What Makes a Star Analyst?

Evidence on Industry Expertise from 1.2 Million Analyst Reports

Alejandro Lopez-Lira, Yuehua Tang, Yuan Wang, Mingyin Zhu*
University of Florida

This version: March 27, 2026

Abstract

We develop novel text-based measures of sell-side analyst industry knowledge and report quality by applying a large language model evaluation pipeline to 1.22 million industry reports from 40 major brokerage houses over 2012 to 2024. Each report is scored across 18 dimensions capturing both general research quality and specific forms of industry expertise. We validate our measures by showing that they are substantially more predictive of *Institutional Investor* All-Star status than traditional forecast accuracy metrics, with estimated coefficients roughly seven times larger. We further show that analysts with higher knowledge and quality scores produce more accurate earnings forecasts, and that their forecast revisions and recommendation changes elicit stronger market reactions. At the firm level, coverage by more knowledgeable analysts is associated with lower information asymmetry, measured through liquidity and price informativeness proxies, and a lower implied cost of capital. In short, analyst industry knowledge is significantly associated with analysts' star status, research output, investor responses, and the information environment of covered firms.

Keywords: Large Language Models, Sell-Side Analysts, Industry Knowledge, Analyst Reports, Information Environment
JEL Classification: C45, G12, G14, G17, G24

*We have benefited from discussions with Jay Ritter and Baolian Wang. We thank An-Ping Lin for sharing the All-Star analyst status data and Baolian Wang for sharing the I/B/E/S analyst code to full name link table. Contact information: Alejandro Lopez-Lira: alejandro.lopez-lira@warrington.ufl.edu; Yuehua Tang: yuehua.tang@warrington.ufl.edu; Yuan Wang: yuan.wang@warrington.ufl.edu; Mingyin Zhu: mingyin.zhu@warrington.ufl.edu. There are no financial disclosures or conflicts of interest.

1 Introduction

What makes a star analyst? A large literature has studied sell-side equity analysts through the lens of their quantitative outputs, including the accuracy of earnings forecasts, the profitability of stock recommendations, and the informativeness of target prices. Yet both practitioners and the consumers of analyst research point to a different answer. In an influential survey of 365 analysts, Brown et al. (2015) find that industry knowledge is the single most useful input to both earnings forecasts and stock recommendations. On the buy side, institutional investors consistently rank industry knowledge as the most important trait of sell-side analysts, above stock selection, earnings estimates, and written reports (Bradshaw (2011), based on *Institutional Investor* survey data). More broadly, Bradshaw, Ertimur, and O'Brien (2017) argue that the literature's focus on quantitative outputs is too narrow and call for research on broader dimensions of analyst work, including the qualitative content of research reports. The message is clear: what distinguishes star analysts is not the precision of their numbers, but the depth of their industry knowledge.

Yet this central dimension of analyst ability has never been measured directly. Prior work infers expertise from outcomes such as forecast accuracy, recommendation profitability, and career advancement, but cannot observe the knowledge content of analyst output. Kadan et al. (2012) measure industry expertise indirectly through relative forecast accuracy across industries, but outcome-based proxies cannot distinguish genuine knowledge from luck or information access. More fundamentally, they tell us little about what analysts actually know: how deeply they understand supply chains, technology shifts, margin dynamics, or the competitive forces shaping the industries they cover. This knowledge, embedded in the textual content of analyst research, has remained unquantified and beyond the reach of archival research. In this paper, we fill that gap.

We develop novel text-based measures of analyst industry knowledge and report quality using a large language model (LLM) evaluation pipeline applied to 1.22 million sell-side industry reports from 40 major brokerage houses over the period 2012 to 2024. Each report is scored against a structured 18-dimension rubric covering both general quality features of research (such as originality, analytical depth, and quantitative rigor) and specific forms of industry expertise (such as knowledge of supply chains, margins, customers, and industry technology). Report-level scores are

then aggregated to the analyst level, producing time-varying measures of industry knowledge and report quality that can be linked to analyst outcomes and the information environment of covered firms.

We begin by asking what drives analyst industry knowledge. Both knowledge and quality scores are highly persistent over time, with lagged coefficients ranging from 0.41 to 0.56. Analysts who produce more reports, write longer reports, cover more firms, and cover larger firms tend to score higher. Covering a greater number of industries is positively associated with knowledge but not quality, consistent with the notion that breadth of industry exposure enhances knowledge specifically. Career length is positively associated with both measures in isolation but becomes insignificant once report characteristics are controlled for, suggesting that the effect of experience operates primarily through report production.

We then examine what predicts All-Star status, the most prominent form of buy-side recognition in the industry. In logit regressions of *Institutional Investor* All-Star ratings on our LLM-based measures and standard analyst controls, knowledge and quality are strongly and significantly associated with All-Star recognition. When entered individually, the estimated coefficients on knowledge (0.612) and quality (0.642) are roughly seven times larger in magnitude than the coefficient on average forecast error (-0.084). Once analyst and brokerage controls are included, average forecast error becomes statistically insignificant, while knowledge and quality remain strongly significant. These results are robust to the inclusion of brokerage, industry, and year fixed effects. The finding that forecast accuracy loses its explanatory power once knowledge and quality are accounted for suggests that the qualitative dimensions captured by our measures are substantially more important for buy-side recognition than traditional numerical performance metrics.

Moving beyond recognition, we test whether our measures are associated with the accuracy of analysts' earnings forecasts. If industry knowledge reflects genuine expertise, analysts with higher scores should produce more accurate forecasts. We find that both knowledge and quality are negatively and significantly associated with absolute forecast error, even after including analyst-firm fixed effects. By contrast, lagged average forecast error is insignificant across all specifications, suggesting that our text-based measures capture dimensions of analyst ability that are not subsumed by past numerical performance.

We next investigate whether investors recognize and price analyst ability. If higher-scoring analysts produce more informative research, the market should react more strongly to their output. We test this prediction for both earnings forecast revisions and stock recommendation changes. For forecast revisions, we find that the interaction between the revision and a high-knowledge (high-quality) indicator is positive and significant, indicating that the stock market reacts more strongly to revisions from analysts with above-median scores. For recommendation changes, we estimate separate regressions for upgrades and downgrades. Both knowledge and quality are positively associated with abnormal returns around upgrades and negatively associated with abnormal returns around downgrades. The effects are notably larger for downgrades than for upgrades, consistent with prior evidence that downgrades are perceived as more credible given analysts' well-documented conflicts of interest.

Finally, we ask whether firms themselves benefit from being covered by analysts with greater industry knowledge. We examine three sets of outcomes: liquidity, price informativeness, and the implied cost of capital. When analyst ability is measured using the top three analyst scores within the firm-year, firms covered by the most knowledgeable analysts exhibit narrower effective and quoted spreads, more balanced order flow, and lower price impact of trades. These associations are strongest among smaller firms, where the marginal informational contribution of a knowledgeable analyst is plausibly largest. We also find that firms covered by higher-scoring analysts face a lower implied cost of equity, with coefficients on knowledge of -0.028 for both cost-of-capital proxies developed by Easton (2004) and Ohlson and Juettner-Nauroth (2005). The consistency of these estimates across alternative cost of capital measures and across multiple dimensions of the information environment is consistent with analyst industry knowledge being associated with real benefits for covered firms through improved information production.

Our paper makes several contributions. First, we contribute to the literature on the "black box" of sell-side financial analysts by providing the first large-scale, text-based measure of industry knowledge derived directly from analyst research reports. Prior work has documented the importance of industry knowledge through surveys (Brown et al. 2015; Bradshaw 2011) and has studied the role of prior industry experience (Bradley, Gokkaya, and Liu 2017) and the informativeness of industry recommendations (Kadan et al. 2012), but no archival measure of the knowledge content

embedded in analyst output has been available. Our measure fills this gap and enables researchers to study a central dimension of analyst ability that has until now remained unquantified.

Second, we contribute to the growing literature that exploits the textual content of analyst research. Early work in this area examines the information content of analyst reports using hand-collected samples (Asquith, Mikhail, and Au 2005) or dictionary-based methods (Huang, Zang, and Zheng 2014). More recent studies use topic modeling (Huang et al. 2018) and other natural language processing techniques to study analyst information discovery and interpretation roles. We extend this literature by using an LLM-based evaluation pipeline that can assess the qualitative attributes of analyst research, such as analytical depth, originality, and industry-specific expertise, at a scale that was previously infeasible. Our approach also differs from prior textual studies in that we focus on analyst industry reports rather than firm-level research, thereby capturing a dimension of analyst output that has not been examined in the literature.

Third, we contribute to the literature on the real effects of analyst research. Prior work has shown that analyst coverage and effort allocation affect firms' information environments (Bowen, Chen, and Cheng 2008; Harford et al. 2019). We extend this evidence by showing that not only the quantity but also the quality and knowledge content of analyst research are associated with firm-level outcomes, including liquidity, price informativeness, and the cost of capital. The finding that these benefits are concentrated among the top analysts covering a firm suggests that the information environment is shaped disproportionately by a firm's most capable analysts rather than by the average quality of all analysts who follow the firm.

Related Literature

Our paper relates to several strands of the literature on sell-side equity analysts.

A large body of research studies analysts through the lens of their quantitative outputs. Early work examines the accuracy of earnings forecasts (O'Brien 1988, 1990; Clement 1999), the investment value of stock recommendations (Womack 1996), and the informativeness of target prices (Brav and Lehavy 2003). A subsequent stream investigates the determinants of these outputs (Stickel 1990; Mikhail, Walther, and Willis 1997) or uses them as proxies for analysts' underlying ability (Sinha, Brown, and Das 1997; Li 2005; Bradshaw, Brown, and Huang 2013). While these

studies have produced substantial insight into analyst behavior, they share a common limitation: the outputs they study represent only a narrow slice of what analysts produce and what their clients value.[1]

A second strand of research has begun to move beyond quantitative outputs by exploiting the textual content of analyst research. Asquith, Mikhail, and Au (2005) catalog the elements of analyst reports and construct a "strength-of-arguments" variable, finding that investors react to qualitative justifications rather than to earnings forecasts and recommendations alone. Huang, Zang, and Zheng (2014) use machine learning to extract textual opinions from analyst reports and show that text plays a greater role in predicting earnings growth than quantitative summary measures. Huang et al. (2018) employ topic modeling to distinguish between analysts' information discovery and interpretation roles. More recently, Décaire and Graham (2024) and K. Li et al. (2026) use textual analysis to study additional dimensions of analyst output at the firm level. We extend this literature in two directions: we focus on analyst industry reports rather than firm-level research, and we use an LLM-based evaluation pipeline that can assess qualitative attributes of research, such as analytical depth, originality, and industry-specific expertise, at a scale that dictionary-based or topic-modeling approaches cannot achieve.

A third strand examines what distinguishes successful analysts. Brown et al. (2015), surveying 365 sell-side analysts, find that industry knowledge is the single most useful input to both earnings forecasts and stock recommendations. Bradshaw (2011), drawing on *Institutional Investor* survey data from 1998 to 2005, documents that buy-side investors consistently rank industry knowledge as the most important trait of sell-side analysts, above stock selection and earnings estimates. Kadan et al. (2012) measure industry expertise indirectly by studying outputs (buy/sell/hold calls on industries) and asking whether they predict returns. Bradley, Gokkaya, and Liu (2017) show that analysts with prior industry experience produce more accurate forecasts, generate greater price impact, and are more likely to achieve All-Star status. Groysberg, Healy, and Maber (2011) find a strong positive association between *Institutional Investor* rankings and analyst compensation, reinforcing the importance of qualitative attributes for analysts' careers. Despite this evidence,

1. Bradshaw, Ertimur, and O'Brien (2017) highlight the "disconnect" between the analyst services that clients value and the outputs that are the subject of most academic research, noting that research that systematically examines the qualitative content of analyst reports remains relatively scarce.

industry knowledge has not been measured directly from analyst output in archival research. We open the black box by studying the content of analyst research using 1.22 million industry reports. Our LLM-based measures provide the first such measurement, enabling a systematic examination of the role that industry knowledge plays in analyst performance and capital market outcomes.

A fourth strand studies the effects of analyst coverage on firms' information environments. Piotroski and Roulstone (2004) show that analysts contribute to the incorporation of industry-level and market-level information into stock prices. Bowen, Chen, and Cheng (2008) find that the cost of raising equity capital decreases with both the quantity and quality of analyst coverage. Crawford, Roulstone, and So (2012) find that analyst coverage initiations reduce stock return synchronicity, consistent with analysts producing firm-specific information. Merkley, Michaely, and Pacelli (2017) examine whether the scope of the sell-side analyst industry affects the bias, accuracy, and information content of analyst reports. Loh and Stulz (2018) show that sell-side research is more valuable during periods of high uncertainty. Amiram, Owens, and Rozenbaum (2016) show that analyst forecasts reduce information asymmetry around announcements, as measured by changes in bid-ask spreads. Harford et al. (2019) show that analyst effort allocation affects firms' information asymmetry, bid-ask spreads, and the implied cost of capital. Our paper extends this evidence by showing that the knowledge content and quality of analyst research, not merely coverage or effort, are associated with liquidity, price informativeness, and the cost of capital at the firm level. Our measure captures the information that analysts actually produce and disseminate, providing a more direct link between analyst ability and firm-level outcomes than proxies based on career concerns or effort allocation.

2 Methodology

Our goal is to construct measures of analysts' industry knowledge and report quality directly from the textual content of sell-side industry reports. To do so, we use an LLM evaluation pipeline based on OpenAI's open-weight gpt-oss-120b model (OpenAI et al. 2025) that scores each report against a structured rubric designed to capture both general quality features of research and more specific forms of industry expertise.

2.1 Scoring Dimensions

The LLM is prompted to evaluate each industry report using a predefined 18-dimension scoring rubric, categorized as 10 quality measures and 8 knowledge measures. The quality dimensions are *Originality*, *Data Richness*, *Analytical Depth*, *Actionability*, *Structure*, *Forward-Looking*, *Technical Expertise*, *Quantitative Rigor*, *Risk Awareness*, and *Timeliness*. Motivated by survey evidence by Brown et al. (2015), the industry knowledge dimensions are *Industry Trends*, *Industry Technology*, *Supply Chains*, *Distribution Models*, *Margins*, *Customers*, *Labor*, and *Management Teams*. We provide prompts and full scoring rubric in Appendix Section A.1.

We choose these dimensions to be thorough as well as consistent with the literature. The 10 quality dimensions capture the findings of a broad literature analyzing various aspects sell-side analyst firm reports. Asquith, Mikhail, and Au (2005) and Huang, Zang, and Zheng (2014) find that the text of analysts' firm reports provide justifications supporting analysts' opinions, provide information beyond what is contained in earnings forecasts, recommendation levels, and price targets. They show that analyst reports provide entirely new information or interpretation of previously released information to the market. We capture these findings in the *Originality* and *Analytical Depth*. Gleason, Bruce Johnson, and Li (2013) find that there are substantial improvements in analysts' price target performance when those analysts use a rigorous valuation technique rather than an heuristic which we capture with *Quantitative Rigor* and *Analytical Depth*. Furthermore, analyst forecast timing matters (Guttman 2010) and herding behaviors exist (Trueman 1994; Clement and Tse 2005) so we include the *Timeliness* dimension.

Several additional quality dimensions are intended to capture features of analyst research that make reports more useful to investors. We include *Data Richness* because the incorporation of richer and less conventional information sets, including alternative data, is associated with more accurate analyst forecasts and stronger market reactions to analyst output (Chi, Hwang, and Zheng 2025). We include *Technical Expertise* because analysts with prior industry work experience produce more accurate forecasts, are more likely to become Institutional Investor All-Stars, and elicit stronger market reactions, consistent with domain-specific expertise being a meaningful component of research quality (Bradley, Gokkaya, and Liu 2017). We include *Structure* because the readability, clarity and conciseness of analyst reports affect how effectively their information can be processed

by investors and other market participants (Huang, Zang, and Zheng 2014; De Franco et al. 2015). We include *Risk Awareness* because high-quality financial analysis should identify thesis-specific risks, uncertainties, and downside scenarios rather than focus solely on the baseline case. Bastianello, Decaire, and Guenzel (2025) finds that in more recent years, analyst firm reports more frequently include scenario analysis both in discussion and modeling indicating the importance of this dimension for consumers of those reports. We extend this focus on risk to the industry reports. Finally, we include *Forward-Looking* and *Actionability* because one of the central functions of sell-side research is not merely to summarize current conditions, but to form a view about future industry developments and valuation implications that lead to investable conclusions.

We derive the 8 knowledge dimensions from Brown et al. (2015) who conduct surveys finding that industry knowledge is the single most useful input to both earnings forecasts and stock recommendations. In these surveys, an analyst has industry knowledge is she has an understanding of (i) industry trends, (ii) technologies, (iii) supply chains, (iv) distribution models, (v) margins, (vi) customers, (vii) labor, and (viii) management teams. We use these survey-based dimensions directly, allowing us to measure the extent to which the text of an industry report reflects the forms of industry expertise that market participants themselves describe as most valuable.

2.2 Scoring Rubric and Procedures

For each dimension, the LLM assigns a score on a 5-point scale. A score of 5 corresponds to exceptional analysis, reflecting proprietary or especially deep industry insight and mechanism-based reasoning. A score of 4 reflects strong and differentiated analysis that is well supported by the underlying discussion. A score of 3 corresponds to solid and competent research that provides some value added beyond a basic summary. A score of 2 reflects only minimal interpretation, while a score of 1 indicates weak analysis that largely rehashes existing information without generating meaningful insight.

The scoring rubric also guides the model's assessment of source quality and evidentiary support. Reports receive stronger evaluations in dimensions such as *Data Richness*, *Quantitative Rigor*, and *Originality* when they draw on more specialized or proprietary sources, such as channel checks, surveys, expert networks, or industry bodies, rather than relying exclusively on company filings,

general news, or press releases. Similarly, reports are rewarded when they provide explicit calculations, testable claims, thesis-specific risk analysis, and a multi-dimensional analytical framework. By contrast, reports that consist largely of tables without interpretation, boilerplate discussion of risks, recycled standing views, or high page counts with limited informational density are scored less favorably.

The LLM is prompted to assign scores for each dimension based only on the content of the report and to ground its evaluation in report-specific evidence. This evidence-based design of the rubric is important because it disciplines the model to evaluate what is actually present in the text rather than relying on stylistic cues or generic associations with the brokerage house, sector, or analyst name. In the case of chunked reports, the model first produces chunk-level evaluations, after which we apply a synthesis step to form a single report-level assessment. This synthesis stage aggregates the evidence across chunks and is intended to preserve information from the full document while avoiding over-weighting any single section. As a result, each report is ultimately mapped into one set of dimension-level scores regardless of whether it was evaluated in one pass or in multiple chunks.

Analyst reports vary substantially in length and we have to accommodate short to very long reports. For reports whose extracted text fits comfortably within the model context window, we submit the full document for evaluation in a single pass. For longer reports, however, we implement a chunking procedure. Specifically, when a report exceeds 50,000 tokens, we split the document into smaller sequential chunks and evaluate each chunk separately. This procedure ensures that the model can assess the full informational content of long reports without truncating important sections.

A key feature of the procedure is reproducibility. All evaluations are conducted under deterministic model settings, with temperature set to zero and the random seed fixed at 42. These settings reduce stochastic variation in LLM outputs and ensure that a given report receives the same evaluation across repeated runs. This is particularly important in our setting because the measures are later aggregated across reports and linked to analyst-level outcomes in panel regressions. A deterministic pipeline helps ensure that the resulting variation reflects differences in report content rather than random variation in LLM responses.

The output of this process is a set of report-level scores on each rubric dimension. These report-level evaluations serve as the input for our LLM-based analyst measures. Because analysts often publish multiple reports within a month and individual reports may emphasize different dimensions depending on the topic and timing, we subsequently aggregate these report-level scores to the analyst-period level, as described in Section 3.2. This aggregation step allows us to capture persistent differences in analysts' underlying industry knowledge and research quality while still permitting the measures to vary over time.

This methodology offers several advantages for our setting. First, it makes it feasible to evaluate a very large corpus of sell-side research reports in a consistent manner. Second, by relying on a common rubric applied uniformly across brokerage houses, sectors, and time, it reduces concerns that analyst ability is being measured using ad hoc or institution-specific conventions. The resulting measures provide a way to quantify otherwise difficult to measure dimensions of analyst ability.

3 Data

3.1 Data Sources

Our dataset is constructed from multiple sources, including LSEG Workspace (formerly Refinitiv Eikon), Institutional Brokers' Estimate System (I/B/E/S), *Institutional Investor (II)*, Center for Research in Security Prices (CRSP), and Compustat.

Sell-side equity research analyst reports are obtained from LSEG Workspace. We focus on English-language industry reports spanning 2012 to 2024 and limit our sample to the 40 major equity research firms, similar to Décaire and Graham (2024).[2] This results in a total dataset of 1.3 million reports. We then exclude reports attributed to research teams or divisional entities rather than individual analysts, resulting in a final sample of 1.22 million reports. Table IA1 presents the distribution of reports across brokerage houses.

We also obtain metadata for the reports directly from LSEG Workspace. Each report is identi-

2. While Décaire and Graham (2024) identifies 42 brokerage houses, one of them, Bear Stearns, has no industry reports during our sample period. We further treat SunTrust Robinson and Truist Securities as a single entity because Truist Securities is the successor to SunTrust Robinson, following the SunTrust–BB&T merger. This results in 40 distinct brokerage houses in our sample.

fied by a unique ID and includes metadata such as the title, contributor (brokerage house), analyst name, and page count, as well as tags assigned by LSEG Workspace, including industry and region classifications.

Analyst recommendations and earnings per share (EPS) forecasts are collected from the Institutional Brokers' Estimate System (I/B/E/S). Consistent with prior literature (e.g. Bradley, Gokkaya, and Liu 2017; Harford et al. 2019; Hirshleifer et al. 2019), we focus on US common stocks and one-year-ahead EPS forecasts, yielding 1.7 million forecasts over the 2012–2024 period. Restricting to brokerages present in our report sample reduces this to 0.9 million forecasts issued by 3,247 unique analysts. Panel A of Table IA2 compares key characteristics of brokerages within our sample and forecasts made by analysts at these brokerages to those outside it. On average, brokerages in our sample are significantly larger and cover more firms than those excluded, and EPS forecasts issued by their analysts exhibit significantly lower forecast errors.

For this sample, we further match I/B/E/S records to analyst reports using analysts' last name, first initial, and brokerage affiliation, verifying match accuracy by cross-checking stock coverage between Refinitiv Eikon and I/B/E/S. Of these 3,247 analysts, 2,662 (over 80%) authored at least one industry report during the sample period, and their forecasts account for 0.82 million observations (more than 90% of the selected sample).[3] Panel B of Table IA2 presents key characteristics of analysts who have authored industry reports alongside those who have not. Analysts who have authored at least one industry report tend to cover more firms and larger firms, relative to their non-authoring peers, while forecast errors are not significantly different between the two groups.

3.2 LLM Measures

Given that brokerage houses differ in reporting conventions, such as report length and publication frequency, individual reports may only cover a subset of dimensions, and an analyst's proficiency in each dimension is unlikely to change substantially within a short period, we aggregate to the analyst-month level by taking the maximum report-level score within each dimension for a given analyst-month. This ensures that our analyst-level measures capture consistency within a period

3. We thank Wang, Li, and Yu (2025) for generously sharing the I/B/E/S analyst code (ANALYS) to full name link table for the period 2020–2023.

while remaining flexible enough to vary over time. For each analyst-month, we then aggregate the dimension-level scores by taking the average across all knowledge and quality dimensions to construct our analyst-month knowledge and quality measures. Panel A of Table 1 presents the summary statistics for the analyst-month *Knowledge* and *Quality* measures and each individual dimension, as well as the total number of reports per analyst-month and average report length. The mean *Knowledge* score is 2.59 with a standard deviation of 0.73, while the mean *Quality* score is 3.37 with a standard deviation of 0.56, suggesting that analysts tend to score higher on report quality dimensions than on industry knowledge dimensions. On average, analysts publish 3.7 reports per month, with an average report length of 20 pages, reflecting considerable variation in publication frequency and report length across analysts.

As the appropriate level of aggregation varies across empirical settings, we also construct LLM-based measures at alternative levels – including the analyst-year, event-date, and firm levels – as dictated by each specification. The construction of these alternative measures follows the same procedure described above and is detailed in the context of each respective analysis. Across all specifications, the resulting measures are standardized to have a zero mean and unit variance to facilitate the interpretation of the estimated coefficients.

3.3 Key Dependent Variables

3.3.1 *II* All-Star Ratings

Since 1972, *Institutional Investor (II)* has conducted annual surveys in which buy-side institutions rate sell-side analysts who "have been most helpful to them in researching U.S. equities over the past twelve months" (*Institutional Investor* 1996, 1997). Based on these surveys, a list of the top-three analysts and runners-up by industry is published in the magazine's October issue. These *II* All-Star ratings carry significant weight in the sell-side industry, as they are closely tied to analyst compensation and career outcomes (Groysberg, Healy, and Maber 2011). Moreover, Bradshaw (2011) mentions that the most important trait valued by institutional investors is industry knowledge. A growing body of research has since adopted All-Star recognition as a prominent measure of analyst career outcomes to study the determinants of analyst success (Bradley, Gokkaya, and

Liu 2017; Harford et al. 2019; C. Li et al. 2020; Huang, Lin, and Zang 2022). We therefore use All-Star recognition as an external benchmark to validate whether our LLM-based knowledge and quality measures capture meaningful variation in analyst ability.[4]

Specifically, we construct an *AllStar* indicator variable equal to one if an analyst is named among the top three or as a runner-up in their industry category in a given year, and zero otherwise.

3.3.2 Earnings Forecast Error

Brown et al. (2015), in a survey-based study, documents that industry knowledge is the single most useful input to analysts' earnings forecasts and stock recommendations. Motivated by this evidence, we examine whether our LLM-based knowledge and quality measures are associated with analyst forecast accuracy.

Following prior literature, we measure forecast accuracy using *Forecast error*. For analyst i forecasting earnings per share (EPS) for company c issued at time t, $Forecast\ error_{i,c,t}$ is computed as the absolute difference between the actual EPS and the forecasted EPS, scaled by the stock price one year prior to the annual earnings announcement at time T, and multiplied by 100:

$$Forecast\ error_{i,c,t} = 100 \times \frac{|Actual\ EPS_{c,T} - Forecasted\ EPS_{i,c,t}|}{Price_{c,T-1}} \tag{1}$$

3.3.3 Market Reaction

We next investigate whether investors recognize analysts' industry knowledge and perceive their research as more informative. If so, we would expect investors to react more strongly to earnings forecast revisions and stock recommendation changes issued by analysts with higher LLM-based knowledge and quality scores. The key dependent variable in this set of tests is the cumulative market-model-adjusted abnormal stock returns over a three-day event window $[-1, +1]$ around a forecast revision or recommendation change.

4. We thank An-Ping Lin for generously sharing the All-Star data through 2022.

3.3.4 Information Asymmetry

Additionally, we test whether analysts' industry knowledge is associated with a more favorable firm-level information environment. Our first set of dependent variables captures liquidity. Specifically, we use *Illiquidity*, *Effective Spread*, and *Quoted Spread*. *Illiquidity* is the averaged daily Amihud illiquidity (Amihud 2002) in a given year with higher values indicating the stock is less liquid. *Effective Spread* measures the realized round-trip trading cost faced by investors relative to the prevailing midpoint quote, while *Quoted Spread* captures the ex ante bid-ask spread posted by market makers. Higher values of both spread measures indicate greater trading frictions and a less favorable information environment. Accordingly, lower values of *Illiquidity*, *Effective Spread*, and *Quoted Spread* reflect lower information asymmetry and better market liquidity.

Our second set of dependent variables captures price informativeness. We use *BS Ratio Vol* and *Lambda*. *BS Ratio Vol* is the absolute value of the buy-sell imbalance measure constructed from trading volume, with higher values indicating that trading is more one-sided and therefore more likely to reflect asymmetric information or order-flow imbalances rather than balanced incorporation of information into prices. *Lambda* is a price impact measure that captures the sensitivity of prices to order flow. Higher values imply that trades move prices more strongly, consistent with a less liquid and less informative market. For both variables, lower values indicate a more favorable information environment in which firm-specific information is incorporated into prices with fewer frictions.

3.3.5 Price Efficiency

We use *Variance Ratio* which is the absolute value of the difference between the ratio of 60-to-300 second stock return variance and one (var_ratio3 from WRDS Millisecond Intraday Indicators). We use this measure to assess the price efficiency of the firm's stock. If analysts' knowledge and quality of reports for a firm's industry enhance the price efficiency of the firm's stock, we should see a reduction in the variance ratio.

3.3.6 Cost of Capital

If analysts with greater industry knowledge and higher quality research improve the firm's information environment, then these benefits should be reflected not only in trading-based outcomes, but also in a lower implied cost of capital. To test this prediction, we use two widely employed implied cost of capital measures: *ICC PEG* and *ICC OJ*. Both variables are forward-looking estimates of the firm's expected cost of equity derived from stock prices and analyst earnings forecasts, but they differ in the valuation models used to infer the discount rate. *ICC PEG* is based on the price-earnings-growth approach as in Easton (2004), while *ICC OJ* is based on the Ohlson-Juettner framework in Ohlson and Juettner-Nauroth (2005). In both cases, lower values of *ICC PEG* and *ICC OJ* indicate that the firm enjoys a lower implied cost of equity financing.

3.4 Other Variables

We include a set of control variables capturing analyst-, firm-, and event-specific characteristics following prior literature (e.g. Harford et al. 2019; Gibbons, Iliev, and Kalodimos 2021). *Analyst career length* is the total number of years the analyst has appeared in I/B/E/S. *Time covering company* is the total number of years since the analyst first issued an earnings forecast for the firm. *No. of industries* is the number of two-digit SIC industries covered by the analyst. *No. of firms* is the number of unique firms followed by the analyst. *Average firm size* is the average size of all firms covered by the analyst. *Average forecast error* is the average forecast error across all forecasts issued by the analyst in the year. *Brokerage size* is the total number of analysts employed at the brokerage house. *Market value* is the firm's year-end market capitalization, measured as share price multiplied by total shares outstanding. *Institutional ownership* is the percentage of a firm's equity held by institutional investors at year-end. *No. of analysts* is the number of unique analysts issuing earnings forecasts for the firm. *Forecast revision* is the difference between the analyst's revised forecast and the prior forecast, scaled by the one-year lagged stock price and multiplied by 100. *Forecast horizon* is the number of days between the forecast issuance date and the associated earnings announcement date. We further control for firm-level stock return characteristics measured prior to the forecast revision or recommendation change. *Lag monthly return* is the firm's stock

return in the prior month. *Volatility* is the standard deviation of daily stock returns over the two months preceding the event. *Momentum* is the cumulative adjusted stock return from six months to one month before the event. Panel B of Table 1 presents the descriptive statistics of all key variables used in the empirical analysis.

In the liquidity and price-informativeness specifications, we control for a broader set of firm and market characteristics following the literature. *Log analyst coverage* is the logarithm of the number of analysts covering the firm. *Size* is firm size. *ROA* is return on assets. *Leverage* is the firm's leverage ratio. *Market-to-book* is the ratio of market value to book value. *Past return* captures prior stock performance, depending on the specification. *Log price* is the logarithm of stock price. *Log volume* is the logarithm of trading volume. We also include the lagged dependent variable to capture persistence in the information environment. In the implied cost of capital tests, we also include *MAE*, *Forecast dispersion*, *Beta*, *Earnings variability*, and *LT earnings growth*.

4 Empirical Results

4.1 What Drives Analyst Industry Knowledge?

In this section, we investigate what drives analyst industry knowledge and report quality. We construct analyst-year LLM-based knowledge and quality measures by averaging the corresponding monthly scores across all months in a given year and standardizing them to have zero mean and unit variance. We further restrict the sample to analysts who have issued at least one report tagged with a US country or region in Refinitiv Eikon in that year, ensuring our analysis focuses on analysts covering US markets. Then we regress these measures on three sets of variables. The first is the lagged dependent variable, included to test the persistence of our measures over time. The second set captures report and brokerage characteristics, including the number of reports issued (*No. of reports*), average report length (*Average report length*), and the number of analysts at the brokerage who author industry reports (*Brokerage size*). [5] The third set captures analyst characteristics from I/B/E/S, including *Analyst career length*, *No. of industries*, *No. of firms*, and

5. In regression analyses, *No. of reports* and *Average report length* are each divided by 10 to facilitate interpretation of the estimated coefficients.

Average firm size. With the exception of the lagged dependent variable, all remaining controls are measured contemporaneously. In each specification, we additionally include brokerage, industry, and year fixed effects.

Table 2 presents the results.[6] For both *Knowledge* and *Quality*, the lagged dependent variable enters positively and highly significantly across all specifications, with coefficients ranging from 0.41 to 0.56, suggesting that analyst industry knowledge and report quality are highly persistent over time. Turning to report characteristics, both the *No. of reports* and *Average report length* are positively and significantly associated with both measures, indicating that analysts who produce more reports and longer reports tend to exhibit higher knowledge and quality scores. *Brokerage size* is positively associated with both measures when analyst characteristics are excluded, but becomes insignificant once they are controlled for. Among analyst characteristics from I/B/E/S, the *No. of industries* is positively associated with *Knowledge* but not *Quality*, consistent with the notion that broader industry exposure enhances industry knowledge. The *No. of firms* and *Average firm size* are positively associated with both measures, suggesting that analysts covering more firms and larger firms accumulate greater industry expertise and produce higher quality research. *Analyst career length* is positively associated with both measures when report characteristics are excluded, but becomes insignificant once they are controlled for, suggesting that the effect of experience may operate through report production — more experienced analysts tend to produce more and longer reports, which in turn drive higher knowledge and quality scores.

4.2 What Makes a Star Analyst?

We next examine what makes a star analyst, providing a further external validation of our LLM-based measures. Specifically, we estimate a logit regression to investigate how analyst characteristics affect the probability of receiving an *Institutional Investor* All-Star rating. Our dependent variable is *AllStar*, an indicator equal to one if the analyst is named among the top three or as a runner-up in their industry category in a given year, and zero otherwise. The independent variables include

6. The difference in sample size across specifications is due to two factors. First, not all analysts who author industry reports necessarily issue EPS forecasts or stock recommendations. Second, I/B/E/S coverage is incomplete; notably, one of the largest brokerages in our sample (UBS Equities) was removed from the I/B/E/S Detail History database.

our LLM-based *Knowledge* and *Quality* measures, alongside a set of analyst and brokerage controls: *Average forecast error*, *Analyst career length*, *Brokerage size*, *No. of industries*, *No. of firms*, and *Average firm size*. The analysis is conducted at the analyst-year level over the period 2012–2022, with all independent variables lagged by one period. Our model is specified as follows: [7]

$$\mathbb{P}(AllStar_{i,t} = 1) = \Lambda\Big(\beta_1 Z(Measure_{i,t-1}) + \beta_2 Z(Average\ forecast\ error_{i,t-1}) + \gamma Control_{i,t-1}\Big) \tag{2}$$

The results are reported in Table 3. Following Harford et al. (2019), we initially include only year fixed effects before adding brokerage and industry fixed effects. Columns (1) and (2) show that both *Knowledge* and *Quality* are individually positively and significantly associated with the probability of receiving an All-Star rating, with coefficients of 0.612 and 0.642 respectively. Column (3) shows that *Average forecast error*, when included on its own, is negatively associated with All-Star status, indicating that forecast accuracy also affects the likelihood of being recognized. Notably, once analyst and brokerage controls are included in columns (4) and (5), *Average forecast error* loses its significance, while *Knowledge* and *Quality* remain positively and significantly associated with All-Star status with coefficients of 0.529 and 0.594 respectively. These results further hold when brokerage and industry fixed effects are additionally included in columns (6) and (7), confirming that the association between our LLM-based measures and All-Star recognition is robust to a broad set of controls and more stringent fixed effect structures.

In terms of magnitude, since all three measures are standardized, their coefficients are directly comparable; the coefficients on *Knowledge* and *Quality* (0.612 and 0.642) are roughly seven times larger than that on *Average forecast error* (-0.084), suggesting that the qualitative dimensions of analyst output captured by our LLM-based measures are substantially more important for buy-side recognition than traditional forecast accuracy metrics.

7. Over the 2012–2022 period, there are 3,286 annual All-Star designations, of which 2,681 are affiliated with brokerage houses in our sample, corresponding to 569 unique analysts. Of these, 545 have authored at least one industry report in our sample, suggesting that our data captures the vast majority of prominent sell-side analysts recognized as All-Stars.

4.3 Effects on EPS Forecasts

Having established that our LLM-based measures serve as a valid proxy for analyst ability, we next examine whether they are associated with analyst forecast accuracy. We expect forecasts issued by analysts with higher *Knowledge* and *Quality* scores to be more accurate. To test this prediction, we estimate the following regression model:

$$Forecast\ error_{i,c,t} = \beta_1 Z(Measure_{i,c,t-1}) + \gamma Control_{t-1} + \mu_i + \delta_c + \theta_t + \varepsilon_{i,c,t} \tag{3}$$

Here, the dependent variable *Forecast error* is defined as in the Section 3.3.2. The key independent variables are analyst-event-specific *Knowledge* and *Quality* measures, constructed using reports issued within a one-month window prior to the forecast issuance date. Specifically, for each dimension, we first take the maximum score across all reports within this window, and then average across the relevant dimensions to obtain the *Knowledge* and *Quality* measures, respectively. We also control for a set of analyst-, firm-, and forecast-level characteristics, including *Average forecast error*, *Ln(Market value)*, *Institutional ownership*, *Ln(No. of analysts)*, *Ln(Brokerage size)*, *Ln(Analyst career length)*, *Ln(Time covering company)*, and *Ln(Forecast horizon)*. With the exception of *Ln(Analyst career length)*, *Ln(Time covering company)*, and *Ln(Forecast horizon)*, which are measured based on the forecast date, all other control variables are lagged by one period. We estimate specifications with varying combinations of analyst, firm, and year-month fixed effects.

Table 4 presents the results. Both *Knowledge* and *Quality* are negatively and significantly associated with *Forecast error* across all specifications, indicating that analysts with higher knowledge and quality scores produce more accurate EPS forecasts. These results are robust to the inclusion of analyst-firm fixed effects in columns (2) and (4). As our LLM-based measures are constructed on a bounded scale, we focus on the sign and statistical significance of the coefficients rather than their economic magnitude. By contrast, *Average forecast error* – a proxy for past forecast accuracy – loses its significance across all of our specifications, while our LLM-based measures remain significant throughout, suggesting that *Knowledge* and *Quality* capture important dimensions of analyst ability that are not subsumed by past numerical performance.

4.4 Market Reactions to Analyst Research

Building on our evidence that LLM-based *Knowledge* and *Quality* measures are associated with both All-Star recognition and forecast accuracy, we next investigate whether investors recognize analysts' industry knowledge and perceive their research as more informative. The All-Star validation results suggest that buy-side clients directly value analyst knowledge and report quality through their voting behavior. Meanwhile, investors may also learn about analyst ability indirectly over time by observing the accuracy of their forecasts. If investors recognize higher-ability analysts, as proxied by our LLM-based measures, we would expect them to react more strongly to earnings forecast revisions and stock recommendations issued by those analysts.

4.4.1 Earnings Forecast Revisions

We first examine market reactions to forecast revisions, expecting more pronounced reactions to revisions issued by analysts with higher *Knowledge* and *Quality* scores. We estimate the following regression model to test our prediction:

$$\begin{aligned} CAR_{i,c,t} = {} & \beta_1 Z(Forecast\ revision_{i,c,t-1}) + \beta_2 D(High\ Measure_{i,t-1}) \\ & + \beta_3 (Z(Forecast\ revision) \times D(High\ Measure)_{i,t-1}) \\ & + \gamma Control_{t-1} + \mu_i + \delta_c + \theta_t + \varepsilon_{i,c,t} \end{aligned} \tag{4}$$

Here, the dependent variable CAR is the cumulative market-model-adjusted abnormal stock returns over a three-day event window $[-1, +1]$ around a forecast revision. The key independent variables are *High Knowledge* and *High Quality*, indicator variables constructed as follows. We first compute analyst-event-specific *Knowledge* and *Quality* measures following the same procedure as described in Section 4.3. We then define *High Knowledge* (*High Quality*) as an indicator equal to one if the analyst's *Knowledge* (*Quality*) measure exceeds the median across all analysts within the same firm-fiscal year, and zero otherwise. The key variables of interest are the interaction terms *Forecast revision* × *High Knowledge(Quality)*, which capture whether the market reaction to forecast revisions is more pronounced for analysts with higher knowledge and quality scores. Control variables include a set of analyst- and firm-level characteristics, as well as stock return character-

istics measured in the month prior to the forecast revision: *Lag monthly return*, *Volatility*, and *Momentum*. All control variables are either lagged by one period or measured using information available prior to the forecast revision date. All specifications include analyst, firm, and year-month fixed effects, and standard errors are clustered at the firm level.

Table 5 presents the results. The coefficient on *Forecast revision* is positive and highly significant in both specifications, indicating that the stock market responds positively to upward revisions and negatively to downward revisions, with larger revisions eliciting greater price reactions. More relevant for our purposes are the interaction terms between *Forecast revision* and the *High Knowledge* and *High Quality* indicators. We find that the interaction coefficients are positive and significant in both specifications, indicating that conditional on the direction and magnitude of a forecast revision, the stock market reacts more strongly to revisions issued by analysts with higher knowledge and quality scores. This further suggests that investors recognize and incorporate analyst ability into their trading responses, treating research from higher-ability analysts as more informative. [8]

4.4.2 Stock Recommendation Changes

We next turn to market reactions to stock recommendation changes. Prior work shows that analysts with superior forecast accuracy also issue more informative stock recommendations (Loh and Mian 2006), and that accurate forecasts are primarily used as inputs for corresponding stock recommendations (Brown et al. 2015). Motivated by this evidence, we expect more pronounced market reactions to recommendation changes issued by analysts with higher *Knowledge* and *Quality* scores. We estimate the following regression model:

$$CAR_{i,c,t} = \beta_1 Z(Measure_{i,c,t-1}) + \gamma Control_{t-1} + \mu_i + \delta_c + \theta_t + \varepsilon_{i,c,t} \tag{5}$$

Here, the key explanatory variables are analyst-event-specific *Knowledge* and *Quality* measures, constructed following the same procedure as described in Section 4.3. All other variables and fixed

8. Our results are robust to an alternative definition of *Forecast revision* following Ivković and Jegadeesh (2004), measured as the difference between the analyst's revised forecast and the prior forecast scaled by the absolute value of the prior forecast.

effects are defined and constructed as in Section 4.4.1. Given the asymmetric market reactions to recommendation upgrades and downgrades documented in the literature (e.g., Liu, Smith, and Syed 1990; Kecskés, Michaely, and Womack 2017), we estimate separate regressions for each.

Table 6 presents the results. Columns (1) and (2) report results for recommendation upgrades, and columns (3) and (4) for downgrades. Consistent with our prediction, the coefficients on *Knowledge* and *Quality* are positive and significant for upgrades, and negative and significant for downgrades. These results indicate that the market reacts more strongly to recommendation changes issued by analysts with higher knowledge and quality scores. The informativeness of stock recommendations, as perceived by the market, thus appears to be increasing in analysts' knowledge and quality scores.

Comparing the magnitudes of the coefficients across upgrades and downgrades, the effects are notably larger for downgrades – with coefficients of -0.245 and -0.275 – than for upgrades – with coefficients of 0.175 and 0.217. This asymmetry is consistent with prior literature suggesting that investors consider downgrades more credible and informative than upgrades, as the latter may be driven by analysts' conflicts of interest, namely their incentive to maintain relationships with firm management and generate order flow.

Taken together, the above two sets of analyses suggest that investors recognize and incorporate analyst ability into their response to analyst research. Specifically, we find that earnings forecast revisions and stock recommendation changes issued by analysts with higher *Knowledge* and *Quality* scores elicit stronger market reactions, indicative of research produced by higher-ability analysts conveying greater information content.

4.5 Do Firms Benefit from Analysts' Industry Knowledge?

Having shown that our LLM-based *Knowledge* and *Quality* measures are positively associated with analyst recognition, forecast accuracy, and market reactions, we next ask whether firms themselves benefit when they are followed by analysts with greater industry knowledge and higher quality research. If analyst ability improves the firm's information environment, then stocks followed by stronger analysts should exhibit lower information asymmetry and a lower cost of capital. Similar phenomena have been shown in analyst coverage (Bowen, Chen, and Cheng 2008) and analyst

effort allocation (Harford et al. 2019). These outcomes represent economically meaningful channels through which analyst industry expertise may create value beyond the immediate trading response to individual research events.

We investigate this broader question along two dimensions. First, we examine whether analyst ability is associated with lower information asymmetry as reflected in liquidity measures and price informativeness proxies. Second, we examine whether these improvements in the information environment translate into a lower implied cost of capital. Across these tests, our central prediction is that firms followed by analysts with higher *Knowledge* and *Quality* scores should face a more favorable information environment and, consequently, lower financing frictions.

4.5.1 Information Asymmetry

We begin by examining whether firms followed by more knowledgeable analysts exhibit lower information asymmetry. Analysts with deeper industry knowledge and higher quality research should produce more informative outputs, reducing uncertainty among market participants. If so, stocks followed by such analysts should display better liquidity and greater price informativeness. To test this prediction, we estimate regressions of the following form:

$$\begin{aligned} Y_{f,t+1} = \beta_1 Z(Measure_{f,t}) + \beta_2 Z(Size_{f,t}) + \beta_3 \big(Z(Measure_{f,t}) \times Z(Size_{f,t})\big) \\ + \gamma Control_{f,t} + \mu_f + \delta_{j,t} + \varepsilon_{f,t}, \end{aligned} \tag{6}$$

where $Y_{f,t+1}$ denotes a next-period measure of the firm's information environment. In Table 7, the dependent variables are *Illiq*, *Effective Spread*, and *Quoted Spread*. In Table 8, the dependent variables are *BS Ratio Vol*, *Lambda*, and the 5-minute *Variance Ratio*. The key explanatory variables are firm-year analyst *Knowledge* and *Quality* measures. We construct these firm-year measures using the mean of the top three analyst scores within each firm-year, with the goal of capturing the influence of the firm's strongest analysts rather than the average quality of all analysts following the firm. We also interact the analyst measure with firm size to allow the relation between analyst expertise and the firm's information environment to vary systematically across firms. All specifications include firm fixed effects, industry-year fixed effects, and the lagged dependent variable. Standard errors are clustered at the analyst and firm levels.

Table 7 reports the liquidity results. The coefficients on *Z(Measure Mean Top3)* are negative across all three liquidity outcomes for both knowledge and quality, indicating that firms followed by stronger analysts tend to exhibit lower future trading frictions. The evidence is strongest for analyst *Knowledge*. In particular, knowledge is negatively and significantly associated with *Illiq*, *Effective Spread*, and *Quoted Spread*. The corresponding quality coefficients are also negative in all three cases, although they are estimated less precisely and are not statistically significant in this table. Overall, these results suggest that analyst industry expertise, especially knowledge, is associated with better future liquidity.

The interaction terms between analyst expertise and firm size are positive and statistically significant across all liquidity specifications. This pattern indicates that the beneficial association between analyst expertise and liquidity is weaker for larger firms. Economically, this is plausible: large firms already tend to have richer information environments, broader investor attention, and greater trading activity, leaving less scope for analyst expertise to further reduce information asymmetry. The marginal informational contribution of highly knowledgeable analysts therefore appears to be more important for smaller firms.

Table 8 turns to price informativeness and market efficiency. We examine three proxies: *BS Ratio Vol*, *Lambda*, and the 5-minute *Variance Ratio*, all measured in the next period. Lower values of *BS Ratio Vol* and *Lambda* indicate a more favorable information environment, reflecting more balanced order flow and lower price sensitivity to trading pressure. We discuss the results of the Variance Ratio test in section 4.5.2.

The results again point to a favorable role for analyst expertise. For *BS Ratio Vol*, both *Knowledge* and *Quality* enter negatively, with the knowledge coefficient significant at the 1% level and the quality coefficient significant at the 10% level. For *Lambda*, the coefficient on *Knowledge* is negative and statistically significant, while the coefficient on *Quality* is also negative but estimated less precisely. These results suggest that firms followed by more knowledgeable analysts and, to a lesser extent, higher-quality analysts exhibit more informative prices in subsequent periods.

As in the liquidity regressions, the interaction between analyst expertise and firm size is positive and strongly significant across all price-informativeness measures. This again implies that the beneficial relation between analyst expertise and the firm's information environment is concentrated

among smaller firms, where analyst information production is likely to be more valuable.

4.5.2 Price Efficiency

Next, we test whether firms covered by more knowledgeable analysts have better subsequent price efficiency. For this test, we use the 5-minute Variance Ratio which is the absolute value of the absolute value of the difference between the ratio of 60-to-300 second stock return variance and one. If more knowledgeable analysts improve the firm's information environment incorporating information into prices more efficiently, then firms covered by such analysts should have lower variance ratios.

The results reported in columns (5) and (6) of Table 8 provide support for this prediction. The coefficient on both *Knowledge* and *Quality* are negative and statistically significant, with *Knowledge* exhibiting a stronger relationship than *Quality*. Similar to the liquidity and price informativeness results, we let the main measures interact with firm size since size is a prominent determinant of price efficiency and find that the interaction between analyst expertise and firm size is positive and statistically significant for both the measures. This pattern provides additional support that the benefits of analyst expertise on price efficiency is weaker for larger firms, likely since information frictions are more severe for smaller firms.

The results in Tables 7 and 8 show that firms followed by analysts with stronger LLM-based expertise scores experience a better subsequent information environment. They exhibit lower trading frictions, more balanced trading, lower price impact, and more efficient price dynamics. The results are especially strong for the knowledge measure and are most pronounced among smaller firms, consistent with the view that analyst industry expertise improves the firm's broader information environment.

4.5.3 Cost of Capital

We next ask whether the improvements in firms' information environments documented above translate into a lower cost of capital. A large literature argues that better information environments reduce required returns by lowering estimation risk, mitigating information asymmetry, and improving the precision with which investors assess firm fundamentals. If analysts with greater

industry knowledge and higher quality research reduce these frictions, then firms followed by such analysts should face a lower implied cost of capital. To test this prediction, we estimate regressions in which the dependent variable is an implied cost of capital measure:

$$ICC_{f,t} = \beta_1 Z(Measure_{f,t}) + \gamma Control_{f,t} + \mu_f + \theta_t + \varepsilon_{f,t}. \tag{7}$$

The dependent variables are two standard implied cost of capital measures, *ICC PEG* and *ICC OJ*. The key explanatory variables are *Z(Measure Mean)* for *Knowledge* and *Quality*. Control variables include analyst coverage, firm size, trading activity, institutional ownership, past returns, return volatility, leverage, market-to-book, forecast error, forecast dispersion, beta, earnings variability, and long-term earnings growth. All specifications include firm fixed effects and year fixed effects, and standard errors are clustered at the firm level. Table 9 reports the results. Across all four specifications, the coefficients on analyst *Knowledge* and *Quality* are negative and statistically significant. For *ICC PEG*, the coefficients are -0.028 for *Knowledge* and -0.025 for *Quality*. For *ICC OJ*, the corresponding coefficients are -0.028 and -0.024. The consistency of these estimates across alternative implied cost of capital measures suggests that the relation is robust and not driven by a particular measurement approach.

These findings indicate that firms followed by analysts with greater industry knowledge and higher quality research face a lower cost of equity capital. This result is economically intuitive and complements the evidence from the previous subsection. If stronger analysts reduce information asymmetry and improve price informativeness, investors should require a lower risk premium to hold the firm's equity. The implied cost of capital results are consistent with analyst ability generating real financing benefits for firms through an improved information environment. Overall, the evidence in this subsection suggests that analyst industry knowledge has benefits to firms. Firms followed by more capable analysts appear to enjoy a more favorable information environment, characterized by better liquidity, more informative prices, and a lower implied cost of capital.

5 Conclusion

Sell-side analysts are among the most important information intermediaries in capital markets, yet the academic literature has predominantly evaluated their contributions through the lens of quantitative outputs such as earnings forecasts, stock recommendations, and target prices. Practitioners and institutional investors, by contrast, have long emphasized that industry knowledge is the most valuable service that sell-side analysts provide. We bridge this gap by developing the first large-scale, text-based measures of analyst industry knowledge and report quality, constructed from 1.22 million industry reports using an LLM-based evaluation pipeline. By doing so, we provide a direct way to study analyst ability, which is highly important but previously unobservable in the literature.

Our findings offer a direct answer to the question posed by the paper's title. Industry knowledge and report quality are substantially more predictive of All-Star recognition than forecast accuracy, with estimated coefficients roughly seven times larger. Average forecast error becomes statistically insignificant once our measures are included. Beyond recognition, analysts with higher knowledge and quality scores produce more accurate earnings forecasts, and their forecast revisions and recommendation changes elicit stronger market reactions. At the firm level, coverage by more knowledgeable analysts is associated with narrower spreads, more informative prices, and a lower implied cost of capital, with these associations concentrated among smaller firms and driven by the top analysts covering the firm. These findings suggest that what distinguishes top analysts is not simply their forecast accuracy, but the depth of their industry understanding manifested in their research output.

References

Amihud, Yakov. 2002. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of Financial Markets* 5 (1): 31–56.

Amiram, Dan, Edward Owens, and Oded Rozenbaum. 2016. "Do information releases increase or decrease information asymmetry? New evidence from analyst forecast announcements." *Journal of Accounting and Economics* 62 (1): 121–138.

Asquith, Paul, Michael B Mikhail, and Andrea S Au. 2005. "Information content of equity analyst reports." *Journal of Financial Economics* 75 (2): 245–282.

Bastianello, Francesca, Paul Decaire, and Marius Guenzel. 2025. "Valuation Models." *SSRN Electronic Journal.*

Bowen, Robert M., Xia Chen, and Qiang Cheng. 2008. "Analyst Coverage and the Cost of Raising Equity Capital: Evidence from Underpricing of Seasoned Equity Offerings." *Contemporary Accounting Research* 25 (3): 657–700.

Bradley, Daniel, Sinan Gokkaya, and Xi Liu. 2017. "Before an analyst becomes an analyst: Does industry experience matter?" *The Journal of Finance* 72 (2): 751–792.

Bradshaw, Mark, Yonca Ertimur, and Patricia O'Brien. 2017. "Financial Analysts and Their Contribution to Well-Functioning Capital Markets." *Foundations and Trends in Accounting* 11, no. 3 (December 12, 2017): 119–191.

Bradshaw, Mark T. 2011. "Analysts' forecasts: what do we know after decades of work?" *Available at SSRN 1880339.*

Bradshaw, Mark T, Lawrence D Brown, and Kelly Huang. 2013. "Do sell-side analysts exhibit differential target price forecasting ability?" *Review of Accounting Studies* 18 (4): 930–955.

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 (5): 1933–1967.

Brown, Lawrence D, Andrew C Call, Michael B Clement, and Nathan Y Sharp. 2015. "Inside the "black box" of sell-side financial analysts." *Journal of Accounting Research* 53 (1): 1–47.

Chi, Feng, Byoung-Hyoun Hwang, and Yaping Zheng. 2025. "The Use and Usefulness of Big Data in Finance: Evidence from Financial Analysts." *Management Science* 71 (6): 4599–4621.

Clement, Michael B. 1999. "Analyst forecast accuracy: Do ability, resources, and portfolio complexity matter?" *Journal of Accounting and Economics* 27 (3): 285–303.

Clement, Michael B., and Senyo Y. Tse. 2005. "Financial Analyst Characteristics and Herding Behavior in Forecasting." *The Journal of Finance* 60 (1): 307–341.

Crawford, Steven S., Darren T. Roulstone, and Eric C. So. 2012. "Analyst Initiations of Coverage and Stock Return Synchronicity." *The Accounting Review* 87, no. 5 (September): 1527–1553.

De Franco, Gus, Ole-Kristian Hope, Dushyantkumar Vyas, and Yibin Zhou. 2015. "Analyst Report Readability." *Contemporary Accounting Research* 32 (1): 76–104.

Décaire, Paul, and John R Graham. 2024. "Valuation fundamentals." *HKU Jockey Club Enterprise Sustainability Global Research Institute-Archive.*

Easton, Peter D. 2004. "PE Ratios, PEG Ratios, and Estimating the Implied Expected Rate of Return on Equity Capital." *The Accounting Review* 79 (1): 73–95.

Gibbons, Brian, Peter Iliev, and Jonathan Kalodimos. 2021. "Analyst information acquisition via EDGAR." *Management Science* 67 (2): 769–793.

Gleason, Cristi A., W. Bruce Johnson, and Haidan Li. 2013. "Valuation Model Use and the Price Target Performance of Sell-Side Equity Analysts." *Contemporary Accounting Research* 30 (1): 80–115.

Groysberg, Boris, Paul M Healy, and David A Maber. 2011. "What drives sell-side analyst compensation at high-status investment banks?" *Journal of Accounting Research* 49 (4): 969–1000.

Guttman, Ilan. 2010. "The Timing of Analysts' Earnings Forecasts." *The Accounting Review* 85 (2): 513–545.

Harford, Jarrad, Feng Jiang, Rong Wang, and Fei Xie. 2019. "Analyst career concerns, effort allocation, and firms' information environment." *The Review of Financial Studies* 32 (6): 2179–2224.

Hirshleifer, David, Yaron Levi, Ben Lourie, and Siew Hong Teoh. 2019. "Decision fatigue and heuristic analyst forecasts." *Journal of Financial Economics* 133 (1): 83–98.

Huang, Allen H, An-Ping Lin, and Amy Y Zang. 2022. "Cross-industry information sharing among colleagues and analyst research." *Journal of Accounting and Economics* 74 (1): 101496.

Huang, Allen H, Amy Y Zang, and Rong Zheng. 2014. "Evidence on the information content of text in analyst reports." *The Accounting Review* 89 (6): 2151–2180.

Huang, Allen H., Reuven Lehavy, Amy Y. Zang, and Rong Zheng. 2018. "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach." *Management Science* 64 (6): 2833–2855.

Ivković, Zoran, and Narasimhan Jegadeesh. 2004. "The timing and value of forecast and recommendation revisions." *Journal of Financial Economics* 73 (3): 433–463.

Kadan, Ohad, Leonardo Madureira, Rong Wang, and Tzachi Zach. 2012. "Analysts' industry expertise." *Journal of Accounting and Economics* 54 (2-3): 95–120.

Kecskés, Ambrus, Roni Michaely, and Kent L Womack. 2017. "Do earnings estimates add value to sell-side analysts' investment recommendations?" *Management Science* 63 (6): 1855–1871.

Li, Congcong, An-Ping Lin, Hai Lu, and Kevin Veenstra. 2020. "Gender and beauty in the financial analyst profession: evidence from the United States and China." *Review of Accounting Studies* 25 (4): 1230–1262.

Li, Kai, Feng Mai, Gabriel Wong, Chelsea Yang, and Tengfei Zhang. 2026. "Female equity analysts and corporate environmental and social performance." *Management Science.*

Li, Xi. 2005. "The persistence of relative performance in stock recommendations of sell-side financial analysts." *Journal of Accounting and Economics* 40 (1-3): 129–152.

Liu, Pu, Stanley D Smith, and Azmat A Syed. 1990. "Stock price reactions to the Wall Street Journal's securities recommendations." *Journal of Financial and Quantitative Analysis* 25 (3): 399–410.

Loh, Roger K, and G Mujtaba Mian. 2006. "Do accurate earnings forecasts facilitate superior investment recommendations?" *Journal of Financial Economics* 80 (2): 455–483.

Loh, Roger K., and René M. Stulz. 2018. "Is Sell-Side Research More Valuable in Bad Times?" *The Journal of Finance* 73 (3): 959–1013.

Merkley, Kenneth, Roni Michaely, and Joseph Pacelli. 2017. "Does the Scope of the Sell-Side Analyst Industry Matter? An Examination of Bias, Accuracy, and Information Content of Analyst Reports." *The Journal of Finance* 72 (3): 1285–1334.

Mikhail, Michael B, Beverly R Walther, and Richard H Willis. 1997. "Do security analysts improve their performance with experience?" *Journal of Accounting Research* 35:131–157.

O'Brien, Patricia C. 1988. "Analysts' forecasts as earnings expectations." *Journal of Accounting and Economics* 10 (1): 53–83.

———. 1990. "Forecast accuracy of individual analysts in nine industries." *Journal of Accounting Research* 28 (2): 286–304.

Ohlson, James A., and Beate E. Juettner-Nauroth. 2005. "Expected EPS and EPS Growth as Determinantsof Value." *Review of Accounting Studies* 10:349–365.

OpenAI, Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, et al. 2025. "Gpt-Oss-120b & Gpt-Oss-20b Model Card." Pre-published, August 8, 2025. Accessed March 13, 2026.

Piotroski, Joseph D., and Barren T. Roulstone. 2004. "The Influence of Analysts, Institutional Investors, and Insiders on the Incorporation of Market, Industry, and Firm-Specific Information into Stock Prices." *The Accounting Review* 79 (4): 1119–1151.

Sinha, Praveen, Lawrence D Brown, and Somnath Das. 1997. "A re-examination of financial analysts' differential earnings forecast accuracy." *Contemporary Accounting Research* 14 (1): 1–42.

Stickel, Scott E. 1990. "Predicting individual analyst earnings forecasts." *Journal of Accounting Research* 28 (2): 409–417.

Trueman, Brett. 1994. "Analyst Forecasts and Herding Behavior." *The Review of Financial Studies* 7 (1): 97–124.

Wang, Baolian, Peixin Li, and Jiawei Yu. 2025. "The Impact of Return-to-Office Mandates on Equity Analysts." *Available at SSRN 5206058.*

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

Tables

Table 1: Summary Statistics

This table presents summary statistics for the main variables used in the analysis. Panel A reports statistics for analyst-monthly LLM-based measures, including the composite *Knowledge* and *Quality* scores, report characteristics (*No. of reports* and *Average report length*), and the individual knowledge and quality dimensions. Panel B reports statistics for variables used in the determinant and All-Star logit analyses. Panel C reports statistics for variables used in the forecast error and market reaction analyses. All variables are defined in Section 3.

Variable	N	Mean	SD	P1	P50	P99
Panel A: Analyst-Monthly LLM Measures						
Knowledge	226,347	2.59	0.73	1.00	2.75	3.88
Quality	226,347	3.37	0.56	1.60	3.50	4.20
No. of reports	226,347	3.70	4.60	1.00	2.00	24.00
Average report length	226,347	20.28	21.82	2.00	14.00	108.00
Industry technology	226,347	2.46	1.03	1.00	2.00	4.00
Industry trends	226,347	3.52	0.72	1.00	4.00	4.00
Supply chain	226,347	2.48	1.05	1.00	2.00	4.00
Distribution model	226,347	2.44	1.00	1.00	2.00	4.00
Margin	226,347	3.10	1.02	1.00	3.00	4.00
Customer	226,347	2.89	0.97	1.00	3.00	4.00
Labor	226,347	1.59	0.90	1.00	1.00	4.00
Management team	226,347	2.28	0.91	1.00	2.00	4.00
Originality	226,347	2.62	0.82	1.00	2.00	4.00
Data	226,347	3.23	0.73	2.00	3.00	5.00
Analytical	226,347	3.10	0.74	1.00	3.00	4.00
Actionability	226,347	3.31	0.87	1.00	4.00	4.00
Structure	226,347	3.83	0.47	2.00	4.00	5.00
Forward	226,347	3.48	0.72	1.00	4.00	4.00
Technical	226,347	3.71	0.65	2.00	4.00	5.00
Quantitative	226,347	3.34	0.71	1.00	3.00	4.00
Risk	226,347	3.18	0.90	1.00	3.00	4.00
Timeliness	226,347	3.95	0.53	2.00	4.00	5.00
Panel B: Determinant and All-Star Analysis						
No. of reports	29,299	30.87	49.52	1.00	13.00	262.00
Average report length	29,299	22.30	20.95	4.00	16.94	109.33
Brokerage size	29,299	135.09	78.66	12.00	135.00	267.00
Analyst career length	11,010	12.92	10.84	0.00	10.00	39.00
No. of firms	11,010	13.20	8.84	1.00	14.00	35.00
Average firm size (*thousands* \$)	11,010	32,727.97	47,711.81	405.76	16,545.98	289,724.59
Average forecast error	10,813	1.95	3.42	0.06	0.88	24.06
Panel C: Forecast Error and Market Reaction Analysis						
Forecast error	474,552	1.48	3.30	0.00	0.40	23.18
Forecast horizon	474,552	203.93	102.71	7.00	200.00	364.00
Maket value (*thousands* \$)	474,552	32,257.85	65,356.78	95.60	8,621.75	400,533.19
Institutional ownership (%)	474,511	0.81	0.20	0.02	0.84	1.17
No. of analysts	474,552	19.52	9.73	3.00	19.00	47.00
Brokerage size	474,552	67.45	24.71	13.00	68.00	117.00
No. of firms	474,552	19.74	8.03	2.00	19.00	43.00
No. of industries	474,552	4.70	2.46	1.00	4.00	12.00
Average firm size (*thousands* \$)	474,552	31,490.19	33,891.30	1,115.23	20,562.31	194,443.36
Average forecast error	474,552	1.97	3.00	0.17	1.03	21.00
Analyst career length	474,552	15.97	10.80	1.00	14.00	39.00
Time covering company	474,552	6.01	6.13	0.00	4.00	29.00
Forecast revision	398,407	-0.04	1.41	-7.19	0.01	6.17
$CAR_{[-1,+1]}$	398,406	-0.07	6.60	-22.79	-0.02	21.67
Lag monthly return	398,402	0.36	11.18	-34.23	0.53	36.63
Volatility	398,407	2.39	1.54	0.71	1.93	9.06
Momentum	398,406	0.04	0.26	-0.60	0.03	1.01

Table 2: Determinants of Analyst Expertise Scores

This table reports OLS regression results examining the determinants of analyst-year LLM-based *Knowledge* and *Quality* measures. The dependent variables are the analyst-year *Knowledge* measure in columns (1)–(3) and the *Quality* measure in columns (4)–(6). The independent variables comprise three sets: the lagged dependent variable, report and brokerage characteristics (*No. of reports*, *Average report length*, and *Brokerage size*), and analyst characteristics from I/B/E/S (*Analyst career length*, *No. of industries*, *No. of firms*, and *Average firm size*). All independent variables are defined in Section 3 and measured contemporaneously, with the exception of the lagged dependent variable. All specifications include brokerage, industry, and year fixed effects. t-statistics in parentheses are based on standard errors clustered at the analyst and year levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable/Measure	*Knowledge*			*Quality*		
	(1)	(2)	(3)	(4)	(5)	(6)
Lag Measure	0.521***	0.562***	0.461***	0.482***	0.532***	0.410***
	(44.27)	(36.05)	(32.18)	(34.02)	(31.49)	(24.77)
No. of reports	0.034***		0.031***	0.034***		0.034***
	(18.44)		(11.41)	(18.25)		(11.48)
Average report length	0.160***		0.163***	0.165***		0.169***
	(21.24)		(13.75)	(20.29)		(15.02)
Ln(Brokerage size)	0.069*		0.052	0.054*		0.037
	(1.97)		(1.41)	(1.86)		(0.75)
Ln(Analyst career length)		0.025*	-0.000		0.023*	-0.002
		(2.13)	(-0.03)		(2.10)	(-0.20)
Ln(No. of industries)		0.047	0.093***		-0.044	-0.017
		(1.76)	(3.63)		(-1.64)	(-0.68)
Ln(No. of firms)		0.057**	0.013		0.084***	0.041**
		(2.78)	(0.65)		(4.65)	(2.31)
Ln(Average firm size)		0.041***	0.031***		0.044***	0.034***
		(5.98)	(4.09)		(8.66)	(6.33)
Brokerage FE	Y	Y	Y	Y	Y	Y
Industry FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
N	21,209	9,288	9,288	21,209	9,288	9,288
Adj. R^2	0.548	0.462	0.546	0.506	0.416	0.511

Table 3: What Makes a Star Analyst?

This table reports logit regression results examining the determinants of *Institutional Investor* All-Star ratings. The dependent variable *AllStar* is an indicator equal to one if the analyst is named among the top three or as a runner-up in their industry category in a given year, and zero otherwise. The independent variables include LLM-based *Knowledge* and *Quality* measures, *Average forecast error*, *Analyst career length*, *Brokerage size*, *No. of industries*, *No. of firms*, and *Average firm size*. All independent variables are defined in Section 3 and lagged by one period. Columns (1)–(5) include year fixed effects only, while columns (6)–(7) additionally include brokerage and industry fixed effects. t-statistics in parentheses are based on standard errors clustered at the analyst and year levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

	Pr(AllStar=1)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Z(Knowledge)	0.612***			0.529***		0.431***	
	(7.17)			(6.23)		(6.05)	
Z(Quality)		0.642***			0.594***		0.463***
		(6.42)			(5.88)		(5.81)
Z(Average forecast error)			-0.084**	0.009	-0.005	0.024	0.014
			(-2.19)	(0.20)	(-0.12)	(0.47)	(0.30)
Ln(Analyst career length)				0.685***	0.686***	0.699***	0.690***
				(9.05)	(8.84)	(8.76)	(8.53)
Ln(Brokerage size)				2.268***	2.224***	0.615*	0.629*
				(6.77)	(6.73)	(1.77)	(1.76)
Ln(No. of industries)				0.159	0.361**	0.247	0.329
				(0.96)	(2.09)	(1.24)	(1.62)
Ln(No. of firms)				1.192***	1.127***	1.415***	1.407***
				(7.83)	(7.45)	(8.12)	(7.97)
Ln(Average firm size)				0.247***	0.260***	0.186***	0.183***
				(4.38)	(4.53)	(2.79)	(2.73)
Brokerage FE						Y	Y
Industry FE						Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y
N	8,987	8,987	8,964	8,964	8,964	6,876	6,876
Pseudo R^2	0.048	0.047	0.004	0.300	0.303	0.334	0.336

Table 4: Analyst Expertise Scores and Effects on EPS Forecast

This table reports OLS regression results examining the relation between LLM-based measures and EPS forecast accuracy. The dependent variable *Forecast error* and the key independent variables, analyst-event-specific *Knowledge* and *Quality* measures, are constructed based on reports issued within a one-month window prior to the forecast issuance date; see Section 4.3 for details. All control variables are defined in Section 3. With the exception of *Ln(Analyst career length)*, *Ln(Time covering company)*, and *Ln(Forecast horizon)*, which are measured based on the forecast date, all other control variables are lagged by one period. Columns (1) and (3) include analyst, firm, and year-month fixed effects, while columns (2) and (4) include analyst-firm and year-month fixed effects. t-statistics in parentheses are based on standard errors clustered at the analyst and firm levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	*Forecast error*			
Measure	*Knowledge*		*Quality*	
	(1)	(2)	(3)	(4)
Z(Measure)	-0.023***	-0.027***	-0.016**	-0.019***
	(-3.44)	(-3.90)	(-2.55)	(-2.94)
Average forecast error	0.009	0.007	0.009	0.007
	(0.95)	(0.73)	(0.95)	(0.72)
Ln(Market value)	-1.524***	-1.581***	-1.524***	-1.582***
	(-19.00)	(-17.73)	(-19.00)	(-17.73)
Institutional ownership	-0.814***	-0.745***	-0.814***	-0.744***
	(-3.31)	(-3.11)	(-3.31)	(-3.10)
Ln(No. of analysts)	0.425***	0.300***	0.425***	0.301***
	(5.33)	(4.07)	(5.33)	(4.07)
Ln(Brokerage size)	0.093	0.099	0.092	0.098
	(1.11)	(0.98)	(1.09)	(0.97)
Ln(Analyst career length)	0.006	-0.256**	0.006	-0.256**
	(0.08)	(-2.34)	(0.07)	(-2.34)
Ln(Time covering company)	0.096***	0.465***	0.096***	0.465***
	(6.14)	(7.46)	(6.14)	(7.45)
Ln(Forecast horizon)	0.446***	0.455***	0.446***	0.455***
	(21.91)	(21.81)	(21.91)	(21.81)
Analyst FE	Y		Y	
Firm FE	Y		Y	
Analyst-Firm FE		Y		Y
Year-Month FE	Y	Y	Y	Y
N	465,533	462,293	465,533	462,293
Adj. R^2	0.440	0.497	0.440	0.497

Table 5: Market Reaction to EPS Forecast Revisions

This table reports OLS regression results examining whether the market reaction to EPS forecast revisions varies with analysts' LLM-based knowledge and quality scores. The dependent variable $CAR_{[-1,+1]}$ is the cumulative market-model-adjusted abnormal return over a three-day event window around the forecast revision date. $D(\textit{High Measure})$ is an indicator equal to one if the analyst's *Knowledge* (*Quality*) measure exceeds the median across all analysts within the same firm–fiscal year, and zero otherwise; see Section 4.4.1 for details. The key variable of interest is the interaction term $Z(\textit{Forecast revision}) \times D(\textit{High Measure})$. All control variables are defined in Section 3 and are either lagged by one period or measured using information available prior to the forecast revision date. We exclude observations with stock prices below \$1. All specifications include analyst, firm, and year-month fixed effects. t-statistics in parentheses are based on standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	$CAR_{[-1,+1]}$	
Measure	*Knowledge*	*Quality*
	(1)	(2)
Z(Forecast revision)	1.030***	1.023***
	(26.18)	(26.84)
D(High Measure)	-0.003	0.027
	(-0.14)	(1.10)
*Z(Forecast revision)*D(High Measure)*	0.076**	0.096***
	(2.38)	(2.95)
Average forecast error	-0.021***	-0.021***
	(-2.70)	(-2.69)
Ln(Market value)	-1.239***	-1.239***
	(-18.43)	(-18.43)
Institutional ownership	-0.015	-0.013
	(-0.08)	(-0.06)
Ln(No. of analysts)	0.078	0.078
	(0.79)	(0.79)
Ln(Brokerage size)	-0.011	-0.010
	(-0.21)	(-0.20)
Ln(Analyst career length)	-0.011	-0.013
	(-0.14)	(-0.17)
Ln(Time covering company)	-0.015	-0.015
	(-0.76)	(-0.77)
Lag monthly return	-0.003	-0.003
	(-1.26)	(-1.27)
Ln(Volatility)	0.194	0.194
	(1.61)	(1.60)
Momentum	-1.716***	-1.715***
	(-14.52)	(-14.51)
Analyst FE	Y	Y
Firm FE	Y	Y
Year-Month FE	Y	Y
N	378,714	378,714
Adj. R^2	0.067	0.067

Table 6: Market Reaction to Stock Recommendation Changes

This table reports OLS regression results examining whether the market reaction to stock recommendation changes varies with analysts' LLM-based knowledge and quality scores. The dependent variable $CAR_{[-1,+1]}$ is the cumulative market-model-adjusted abnormal return over a three-day event window around the recommendation change date. The key independent variables, analyst-event-specific *Knowledge* and *Quality* measures, are constructed based on reports issued within a one-month window prior to the forecast issuance date; see Section 4.4.2 for details. Columns (1) and (2) report results for recommendation upgrades, and columns (3) and (4) for downgrades. All control variables are defined in Section 3 and are either lagged by one period or measured using information available prior to the forecast revision date. We exclude observations with stock prices below $1. All specifications include analyst, firm, and year-month fixed effects. t-statistics in parentheses are based on standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	$CAR_{[-1,+1]}$			
	Rec. Upgrade		Rec. Downgrade	
Measure	*Knowledge*	*Quality*	*Knowledge*	*Quality*
	(1)	(2)	(3)	(4)
Z(Measure)	0.175*	0.217**	-0.245**	-0.275**
	(1.90)	(2.35)	(-1.97)	(-2.19)
Average forecast error	0.101*	0.102*	-0.102	-0.103
	(1.86)	(1.88)	(-1.39)	(-1.41)
Ln(Market value)	-1.579***	-1.575***	-1.164***	-1.161***
	(-7.43)	(-7.42)	(-3.61)	(-3.60)
Institutional ownership	-0.635	-0.615	-0.499	-0.520
	(-0.89)	(-0.87)	(-0.52)	(-0.54)
Ln(No. of analysts)	0.588*	0.591*	0.982**	0.987**
	(1.87)	(1.88)	(2.00)	(2.01)
Ln(Brokerage size)	0.249	0.250	-0.802*	-0.817*
	(0.79)	(0.79)	(-1.87)	(-1.91)
Ln(Analyst career length)	-0.442	-0.439	-1.079*	-1.057*
	(-0.85)	(-0.84)	(-1.68)	(-1.65)
Ln(Time covering company)	-0.177	-0.181	-0.092	-0.096
	(-1.23)	(-1.25)	(-0.52)	(-0.55)
Lag monthly return	-0.003	-0.003	0.020	0.020
	(-0.31)	(-0.33)	(1.61)	(1.62)
Ln(Volatility)	1.371***	1.362***	0.927	0.938
	(3.02)	(3.00)	(1.43)	(1.44)
Momentum	-1.670***	-1.666***	0.280	0.268
	(-3.92)	(-3.91)	(0.46)	(0.44)
Analyst FE	Y	Y	Y	Y
Firm FE	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y
N	10,219	10,219	12,600	12,600
Adj. R^2	0.164	0.164	0.326	0.327

Table 7: Analyst Expertise Scores and Stock Liquidity

This table reports OLS regression results examining whether analysts' LLM-based knowledge and quality scores predict future liquidity. The dependent variables are next-period measures of illiquidity, effective spread, and quoted spread. Columns (1) and (2) report results for illiquidity, columns (3) and (4) for effective spread, and columns (5) and (6) for quoted spread. Within each outcome, odd-numbered columns report knowledge results and even-numbered columns report quality results. All specifications include firm fixed effects and industry-year fixed effects and control for lagged values of the dependent variable. t-statistics in parentheses are based on standard errors clustered at the analyst and firm levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	*Illiq (t+1)*		*Effective Spread (t+1)*		*Quoted Spread (t+1)*	
Measure	*Knowledge*	*Quality*	*Knowledge*	*Quality*	*Knowledge*	*Quality*
	(1)	(2)	(3)	(4)	(5)	(6)
Z(Measure Mean Top3)	-0.019***	-0.006	-0.015***	-0.004	-0.015***	-0.007
	(-3.77)	(-1.06)	(-2.96)	(-0.77)	(-3.13)	(-1.55)
Z(Measure Mean Top3) × Z(Log Size)	0.065**	0.039**	0.060***	0.035***	0.052***	0.035**
	(2.51)	(2.22)	(3.08)	(2.73)	(2.97)	(2.49)
Z(Log Size)	-0.221*	-0.229	-0.167***	-0.173**	-0.214**	-0.220**
	(-1.66)	(-1.59)	(-2.71)	(-2.49)	(-2.58)	(-2.43)
Z(Analyst Coverage)	0.008	0.009	0.011**	0.012**	0.003	0.005
	(0.92)	(0.90)	(2.05)	(2.10)	(0.67)	(0.93)
Institutional Ownership	-0.122	-0.119	-0.153	-0.151	-0.075	-0.074
	(-1.00)	(-0.97)	(-1.53)	(-1.50)	(-1.00)	(-0.98)
Institutional Ownership^2	0.092	0.074	0.087	0.072	0.013	0.001
	(1.09)	(0.91)	(1.27)	(1.06)	(0.24)	(0.02)
Return Volatility	-0.422	-0.456	0.317	0.245	0.196	0.128
	(-0.69)	(-0.70)	(0.37)	(0.27)	(0.18)	(0.11)
ROA	0.034	0.043	0.009	0.019	0.015	0.022
	(1.11)	(1.61)	(0.37)	(0.71)	(0.64)	(0.91)
Leverage	-0.017	-0.020	0.004	0.001	0.005	0.003
	(-1.03)	(-1.13)	(0.25)	(0.07)	(0.31)	(0.21)
Market-to-Book	0.001	0.001	0.002*	0.002*	0.001	0.001
	(1.51)	(1.61)	(1.72)	(1.78)	(1.26)	(1.32)
Annual Return	-0.044***	-0.046***	-0.082***	-0.084***	-0.070***	-0.071***
	(-3.52)	(-3.89)	(-7.29)	(-6.93)	(-7.14)	(-6.78)
Log Price	0.005	0.005	-0.055***	-0.055**	0.003	0.003
	(0.18)	(0.17)	(-2.60)	(-2.56)	(0.13)	(0.13)
Log Volume	-1.493*	-1.526*	-0.783**	-0.791**	-0.741**	-0.752**
	(-1.65)	(-1.67)	(-2.24)	(-2.31)	(-2.24)	(-2.31)
Log Volume^2	0.038	0.038	0.019**	0.020**	0.019**	0.019**
	(1.63)	(1.65)	(2.13)	(2.19)	(2.14)	(2.20)
Firm FE	Y	Y	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y	Y	Y
Lags	Y	Y	Y	Y	Y	Y
N	11,908	11,908	11,888	11,888	11,888	11,888
Adj. R^2	0.698	0.695	0.845	0.843	0.859	0.858

Table 8: Analyst Expertise Scores, Price Informativeness, and Market Efficiency

This table reports OLS regression results examining whether analysts' LLM-based knowledge and quality scores predict future price informativeness. The dependent variables are next-period measures of the absolute buy-sell ratio volume, lambda, and 5-min variance ratio. Columns (1) and (2) report results for buy-sell ratio volume, columns (3) and (4) for lambda, and columns (5) and (6) for variance ratio. Within each outcome, odd-numbered columns report knowledge results and even-numbered columns report quality results. All specifications include firm fixed effects and industry-year fixed effects and control for lagged dependent variables. t-statistics in parentheses are based on standard errors clustered at the analyst and firm levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	*BS Ratio Vol (t+1)*		*Lambda (t+1)*		*Variance Ratio (t+1)*	
Measure	*Knowledge*	*Quality*	*Knowledge*	*Quality*	*Knowledge*	*Quality*
	(1)	(2)	(3)	(4)	(5)	(6)
Z(Measure Mean Top3)	-0.016***	-0.011*	-0.015**	-0.010	-0.020***	-0.010*
	(-2.69)	(-1.89)	(-2.10)	(-1.05)	(-3.49)	(-1.77)
Z(Measure Mean Top3) × Z(Log Size)	0.055***	0.041***	0.065***	0.042***	0.066***	0.052***
	(4.24)	(3.08)	(4.86)	(4.61)	(2.82)	(3.39)
Z(Log Size)	-0.198***	-0.204***	-0.152***	-0.158***	-0.196***	-0.204**
	(-3.08)	(-2.97)	(-3.04)	(-3.11)	(-2.71)	(-2.58)
Z(Analyst Coverage)	0.009	0.011	0.022**	0.025**	0.026***	0.026**
	(0.90)	(1.10)	(2.35)	(2.55)	(2.76)	(2.58)
Institutional Ownership	-0.148	-0.146	-0.080	-0.081	-0.169	-0.166
	(-1.27)	(-1.26)	(-0.61)	(-0.63)	(-1.14)	(-1.12)
Institutional Ownership^2	0.088	0.077	-0.040	-0.052	0.053	0.039
	(1.00)	(0.87)	(-0.53)	(-0.68)	(0.55)	(0.42)
Return Volatility	1.032	0.962	3.771***	3.645***	-0.358	-0.435
	(1.29)	(1.11)	(5.82)	(5.60)	(-0.37)	(-0.42)
ROA	0.049**	0.055***	-0.079**	-0.071**	0.122***	0.129***
	(2.40)	(2.75)	(-2.49)	(-2.20)	(5.15)	(5.10)
Leverage	-0.012	-0.012	0.053**	0.052**	0.020	0.019
	(-0.55)	(-0.56)	(2.52)	(2.57)	(1.05)	(0.98)
Market-to-Book	-0.000	-0.000	0.003***	0.003***	0.001	0.001
	(-0.18)	(-0.13)	(2.66)	(2.72)	(1.43)	(1.49)
Annual Return	-0.036***	-0.037***	-0.088***	-0.090***	-0.067***	-0.068***
	(-3.30)	(-3.50)	(-5.71)	(-5.68)	(-5.10)	(-4.92)
Log Price	-0.055*	-0.055*	-0.266***	-0.265***	-0.053*	-0.053*
	(-1.81)	(-1.79)	(-3.82)	(-3.77)	(-1.75)	(-1.73)
Log Volume	-1.334***	-1.351***	0.425	0.387	-1.299***	-1.321***
	(-5.38)	(-5.45)	(1.43)	(1.28)	(-5.27)	(-5.35)
Log Volume^2	0.031***	0.031***	-0.014*	-0.013*	0.032***	0.033***
	(4.73)	(4.79)	(-1.90)	(-1.75)	(5.06)	(5.13)
Firm FE	Y	Y	Y	Y	Y	Y
Industry-Year FE	Y	Y	Y	Y	Y	Y
Lags	Y	Y	Y	Y	Y	Y
N	11,888	11,888	11,888	11,888	11,888	11,888
Adj. R^2	0.832	0.831	0.847	0.846	0.734	0.733

Table 9: Analyst Expertise Scores and Implied Cost of Capital

This table reports OLS regression results examining whether implied cost of capital varies with analysts' LLM-based knowledge and quality scores. The dependent variables are two implied cost of capital measures. Columns (1) and (2) report results for *ICC PEG*, and columns (3) and (4) for *ICC OJ*. Within each outcome, odd-numbered columns report knowledge results and even-numbered columns report quality results. All specifications include firm fixed effects and industry-year fixed effects. t-statistics in parentheses are based on standard errors clustered at the analyst and firm levels. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Dependent Variable	*ICC PEG*		*ICC OJ*	
Measure	*Knowledge*	*Quality*	*Knowledge*	*Quality*
	(1)	(2)	(3)	(4)
Z(Measure Mean)	-0.028***	-0.025***	-0.028***	-0.024***
	(-2.64)	(-2.61)	(-2.58)	(-2.59)
Z(Log Analyst Coverage)	-0.062**	-0.061*	-0.046	-0.046
	(-2.18)	(-1.94)	(-1.28)	(-1.28)
Z(Log Size)	-0.446***	-0.448***	-0.656***	-0.658***
	(-7.22)	(-7.25)	(-10.42)	(-10.46)
Z(Log Size)^2	-0.012	-0.012	-0.030	-0.030
	(-0.44)	(-0.44)	(-1.07)	(-1.06)
Log Volume	-0.004	-0.012	-0.080	-0.088
	(-0.01)	(-0.03)	(-0.17)	(-0.19)
Log Volume^2	0.004	0.004	0.007	0.007
	(0.34)	(0.35)	(0.57)	(0.59)
Institutional Ownership	0.667**	0.665**	0.872***	0.871***
	(2.28)	(2.28)	(2.83)	(2.82)
Institutional Ownership^2	-0.731***	-0.730***	-0.929***	-0.930***
	(-3.05)	(-3.05)	(-3.73)	(-3.73)
Past Return	0.001	0.002	0.064*	0.066*
	(0.02)	(0.06)	(1.90)	(1.94)
Return Volatility	7.730***	7.721***	5.080*	5.068*
	(3.18)	(3.17)	(1.94)	(1.94)
Leverage	0.099*	0.100*	0.145**	0.147**
	(1.67)	(1.69)	(2.33)	(2.35)
Market-to-Book	-0.004**	-0.004**	-0.002	-0.002
	(-2.36)	(-2.36)	(-1.02)	(-1.01)
MAE	0.070*	0.069	0.048	0.048
	(1.65)	(1.64)	(1.21)	(1.20)
Forecast Dispersion	0.023	0.023	0.096**	0.096**
	(0.47)	(0.47)	(2.00)	(2.01)
Beta	0.131***	0.131***	0.078***	0.078***
	(4.62)	(4.63)	(2.75)	(2.75)
Earnings Variability	0.030***	0.031***	0.029***	0.030***
	(2.86)	(2.86)	(2.58)	(2.61)
LT Earnings Growth	0.140***	0.140***	0.101***	0.100***
	(4.12)	(4.11)	(2.88)	(2.87)
Firm FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
N	10,303	10,303	10,295	10,295
Adj. R^2	0.594	0.594	0.511	0.511

A Appendix

A.1 Evaluating Industry Reports

A.1.1 Prompts

- **System Prompt:**

  You are an expert evaluator of sell-side analyst research reports. Use the following rubric to score reports:

  {`rubric`}

  Score each dimension from 1-5 based on the indicators in the rubric.

- **User Prompt:**

  Evaluate this analyst report:

  Filename: {`filename`}

  Report Content: {`full_text`}

  Output a markdown table with columns: Dimension, Score (1-5), Why (brief evidence from report). Then provide the composite score and quality tier. Nothing else.

- **System Prompt:**

  You are an expert evaluator of sell-side analyst research reports. Use the following rubric to score reports:

  {rubric}

  IMPORTANT: You are evaluating CHUNK {chunk_num} of {total_chunks} of a larger report. Your job is to identify and document evidence for each dimension found in THIS CHUNK. Focus on finding specific evidence: stock picks, price targets, risk sections, data sources, models, forecasts, etc.

- **User Prompt:**

  Review this CHUNK of an analyst report and document evidence for each dimension:

  Filename: {filename}

  This is chunk {chunk_num} of {total_chunks}
  (pages {chunk['start_page']}-{chunk['end_page']} of {chunk['total_pages']})

  Chunk Content:

  {chunk['text']}

  For each dimension, list specific evidence found in this chunk. If no evidence, write "None in this chunk". Output format:

```
| Dimension | Evidence from this chunk |
|-----------|--------------------------|
```

- **System Prompt:**

  You are an expert evaluator of sell-side analyst research reports. Use the following rubric to score reports:

  {`rubric`}

  You will receive evidence collected from multiple chunks of a large report. Your job is to synthesize ALL evidence across chunks and assign final scores.

- **User Prompt:**

  Based on the combined evidence from all chunks of this report, assign final scores:

  Filename: {`filename`}

  {`evidence_combined`}

  Now synthesize the evidence across ALL chunks and score the COMPLETE report. Consider cumulative evidence - e.g., technical expertise shown across multiple sectors in different chunks should score higher than expertise in just one area.

  CALIBRATION GUIDANCE:

  - Score 5 (Exceptional) should be RARE - only 1-2 dimensions per report at most
  - Score 4 (Strong) is appropriate for solid professional work that meets high standards
  - Score 3 (Adequate) is the baseline for competent professional work
  - Be critical: having evidence doesn't automatically mean a high score - evaluate QUALITY not just presence
  - A typical high-quality report should average 3.5-4.0, not 4.5+
  - For Originality: educational/overview reports score lower (3) even if they contain some contrarian views; score 5 requires the report's PRIMARY purpose to be a novel thesis
  - For Timeliness: scheduled events (teach-ins, conferences) score 3; score 4-5 requires event-driven/breaking analysis
  - For Technical Expertise: breadth across multiple sectors (MLPs + drilling + LNG + shipping + refining) warrants 5

  Output a markdown table with columns: Dimension, Score (1-5), Why (cite specific evidence from chunks). Then provide the composite score and quality tier. Nothing else.

A.1.2 Rubric

Analyst Report Quality Rubric

18 Scoring Dimensions (1–5 Scale)

Part A. General Analyst Report Quality (10 Dimensions)

#	Dimension	What to Look For
1	Originality	Contrarian view; differentiated perspective relative to consensus
2	Data Richness	Source tier; use of proprietary data
3	Analytical Depth	Explanation of "why" and "so what"; second-order effects
4	Actionability	Specific investment ideas, catalysts, or entry points
5	Structure	Clear thesis and organization; layered presentation for different audiences
6	Forward-Looking	Multi-year perspective; discussion of inflection points
7	Technical Expertise	Depth of domain knowledge; operational and industry nuances
8	Quantitative Rigor	Explicit calculations; use of specific numerical support
9	Risk Awareness	Thesis-specific risks; scenario analysis
10	Timeliness	Responsiveness to events; insight ahead of consensus

Part B. Industry Knowledge (8 Dimensions)

#	Dimension	What to Look For
11	Industry Trends	Industry cycles; secular versus cyclical forces; implications of structural change
12	Industry Technology	Key technologies; technology shifts; effects on costs, productivity, or competition
13	Supply Chains	Upstream and downstream structure; key inputs; capacity constraints or bottlenecks
14	Distribution Models	Go-to-market structure; role of intermediaries; margin or competitive implications
15	Margins	Margin drivers; cost structure; pricing power or competitive pressure
16	Customers	Customer segments; demand drivers; usage patterns; customer economics
17	Labor	Labor availability; wage dynamics; productivity or skill constraints
18	Management Teams	Observable actions; execution outcomes; capital allocation and strategic decisions

Quick Scoring Guide

Five-Point Scoring Guide

Score	Interpretation
5	Exceptional: proprietary evidence or deep industry-mechanism insight; paradigm-shifting analysis
4	Strong: differentiated and well-supported
3	Solid: competent and provides some value added
2	Basic: limited interpretation and modest analytical content
1	Weak: largely descriptive or repetitive, with little original insight

Source Tier Hierarchy

1. Proprietary surveys, channel checks, and expert networks
2. Specialized industry bodies (e.g., IGU, IEA, OPEC, EIA)
3. General data providers (e.g., Bloomberg, Reuters)
4. Company filings and management guidance
5. News articles and press releases

Quality Tiers

Composite Score Quality Tiers

Composite Score	Tier	Example
4.5–5.0	Exceptional	—
3.5–4.4	High Quality	BNP Clean Pulse (4.05), Kepler GTT (3.5)
2.5–3.4	Average	Kepler OFS (3.3), Truist Darlings (2.8), UBS Refiners (2.6)
1.5–2.4	Below Average	Kepler Political (2.3), Truist Data Comp (1.8)
1.0–1.4	Low Quality	—

Key Patterns Identified from 14 Reports

1. Report type predicts the ceiling: Weekly monitor-style reports tend to cap out around 3.0, whereas thematic reports can exceed 4.0.
2. Originality is the main differentiator: Variation in originality explains a substantial share of differences across quality tiers.
3. Quantitative rigor is not the same as insight: Reports with extensive tables but limited interpretation receive relatively low scores.
4. Actionability is distinct from rigor: A report may contain recommendations without strong analysis, or detailed analysis without explicit recommendations.
5. Flash notes can still be high quality: Short reports can score highly when they incorporate proprietary work.
6. Contrarian and quantitative analysis tends to score highest: Testable claims supported by evidence are a common feature of top-scoring reports.
7. Data compilation alone is not research: Reports with little narrative interpretation score poorly regardless of length.
8. Templates improve consistency: Standardized frameworks increase the likelihood that all relevant dimensions are addressed.

Red Flags

- No narrative discussion (tables only)
- Generic or boilerplate risk language
- Recycled standing views presented as event-driven research
- High page count but low informational density
- No source citations
- No specific numerical support or calculations

Green Flags

- Explicit contrarian estimate relative to consensus
- Proprietary data or proprietary modeling
- Bottom-up calculations shown explicitly
- Multi-dimensional analytical framework
- Testable or falsifiable thesis
- Part of a systematic research series

Table IA1: Number of Industry Reports by Brokerage House

The table presents the distribution of industry reports across brokerage houses in the sample.

Brokerage house	*No. of Reports*	Brokerage house	*No. of Reports*
Bank of America Global Research	164,144	Kepler Cheuvreux	7,667
JP Morgan	140,826	Scotiabank Gmb	7,585
UBS Equities	137,716	Canaccord Genuity	7,261
Morgan Stanley	111,000	Societe Generale	7,228
Deutsche Bank	73,137	Wedbush Securities	6,261
Credit Suisse	64,776	Santander	5,962
Jefferies	52,933	CIBC Capital Markets	5,254
Barclays	50,425	National Bank	4,914
RBC Capital Markets	36,960	Raymond James Ltd.	4,396
Wells Fargo	30,963	Guggenheim	4,170
Macquarie Research	25,515	BNP Paribas	3,957
BMO Capital Markets	21,000	Cantor Fitzgerald	3,601
Evercore ISI	18,789	SVB Leerink	3,494
Cowen and Company	18,722	Berenberg	3,112
HSBC	18,487	NatWest Markets	1,659
Wolfe Research	12,697	Nordea Markets	1,156
SunTrust Robinson / Truist Securities	12,549	Auerbach Grayson	1,082
Piper Sandler	10,340	CA Cheuvreux	343
Nomura	8,372	Equita SIM	159
Oppenheimer	8,302	Warburg Research Gmb	114

Table IA2: Sample Representativeness

This table presents t-test results comparing key characteristics of brokerages and analysts within and outside our sample. Panel A compares brokerage and forecast characteristics between brokerages selected for our sample and those excluded. *Mean(Selected)* and *Mean(Excluded)* report the average characteristics for each group, respectively. Panel B compares analyst characteristics within the selected brokerages, between analysts who have authored at least one industry report during the sample period and those who have not. *Mean(With Reports)* and *Mean(No Reports)* report the average characteristics for each group, respectively. *Mean Diff* is the difference between the two groups, and t-statistics in parentheses are based on standard errors clustered at the firm level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Comparison by Brokerage				
Variable	Mean(Selected)	Mean(Excluded)	Mean Diff	*t*-stat
Forecast error	1.688	2.054	-0.366***	-52.55
Brokerage size	37.243	4.642	32.601***	17.02
No. of firms	392.186	44.849	347.337***	15.01
Average firm size	27.808	37.590	-9.782	-0.80
Panel B: Comparison by Analyst				
Variable	Mean(With Reports)	Mean(No Reports)	Mean Diff	*t*-stat
Forecast error	1.652	1.639	0.013	0.13
No. of firms	9.626	5.325	4.301***	13.71
Average firm size	25.166	17.645	7.521***	4.72