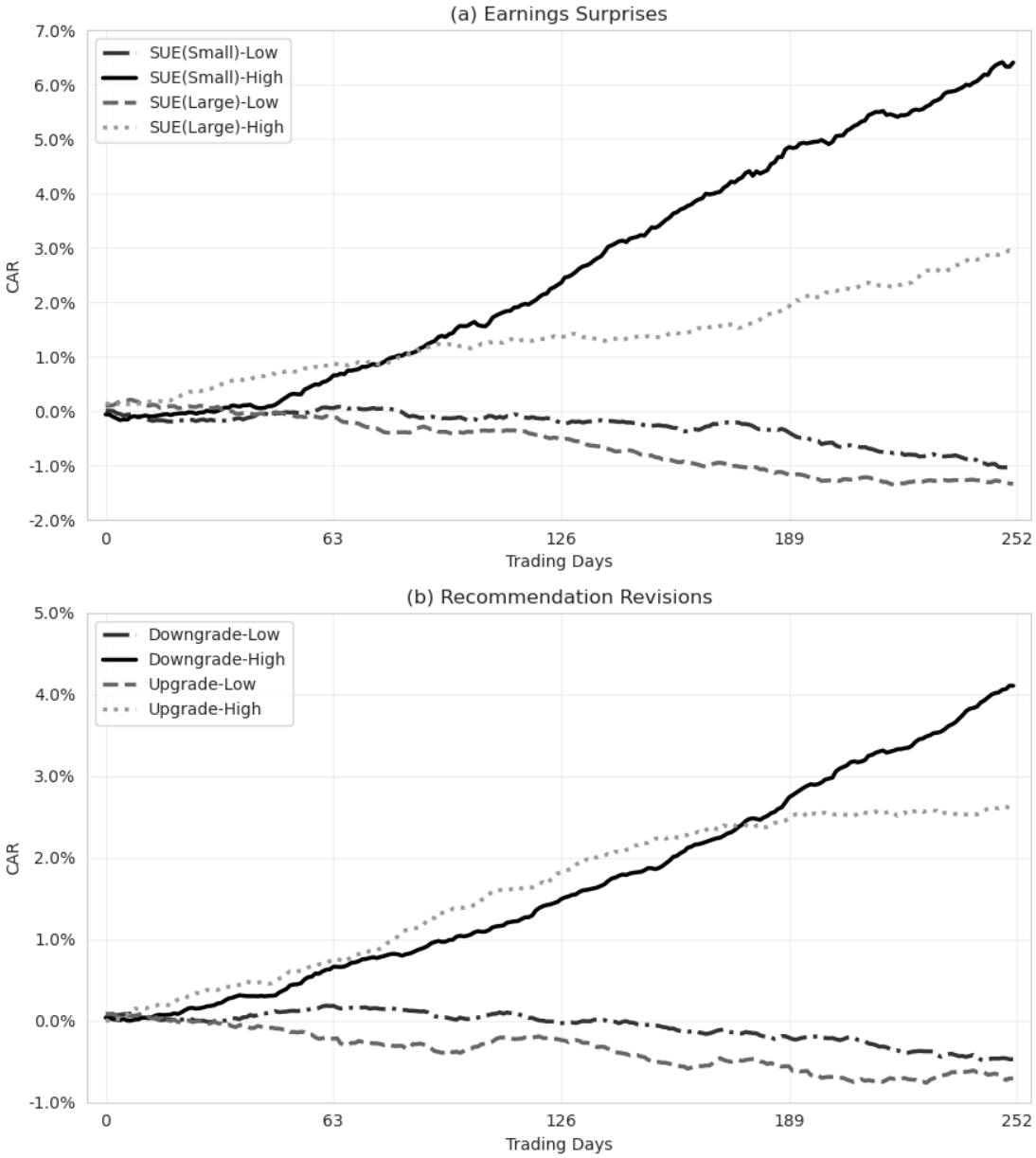
Note: This figure presents DGTW characteristic-adjusted buy-and-hold abnormal returns following research report releases, measured over a [0, +252] trading day window after each announcement. Reports are sorted into deciles based on their predicted 12-month returns. The 'High (H)' group consists of reports predicting returns in the highest decile within each trading day, while the 'Low (L)' group comprises reports predicting returns in the lowest decile. The lines represent value-weighted abnormal returns of the stocks corresponding to these reports, and 'H-L' denotes the return spread between the High and Low groups. The sample spans from 2005 to 2023.

FIGURE 2: Cumulative Log-return of Portfolios Sorted on Report Forecasts



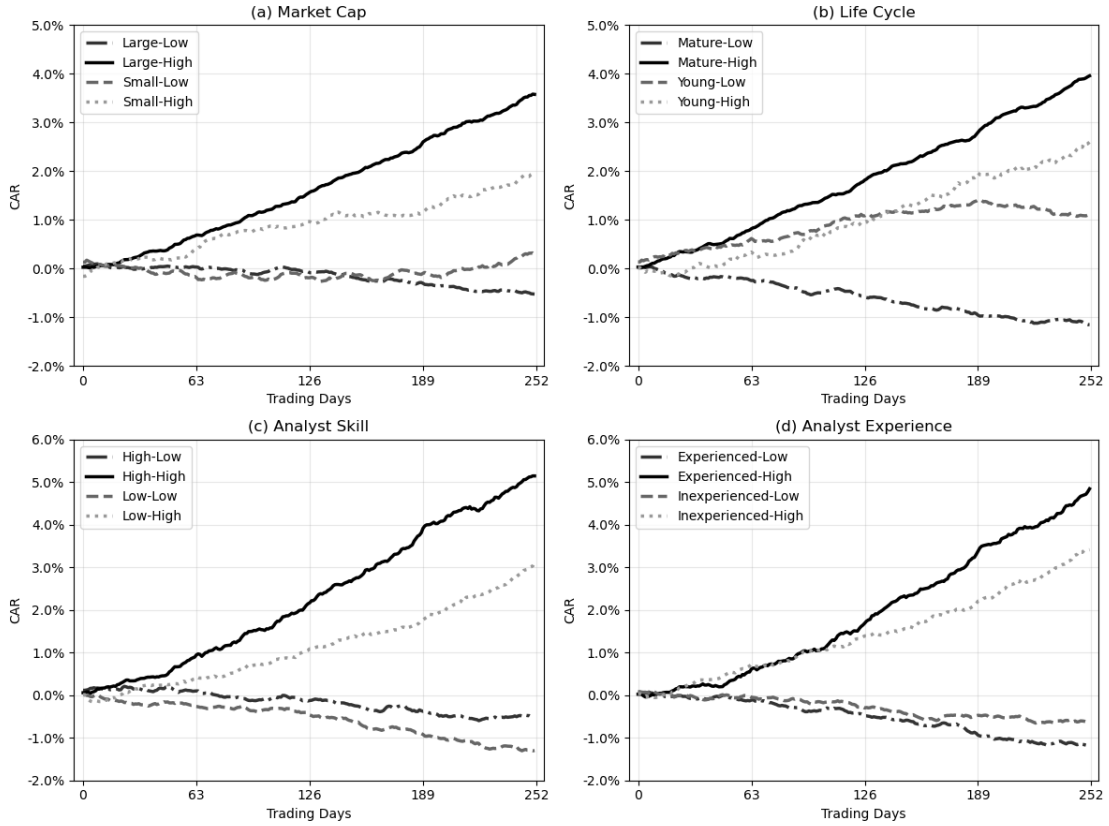
Note: This figure shows cumulative log excess returns for portfolios constructed based on analyst report forecasts from the previous 12 months. For each panel, three portfolios are presented: the highest decile (H), the lowest decile (L), and a long-short portfolio (H-L) that buys the highest and sells the lowest decile. Grey shaded areas represent NBER recession periods. The dashed vertical line marks the LLaMA3-8B model knowledge cutoff. The return time series spans from January 2005 to June 2024.

FIGURE 3: Report Outlooks and Abnormal Stock Returns: Good News versus Bad News



Note: This figure presents DGTW characteristic-adjusted buy-and-hold abnormal returns following research report releases, measured over a [0, +252] trading day window after each announcement. Reports are sorted into deciles based on their predicted 12-month returns. Reports are classified into 'SUE(Small)' or 'SUE(Large)' groups based on whether the stocks' most recent earnings surprise falls below or above the monthly sample median. 'Downgrade' and 'Upgrade' represent the analysts' stock recommendation changes for these groups. The 'High' group consists of reports predicting returns in the highest decile within each month, while the 'Low' group comprises reports predicting returns in the lowest decile. The lines represent value-weighted abnormal returns of the stocks corresponding to these reports. The sample spans from 2005 to 2023.

FIGURE 4: Report Outlooks and Abnormal Stock Returns: Cross-sectional Analysis



Note: This figure presents DGTW characteristic-adjusted buy-and-hold abnormal returns following research report releases, measured over a [0, +252] trading day window after each announcement. Reports are sorted based on two dimensions. First, they are categorized by the monthly median of the following characteristics at the end of the previous month: (a) market capitalization, (b) firm age, (c) analyst skill (measured by average historical EPS forecast accuracy), and (d) analyst experience (measured by years of coverage). Second, within each category, reports are sorted into deciles based on their predicted 12-month returns. The lines represent value-weighted abnormal returns of the stocks corresponding to these reports. The sample spans from 2005 to 2023.

TABLE 1: Summary Statistics of Analyst Reports

This table presents summary statistics for analyst reports covering S&P 1500 firms from 2000 to 2023. For each year, it reports the number of research reports, distinct brokerage firms, unique sell-side analysts, and the average report characteristics (number of pages, sentences, and tokens per report).

Year	Reports	Brokers	Analysts	Pages	Sentences	Tokens
2000	14224	49	555	5	56	1244
2001	22309	50	639	5	57	1256
2002	27555	55	791	5	57	1172
2003	28323	67	859	6	64	1257
2004	33431	70	969	6	66	1210
2005	39365	71	967	6	64	1167
2006	42347	70	951	6	64	1147
2007	43742	69	970	7	69	1217
2008	44676	72	1070	7	73	1288
2009	39425	83	1104	7	73	1281
2010	28998	86	1028	7	75	1312
2011	59367	86	1279	7	76	1263
2012	66576	85	1269	8	76	1205
2013	66960	78	1202	7	72	1143
2014	65748	73	1176	7	69	1086
2015	66699	75	1137	8	71	1138
2016	66891	71	1117	8	74	1170
2017	66846	67	1010	9	78	1234
2018	64310	62	930	9	82	1253
2019	64271	60	972	9	83	1287
2020	64567	63	964	10	85	1297
2021	56793	61	971	9	87	1300
2022	57836	60	981	10	92	1399
2023	63071	59	972	10	91	1353

TABLE 2: Return Predictability

This table reports the results from regressions of the specification $RET_{i,12m} = \alpha_t + \beta' x_{i,t} + \varepsilon_{i,t+12}$, where $RET_{i,12m}$ is the next 12 months' return of stock i. $x_{i,t}$ represents analyst information from each report. $\widehat{RET_{12m}}$ is the predicted next 12 months' stock return from Ridge regressions. REC_{rev} denotes recommendation revision, calculated as the current report's recommendation minus the last recommendation in I/B/E/S issued by the same analyst for the same stock. EF_{rev} refers to earnings forecast revision, calculated as the current report's EPS forecast minus the last EPS forecast in I/B/E/S issued by the same analyst for the same stock, scaled by the stock price 50 days before the report release. TP_{rev} represents target price revision, calculated as the current report's target price minus the last target price in I/B/E/S issued by the same analyst for the same stock, scaled by the stock price 50 days before the report release. I include analyst-year-month fixed effect and industry-year-month fixed effect. Standard errors are two-way clustered by firm and year-month. The t-statistics are shown in parentheses, with ***, **, and * indicating statistical significance at the 1%, 5%, and 10% level, respectively. The sample spans from 2005 to 2023.

	(1) RET_{12m}	(2) RET_{12m}	(3) RET_{12m}	(4) RET_{12m}	(5) RET_{12m}	(6) RET_{12m}
$\widehat{RET_{12m}}$	0.087***	0.060***	0.059***	0.062***	0.065***	0.071***
	(3.57)	(4.37)	(4.27)	(4.45)	(4.31)	(4.79)
REC_{rev}			-0.000			-0.002
			(-0.07)			(-0.17)
EF_{rev}				-0.300		0.003
				(-0.77)		(0.01)
TP_{rev}					0.011	0.016
					(0.47)	(0.78)
N	837233	488151	480523	344742	341049	247872
YM FE	Yes	No	No	No	No	No
Ana × YM FE	No	Yes	Yes	Yes	Yes	Yes
Ind × YM FE	No	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.162	0.491	0.489	0.573	0.475	0.579

TABLE 3: Portfolio Statistics

This table presents the performance of value-weighted decile portfolios sorted by Ridge models' return forecasts, which are based on the past {LB} months of analyst report information. The models predict 12-month ahead stock returns. For each portfolio, I report several performance measures: excess return mean and standard deviation, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the corresponding t-statistics for α. Portfolios are rebalanced monthly using the average of the most recent 12 months of out-of-sample predictions. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) and a short position in the lowest decile (Low). The sample period extends from January 2005 to June 2024.

	LB = 9 months					LB = 12 months				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.53	4.68	0.39	-0.09	-0.45	0.52	4.90	0.37	-0.13	-0.68
2	0.74	4.76	0.54	0.06	0.56	0.63	4.53	0.48	-0.05	-0.44
3	0.65	4.34	0.52	-0.11	-1.38	0.66	4.53	0.50	-0.15	-1.56
4	0.71	4.54	0.54	-0.05	-0.53	0.77	4.45	0.60	0.04	0.39
5	0.75	4.71	0.55	-0.05	-0.60	0.75	4.53	0.58	-0.04	-0.55
6	0.72	4.53	0.55	0.00	0.02	0.84	4.76	0.61	0.07	0.62
7	0.81	4.89	0.57	0.04	0.30	0.73	4.75	0.53	-0.06	-0.58
8	0.95	4.68	0.70	0.13	1.61	1.07	4.83	0.76	0.25	2.07
9	0.94	5.32	0.61	0.06	0.34	0.78	5.32	0.51	-0.08	-0.61
High (H)	1.52	6.38	0.82	0.50	2.66	1.56	6.29	0.86	0.55	2.73
H-L	0.98	5.20	0.66	0.58	2.48	1.04	5.21	0.69	0.68	2.64
	LB = 18 months					LB = 24 months				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.56	4.93	0.40	-0.12	-0.54	0.53	4.98	0.37	-0.16	-0.73
2	0.56	4.56	0.43	-0.11	-0.99	0.49	4.62	0.37	-0.21	-1.98
3	0.70	4.42	0.55	-0.05	-0.74	0.71	4.44	0.55	-0.01	-0.08
4	0.69	4.53	0.53	-0.07	-0.84	0.76	4.28	0.62	0.03	0.30
5	0.77	4.43	0.60	0.04	0.36	0.58	4.75	0.42	-0.19	-1.36
6	0.84	4.69	0.62	-0.00	-0.06	0.97	4.47	0.75	0.14	1.46
7	0.69	4.63	0.52	-0.09	-0.63	0.78	4.58	0.59	-0.02	-0.13
8	1.09	4.82	0.79	0.26	2.02	0.92	5.11	0.63	0.10	0.97
9	1.12	5.56	0.70	0.22	1.42	0.90	5.17	0.60	-0.00	-0.04
High (H)	1.43	6.21	0.80	0.40	2.20	1.70	6.48	0.91	0.63	2.67
H-L	0.87	4.84	0.62	0.52	1.99	1.16	5.13	0.79	0.79	2.48

TABLE 4: Sentiment Analyses

This table presents deciles of analyst reports sorted by Ridge models' return forecasts. For each decile, I report the predicted and realized 12-month ahead returns, and four sentiment measures. $CAR_{[0,+1]}$ is two-day DGTW characteristic-adjusted buy-and-hold abnormal returns following report release dates (in %). REC_{rev} denotes recommendation revision, calculated as the current report's recommendation minus the last recommendation in I/B/E/S issued by the same analyst for the same stock. $Tone_{Head}$ represents the sentiment score of report headlines measured using a fine-tuned BERT model. $Tone_{Body}$ represents the average sentiment score of report body content measured using a fine-tuned BERT model. *** denotes statistical significance at the 1% level.

Decile	$\widehat{RET_{12m}}$	RET_{12m}	$CAR_{[0,+1]}$	REC_{rev}	$Tone_{Head}$	$Tone_{Body}$
Low (L)	-0.19	0.12	0.16	0.03	0.06	0.14
2	-0.05	0.11	0.11	0.02	0.06	0.10
3	0.01	0.12	0.13	0.01	0.04	0.06
4	0.06	0.12	0.10	-0.00	0.03	0.03
5	0.10	0.12	0.09	-0.01	0.01	0.00
6	0.14	0.12	0.07	0.00	-0.00	-0.01
7	0.18	0.13	0.04	-0.00	-0.02	-0.05
8	0.22	0.14	0.02	-0.01	-0.04	-0.07
9	0.28	0.14	0.00	-0.01	-0.05	-0.10
High (H)	0.43	0.17	-0.09	-0.02	-0.08	-0.09
H-L	0.62***	0.05***	-0.25***	-0.05***	-0.14***	-0.23***

TABLE 5: Sources of Investment Value

This table presents an analysis of analyst report topics across five categories, examining both their distribution and investment value contribution. The categories cover Financial Analysis, Company and Industry Overview, Strategic Outlook, Risk and Governance, and Additional Content. Panel A displays aggregate statistics on topic distribution, showing the total number of sentences (in millions) and tokens (in billions) per category, along with their respective percentages of the total content. Panel B quantifies each category's contribution to portfolio performance through Shapley value decomposition, specifically examining both the Sharpe ratios and returns of the value-weighted 'H-L' portfolio.

Panel A: Topic Distribution				
Category	#Sentences	#Tokens	%Sentences	%Tokens
Financial Analysis	19.14	0.61	36.56	36.57
Company and Industry Overview	14.94	0.48	28.53	28.75
Strategic Outlook	7.92	0.26	15.13	15.61
Risk and Governance	7.40	0.24	14.14	14.35
Additional Content	2.95	0.08	5.63	4.71
Panel B: Shapley Value Decomposition				
Category	SHAP(SR)	SHAP(Ret)	%SHAP(SR)	%SHAP(Ret)
Strategic Outlook	0.24	0.28	41.34	31.43
Company and Industry Overview	0.16	0.24	27.61	26.92
Risk and Governance	0.06	0.19	11.21	21.36
Financial Analysis	0.09	0.18	16.39	19.53
Additional Content	0.02	0.01	3.44	0.77

TABLE 6: Sources of Investment Value in Strategic Outlook

This table analyzes analysts' strategic outlook discussions across three dimensions, quantifying their contribution to investment value through Shapley analysis. Panel A examines the temporal dimension, decomposing sentences into long-term, short-term, and combined timeframe categories. Panel B classifies sentences by sentiment orientation (negative, neutral, or positive), while Panel C categorizes content based on discussion focus (risk versus fundamental analysis). For each dimension and category, I report both absolute Shapley values (SHAP(SR) and SHAP(Ret)) and their relative percentages, measuring their contributions to portfolio Sharpe ratio and returns, respectively.

Dimension	SHAP(SR)	SHAP(Ret)	%SHAP(SR)	%SHAP(Ret)
Panel A: Timeframe				
Long-term	0.39	0.52	49.51	50.20
Short-term	0.21	0.28	26.69	26.70
Both	0.19	0.24	23.80	23.10
Panel B: Sentiment				
Negative	0.19	0.21	24.34	20.23
Neutral	0.20	0.30	25.72	28.91
Positive	0.39	0.53	49.94	50.86
Panel C: Focus				
Risk	0.10	0.13	13.00	12.25
Fundamental	0.68	0.92	87.00	87.75

TABLE 7: Category Portfolio Statistics

This table reports the value-weighted decile portfolio performances sorted by Ridge models' forecast using the past 12 months' analyst report category-specific content. Categories include strategic outlook, company and industry overview, financial analysis, and risk and governance. The prediction target is the next 12 months' return of the stock. For each portfolio, I report several performance measures: excess return mean and standard deviation, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the corresponding t-statistics for α. Portfolios are rebalanced monthly using the average of the most recent 12 months of out-of-sample predictions. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) and a short position in the lowest decile (Low). The sample period extends from January 2005 to June 2024.

	Strategic Outlook					Company and Industry Overview				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.46	4.32	0.37	-0.29	-2.18	0.76	5.69	0.46	0.08	0.38
2	0.81	4.26	0.66	0.10	1.03	0.56	4.57	0.43	-0.14	-1.15
3	0.73	4.85	0.52	-0.00	-0.03	0.61	4.28	0.49	-0.09	-1.25
4	0.71	4.62	0.54	-0.10	-1.04	0.59	4.37	0.47	-0.15	-2.08
5	0.58	4.94	0.40	-0.19	-1.43	0.75	4.54	0.57	-0.02	-0.20
6	0.68	4.84	0.49	-0.06	-0.69	0.79	4.83	0.56	0.01	0.14
7	0.79	4.92	0.55	-0.00	-0.03	0.97	4.67	0.72	0.17	1.50
8	0.87	5.07	0.59	0.08	0.63	0.81	4.74	0.59	-0.06	-0.46
9	1.08	4.92	0.76	0.25	1.74	0.94	4.79	0.68	0.08	0.85
High (H)	1.87	6.65	0.97	0.72	2.85	1.33	5.97	0.77	0.29	1.71
H-L	1.41	5.25	0.93	1.00	3.18	0.57	5.23	0.38	0.21	0.92
	Financial Analysis					Risk and Governance				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.82	5.03	0.56	0.14	0.66	0.85	4.97	0.59	0.20	1.01
2	0.53	4.26	0.43	-0.17	-1.71	0.63	4.50	0.48	-0.01	-0.10
3	0.52	4.60	0.39	-0.24	-2.38	0.49	4.49	0.38	-0.23	-2.63
4	0.73	4.65	0.54	0.00	0.03	0.77	4.69	0.57	0.02	0.23
5	0.63	4.52	0.48	-0.12	-1.49	0.70	4.65	0.52	-0.05	-0.59
6	0.93	4.62	0.70	0.19	1.59	0.82	4.55	0.63	0.04	0.41
7	0.81	4.80	0.58	-0.03	-0.28	0.72	4.77	0.52	-0.11	-1.34
8	0.75	4.82	0.54	-0.08	-0.80	0.94	4.82	0.68	0.09	0.84
9	1.07	4.99	0.74	0.17	1.33	1.06	4.89	0.75	0.15	1.34
High (H)	1.39	6.09	0.79	0.34	1.71	1.37	5.90	0.81	0.32	1.97
H-L	0.58	4.81	0.41	0.19	0.74	0.52	4.90	0.37	0.12	0.44

TABLE 8: Incremental Investment Value

This table examines the incremental investment value of analyst reports compared to existing factors. The analysis includes: full report content (RP), strategic outlook sections (SO), 18 analyst-based factors from Chen and Zimmermann (2022) (ANA), and 92 fundamental-based factors from Gu et al. (2020) (ANOM). For each strategy and combination, I report mean returns, Sharpe ratio (SR), and information ratio (IR), with corresponding t-statistics in parentheses. I also report the alphas of long-short portfolios relative to four factor models: Fama and French (2015) five factors (α_{F5}), Fama and French (2015) five factors plus momentum (α_{F6}), Hou et al. (2015) factors (α_{HXZ}), and Daniel et al. (2020) factors (α_{DHS}). Panel A presents individual factor performances. Panel B evaluates combinations of full report content with existing factors, while Panel C analyzes combinations of strategic outlook content with existing factors. The sample spans from January 2005 to December 2023.

	Mean	SR	α_{F5}	α_{F6}	α_{HXZ}	α_{DHS}	IR
Panel A: Factor Performances							
RP	1.07	0.71	0.73	0.75	0.86	1.21	-
	(3.07)	(3.04)	(2.73)	(2.86)	(3.52)	(3.90)	-
SO	1.30	0.87	0.95	0.96	1.04	1.46	-
	(3.79)	(3.74)	(3.03)	(3.01)	(3.50)	(3.89)	-
ANA	0.27	0.48	0.28	0.25	0.22	0.16	-
	(2.08)	(2.07)	(2.80)	(3.01)	(2.08)	(1.26)	-
ANOM	1.34	1.03	1.40	1.31	1.28	1.15	-
	(4.47)	(4.38)	(5.05)	(4.65)	(4.43)	(3.84)	-
ANA + ANOM	0.80	1.02	0.84	0.78	0.75	0.66	-
	(4.44)	(4.35)	(5.30)	(5.16)	(4.36)	(3.58)	-
Panel B: Report versus Factors							
RP + ANA	0.67	0.90	0.51	0.50	0.54	0.68	0.73
	(3.91)	(3.85)	(3.94)	(3.92)	(4.81)	(5.34)	(2.81)
RP + ANOM	1.21	1.55	1.06	1.03	1.07	1.18	1.16
	(6.75)	(6.44)	(7.79)	(7.37)	(7.13)	(8.01)	(3.91)
RP + ANA + ANOM	0.89	1.60	0.80	0.77	0.78	0.84	1.23
	(6.97)	(6.62)	(8.35)	(8.17)	(7.32)	(8.66)	(4.05)
Panel B: Strategic Outlook versus Factors							
SO + ANA	0.78	1.07	0.62	0.61	0.63	0.81	0.96
	(4.68)	(4.57)	(4.34)	(4.26)	(5.30)	(4.95)	(3.45)
SO + ANOM	1.32	1.69	1.17	1.14	1.16	1.30	1.34
	(7.37)	(6.97)	(7.92)	(7.79)	(8.21)	(7.92)	(4.24)
SO + ANA + ANOM	0.97	1.74	0.88	0.84	0.84	0.92	1.41
	(7.57)	(7.14)	(9.05)	(9.55)	(9.51)	(8.64)	(4.35)

TABLE 9: Post-Knowledge Cutoff Portfolio Performance

This table presents value-weighted decile portfolio performance during post-language model training knowledge cutoff periods, comparing four different models: BERT, RoBERTa, LLaMA2-13B, and LLaMA3-8B. Each portfolio is constructed using the corresponding model's embeddings. For each portfolio, I report three key measures: mean and standard deviation of returns, and the Sharpe ratio (SR). The analysis spans different time periods based on each model's knowledge cutoff: BERT (January 2019 to June 2024), RoBERTa (March 2019 to June 2024), LLaMA2 (October 2022 to June 2024), and LLaMA3 (April 2023 to June 2024). Panel A evaluates portfolios constructed using full report content, while Panel B focuses specifically on strategic outlook sections. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) while shorting the lowest decile (Low).

	BERT			RoBERTa			LLaMA2			LLaMA3		
	Mean	SD	SR	Mean	SD	SR	Mean	SD	SR	Mean	SD	SR
Panel A: Full Report												
Low (L)	0.94	5.30	0.61	0.63	5.16	0.42	1.67	5.03	1.15	1.13	4.06	0.96
2	0.95	5.53	0.60	0.78	5.71	0.48	0.88	3.65	0.83	0.63	3.88	0.57
3	1.12	5.31	0.73	1.02	5.02	0.71	1.01	4.61	0.76	1.88	3.56	1.84
4	1.00	4.93	0.70	0.78	4.97	0.54	1.25	4.52	0.96	0.93	4.94	0.65
5	1.30	5.02	0.89	0.76	5.36	0.49	1.15	4.91	0.81	2.32	5.48	1.47
6	1.27	5.17	0.85	0.90	5.52	0.57	1.42	5.02	0.98	2.14	4.48	1.65
7	1.09	5.25	0.72	1.25	5.35	0.81	1.66	4.74	1.22	1.00	5.53	0.63
8	1.15	5.47	0.73	1.53	5.66	0.93	1.56	5.22	1.04	0.75	5.15	0.51
9	1.53	5.92	0.89	1.89	6.11	1.07	3.10	5.09	2.11	1.44	5.75	0.87
High (H)	2.70	7.41	1.26	2.65	7.93	1.16	3.26	7.08	1.60	3.83	7.71	1.72
H-L	1.76	6.06	1.01	2.01	6.67	1.05	1.59	7.39	0.74	2.70	7.30	1.28
Panel B: Strategic Outlook												
Low (L)	0.82	5.69	0.50	0.97	5.36	0.63	1.08	4.23	0.89	0.01	3.48	0.01
2	1.10	4.92	0.78	0.77	5.04	0.53	1.21	4.37	0.96	1.19	3.77	1.09
3	0.92	5.16	0.61	0.98	5.22	0.65	1.15	4.51	0.88	1.05	4.19	0.87
4	1.39	5.04	0.95	0.99	4.96	0.69	1.39	4.49	1.07	2.05	4.97	1.43
5	1.20	5.15	0.81	0.68	5.37	0.44	1.82	5.01	1.26	1.69	3.60	1.63
6	0.66	5.52	0.42	0.95	5.34	0.62	1.91	4.85	1.36	0.66	4.56	0.51
7	1.14	5.48	0.72	1.35	5.04	0.93	1.45	4.73	1.06	0.67	5.09	0.46
8	1.53	5.77	0.92	1.34	5.56	0.84	2.35	5.64	1.44	1.77	4.81	1.28
9	2.17	6.67	1.13	1.94	7.33	0.92	2.30	6.14	1.30	2.26	5.13	1.53
High (H)	3.19	8.09	1.37	2.78	8.28	1.16	3.73	7.10	1.82	5.29	7.88	2.32
H-L	2.37	6.51	1.26	1.81	7.37	0.85	2.64	6.95	1.32	5.28	7.73	2.37

Appendix

A Tables and Figures



(1) Company and Industry Overview



(2) Financial Analysis



(3) Strategic Outlook



(4) Risk and Governance

FIGURE A1: Word Clouds of Topics

This figure presents word clouds for 4 categories frequently discussed in analyst reports. Each word cloud visually represents the most common terms associated with the topic, with word size indicating term frequency. The categories include Company and Industry Overview, Financial Analysis, Strategic Outlook and Risk and Governance. The "Additional Content" category is excluded from the visualization.

Table A1: Anomaly Variables

This table reports the market and firm-related fundamental factors considered in this paper. I follow Gu et al. (2020) to construct the dataset. Gu et al. (2020) provide a detailed description of the dataset in their appendix.

Acronym	Description	Acronym	Description
absacc	Absolute accruals	mom36m	36-month momentum
acc	Working capital accruals	mom6m	6-month momentum
aeavol	Abnormal earnings announcement volume	ms	Financial statement score
age	# years since first Compustat coverage	mvel1	Size
agr	Asset growth	mve_ia	Industry-adjusted size
baspread	Bid-ask spread	niincr	Number of earnings increases
beta	Beta	operprof	Operating profitability
betasq	Beta squared	orgcap	Organizational capital
bm	Book-to-market	pchcapx_ia	Industry adjusted % change in capital expenditures
bm_ia	Industry-adjusted book to market	pchcurrat	% change in current ratio
cash	Cash holdings	pchdepr	% change in depreciation
cashdebt	Cash flow to debt	pchgm_pchsale	% change in gross margin - % change in sales
cashpr	Cash productivity	pchquick	% change in quick ratio
cfp	Cash flow to price ratio	pchsale_pchinvt	% change in sales - % change in inventory
cfp_ia	Industry-adjusted cash flow to price ratio	pchsale_pchrect	% change in sales - % change in A/R
chatoia	Industry-adjusted change in asset turnover	pchsale_pchxsga	% change in sales - % change in SG&A
chcsho	Change in shares outstanding	pchsaleinv	% change in sales-to-inventory
chempia	Industry-adjusted change in employees	pctacc	Percent accruals
chinv	Change in inventory	pricedelay	Price delay
chmom	Change in 6-month momentum	ps	Financial statements score
chpmia	Industry-adjusted change in profit margin	quick	Quick ratio
chtx	Change in tax expense	rd	R&D increase
cinvest	Corporate investment	rd_mve	R&D to market capitalization
convind	Convertible debt indicator	rd_sale	R&D to sales
currat	Current ratio	realestate	Real estate holdings
depr	Depreciation / PP&E	retvol	Return volatility
divi	Dividend initiation	roaq	Return on assets
divo	Dividend omission	roavol	Earnings volatility
dolvol	Dollar trading volume	roeq	Return on equity
dy	Dividend to price	roic	Return on invested capital
ear	Earnings announcement return	rsup	Revenue surprise
egr	Growth in common shareholder equity	salecash	Sales to cash
ep	Earnings to price	saleinv	Sales to inventory
gma	Gross profitability	salerec	Sales to receivables
grCAPX	Growth in capital expenditures	secured	Secured debt
grltnoa	Growth in long term net operating assets	securedind	Secured debt indicator
herf	Industry sales concentration	sgr	Sales growth
hire	Employee growth rate	sin	Sin stocks
idiovol	Idiosyncratic return volatility	sp	Sales to price
ill	Illiquidity	std_dolvol	Volatility of liquidity (dollar trading volume)
indmom	Industry momentum	std_turn	Volatility of liquidity (share turnover)
invest	Capital expenditures and inventory	stdacc	Accrual volatility
lev	Leverage	stdcf	Cash flow volatility
lgr	Growth in long-term debt	tang	Debt capacity/firm tangibility
maxret	Maximum daily return	tb	Tax income to book income
mom12m	12-month momentum	turn	Share turnover
mom1m	1-month momentum	zerotrade	Zero trading days

Table A2: Analyst Variables

This table reports the analyst factors considered in this paper. I compile the data from Chen and Zimmermann (2022) website.

Acronym	Journal (Publish Year)	Description
AnalystRevision	FAJ (1984)	EPS forecast revision
AnalystValue	JAE (1998)	Analyst Value
AOP	JAE (1998)	Analyst Optimism
ChangeInRecommendation	JF (2004)	Change in recommendation
ChForecastAccrual	RAS (2004)	Change in Forecast and Accrual
ChNAnalyst	ROF (2008)	Decline in Analyst Coverage
ConsRecomm	JF (2001)	Consensus Recommendation
CredRatDG	JF (2001)	Credit Rating Downgrade
DownRecomm	JF (2001)	Down forecast EPS
EarningsForecastDisparity	JFE (2011)	Long-vs-short EPS forecasts
ExclExp	RAS (2003)	Excluded Expenses
FEPS	WP (2006)	Analyst earnings per share
fgr5yrLag	JF (1996)	Long-term EPS forecast
ForecastDispersion	JF (2002)	EPS Forecast Dispersion
Recomm_ShortInterest	AR (2011)	Analyst Recommendations and Short-Interest
REV6	JF (1996)	Earnings forecast revisions
sfe	AR (2001)	Earnings Forecast to price
UpRecomm	JF (2001)	Up Forecast

TABLE A3: Topic Categories and Descriptions

Topic	Desriptions
Company and Industry Overview	
Executive Summary	Provides a high-level overview of the report's key findings and conclusions; includes a brief description of the company, its industry, and the purpose of the report; highlights the most important points from the analysis, such as the company's financial performance, competitive position, and growth prospects.
Company Overview	Offers a comprehensive description of the company, including its history, business model, and key products or services; discusses the company's organizational structure, management team, and corporate governance; analyzes the company's mission, vision, and strategic objectives.
Industry Analysis	Provides an in-depth analysis of the industry in which the company operates; includes information on market size, growth trends, and key drivers; discusses the regulatory environment, technological advancements, and other external factors affecting the industry; analyzes the industry's competitive dynamics and the company's position within the industry.
Competitive Landscape	Identifies the company's main competitors and their market share; compares the company's products, services, and pricing strategies with those of its competitors; analyzes the strengths and weaknesses of the company and its competitors; discusses potential new entrants and substitutes that could disrupt the competitive landscape.
Business Segments	Provides a detailed analysis of the company's various business segments or divisions; discusses the financial performance, growth prospects, and challenges of each segment; analyzes the contribution of each segment to the company's overall revenue and profitability.
Growth Strategies	Discusses the company's strategies for driving future growth, such as organic growth initiatives, product innovations, and geographic expansions; analyzes the company's mergers and acquisitions (M&A) strategy and potential targets; examines the company's investments in research and development (R&D) and marketing.
Financial Analysis	
Income Statement Analysis	Analyzes the company's revenue, expenses, and profitability.
Balance Sheet Analysis	Examines the company's assets, liabilities, and shareholders' equity.
Cash Flow Analysis	Analyzes the company's cash inflows and outflows to evaluate liquidity.
Financial Ratios	Discusses key ratios like profitability, liquidity, and solvency ratios.
Strategic Outlook	
Investment Thesis	Summarizes the key reasons for investing (or not investing) in the company's shares; discusses the potential catalysts and risks that could impact the company's valuation and stock price performance; provides a target price or price range for the company's shares based on the valuation analyses and investment thesis.
Valuation	Estimates the intrinsic value of the company's shares using various valuation methodologies, such as discounted cash flow (DCF) analysis, relative valuation multiples, and sum-of-the-parts analysis; compares the company's valuation with that of its peers and historical benchmarks; discusses the key assumptions and sensitivities underlying the valuation analyses.
Risk and Governance	
Risk Factors	Identifies and analyzes the key risks facing the company, such as market risks, operational risks, financial risks, and legal/regulatory risks; discusses the potential impact of these risks on the company's financial performance and growth prospects; examines the company's risk management strategies and mitigation measures.

Table A3 – continued from previous page

Topic	**Desriptions**
Management and Governance	Provides an overview of the company's management team, including their experience, expertise, and track record; analyzes the company's corporate governance practices, such as board composition, executive compensation, and shareholder rights; discusses the company's succession planning and key person risks.
ESG	Analyzes the company's performance and initiatives related to environmental sustainability, social responsibility, and corporate governance; discusses the potential impact of ESG factors on the company's reputation, risk profile, and financial performance; examines the company's compliance with ESG regulations and industry standards.
Additional Content	
Appendices and Disclosures	Includes additional supporting materials, such as financial statements, ratio calculations, and detailed segment data; provides important disclosures, such as the analyst's rating system, potential conflicts of interest, and disclaimers; discusses the limitations and uncertainties of the analysis and the need for further due diligence by investors.
None of the Above	Covers any content that does not fall into the specified topics.

TABLE A4: Sentiment Portfolio Statistics

This table reports the value-weighted decile portfolio performances sorted by the average of four measures of the past 12 months' analyst reports. Contemporaneous Return denotes two-day DGTW characteristic-adjusted buy-and-hold abnormal returns following report release dates. Recommendation Revision denotes recommendation revision, calculated as the current report's recommendation minus the last recommendation in I/B/E/S issued by the same analyst for the same stock. Headline Sentiment represents the sentiment score of report headlines measured using a fine-tuned BERT model. Body Sentiment represents the average sentiment score of report body content measured using a fine-tuned BERT model. For each portfolio, I report several performance measures: excess return mean and standard deviation, Sharpe ratio, α relative to Fama and French (2015) five factors with momentum (six factors in total), and the corresponding t-statistics for α. Portfolios are rebalanced monthly. The 'H-L' row represents a long-short strategy that takes a long position in the highest decile (High) and a short position in the lowest decile (Low). The sample period extends from January 2005 to June 2024.

	Contemporaneous Return					Recommendation Revision				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	1.10	6.66	0.57	0.26	1.42	0.84	4.70	0.62	0.01	0.11
2	0.93	5.99	0.54	0.14	0.86	0.87	4.81	0.62	0.09	0.95
3	0.71	5.19	0.47	-0.01	-0.10	0.88	4.61	0.66	0.10	0.75
4	0.82	4.45	0.64	0.11	1.21	0.46	4.30	0.37	-0.33	-3.63
5	0.80	4.53	0.61	-0.08	-0.93	0.54	4.90	0.38	-0.20	-1.73
6	0.79	4.41	0.62	0.00	0.03	0.69	5.31	0.45	-0.12	-1.07
7	0.84	4.56	0.64	0.00	0.05	0.81	5.61	0.50	-0.04	-0.29
8	0.94	4.78	0.68	0.08	0.75	1.02	4.90	0.72	0.15	1.32
9	1.12	5.43	0.72	0.27	2.02	0.77	4.54	0.59	0.01	0.08
High (H)	0.78	6.06	0.44	-0.10	-0.72	1.11	4.73	0.81	0.25	1.94
H-L	-0.33	4.20	-0.27	-0.36	-1.57	0.27	2.68	0.35	0.24	1.56
	Headline Sentiment					Body Sentiment				
	Mean	SD	SR	α	t_α	Mean	SD	SR	α	t_α
Low (L)	0.48	5.57	0.30	-0.23	-2.02	0.98	5.75	0.59	0.22	1.25
2	0.63	5.60	0.39	-0.09	-0.97	0.48	5.47	0.30	-0.17	-1.83
3	0.81	5.16	0.55	0.07	0.60	0.84	5.18	0.56	0.14	1.41
4	0.98	4.58	0.74	0.25	2.65	0.65	4.80	0.47	-0.09	-0.88
5	1.02	4.57	0.78	0.18	1.84	1.02	4.96	0.71	0.18	1.63
6	0.80	4.55	0.61	-0.02	-0.17	0.87	4.71	0.64	0.05	0.41
7	0.78	4.77	0.57	-0.08	-0.97	0.81	4.75	0.59	-0.10	-1.35
8	0.67	4.55	0.51	-0.19	-1.68	0.82	4.59	0.62	-0.03	-0.31
9	0.85	4.93	0.60	-0.03	-0.19	0.90	4.50	0.70	0.03	0.29
High (H)	0.74	5.11	0.50	-0.15	-1.50	0.82	4.56	0.63	-0.01	-0.16
H-L	0.27	3.31	0.28	0.08	0.58	-0.16	4.17	-0.13	-0.24	-0.97